

CARS 2022—Computer Assisted Radiology and Surgery Proceedings of the 36th International Congress and Exhibition Tokyo, Japan, June 7–11, 2022

© CARS 2022

Preface



Reflections on the CARS 2022 Theme of “Intelligent Technologies for Precision Diagnosis and Therapy” and on Model Guided Medicine

Every year on occasion of the CARS Congress, we are facing very similar questions, such as:

1. What are the issues in health care and specifically in the domain of CARS and what previous work can be built on?
2. Why are new concepts needed?
3. How can possible new concepts, for example, towards a Model Guided Medicine (MGM), be realised?
4. Who are the stakeholders and beneficiaries of possible MGM realisations?
5. When can the different stages of MGM solution concepts be realized?

In the following Preface of the IJCARS proceedings, an attempt is being made to address and to shine some light on the above questions in the context of the scope of IJCARS and the CARS 2022 Congress in Tokyo and beyond.

1. What are the issues in health care and specifically in the domain of CARS and what previous work can be built on?

More specifically, will the CARS concepts and themes, such as the

- Digital Operating Room (DOR)
- Hospital of the Future (HoF)
- Multi-disciplinary Computational Anatomy (MCA)

- Patient Specific Model (PSM)
- diagnostic and therapeutic Process Model (PM)
- Model Guided Medicine (MGM)
- Medical Information Modelling and Management System (MIMMS), etc.

still be part of a greater picture in a holistic medicine based health care and what other concepts are appearing on the horizon? What are the R&D questions that need to be addressed now and in the not too distant future, and how do they relate to what has been the focus of CARS in the past?

It must be observed, for example, even though patient specific modelling and the concept of a model guided medicine, based on an ICT architecture MIMMS are not new, they still do not represent part of “mainstream” R&D in CARS or similar domains.

2. Why are new concepts needed?

With the overwhelming amount of data, information, knowledge and occasional wisdom which circulate around health care and its institutions, the search for new concepts to handle this complex situation is mind blowing. Not least because of the fact that these and related new methods and technologies are expected to improve the quality of healthcare while at the same time are supposed to bring healthcare costs to lower levels. They, therefore, need to be addressed from a wider perspective by including many view points from different stakeholders, particularly when looking at complex healthcare situations and processes.

The wide spectrum of possible mathematical models is, why the challenges facing the CARS R&D community implies a continuous focus on new methods and tools, exemplified at CARS 2022 under the overall theme of “Intelligent Technologies for Precision Diagnosis and Therapy”.

3. How can possible new concepts, for example, towards a Model Guided Medicine (MGM) be realised?

In general terms it can be observed that a health care delivery with a wider view of the patient in mind implies a thorough if possible holistic understanding of the patient. The main question of course is,

how to build such a holistic model represented as a PSM and how can present and future methods and technical tools support integrated model building activities?

Recent advances in ICT-based methods and tools allow, in principle, only for a somewhat comprehensive but still very modest presentation of the individual patient and corresponding medical processes within a given domain of discourse, for example, in cardiovascular, neurological, orthopaedic or oncological disorders.

4. Who are the stakeholders and beneficiaries of possible MGM realisations?

Increasingly, the teams working in R&D projects in academia, healthcare settings and industry come from a wide variety of disciplines. In consequence, the synergy between the various active groups and individuals in medical technology developments give rise to the hope that integrated views of the patient will be facilitated when supported by appropriate tools. It is the objective of CARS to provide a platform and framework to enable this type of cooperation and synergies between physicians, engineers, scientists, providers and decision makers.

A possible concept on how these stakeholders for MGM realisations may cooperate is reflected in the horizontal and vertical

integration of MIMMS-like based MGM centres, see Fig. 1. Two infrastructures could be envisaged to support MGM research, services and management. One infrastructure is an MGM network of research centres with focus on modelling, modelling tools, mathematics of patient specific and process models, validation and validation tools (lower part of Fig. 1). The other is an MGM network of clinical centres focussing on health care domain selection with associated first and second order information entities as well as Model-Based Medical Evidence (MBME) with associated tools for MBME criteria definitions and selection (upper part of Fig. 1).

Both types of MGM networks are vertically connected via appropriate firewall-like security arrangements, in order to enable the exchange of PSMs and PMs as well as related algorithms while taking care about security and protection against cyber-attack. Specific configurations of firewalls need to be designed, as different research and clinical situations may demand.

Who will be part of the proposed horizontal MGM network centres should be subject to exhaustive discussions and evaluations. Multidisciplinary work, however, should not be impeded by different focusses of the two horizontal network centres but enhanced by means of think tanks, workshops and joint publications from members of cooperating and complimentary network centres. Active steps towards these multidisciplinary working modes have been taken by IFCARS in the past and are judged to be of prime importance in the future.

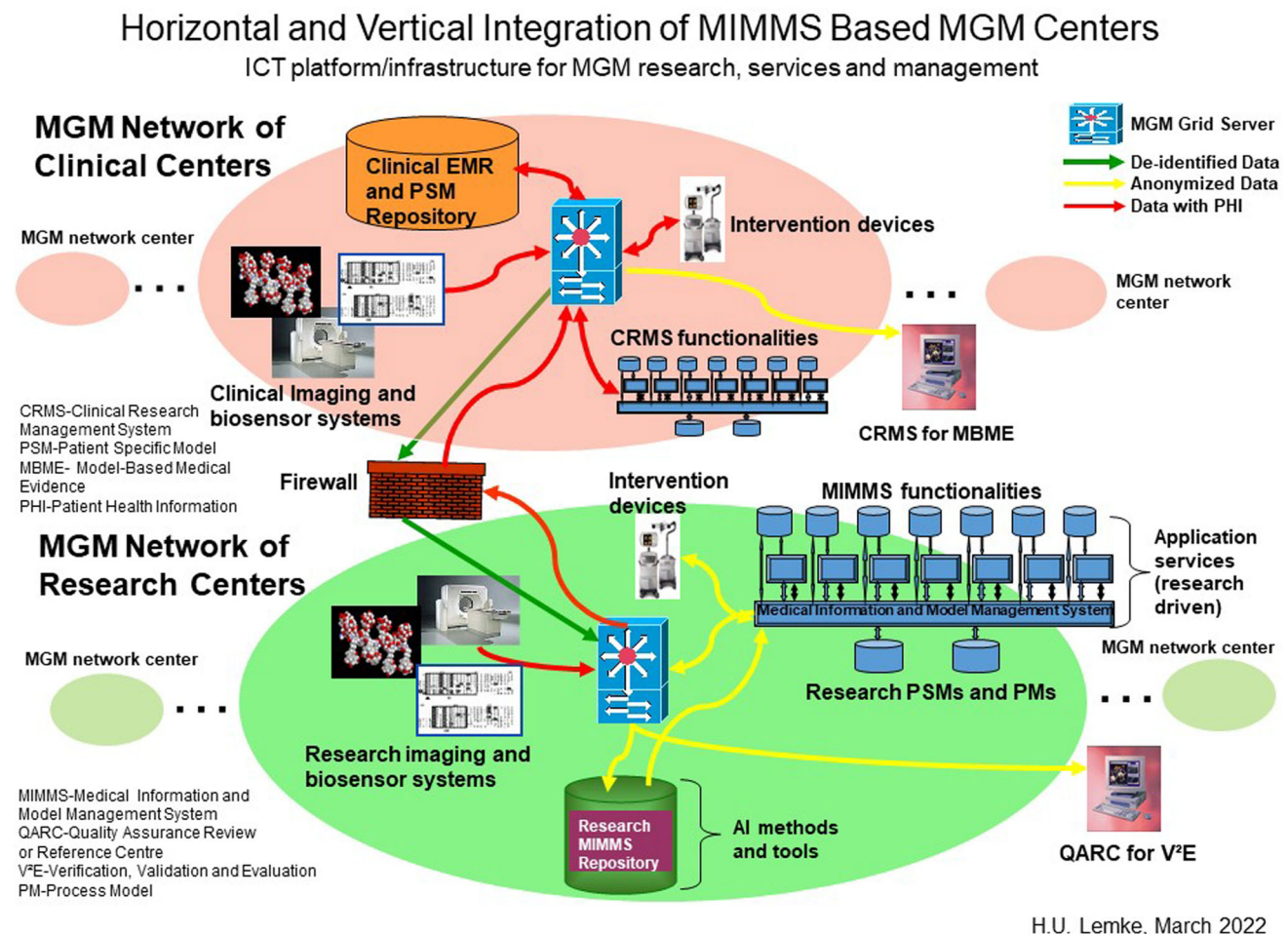


Fig. 1 Horizontal and vertical integration of MIMMS based MGM centres

5. When can the different stages of the MGM solution concepts be realized?

Compared to the four questions posed in the previous sections of this preface, the question about timing is the most difficult to answer. As far as short and mid-term future CARS congresses are concerned, it is likely that MGM will become a focus in selected healthcare domains. In the long term, however, this depends on the view points of the different stakeholders in the CARS community and their strategic decisions on the importance rating assigned to different R&D themes, including MGM.

Endeavours to understand the digital transformation in radiology and surgery are an important step in the direction to increase awareness to a balanced selection of future desirable “mainstream” R&D activities. The domain of CARS would therefore benefit from an introspection and idea generation, for example, in think tanks and workshops, on how it can best serve radiology and surgery specifically and medicine generally, when pursuing the path towards an MGM.

Acknowledgment

Finally, we should like to thank the enablers of the hybrid CARS 2022 Congress, in particular our Tokyo colleagues Ken Masamune, Kitaro Yoshimitsu, Shuji Kitahara, Manabu Tamura and all their assistants, but also the authors who took great effort to attend CARS 2022 personally in these difficult times and those who submitted a video version of their presentations. As we expect a stimulating discussion on the aforementioned topics also during the virtual part of the CARS 2022 Congress, we look forward continuing the discussion (in-presence) on the “Intelligent Technologies for Precision Diagnosis and Therapy” and “Model Guided Medicine” in subsequent workshops (exploratory) later in the year 2022 and beyond.

Heinz U. Lemke, PhD and Yoshihiro Muragaki, MD
Tokyo, June 2022

CARS 2022

Computer Assisted Radiology and Surgery

Proceedings of the 36th International Congress and Exhibition
Tokyo, Japan, June 7–11, 2022

Editors:

Professor Heinz U. Lemke, PhD

International Foundation for
Computer Assisted Radiology and Surgery
Klettgaustrasse 20
79790 Kuessaberg, Germany

Professor Leonard Berliner, MD

Staten Island University Hospital—Northwell Health
Department of Radiology
Staten Island, New York
475 Seaview Avenue
Staten Island, NY, 10305, USA

Professor Cristian A. Linte, PhD

Rochester Institute of Technology
Biomedical Modeling, Visualization and
Image-guided Navigation Laboratory
160 Lomb Memorial Dr
Rochester, NY 14623, USA

Professor Kensaku Mori, PhD

Nagoya University
Graduate School of Information Science
Information and Communications Headquarter
1 Furo-cho, Chikusa-ku
Nagoya 464-8601, Japan

Professor Yoshihiro Muragaki, MD

Tokyo Women's Medical University
Institute of Advanced Biomedical Engineering & Science
Faculty of Advanced Techno-Surgery
Department of Neurosurgery
8-1 Kawada-cho Shinjuku-ku
Tokyo 162-8666, Japan

Professor Hiroyuki Yoshida, PhD

Harvard Medical School
Massachusetts General Hospital
Department of Radiology
Suite 400C, 25 Chardon Street
Boston, MA 02,114, USA

Managing Editor (Editorial Office):

Franziska Schweikert

International Foundation for
Computer Assisted Radiology and Surgery
Klettgaustrasse 20
79790 Kuessaberg, Germany

CARS 2022 President**Yoshihiro Muragaki, MD, PhD**

Tokyo Women's Medical University (JP)

CARS Congress Organizing Committee**Christos Angelopoulos, DDS**

Columbia University, New York, NY (US)

Elizabeth Beckmann, BSc

Lanmark, Beaconsfield (GB)

Leonard Berliner, MD

Staten Island University Hospital, New York, NY (US)

Ulrich Bick, MD**Charité—University Medicine Berlin (DE)****Davide Caramella, MD**

University of Pisa (IT)

Kevin Cleary, PhD

Children's National Medical Center, Washington, DC (US)

Mario A. Cypko, PhD

Free University Berlin (DE)

Takeyoshi Dohi, PhD

Tokyo Denki University (JP)

Kunio Doi, PhD

Chicago, IL (US)

Volkmar Falk, MD, PhD

German Heart Institute Berlin (DE)

Allan G. Farman, PhD, DSc

Chicago, IL (US)

Hubertus Feussner, MD

Klinikum rechts der Isar, Munich (DE)

Guy Frija, MD

Georges Pompidou European Hospital, Paris (FR)

Miguel Á. González Ballester, PhD

ICREA-Pompeu Fabra University, Barcelona (ES)

Makoto Hashizume, MD, PhD, FACS

Kitakyushu Koga Hospital, Fukuoka (JP)

Yoshihiko Hayakawa, PhD

Kitami Institute of Technology, Hokkaido (JP)

Javier Herrero Jover, MD, PhD

Centro Médico Teknon, Barcelona (ES)

David Hilderbrand

Lakeway, TX (US)

Pierre Jannin, PhD

University of Rennes (FR)

Leo Joskowicz, PhD

The Hebrew University of Jerusalem (IL)

Shoji Kido, MD, PhD

Osaka University Graduate School of Medicine (JP)

Heinz U. Lemke, PhD (Chair)

International Foundation for CARS, Kuessaberg (DE)

Ken Masamune, PhD

Tokyo Women's Medical University (JP)

Kensaku Mori, PhD

Nagoya University (JP)

Yoshihiro Muragaki, MD, PhD

Tokyo Women's Medical University (JP)

Nassir Navab, PhD

Technical University Munich, Garching (DE)

Terry M. Peters, PhD

Robarts Research Institute, London, ON (CA)

Osman M. Ratib, MD, PhD, FAHA

University Hospital of Geneva (CH)

Hans G. Ringertz, MD, PhD

Karolinska Hospital, Stockholm (SE)

Yoshinobu Sato, PhD

Nara Institute of Science and Technology (JP)

Ramin Shahidi, PhD

Stanford University Medical Center (US)

Akinobu Shimizu, PhD

Tokyo University of Agriculture and Technology (JP)

Gabriele von Voigt, PhD

Leibniz University Hannover (DE)

Dirk F. Wilhelm, MD

Klinikum rechts der Isar der TUM, Munich (DE)

Hiroyuki Yoshida, PhD.

Harvard Medical School, Boston, MA (US)

Past Presidents

CAR '89	Heinz Oeser, MD Berlin (DE)	CARS 2008	Javier Herrero Jover, MD, PhD Barcelona (ES)
CAR '91	Auguste Wackenheim, MD Strasbourg (FR)	CARS 2009	Bernard L. Crowe, BA, MPH Canberra, NSW (AU)
CAR '93	J. Oscar M.C. Craig, MD London (GB)	CARS 2010	Osman M. Ratib, MD, PhD Geneva (CH)
CAR '95	Alexander R. Margulis, MD San Francisco (US)	CARS 2011	Michael M. Vannier, MD Chicago, IL (US)
CAR '97	Takahiro Kozuka, MD Osaka (JP)	CARS 2012	Davide Caramella, MD Pisa (IT)
CAR '98	Herbert Kaufmann, MD Berlin (DE)	CARS 2013	Nassir Navab, PhD Munich (DE)
CARS '99	Raffaella de Dominicis, MD Florence (IT)	CARS 2014	Makoto Hashizume, MD, PhD, FACS Fukuoka (JP)
CARS 2000	Pierre Rabischong, MD, PhD Montpellier (FR)	CARS 2015	Javier Herrero Jover, MD, PhD Barcelona (ES)
CARS 2001	Toyomi Fujino, MD, PhD Tokyo (JP)	CARS 2016	Davide Caramella, MD Pisa (IT)
CARS 2002	Kiyonari Inamura, PhD Hyogo (JP)	CARS 2017	Javier Herrero Jover, MD, PhD Barcelona (ES)
CARS 2003	H.K. Huang, DSc, FRCR (Hon.) Los Angeles, CA (US)	CARS 2018	Volkmar Falk, MD, PhD Berlin (DE)
CARS 2004	Karl-Heinz Höhne, PhD Hamburg (DE)	CARS 2019	Pierre Jannin, PhD Rennes (FR)
CARS 2005	Martin Lipton, MD, FACR, FAHA Los Angeles, CA (US)	CARS 2020	Hubertus Feussner, MD Munich (DE)
CARS 2006	Hans G. Ringertz, MD, PhD Stockholm (SE)	CARS 2021	Dirk Wilhelm, MD Munich (DE)
CARS 2007	Madjid Samii, MD Hannover (DE)		

Program Committee

Purang Abolmaesumi, PhD University of British Columbia, Vancouver, BC (CA)	Ulrich Bick, MD Charité—University Medicine Berlin (DE)
Asma'a Al-Ekrish, MD King Saud University, Riyadh (SA)	Uldis Bite, MD Mayo Clinic, Rochester, MN (US)
Paul R. Algra, MD, PhD Medical Center Alkmaar (NL)	Mario Ceresa, PhD Pompeu Fabra University, Barcelona (ES)
Russell J. Andrews, MD NASA Ames Research Center, Los Gatos, CA (US)	Ruey-Feng Chang, PhD National Taiwan University, Taipei (RC)
Christos Angelopoulos, DDS Columbia University, New York, NY (US)	Kiyoyuki Chinzei, PhD AIST, Tsukuba (JP)
Yoshinori Arai, PhD, DDS Nihon University, Atsugi (JP)	Kevin Cleary, PhD Children's National Medical Center, Washington, DC (US)
Jumpei Arata, PhD Kyushu University, Fukuoka (JP)	D. Louis Collins, PhD McGill University, Montreal, QC (CA)
Eiichiro Arij, DDS, PhD Aichi-Gakuin University School of Dentistry, Nagoya (JP)	Mario A. Cypko, PhD Free University Berlin (DE)
Paulo M. Azevedo-Marques, PhD University of Sao Paulo, Ribeirao Preto (BR)	Andreas Dietz, MD, PhD University Hospital Leipzig (DE)
Frits H. Barneveld Binkhuysen, MD, PhD Meander Medical Center, Soest (NL)	Adrienn Dobai, MD Semmelweis University, Budapest (HU)
John S.H. Baxter, PhD University of Rennes (FR)	Kunio Doi, PhD Chicago, IL (US)
Erwin Bellon, PhD University Hospitals Leuven (BE)	André J. Duerinckx, MD, Ph. Howard University College of Medicine, Washington, DC (US)
Maximilian Berlet, MD Klinikum rechts der Isar der TUM, Munich (DE)	Georg Eggers, MD, DMD, PhD Weinheim Head Center (DE)
Dirk Beyersdorff, MD University Hospital Hamburg-Eppendorf (DE)	Randy E. Ellis, PhD Queen's University, Kingston, ON (CA)
Silvio Diego Bianchi, MD University of Torino (IT)	Masahiro Endo, MD, PhD Shizuoka Cancer Center (JP)

- Uwe Engelmann, PhD**
CHILI GmbH, Dossenheim/Heidelberg (DE)
- Giuseppe Esposito, MD**
Georgetown University Hospital, Washington, DC (US)
- Caroline Essert, PhD**
University of Strasbourg (FR)
- Allan G. Farman, PhD, DSc**
Chicago, IL (US)
- Mansoor Fatehi, MD**
Federico II University, Tehran (IR)
- Josep Fernández-Bayó, PhD**
UDIAT Diagnostic Center, Sabadell (ES)
- Gabor Fichtinger, PhD**
Queen's University, Kingston, ON (CA)
- Michael H. Friebe, PhD**
Otto-von-Guericke-University, Magdeburg (DE)
- Masakatsu G. Fujie, PhD**
Waseda University, Tokyo (JP)
- Hiroshi Fujita, PhD**
Gifu University Graduate School of Medicine (JP)
- Shigekazu Fukuda, PhD**
National Institute of Radiological Sciences, Chiba (JP)
- Pablo Garcia Tahoces, PhD**
University of Santiago de Compostela (ES)
- Enrique J. Gomez-Aguilera, PhD**
Polytechnic University of Madrid (ES)
- Miguel Á. González Ballester, PhD**
ICREA-Pompeu Fabra University, Barcelona (ES)
- Lixu Gu, PhD**
Shanghai Jiaotong University (CN)
- Horst K. Hahn, PhD**
Fraunhofer MEVIS, Bremen (DE)
- Hideaki Haneishi, PhD**
Chiba University (JP)
- Stefan Habfeld, MD, MDS**
University Hospital Dortmund (DE)
- Yoshihiko Hayakawa, PhD**
Kitami Institute of Technology, Hokkaido (JP)
- Vicente Hernandez-Soler, MD, DDS**
Valencia University, Alicante (ES)
- Hiroshi Honda, MD, PhD**
Kyushu University, Fukuoka (JP)
- Jaesung Hong, PhD**
Daegu Gyeongbuk Inst. of Science and Technology (KR)
- Hidetaka Hontani, PhD**
Nagoya Institute of Technology (JP)
- Juan A. Hueto Madrid, MD**
Vall d'Hebron University Hospital, Barcelona (ES)
- Koji Ikuta, PhD**
The University of Tokyo (JP)
- Hiroshi Iseki, MD, PhD**
Waseda University, Tokyo (JP)
- Masayori Ishikawa, PhD**
Hokkaido University, Sapporo (JP)
- Hiroyasu Iso, MD**
Osaka University Graduate School of Medicine (JP)
- Stephan Jacobs, MD**
German Heart Institute Berlin (DE)
- Alissa Jell, MD**
Klinikum rechts der Isar der TUM, Munich (DE)
- Nobuyuki Kanematsu, PhD**
National Institute of Radiological Sciences, Chiba (JP)
- Akitoshi Katsumata, DDS, PhD**
Asahi University, Gifu (JP)
- Shigehiko Katsuragawa, PhD**
Teikyo University, Omuta (JP)
- Zafrulla Khan, DDS, MS**
James Graham Brown Cancer Center, Louisville, KY (US)
- Bob B. Kiaii, MD, FRCS**
London Health Sciences Centre, ON (CA)
- Shoji Kido, MD, PhD**
Osaka University Graduate School of Medicine (JP)
- Ron Kikinis, MD**
Harvard Medical School, Boston, MA (US)
- Young Soo Kim, MD, PhD**
Hanyang University Medical Center, Seoul (KR)
- Yasutomi Kinosada, MD, PhD**
Gifu University Graduate School of Medicine (JP)
- Masaki Kitajima, MD, PhD**
International University of Health and Welfare, Tokyo (JP)
- Etsuko Kobayashi, PhD**
Tokyo Women's Medical University (JP)
- Yo Kobayashi, PhD**
Waseda University, Tokyo (JP)
- Cornelia Kober, PhD**
Hamburg University of Applied Sciences (DE)
- Yoshie Kodera, PhD**
Nagoya University Graduate School of Medicine (JP)
- Masahiko Koizumi, MD**
Osaka University Hospital (JP)
- Vassili Kovalev, PhD**
United Institute of Informatics, Minsk (BY)
- Christian Krettek, MD**
Medical School Hannover (DE)
- Tomohiro Kuroda, PhD**
Kyoto University Hospital (JP)
- Florent Lalys, PhD**
Therenva SAS, Rennes (FR)
- Heinz U. Lemke, PhD**
International Foundation for CARS (DE)
- Lilian L.Y. Leong Fung, MBBS**
Hong Kong College of Radiologists (HK)
- Hongen Liao, PhD**
Tsinghua University School of Medicine, Beijing (CN)
- Christian A. Linte, PhD**
Rochester Institute of Technology (US)
- Brent J. Liu, PhD**
University of Southern California, Los Angeles, CA (US)
- Xiongbiao Luo, PhD**
Xiamen University (CN)
- Iván Maciá Oliver, PhD**
VICOMTech, San Sebastián (ES)
- Francisco Maculé Beneyto, MD**
Hospital Clinic, Barcelona (ES)
- Koutaro Maki, DDS, PhD**
Showa University School of Dentistry, Tokyo (JP)
- Rüdiger Marmulla, MD**
University Hospital of Heidelberg (DE)
- Ken Masamune, PhD**
Tokyo Women's Medical University (JP)
- Yoshitaka Masutani, PhD**
Hiroshima City University (JP)
- Yasushi Matsumura, MD, PhD**
Osaka University Hospital (JP)
- Andreas Melzer, MD, DDS**
Universities Dundee and St. Andrews (GB)
- Mamoru Mitsuishi, PhD**
The University of Tokyo (JP)
- Carlos Monserrat-Aranda, PhD**
Polytechnic University of Valencia (ES)
- Xavier Morandi, MD, PhD**
University of Rennes (FR)

- Kensaku Mori, PhD**
Nagoya University (JP)
- Parvin Mousavi, PhD**
Queen's University, Kingston, ON (CA)
- Seong K. Mun, PhD**
Arlington Innovation Center Health Research (US)
- Yoshihiro Muragaki, MD, PhD**
Tokyo Women's Medical University (JP)
- Kikkeri S. Nagesh, MDS**
R.V. Dental College, Bangalore (IN)
- Janne J. Näppi, PhD**
Harvard Medical School, Boston, MA (US)
- Shigeru Nawano, MD**
International University of Health & Welfare, Tokyo (JP)
- Valery Naranjo Ornedo, PhD**
Polytechnic University of Valencia (ES)
- Kwan-Hoong Ng, PhD**
University of Malaya, Kuala Lumpur (MY)
- Christopher Nimsky, MD**
University Hospital Giessen and Marburg (DE)
- Robert M. Nishikawa, PhD**
University of Pittsburgh (US)
- Fridtjof Nüsslin, PhD**
Technical University Munich (DE)
- Steffen Oeltze-Jafra, PhD**
Otto-von-Guericke-University, Magdeburg (DE)
- Isao Ohnishi, MD, PhD**
International University of Health and Welfare, Chiba (JP)
- Yoshiharu Ohno, MD, PhD**
Kobe University Graduate School of Medicine (JP)
- Minoru Ono, MD, PhD**
The University of Tokyo (JP)
- Kaan Orhan, DDS, PhD**
Ankara University Faculty of Dentistry (TR)
- Daniel Ostler, MSc**
Klinikum rechts der Isar der TUM, München (DE)
- José B. Pagador Carrasco, PhD**
Minimally Invasive Surgery Centre Jesus Usón (ES)
- Javier Pascua, PhD**
University Carlos III, Madrid (ES)
- Ewa Pietka, PhD**
Silesian University of Technology, Gliwice (PL)
- Andreas Pommert, PhD**
University Hospital Hamburg-Eppendorf (DE)
- Henri Primo, MS**
Primo Medical Imaging Informatics, Chicago, IL (US)
- Jörg Raczkowski, PhD**
University of Karlsruhe (DE)
- Klaus Radermacher, PhD**
Helmholtz-Institute for Biomedical Engineering, Aachen (DE)
- Stephen J. Riederer, PhD**
Mayo Clinic, Rochester, MN (US)
- Maria J. Rupérez-Moreno, PhD**
Polytechnic University of Valencia (ES)
- José M. Sabater Navarro, PhD**
University Miguel Hernandez, Elche (ES)
- Hidetoshi Saitoh, PhD**
Tokyo Metropolitan University of Health Sciences (JP)
- Luisa F. Sánchez Peralta, PhD**
Minimally Invasive Surgery Centre Jesús Usón (ES)
- Patricia Sánchez-González, PhD**
Polytechnic University of Madrid (ES)
- Warren S. Sandberg, MD, PhD**
Vanderbilt University Medical Center, Nashville, TN (US)
- Duygu Sarikaya, PhD**
University of Rennes (FR)
- Richard M. Satava, MD**
University of Washington Medical Center, Seattle, WA (US)
- Yoshinobu Sato, PhD**
Nara Institute of Science and Technology (JP)
- Kurt Alexander Schicho, MD, PhD**
Vienna General Hospital (AT)
- Wolfgang Schlegel, PhD**
German Cancer Research Center, Heidelberg (DE)
- Armin Schneider, PhD**
Klinikum rechts der Isar, Munich (DE)
- Alexander Schramm, MD, DDS, PhD**
Bundeswehrkrankenhaus Ulm (DE)
- Rainer Schubert, MD**
University for Health Sciences, Hall (AT)
- Dirk Schulze, MD, DMD**
Digitales Diagnostikzentrum Breisgau, Freiburg (DE)
- Achim Schweikard, PhD**
University of Lubeck (DE)
- Ramin Shahidi, PhD**
Stanford University Medical Center (US)
- Mitsuo Shimada, MD, PhD**
Tokushima University (JP)
- Akinobu Shimizu, PhD**
Tokyo University of Agriculture and Technology (JP)
- Amber Simpson, PhD**
Queen's University, Kingston, ON (CA)
- Eduardo Soudah, PhD**
Polytechnic University of Catalonia, Barcelona (ES)
- Stefanie Speidel, PhD**
National Center for Tumor Diseases, Dresden (DE)
- Jana Steger, MSc**
Klinikum rechts der Isar der TUM, Munich (DE)
- Matthäus Stöhr, MD**
University Hospital Leipzig (DE)
- Geert J. Streekstra, PhD**
Academic Medical Center, Amsterdam (NL)
- Nobuhiko Sugano, MD, PhD**
Osaka University Graduate School of Medicine (JP)
- Kazuro Sugimura, MD**
Kobe University Graduate School of Medicine (JP)
- Tae Suk Suh, PhD**
The Catholic University of Korea, Seoul (KR)
- Iori Sumida, PhD**
Osaka University Graduate School of Medicine (JP)
- Russell H. Taylor, PhD**
Johns Hopkins University, Baltimore, MD (US)
- Ulf K.-M. Teichgräber, MD, PhD**
University Hospital Jena (DE)
- Bart M. ter Haar Romeny, PhD**
Eindhoven University of Technology (NL)
- Klaus D. Toennies, PhD**
Otto-von-Guericke-University, Magdeburg (DE)
- Thomas Tolxdorff, PhD**
Charité—University Medicine Berlin (DE)
- Noriyuki Tomiyama, MD, PhD**
Osaka University Graduate School of Medicine (JP)
- Fukai Toyofuku, PhD**
Kyushu University, Fukuoka (JP)
- Jocelyne Troccaz, PhD**
University of Grenoble, La Tronche (FR)
- Jayaram K. Udupa, PhD**
University of Pennsylvania, Philadelphia, PA (US)
- Tamas Ungi, PhD**
Queen's University, Kingston, ON (CA)
- Maarten J.A. van Alphen, PhD**
The Netherlands Cancer Institute, Amsterdam (NL)

Eric vanSonnenberg, MD

St. Joseph's Hospital and Medical Cent., Phoenix, AZ (US)

Gabriele von Voigt, PhD

Leibniz University Hannover (DE)

Kirby G. Vosburgh, PhD

CIMIT, Boston, MA (US)

Shinichi Wada, PhD

Niigata University Faculty of Medicine (JP)

Andreas Wahle, PhD

The University of Iowa (US)

Stefan Weber, PhD

University of Berne (CH)

Michael Wich, MD

Unfallkrankenhaus Berlin (DE)

Dirk F. Wilhelm, MD

Klinikum rechts der Isar der TUM, Munich (DE)

Takaharu Yamazaki, PhD

Saitama Institute of Technology (JP)

Guang-Zhong Yang, PhD

Shanghai Jiao Tong University (CN)

Byung-Ju Yi, PhD

Hanyang University, Ansan (KR)

Hiroyuki Yoshida, PhD

Harvard Medical School, Boston, MA (US)

Toshiki Yoshimine, MD

Osaka University Graduate School of Medicine (JP)

Kitaro Yoshimitsu, PhD

Tokyo Women's Medical University (JP)

Computer Assisted Radiology and Surgery—36th International Congress and Exhibition

Chairs: Ulrich Bick, MD (DE)

Deep clustering of skin cancer from mass spectrometry imaging

R. Ellis¹, T. Mainguy¹, B. Everitt¹, R. Theriault¹, K. Y. M. Ren¹

¹Queen's University, Kingston, Canada

Keywords Skin Cancer, Mass Spectrometry, Variational Autoencoder, Digital Pathology

Purpose

Skin cancer is the most common malignancy reported in the United States, with basal cell carcinoma (BCC) accounting for approximately 80% of non-melanoma skin cancers [1]. The most common approach to treating BCC is surgical excision including Mohs micrographic surgery (MMS). MMS is associated with lower recurrence rates compared to conventional surgical excision [1] as it includes intra-operative histologic margin assessment to detect microscopic tumour foci at surgical margins during the procedure. A challenge of MMS is that it requires specialized personnel and the addition of a frozen section laboratory setup incorporated into the operating room making MMS more expensive and time-consuming than conventional excision methods [2]. Methods from mass spectrometry (MS), such as desorption electrospray ionization (DESI), may provide a less costly and time-intensive means of detecting malignancy at margins. Spatially sampled MS, called mass spectrometry imaging (MSI), can detect distributions of ions in a physical tissue sample. Recent developments have shown the effectiveness of MSI in discovering disease biomarkers, but clinical applications of MSI are still limited. MSI can be vexing to analyze; background signals are spectrally rich, and as a result distinguishing tissue pixels from background is often non-trivial. Because MSI data are typically unlabelled, analysis is often unsupervised and identifies clusters in MS data based on a similarity metric. Most works dealing in multivariate analysis of MSI data divide their approach into a two-stage process consisting of dimensionality reduction followed by clustering. While the results are impressive, models optimized to learn representations exclusively for the task of clustering have not yet been explored. Our aim was to use a deep clustering model to jointly optimize for the dimensionality reduction and clustering tasks, and to perform non-linear multivariate malignant versus benign tissue analysis. We proposed a two-step process that first segmented an MSI dataset into foreground and background, and then clustered the MSI foreground pixels. This work may be able to guide a clinician in determining whether positive margins are present that might require further tissue removal, specifically with applications to skin cancer such as BCC.

Methods

This study included 7 samples of excised skin tissue from 6 patients who underwent surgical removal of BCC. Each sample contained both malignant and benign tissue. Tissue was cryotomed onto a plain glass slide, scanned for MSI acquisition, formalin fixed, then H&E stained for routine pathological analysis. Each MS image pixel originally contained 2000 mass-charge abundances. Alignment was performed across all samples, resulting in a unified feature space with dimensionality 3120.

Our method used the variational autoencoder (VAE)-based variational deep embedding (VaDE) architecture to perform clustering of MSI data. Two separate VAE models were used for the segmentation and clustering tasks, each with the same architecture. For segmentation, the model was trained to search for three clusters; two background clusters and one foreground cluster. After model convergence, the two background clusters were combined to create one background cluster. We then applied morphological operators to the clustering mask; these retained and filled in only the largest connected

Table 1 Dice scores of the segmentation of histology images registered to segmentations of the MS images. The first row is the alphabetic code of the tissue sample. A Dice score of 0.9 or greater is usually interpreted as an excellent segmentation correspondence

Sample code	A	B	C	D	E	F	G	$\mu \pm \sigma$
Dice score	0.95	0.95	0.93	0.93	0.86	0.85	0.9	0.93 ± 0.03

component in the mask. Histology masks were affinely registered to each corresponding VAE mask, and the degree of overlap was assessed using Dice scores. The resulting foreground pixels were then processed using a VAE into 3 distinct clusters. The cluster association of each MS image pixel was visualized. VAE clustering models were trained and assessed using two methods. The first method was a leave-one-out method in which the foreground pixels in 6 MSI tensors were merged and used to train the model and determine VAE clusters; the remaining MSI tensor was clustered using this trained model. Results were visually compared to histological annotations performed by a board-certified pathologist. Because cancer is a highly heterogeneous disease, a second assessment was performed by training a distinct MSI model for each sample. Visual assessment was repeated.

Results

Image registration was visually highly accurate because the processing produced negligible tissue deformation. Our background segmentation method was able to achieve an average Dice score of 0.93 with a standard deviation of 0.03 (Table 1). The leave-one-out clustering technique was compared to the annotated histology images with disappointing results; most tissue samples had poor visual correspondence. The second method, which trained a model for each MS image, produced clusters that were in close correspondence with the pathologist's annotations (Fig. 1).

Conclusion

This work demonstrated an automated segmentation and clustering pipeline for MS images that jointly optimized for both the dimensionality reduction and clustering steps without requiring prior feature selection. One finding of this work was that for our dataset, cutaneous BCC was sufficiently heterogeneous such that a patient-specific

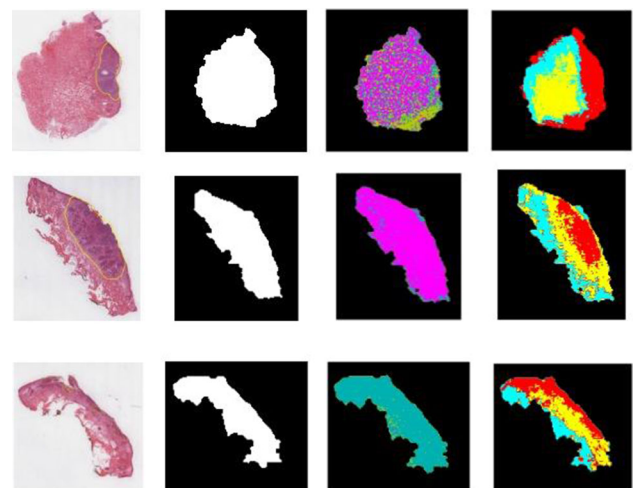


Fig. 1 For three MS images: the first column contains the corresponding histology image, the second column contains the foreground MS image segmentation, the third column contains the leave-one-out tissue clustering, and the final column contains the individual sample tissue clustering

model was required. MS methods such as DESI may provide a supplementary diagnostic tool for margin detection in surgeries such as MMS. Real-time microscopic analysis of resection margins is a time-consuming step that requires considerable expertise. An independent source of information—analysis of MS signals—could reduce the time necessary to determine whether the excised cutaneous tissue contains BCC or not. This work demonstrated the feasibility of using mass spectrometry for margin detection in skin cancer.

This work is a first and promising step in the use of MS to provide computer-aided diagnosis of malignant tissue, including at the margins, by detecting and analyzing small-molecules using metabolites. Future work could include the physical resection of marginal tissue for independent analysis, which might mimic a reasonable clinical application, and intraoperative testing using real-time DESI-MS acquisition. Because a VAE is effectively a Gaussian mixture model (GMM) in latent variables, MS methods may provide a pathologist with a greater understanding of the relations between cellular structure and cellular metabolism in cancer.

References

- [1] Rubin AI, Chen EH, Ratner D (2005) Basal-cell carcinoma. *N Engl J Med* 353(21):2262–2269.
- [2] Smeets NW, Krekels GA, Ostertag JU, Essers BA, Dirksen CD, Nieman FH, Neumann HM (2004) Surgical excision vs Mohs' micrographic surgery for basal-cell carcinoma of the face: randomised controlled trial. *Lancet* 364(9447):1766–1772, 2004.

Bone radiomics using Ultra-High Resolution CT: effects of imaging chain on texture biomarkers of bone quality

G. Shi¹, Y. Tai¹, F. Quevedo Gonzalez², R. Breighner², J. Siewerdsen¹, W. Zbijewski¹

¹Johns Hopkins University, Biomedical Engineering, Baltimore, United States ²Hospital for Special Surgery, New York, United States

Keywords CT quantitative biomarkers, bone quality, CT system models, radiomics.

Purpose

The increasing prevalence and healthcare burden of chronic skeletal disease such as osteoporosis (OP) and osteoarthritis (OA) necessitates the development of new quantitative biomarkers of bone health. In particular, there is a need for imaging technologies to assess trabecular microstructure ("bone quality") to augment the well-established bulk measurement of bone mineral density (BMD, "bone quantity"). Since the cancellous details are typically $< 100 \mu\text{m}$ size and thus beyond the resolution limit of current diagnostic modalities, there has been an increasing interest in utilization of image texture as an indirect marker of the underlying trabecular microarchitecture. Rigorous assessment of the confounding imaging chain effects on the quantitative texture features is necessary to develop robust radiomic models of bone health incorporating these metrics. Potential physical factors that might affect image texture include: noise magnitude and correlations (related to patient size), reconstruction algorithm, and variable spatial resolution throughout the field-of-view (FOV) due to e.g. focal spot blur.

Recently, a new generation of multi-detector CT systems with $\sim 2 \times$ improved spatial resolution compared to current clinical standard has been introduced. These emerging Ultra-High Resolution CT (UHR CT) devices, such as Canon Precision CT, can visualize

details to $\sim 150 \mu\text{m}$. UHR CT might thus yield textural biomarkers that better reflect the underlying cancellous microarchitecture than conventional CT [1]. However, many of the physical factors affecting image texture have different impact in UHR CT compared to conventional systems. For example, the variability of spatial resolution throughout the FOV is more pronounced in UHR imaging compared to standard resolution CT [2]. Therefore, this study employs experimentally calibrated, physics-based CT system models to evaluate the impact of UHR imaging chain on texture features of bone. The aim is to develop reproducible trabecular biomarkers that fully exploit the potential of UHR CT in bone radiomics. Here, we report specifically on the impact of the location-dependent azimuthal blur associated with CT gantry motion during x-ray exposures.

Methods

UHR CT: The Canon Precision UHR CT is based on a 160 slice detector with 0.25 mm slice thickness (at the isocenter) and 0.25 mm axial pixel size. The x-ray source provides focal spot sizes ranging from $0.4 \times 0.5 \text{ mm}$ (the finest setting) to $1.6 \times 1.4 \text{ mm}$. These novel components enable potential $\sim 2 \times$ improvement in spatial resolution over current conventional CT (0.5 mm slice thickness). We compare UHR CT acquisition protocols involving the finest focal spot and scan rotation times from 0.35 s to 1.5 s. Reduction in scan time implies longer distance traveled by the source during a single x-ray exposure, leading to worsening axial-plane azimuthal blur that increases with the distance from the isocenter. Even for the relatively slow 1.5 s rotation, previous experimental studies [2] estimated that this blur results in $\sim 25\%$ reduction in the frequency of 10% modulation (f10) over a radial distance of 150 mm from the isocenter. Note that in conventional resolution CT, this resolution loss is much less pronounced due to the more prominent role of detector pixel blur ($< 10\%$ reduction at 150 mm from isocenter).

Blur simulation: A digital trabecular bone phantom was obtained from a micro-CT scan of a human hamate. Voxel size was 0.1 mm. In this initial investigation, azimuthal blur was modeled as an image-domain rectangular convolution kernel. The width of the kernel was adjusted based on gantry speed and distance from the isocenter (ranged 0 mm–150 mm) using a parametrization obtained from experimental UHR CT measurements of a $150 \mu\text{m}$ tungsten wire at 0 mm and 65 mm from the isocenter. Ongoing studies employ a realistic forward projector with a cascaded systems detector model to capture a broader range of physical effects affecting imaging performance.

Texture metrics and performance assessment: Gray Level Co-occurrence Matrix (GLCM) and Gray Level Run Length Matrix (GLRM) features were obtained in 21 square regions-of-interest (ROIs, 2.2 mm side length) in each of the blurred trabecular bone images. The variability of texture features across the FOV imparted by the location-dependent azimuthal blur, concordance correlation coefficient (CCC) was computed between the ROI texture metric values at the isocenter and the values measured at each of the simulated distances from the isocenter.

Results

Figure 1 illustrates typical CCC results for representative texture features: GLCM entropy and GLRM gray level non-uniformity (denoted GLNU). For both scan speeds, the concordance with measurements obtained at the isocenter worsens with radial distance. However, the increasing disagreement is more pronounced for the fast 0.35 s scan. For the 1.5 s rotation and both texture metrics, the CCC at 105 mm is ~ 0.9 , indicating good agreement with the 0 mm reference. For the 0.35 s rotation, the CCC at 105 mm is ~ 0.7 for GLNU and ~ 0.5 for entropy, suggesting a substantial lack of reproducibility between measurements obtained at different locations in the FOV. This change in image texture is consistent with the visual impression of increasing azimuthal blur evident in the trabecular bone images for the 105 mm location and 0.35 s scan (shown above the

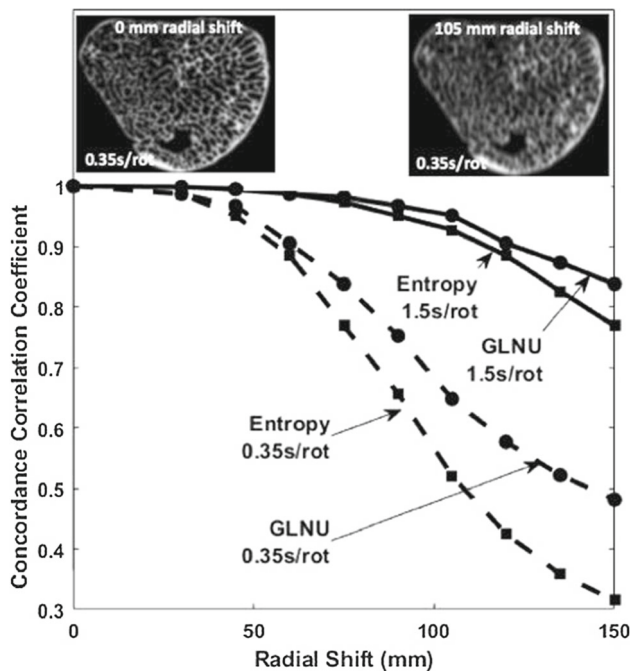


Fig. 1 Concordance correlation coefficient between texture features of trabecular bone (GLCM entropy, squares, and GLRM gray level non-uniformity, GLNU, circles) measured at the isocenter of an UHR CT system and at a range of radial distances from isocenter. Azimuthal blur due to gantry motion increases with radial distance. The resulting decrease in concordance with the reference isocenter value is more pronounced for the faster 0.35 s rotation scan (dashed line) compared to 1.5 s scan (solid line). The effects of the blur are illustrated with images at 0 mm and 105 mm for 0.35 s rotation

CCC plot). Similar trends were observed for other texture features: for the majority of the metrics, the CCC at 105 mm was ≥ 0.8 for the 1.5 s rotation and ≤ 0.7 for the 0.35 s rotation.

Conclusion

The results illustrate that texture features of trabecular bone extracted from UHR CT exhibit appreciable dependence on the location within the FOV due to a variety of non-stationary CT system blurs. Rigorous understanding of such variability is essential to ensure that radiomic models of bone health (e.g. to predict OP fracture risk) are robust to changes in the acquisition parameters, patient size, and positioning. The preliminary study presented here suggest that scan rotation speed must be kept relatively slow to obtain reproducible biomarkers of bone quality.

Ongoing work using realistic CT system models further investigates the link between trabecular microarchitecture, image texture, and imaging system physics across a broader range of settings. The results will likely be applicable for other emerging high-resolution technologies where the minimization of detector blur may exacerbate the effects of source-side blurs, such as photon-counting CT.

References

- [1] G. Shi, S. Subramanian, Q. Cao, S. Demehri, J.H. Siewerdsen, W. Zbijewski, “Application of a novel ultra-high resolution multi-detector CT in quantitative imaging of trabecular microstructure,” SPIE Medical Imaging 2020, Houston, TX, 113171E, 2020.

- [2] A. Hernandez, P. Wu, M. Mahesh, J.H. Siewerdsen, J.M. Boone. Location and direction dependence in the 3D MTF for a high-resolution CT system. *Med Phys.* 48(6):2760–2771, 2021.

Deep extracted features to support Content-Based Image Retrieval systems in the diagnosis of Covid-19 and Interstitial diseases

M. Bedo^{1,2}, L. Lima¹, M. Koenigkam-Santos¹, A. Traina³, P. Azevedo-Marques¹

¹USP, FMRP, Ribeirão Preto—SP, Brazil ²Fluminense Federal University, S. A. Padua, Brazil ³USP, ICMC, Sao Carlos, Brazil

Keywords Covid-19, CBIR, kNN, Deep Learning.

Purpose

Coronavirus disease 2019 (Covid-19) may cause dyspnoea, whereas Interstitial Lung Diseases (ILD) may lead to the loss of breathing ability. In both cases, chest X-Ray is typically one of the initial studies to identify the diseases as they are simple and widely available scans, especially in under-development countries. However, the assessment of such images is subject to a high intraobserver variability because it depends on the reader’s expertise, which may expose patients to unnecessary investigations and delay the diagnosis. Content-based Image Retrieval (CBIR) tools can bridge such a variability gap by recovering similar past cases to a given reference image from an annotated database and acting as a differential diagnosis CAD-IA system [1]. The main CBIR components are the feature extraction and the query formulation. The former represents the compared images into a space where a distance function can be applied, and the latter relies on the k-Nearest Neighbor (kNN) method to fetch the most similar cases by their distances to the query reference.

In this study, we examine the quality of Covid-19 and ILD deep features extracted by a modified VGG-19 Convolutional Neural Network (CNN) [2] following the perspective of the Voronoi frontiers induced by kNN, which is at the core of the CBIR query formulation component.

Methods

We curated a dataset of annotated chest X-Rays from our PACS/HIS systems following a retrospective study approved by the institutional board. A set of 185 Covid-19 and 307 ILD cases from different patients was selected, being Covid-19 cases confirmed by RT-PCR tests and ILD images included after the analysis of two thoracic radiologists. We also added 381 images of “Healthy” lungs (without Covid-19 or ILD) to enrich the dataset. The resulting set includes 873 X-Rays (mean age 60.49 ± 15.21 , and 52.58% females).

We cast the DICOM images into PNG files by using the Hounsfield conversion and a 256 Gy-scale window. The files were scaled to 224×224 images and fed into a modified VGG-19 version we implemented [2]. Our version includes the stack of convolutional layers and five new layers after the block5_pool, namely: GlobalAveragePooling2D, BatchNormalization, a dense layer with 128 units and ReLU, a dropout layer with ratio 0.6, and a final dense layer with three neurons for classification. The Adam function was used to minimize cross-entropy, whereas batch size and epochs were set to 36 and 100, respectively. All layers start with ImageNet weights that were frozen until block4_pool so that only the remaining layers were updated. We fed the CNN with images and labels (*i.e.*, {Covid-19, ILD, Healthy}) so that our feature extraction procedure was oriented towards those classes rather than autoencoders.

The flattened outputs of the last max-pooling layer were collected as feature vectors of dimensionality $d = 512$. We clean and

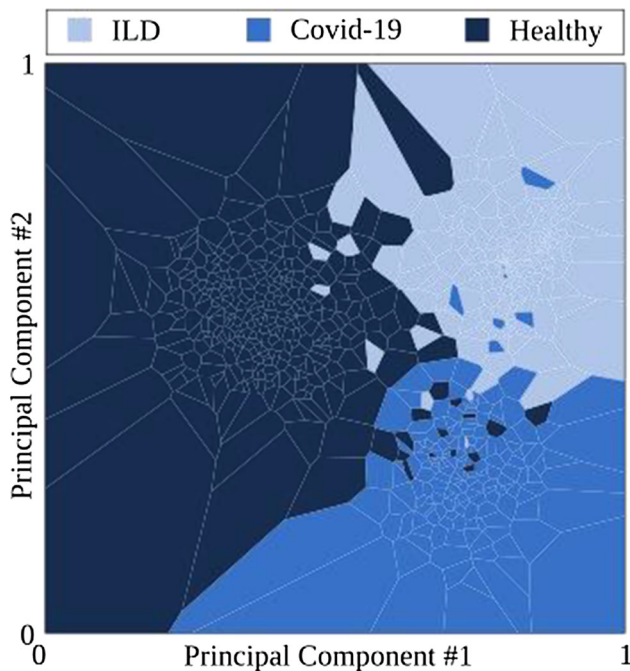


Fig. 1 Voronoi frontiers induced by kNN

preprocess those vectors before applying the kNN-based search mechanism. First, we scaled the dimensions into the [0,1] interval. Then, we perform a reduction by using the Principal Component Analysis (PCA). The number of reduced dimensions was determined by the intrinsic dimensionality of the features, estimated by the mean (μ) and standard deviation (σ) of the pairwise distance distribution as the value $\mu^2/2 \cdot \sigma^2$. Finally, the reduced vectors were also scaled into the [0,1] interval. The experiments were performed in a 3854 core 1.5 GHz GPU NVidia TitanX 12 GB RAM, and an Intel(R) Xeon(R) CPU 2.00 GHz, 96 GB RAM. The code was implemented under Tensorflow (v.2.1.0) and R (v4.1.2).

Results

We used two Principal Components to reduce the vectors according to the estimated intrinsic dimensionality. Figure 1 shows the Voronoi frontiers induced by kNN with a smooth separation between the three classes, which creates a search space in which CBIR searches are expected to be accurate. We quantify such behavior through a kNN-based classification on the two experimental settings (*i.e.*, *10-folds* and *Holdout*) by using the scaled features with and without dimensionality reduction. Table 1 summarizes the results with the following findings:

- The accuracy measures increased with the neighborhood ($k = 1$ vs. $k = 5$) in all experimental cases,
- Covid-19 cases were more difficult to label than ILD according to F1 and RC,
- The kNN hit-ratio (TP) for Covid-19 was comparable to the very first diagnosis stored into the PACS/HIS systems by readers on duty regarding the *Holdout* cases (readers' mean $\sim 63\%$ vs. KNN $\sim 59\%$),
- Searches over the reduced data were $\sim 4 \times$ faster, and
- While dimensionality reduction was just as suitable as non-reduced data in the *10-folds* evaluation, it expressively enhanced the kNN performance for the *Holdout* test (*e.g.*, 0.68 vs. 0.82, $k = 1$ and F1). This result shows the side-effects of searching high-dimensional spaces with kNN (the "curse of dimensionality"), which requires pre-processing the vectors or defining other query criteria to browse the data.

Table 1 Accuracy measures of test cases

Accuracy measures for both tests: 10-folds cross-validation (10-folds) And external instances (Holdout). Average values are presented outside the brackets, while inner brackets' entries represent classes ILD/Covid-19, *i.e.*, lines in the format of values AVG[ILD/Covid-19] on a $\times 10^2$ scale

	$k=5$			
	$d=512$		$d=2$	
	<i>10-folds</i>	<i>Holdout</i>	<i>10-folds</i>	<i>Holdout</i>
TP	91[93/83]	66[83/50]	91[92/84]	75[100/50]
PR	91[92/89]	73[71/75]	91[92/84]	87[75/1]
RE	91[93/83]	66[83/50]	91[92/84]	75[100/50]
RC	93[94/90]	70[75/66]	93[94/89]	79[83/75]
F1	91[92/86]	68[76/60]	91[92/84]	76[85/66]
CK	87	38	86	53

	$k=5$			
	$d=512$		$d=2$	
	<i>10-folds</i>	<i>Holdout</i>	<i>10-folds</i>	<i>Holdout</i>
TP	94[95/89]	83[100/66]	94[95/90]	83[100/66]
PR	94[94/89]	87[75/100]	94[95/88]	87[75/100]
RE	94[95/89]	80[100/67]	94[95/90]	83[100/66]
RC	97[98/96]	80[80/80]	97[97/96]	80[80/80]
F1	94[95/89]	82[85/80]	94[95/89]	82[85/80]
CK	90	66	91	66

Conclusion

This study has discussed feature extraction for Covid-19 and ILD images from the perspective of kNN queries, the query formulation component within CBIR systems. Although we used cross-validation and one external batch to mitigate overfitting, a practical limitation was the size of the CNN training set. Still, our approach showed promising results in the extraction of suitable features for CBIR environments.

References

- [1] Choe, J., Hwang, H. J., Seo, J. B., Lee, S. M., Yun, J., Kim, M.-J., Jeong, J., Lee, Y., Jin, K., Park, R., Kim, J., Jeon, H., Kim, N., Yi, J., Yu, D., and Kim, B. (2021). Content-based Image Retrieval by Using Deep Learning for Interstitial Lung Disease Diagnosis with Chest CT. *Radiology*, 1(1):1–12.
- [2] Lima, L. L., Nather Jr., J. C., Koenigkam-Santos, M., Peron, L., Lopes, K. L. F., Machado, C., Sarmiento-Neto, F., Wada, D., Koenigkam-Santos, M., and Azevedo-Marques, P. M. (2021). Predicting interstitial and Covid-19 diseases on chest X-Ray using a convolutional neural network with transfer learning. *Supplement of the International Journal for Computer Assisted Radiology and Surgery*, 16(S1):68–69.

Fusion of high resolution manometry and cinematography in the diagnosis of dysphagia

A. Jell¹, L. Bernhard², D. Wilhelm¹, H. Feußner¹

¹Klinikum rechts der Isar, Surgical Department, München, Germany

²Klinikum rechts der Isar, TU, Munich, Germany

Keywords dysphagia, cinematography, data fusion, esophageal manometry

Purpose

Dysphagia, the inability or difficulty to swallow, has a lifetime prevalence of about 5%. Especially in elderly people dysphagia occurs in up to 33% above the age of 65 years. Dysphagia occurs both in benign and malignant diseases of the oropharyngeal tract and the esophagus as well as in neuromuscular diseases. Being unable to swallow food, drinks or even saliva has a decisive influence on the quality of life if patients are no longer able to eat in public or at all, which is leading to malnutrition.

The gold standard in the diagnosis of dysphagia is, after ruling out malignancy by endoscopy and CT, functional assessment, esophageal manometry and X-ray cinematography. These diagnostic means are performed consecutively, mostly not even on the same day and especially not by the same physicians or examiners. Therefore they cover different acts of swallowig and (dys)motility events, which cannot be compared one on one. Especially oropharyngeal dysphagia sill poses major problems in comparison different diagnostic means due to swallowing in a highly dynamic area with the interaction of a large number of muscles and cranial nerves [1].

Furthermore different diagnostic means in one single patient are rarely presented to the physician side-by-side, non-speaking simultaneously. Therefore we propose a novel approach for fusing time-synchronized high-resolution manometric pressure data and X-ray imaging to a single picture visualization method as a first step.

In the second step the exposure to X-ray radiation during cinematography can be reduced by synchronizing these two diagnostic modalities to such an extent that the X-ray radiation field and time of radiography are cut down to minimum necessary time and dose [2].

Methods

We used a state-of-the-art high resolution esophageal manometry probe with the mobile data logger MALT (Standard Instruments, Karlsruhe, Germany), the software ViMeDat (Standard Instruments, Karlsruhe, Germany) and an extended digital imaging X-ray machine (Philips Medical Systems, Hamburg, Germany) at the Klinikum rechts der Isar (Technical University of Munich, Germany). Both diagnostic means were recorded simultaneously and time-synchronized in study participants who were asked to swallow multiple times 10 ml of fluid contrast agent.

In the next step positions of the pressure sensors on the manometric probe had to be detected for each frame of the cinematography. Subsequently we fitted a rectangular b-spline through these detected pressure sensors to estimate the path of the manometric probe and its movement, Fig. 1.

The most complex step was then to integrate the manometric pressure data into the cinematographic video as high-resolution manometry is sampled at 50 Hz in comma separated XML whilst dynamic X-ray is recorded at about 30 Hz in DICOM-Standard. Especially with our primary aim to make diagnosis of oropharyngeal dysphagia more intuitive and feasible for physicians we were finally able to implement a graphical user interface application based on Python (PyQT, Riverbank Computing Limited, UK).

Results

We tested our sensor detection algorithm on a small data set in patients with oropharyngeal dysphagia with very positive outcome. Both experienced examiners and young doctors appreciated our program as very intuitive and as an enrichment in the diagnosis of benign forms of oropharyngeal or esophageal dysphagia.

The template matching approach for the detection of manometric pressure sensors alone yields rather poor results in underexposed areas of the X-ray frames (e.g. mandibular region). However, when combined with spline-fitting, sensor positions can be estimated very reliably and with high accuracy even beyond the sensor visibility for human eyes. This makes it for the first time feasible to reduce X-ray radiation field and time of radiography without significant loss of information.

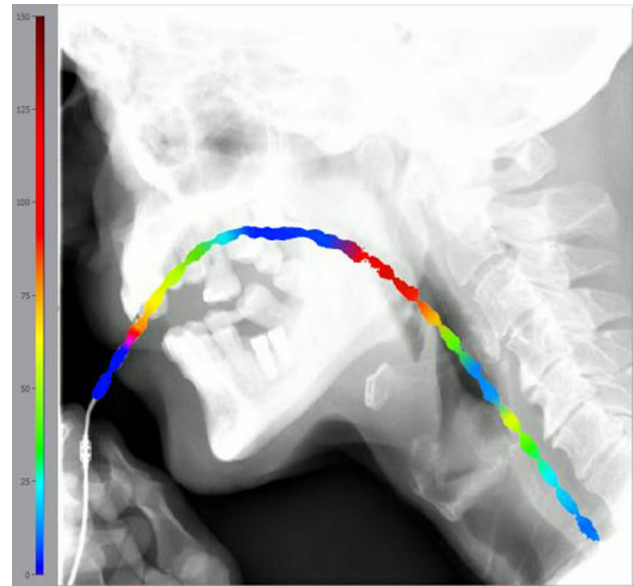


Fig. 1 Fused augmented simultaneously performed high resolution manometry and cinematography

Our new tool for fusing manometric and cinematographic data can be used to create patient-individual, but standardized examinations or to create specific project files associated with different clinical cases that patients' examinations can be integrated in. After time-synchronized manometric and X-ray (video) data has been loaded into a project, the sensor detection algorithm is started. Once the analysis has terminated, the detected sensor positions are shown in the X-ray cinematography and the pressure values are overlaid as spatio-temporal color plots which are state of the art in esophageal high resolution manometry.

Conclusion

Fusion of multiple diagnostic means of different origins is despite today's technical possibilities still often done by medical experts by viewing diagnostics one by one or in a non-intuitive and insufficient way side by side.

In the diagnosis of oropharyngeal dysphagia manometric and cinematographic data is most commonly used. The goal of our work was to develop a tool for fusing manometric and cinematographic information into augmented, patient-individual examinations. The great feedback by medical experts shows the huge potential of such fused imaging applications, which directly support medical doctors in daily clinical routine and reduce time and dosage of exposure to X-ray in patients.

References

- [1] C. A. Jones, E. L. Meisner, C. K. Broadfoot, S. P. Rosen, C. R. Samuelsen, and T. M. McCulloch, "Methods for measuring swallowing pressure variability using high-resolution manometry," *Frontiers in applied mathematics and statistics*, vol. 4, p. 23, 2018.
- [2] H. Y. Kim and S. A. De Araújo, "Grayscale template-matching invariant to rotation, scale, translation, brightness and contrast," in *Pacific-Rim Symposium on Image and Video Technology*, pp. 100–113, Springer, 2007.

Deep learning based radiological longitudinal volumetric evaluation of brain metastases after Stereotactic Radiosurgery

L. Joskowicz¹, A. Offenbacher¹, L. Kahanov^{1,2}, I. Ramirez^{1,2}, Y. Shoshan^{1,2}

¹The Hebrew University of Jerusalem, School of Computer Science and Engineering, Jerusalem, Israel ²Hadassah University Medical Center, Department of Neurosurgery, Jerusalem, Israel

Keywords Longitudinal analysis, Machine learning, Brain metastases, Stereotactic neurosurgery

Purpose

Stereotactic radiosurgery (SRS) is an established and increasingly used indication for the treatment of brain tumors and metastases. It requires the annotation of the brain metastases to be irradiated in a contrast-enhanced T1-Gd brain MRI scan acquired prior to the treatment. After SRS, radiological monitoring of the irradiated brain metastases is performed with consecutive MRI scans. While SRS may not fully eliminate these lesions, it may slow their growth, reduce their volume, and thus decrease their potential malignancy.

Clinical decisions on SRS treatment efficacy and treatment continuation are based on the comparison of the size of the lesions in the pre- (baseline) and post- (follow-up) treatment MRI scans. While volumetric measurements are considered superior to the currently used RECIST linear measurements, they are not performed routinely since they require the manual/interactive delineation of the brain metastases, which is time-consuming, error prone, and subject to significant observer variability. In previous works, we have developed methods for longitudinal follow-up of various types brain lesions [1, 2].

The goal of our work is to develop an automatic method for the accurate and reliable volumetric radiological evaluation and visualization of brain metastases after SRS on follow-up brain T1-Gd MRI scans using the available baseline metastases delineation and a deep learning network trained on a variety of annotated datasets.

Methods

Our method consists of a pipeline of five steps: (1) resolution matching between the baseline and the follow-up scans; (2) registration of the baseline and follow-up scans by Normalized Mutual Information; (3) Region of Interest (ROI) computation for each brain metastasis in the follow-up scan based on its segmentation in the baseline scan; (4) automatic segmentation of the brain metastases in the follow-up scan; (5) simultaneous analysis and visualization of the brain metastases segmentations in the baseline and follow-up scans.

The brain metastases segmentation is performed with two custom 3D U-Net classifiers: one for small metastases (diameter ≤ 10 mm) and one for large metastases (diameter > 10 mm). The reason to use two classifiers instead of one is the significant difference in the appearance of the brain metastases: small metastases tend to be homogeneous, while larger ones are more inhomogeneous and may have different sub-components. Additionally, fewer annotated datasets are needed to train each network separately. Both networks use non-isotropic sampling and batch normalization; they are trained with training time augmentations consisting of flipping, rotation, scaling and shearing. The loss function is the Dice similarity coefficient. The patch size is $64 \times 64 \times 16$. We applied transfer learning from the large metastases network to the small metastases network to compensate for class imbalance.

The results of the analysis are presented in a custom viewer which consists of three main windows (Fig. 1): a main window in axial slices of the registered baseline and follow-up scans are visualized, and a

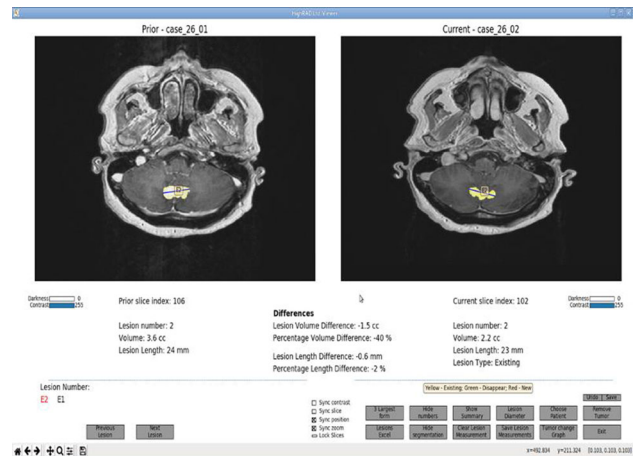


Fig. 1 Screenshot of the viewer. Registered axial slices of the baseline (left) and follow-up (right) scans. The brain metastasis is segmented (yellow) with the linear measurement superimposed on it (blue line). The metastasis volume information is displayed below. Information regarding change over time is displayed in the middle. In this example, the lesion length (diameter) changed by -2% while volume decreased by -40% , indicating partial response to SRS

summary window that details the brain metastases number, new, existing and disappeared metastases, and the total tumor burden and tumor burden difference between the two timepoints of the scans.

Two clinical datasets of patients from the Hadassah University Medical Center, Jerusalem, Israel between 2016 and 2020 were retrospectively collected. The first consists of 77 pairs of pre-SRS and post SRS scans with a mean time of 14 months between them. A total of 663 brain tumors were manually delineated: 202 baseline brain metastases, 123 existing follow up brain metastases and 211 new follow-up brain metastases not present in the pre-SRS scan. The second set consists of 81 scans with a total of 127 brain metastases. The brain metastases volumes are 0.06–22.5 cc. The ground truth brain metastases delineations in both datasets were obtained from the SRS pre-operative planning file generated with the BrainLab software by the three neurosurgeon co-authors. For the study, fivefold cross validation was performed with 639/8/17 brain metastases training/validation/testing split. We also quantified the manual delineation observer variability between two neurosurgeon authors on 52 brain metastases.

Results

Our method achieves a mean Dice score of 0.81 (std = 0.12) and Average Symmetric Surface Distance (ASSD) of 1.15 (std = 0.54) mm on the fivefold cross validation study, surpassing the observer variability of 0.77 (std = 0.14). For large (small) brain metastases, the Dice score is 0.84 (0.77), with a significant improvement of 0.18 in the Dice score when transfer learning was used. It is important to note that these results were achieved with relatively few annotated datasets.

Conclusion

We have presented a novel method for the quantitative evaluation of brain metastases after SRS on follow-up brain T1-Gd MRI scans. The novelties of our method are that incorporates expert prior knowledge from the pre-treatment brain metastases delineation, that it uses relatively few brain metastases manual segmentations to train the classification deep network, and that it provides a scans viewer that enables the simultaneous visualization of the metastases segmentations and their analysis for longitudinal volumetric scan studies.

References

- [1] Ben Shimol E, Joskowicz L, Eliahou R, Shoshan Y, 2018. Computer-based radiological longitudinal evaluation of meningiomas following stereotactic radiosurgery. *Int. Journal Computer Assisted Radiology and Surgery* 13(2):215–228.
- [2] Weizman L, Joskowicz L, Ben-Sira L, Shofty B, Constantini S, Ben-Bashat D, 2011. Longitudinal assessment of brain tumors using a repeatable prior-based segmentation. *Proc. IEEE Int. Symposium on Biomedical Imaging*, pp. 1733–1736.

Body fat compartments segmentation on abdominal T1w MRI using CNN

I. Vernikouskaya¹, H. P. Müller², D. Felbel¹, J. Kassubek², V. Rasche¹

¹Ulm University Medical Center, Dept. of Internal Medicine II, Ulm, Germany ²University of Ulm, Dept. of Neurology, Ulm, Germany

Keywords segmentation, body fat, convolutional neural network, ALS

Purpose

The total amount of abdominal adipose tissue is commonly separated into the two main components of visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT), with the former being more closely associated with health risks and linked to a variety of medical conditions such as metabolic, neurodegenerative and cardiovascular diseases. It has been shown that patients with neurodegenerative motor neuron disorder amyotrophic lateral sclerosis (ALS) display increased visceral fat and an increased ratio of visceral to subcutaneous fat [1]. Accurate quantification of visceral fat volume by MRI is a challenging task, often demanding manual segmentation, which results in non-reproducible measurements due to the high variation between manual delineations. Further, MRI related image distortions demand highly trained experts for proper segmentation. For automatization of the semantic segmentation in the computer vision community machine-learning strategies such as convolutional neural networks (CNN) and its derived architectures have been introduced and proven to outperform conventional automated and semi-automated techniques as clustering, thresholding, histogram-based region growing etc. Based on the concept of fully convolutional networks, we investigate the application of CNN of U-Net like architecture to automated segmentation of SAT and VAT from T1-weighted (T1w) whole body MRI scans and evaluate its robustness and accuracy in a cohort of ALS patients and healthy controls, in which fat characterization references have been derived and manually checked prior.

Methods

155 T1w spin-echo MRI scans were obtained from 74 patients with ALS at a disease duration of 23 ± 15 months and 81 healthy controls without any neurological/psychiatric disease or other medical condition. The whole body MRI volumes were merged from 6 to 8 subsequent acquisitions, each consisting of 36 2D slices using in-house developed software package ATLAS (Automatic Tissue Labelling Analysis Software) [2]. Resulting images of 384×384 pixels were supersampled to isotropic voxels with size 1.2 mm^3 . The arms of the subjects were manually removed from all scans prior to data pre-processing. For each single image slice a corresponding reference mask segmented using ARTIS algorithm (Adapted Rendering for Tissue Intensity Segmentation) within ATLAS software was created. All available data from both groups were split into

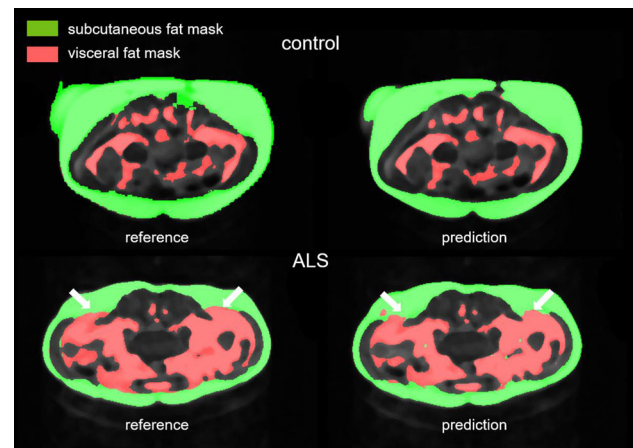


Fig. 1 Comparison of prediction results vs. reference segmentation from test dataset of control and ALS patient image slices

training (70%), validation (6%), and test (24%) data, based on age and BMI strata to ensure a balanced population distribution.

The problem was addressed with a CNN of encoder-decoder architecture, consisting of a convolutional part down-sampled with the maxpool layer and strided transposed convolutional up-sampling part in combination with drop out regularization. The network was implemented in Keras and trained based on the given reference segmentations on a GeForce GTX 1060 6 GB graphics card for 10 epochs with batch size of 16 using the Adam optimizer and categorical cross-entropy as loss function for multi-class semantic segmentation.

Results

Figure 1 shows the comparison of the segmentation performed by U-Net with the reference segmentation for a control and ALS patient of the test dataset. The dice similarity coefficients between the predicted masks and reference segmentations averaged among both groups achieved 0.91 ± 0.03 for SAT and 0.76 ± 0.12 for VAT. The most commonly observed errors in segmentation with the U-Net were erroneous interpretation of the hip bones as VAT and errors in discrimination between VAT and SAT in ambiguous areas (Fig. 1, white arrows); moreover, differences appeared at the edges of structures suggesting that in comparison to the reference technique the neural network tend to a smoothed segmentation without pixelated edges. Fig. 1 also clearly visualizes the differences in segmented SAT and VAT volumes between a control and ALS patient.

A significant linear correlation between the SAT and VAT volumes calculated from the reference segmentation and predicted segmentation was obtained (Fig. 2). Respective Pearson coefficients were $r = 0.953$ and $r = 0.934$.

Conclusion

Convolutional neural network of U-Net architecture has been applied to medical whole body MRI images for differentiation and quantification of body fat compartments into SAT and VAT. The network was robust enough to generate accurate SAT and VAT segmentations from both control and ALS patient groups. There was a clear correlation between the prediction and reference in volume quantification, suggesting that CNN-based quantification might serve as a potential biological marker to monitor specific body composition changes.

Acknowledgements

This study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG Grant Nu.

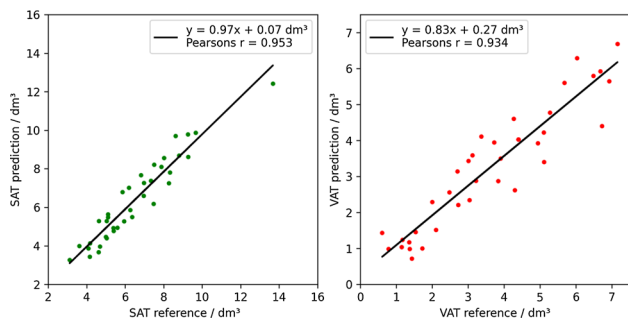


Fig. 2 Correlation between volumes of subcutaneous (SAT) and visceral adipose tissue (VAT) calculated based on the segmentation predicted by U-Net vs. reference in test dataset

References

[1] Lindauer E, Dupuis L, Müller HP, Neumann H, Ludolph AC, Kassubek J Adipose (2013) Tissue Distribution Predicts Survival in Amyotrophic Lateral Sclerosis. *PLoS One* 8(6): e67783.
 [2] Müller HP, Raudies F, Unrath A, Neumann H, Ludolph AC, Kassubek J (2011) Quantification of human body fat tissue percentage by MRI. *NMR Biomed* 24(1):17–24.

A novel curriculum based paradigm for self-training deep learning networks for brain glioma segmentation

A. Liew¹, C. C. Lee², B. L. Lan¹, M. Tan¹

¹Monash University Malaysia, School of Engineering, Subang Jaya, Selangor, Malaysia ²Sunway Medical Centre, Radiology Department, Selangor, Malaysia

Keywords Brain Tumor Segmentation, Self-Training, Curriculum Based Learning, Noisy Student

Purpose

The success of fully supervised deep learning methods for computer vision tasks hinges on the availability of large datasets. These datasets can be easily annotated through crowdsourcing efforts. In contrast, annotating biomedical images require substantial knowledge of subject matter experts.

To alleviate the shortage of labeled data, self-training uses unlabeled data as pseudo labels with great success. Furthermore, the Noisy Student (NS) method found that noise incorporated with unlabeled data improves the accuracy of object classification [1]. This suggests that unlabeled data can boost the availability of data in biomedical image tasks. This study introduces a teacher student training paradigm through a curriculum which incrementally increases the difficulty of samples being trained by the model, known as Noisy Student Curriculum Learning (NSCL).

The biomedical dataset selected for this study consists of 285 patients containing primary brain gliomas obtained from the BraTS platform [2]. Gliomas have poor prognosis, with average survival of 12–18 months. Our motivation for evaluating the segmentation over a classification task is that the outcome of a patient largely depends on the accuracy of segmentation of glioma and healthy tissue for treatment planning.

To our best knowledge, this study is the first of its kind to assess feasibility of the self-training method specifically in glioma images, and the adaptability of this method to glioma segmentation.

Methods

Our new NSCL method is evaluated using data from 285 fully-annotated volumetric brain scans and 209 unlabeled scans. Segmentations are quantitatively assessed using Dice Scores and Hausdorff distance computed on the BraTS platform (<https://ipp.cbica.upenn.edu/>).

A teacher model is trained for 300 epochs on the full set of labelled brain scans and used to synthesize pseudo-labels on the remaining 209 unlabeled scans. Pseudo labels and labeled data are used together to train the student model for 3 iterations. In the NS method, every iteration is trained with noise applied onto the data and model. Our NSCL method implements a curriculum approach to applying noise into training the student model with ascending levels of difficulty.

To compare the efficacy of the proposed NSCL, we compare the Dice Scores and Hausdorff distance of the Teacher Model and NS method. The Teacher Model was trained on only the labeled data. The results for a 3-iteration (NSCL3) and 4-iteration (NSCL4) scheme are reported.

Results

In Table 1, we compare the teacher model, which is trained only on labeled data with the original approach i.e., NS and our proposed approach, NSCL. Interestingly, the training for all 3 methods was performed with ground truth from BraTS2018, and then used to predict data on BraTS2020. Comparing NSCL3 with the winning model of BraTS2020, the method can perform sufficiently well, registering higher segmentation accuracies for TC regions (+ 0.56%) and lower accuracy of − 0.01%, − 0.08%, respectively for ET and WT. This suggests that the curriculum learning self-training method on unlabelled data was sufficient to produce good accuracy without any additional training. On NSCL4, we have a much better TC score by incorporating Manifold MixUp. This result is clinically significant to improving prognosis of brain gliomas as stereotactic radiosurgery is applied onto the TC region.

Through experimentation, we have used a curriculum which consists of model noise in the form of stochastic depth and dropout, and data noise in the form of MixUp and Copy Paste in the following sequence:

1. MixUp
2. Copy-Paste + Dropouts
3. MixUp + Copy-Paste + Dropouts
4. Manifold MixUp + Copy-Paste + Dropouts

Also, we found that model refinement did not improve the accuracy of the segmentations after NSCL. In NS, we found that by

Table 1 Dice score and Hausdorff distance for tumor regions on the BraTS2020 validation dataset. Three tumor regions are evaluated: Enhancing Tumor (ET), Whole Tumor (WT) and Tumor Core (TC)

Model	Dice Scores (%)			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Teacher Model	77.37	89.26	81.56	27.13	7.22	9.45
Noisy Student Training (NS)	77.29	87.84	83.55	29.80	14.37	8.16
Noisy Student Curriculum Learning 3 Iterations (NSCL3)	78.49	90.62	84.26	23.62	6.06	5.20
Noisy Student Curriculum Learning 4 Iterations (NSCL4)	78.16	90.24	85.20	24.43	5.82	4.85

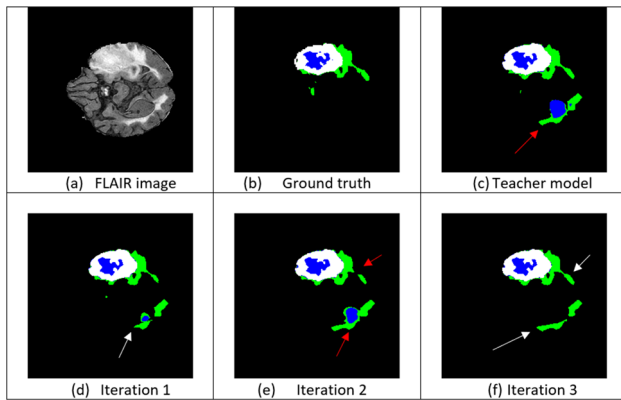


Fig. 1 The series of images above depict the predictions for Brats18_TCIA02_491_1 (Slice 60) for FLAIR image in (a) with its corresponding ground truth in (b). Image (c) is the teacher model prediction, while (d), (e) and (f) represent the prediction of each iteration of NSCL3. White arrows denote regions of improved segmentation in each scheme, red arrows denote regions which did not perform well

injecting all forms of noise at the onset of training caused deterioration in two of the three segmentation regions, while improving the tumor core region. By the third iteration of NS, we found that the segmentation continued to degrade, and the training results were discarded.

Figure 1 shows the progress of the different iterations in NSCL3. In the teacher model in (c), the red arrow indicates substantial error in segmentation. Tracking this area in the first iteration (d), the white arrow indicates a reduction in error. As we progress to (e), the progressive noise caused some instability, resulting in missing segmentations and enlarged erroneous area shown by the two arrows. However, the overall validation scores have improved from each iteration. In (f), segmentations appear to correct itself—with the recovery of the missing tail region and a much smaller erroneous area.

Conclusion

This study shows that the application of noise in a systematic manner can outperform the NS method which applies as much noise as possible during self-training. We have shown that the multiclass segmentation task is able to generalize to unseen data by applying the NSCL method to unlabeled data in a small dataset such as the BraTS dataset. Further work is needed to expand the curriculum for the NSCL to prevent catastrophic forgetting in later iterations.

References

- [1] Xie Q, Luong M-T, Hovy E, Le QV (2020) Self-Training With Noisy Student Improves ImageNet Classification, Nov. 2020, pp. 10684–10695, <https://doi.org/10.1109/cvpr42600.2020.01070>.
- [2] Bakas S, et al. (2019) Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, Nov. 2018, Accessed: Nov. 08, 2019. [Online]. Available: <http://arxiv.org/abs/1811.02629>.

Extraction of mediastinal great vessels from non-contrast CT images using 3D U-Net and its application to CTEPH

H. Suzuki¹, Y. Kawata¹, M. Matsuhiro¹, T. Sugiura², N. Tanabe², M. Kusumoto³, M. Kaneko⁴, N. Niki⁵

¹Tokushima University, Graduate School of Technology, Industrial and Social Sciences, Tokushima, Japan ²Chiba University, Department of Respiratory Medicine, Chiba, Japan ³National Cancer Center Hospital, Diagnostic Radiology, Tokyo, Japan ⁴Tokyo Health Service Association, Tokyo, Japan ⁵Medical Science Institute Inc., Tokushima, Japan

Keywords Computed tomography, Computer aided detection, U-Net, Artificial Intelligence

Purpose

Quantitative evaluation of mediastinal great vessels is a useful information for detection of pulmonary hypertension. A major problem for segmentation of mediastinal great vessels from non-contrast CT images is contact with blood vessels. In this study, we apply 2D and 3D U-Net [1] to segmentation of the aorta, vena cava (VC), main pulmonary artery (MPA), and main pulmonary vein (MPV) of normal and chronic thromboembolic pulmonary hypertension (CTEPH). Those segmentation performances are compared with a conventional method [2]. In addition, the robustness to the contacts between blood vessels is evaluated.

Methods

This study used two datasets of non-contrast chest CT images, normal cases (dataset A) and CTEPH cases (dataset B). The CT images of dataset A were acquired on Aquilion Lightning scanner with 30 mA at 120 kVp, plane resolution: 0.625 mm, reconstruction matrix: 512 × 512, convolution kernel: FC01, slice thickness: 1.0 mm, and reconstruction interval: 1.0 mm. The CT images of dataset B were acquired on Aquilion One scanner with 112–295 mA at 120 kVp, plane resolution: 0.570–0.698 mm, reconstruction matrix: 512 × 512, convolution kernel: FC07, slice thickness: 0.5 mm, and reconstruction interval: 0.5 mm. The number of cases in dataset A and B were 100 and 24. Manual annotations of ground truth of the mediastinal great vessels were performed in the axial plane. The vertical interval of this annotation was 10 slices. The window level and width for the annotation were adjusted to 50 HU and 300 HU. The boundary voxels of the ground truth were manually classified into five classes, (a) Non-contact, (b) contact of aorta with MPA, (c) contact of aorta with VC, (d) contact of MPA with VC, and (e) contact of aorta with esophagus.

Our approach consists of three steps. First, a branch point of trachea is detected. Then, a region of interest (256 × 256 × 256) centered on the branch point is cut out from original CT images. Finally, the mediastinal great vessels in the ROI are segmented by 2D and 3D U-Net. Training and testing were achieved by tenfold cross validation. The network was trained using the Adam optimization algorithm, with Dice loss function, using a single graphical processing unit (NVIDIA GeForce RTX 3090). The network was implemented using TensorFlow.

We defined contact surface ratio (CSR) to evaluate the robustness to contacts with blood vessels. CSR for each axial slice was

calculated as a ratio of contact boundary pixels per whole boundary pixels. Relation of the CSR and segmentation performance of U-Net, 3D U-Net, and conventional method is evaluated by Dice similarity coefficient (DSC).

Results

Extraction performance of ascending aorta of dataset A (and dataset B) using 2D and 3D U-Net, and the conventional method were 0.977 (0.972), 0.977 (0.971), and 0.959 (0.960), respectively. Extraction performance of MPA of dataset A (and dataset B) using 2D and 3D U-Net, and the conventional method were 0.965 (0.967), 0.965 (0.964), and 0.934 (0.953), respectively. The DSCs of 2D and 3D U-Net were significantly higher than the conventional method (p values < 0.05) for both of normal and CTEPH cases. We divided the ascending aorta of dataset A into two groups: those with less than 30% CSR and those with more than 30% CSR. For dataset A, the extraction performance of ascending aorta with more than 30% CSR was slightly decreased in all methods (p values < 0.05). For the dataset B, the extraction performance of ascending aorta using 3D U-Net was not decreased by CSR.

Conclusion

Extraction of mediastinal great vessels using U-Net and 3D U-Net achieved higher performance than the conventional method regardless the contacts with blood vessels, whereas the extraction performance with more than 30% CSR was slightly decreased.

References

- [1] Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention 234–241.
- [2] Suzuki H, Kawata Y, Niki N, Sugiura T, Tanabe N, Kusumoto M, Eguchi K, Kaneko M (2018) Automated assessment of aortic and main pulmonary arterial diameters using model-based blood vessel segmentation for predicting chronic thromboembolic pulmonary hypertension in low-dose CT lung screening. Proc SPIE 10575:105750X.

Two-stage convolutional neural network-based approach for improved segmentation of basal ganglia on MR images

T. Sugino¹, T. Kin², N. Saito², Y. Nakajima¹

¹Institute of Biomaterials and Bioengineering, Tokyo Medical and Dental University, Department of Biomedical Information, Tokyo, Japan ²Graduate School of Medicine, The University of Tokyo, Department of Neurosurgery, Tokyo, Japan

Keywords Segmentation, Convolutional neural networks, Object detection, Basal ganglia

Purpose

This paper presents a multi-stage convolutional neural network (CNN) framework to segment the basal ganglia from magnetic resonance (MR) images more accurately. In stereotactic neurosurgery, where electrodes are implanted in the basal ganglia for electrical stimulation, accurate segmentation of the basal ganglia on MR images is important for perioperative planning and simulation to accurately

implant electrodes into the basal ganglia located deep in the brain. Thus, CNNs have recently been applied to automated segmentation of the basal ganglia. However, because the basal ganglia are relatively small in the brain and consist of different sizes of nuclei, the distributions between background and foreground classes or among different foreground classes are imbalanced. The class imbalance problem makes it difficult for CNNs to accurately segment the basal ganglia from MR images. Although one way to solve the problem of class imbalance is to use a weighted loss function, that is not sufficient to deal with the extreme imbalance between background and foreground classes. Therefore, we aim to reduce the negative influence of class imbalance and improve the performance of CNNs on segmentation of the basal ganglia by introducing a framework that consists of CNNs for object detection and segmentation.

Methods

We designed a two-stage CNN-based approach for segmentation of the basal ganglia, including caudate nucleus, globus pallidus, putamen, and thalamus, from T1-weighted MR images. In this study, we adopted two-dimensional (2D) CNNs for our framework to support MR images with various slice thicknesses. The first and second stages of our framework consisted of 2D CNNs for coarse region detection and fine segmentation of the basal ganglia, respectively. In the first stage, we used YOLOv4-CSP [1], which is one of state-of-the-art object detectors, to detect bounding boxes of rough areas including the basal ganglia and surrounding margins on each image slice. Subsequently, by cropping the image slices based on the detected regions, we excluded image regions that are far from the basal ganglia, so that the second stage CNN for segmentation can focus on image features that are more relevant to the basal ganglia. In the second stage, we used U-net [2], which can be widely applied to medical image segmentation tasks, to segment detailed regions of basal ganglia from the cropped images detected in the first stage. Moreover, because there is still an imbalance between foreground classes (i.e., caudate nucleus, globus pallidus, putamen, and thalamus) although the background-foreground class imbalance can be alleviated by cropping the images with the first-stage CNN, we adopted a weighted loss function, which is known as Focal loss, to alleviate the foreground-foreground class imbalance. With the above procedure, we obtained fine segmentation results of the basal ganglia from MR images.

Results

To validate segmentation performance of the proposed method, we performed an experiment for segmentation of the basal ganglia on T1-weighted MR head images of 79 cases, which were imaged for medical examination in the Department of Neurosurgery of the University of Tokyo Hospital (The use of medical images for this study was approved by the ethics committees of the University of Tokyo and Tokyo Medical and Dental University). The size and resolution of MR images were $256 \times 256 \times 104\text{--}228$ [voxels] and $0.94\text{--}1.00 \times 0.94\text{--}1.00 \times 0.80\text{--}1.60$ [mm3/voxel], respectively. In this experiment, we compared segmentation results between U-net only and the proposed method (i.e., YOLOv4-CSP + U-net) to evaluate the effect of adding the procedure of coarse region detection before segmentation. The CNNs were trained and tested using four-fold cross-validation, in which we randomly divided 76 cases into training (57 cases) and test (19 cases) sets, and utilized the remaining 3 cases as a validation set. In each trial, YOLOv4-CSP pre-trained on ImageNet was retrained at 100 epochs, while U-net was trained from scratch at 50 epochs. For U-net, the Adam optimizer was applied with the following parameters: α (learning rate) = 10^{-3} (in U-net only) and 10^{-4} (in the proposed method), $\beta_1 = 0.9$, $\beta_2 = 0.999$, and epsilon = 10^{-7} .

Table 1 indicates the segmentation results with Dice similarity coefficients. The proposed method achieved better segmentation performance than U-net only in all the classes. Moreover, Fig. 1 shows the visualization results of basal ganglia segmented by U-net only and the proposed method. From the results, we found that the proposed method improved segmentation accuracies of smaller objects such as globus pallidus. These results suggest that the procedure of cropping with the first-stage CNN can be useful to alleviate the class imbalance problem and allow the second-stage CNN to capture the image features of not only large targets but also small targets.

Table 1 Segmentation results with Dice similarity coefficients (Ave. \pm Std.)

Classes	U-net	Proposed method (YOLOv4-CSP + U-net)
Caudate nucleus	0.721 \pm 0.099	0.792 \pm 0.038
Globus pallidus	0.487 \pm 0.201	0.632 \pm 0.105
Putamen	0.756 \pm 0.102	0.815 \pm 0.051
Thalamus	0.808 \pm 0.078	0.831 \pm 0.045

Conclusion

In this study, we proposed a two-stage CNN framework for improved segmentation of the basal ganglia on MR images. The experimental results showed that the proposed method could outperformed the conventional method in all the target classes. Especially in globus pallidus class, the proposed method improved segmentation accuracy by about 15% compared with the conventional method. It is

considered that segmentation performance of CNNs on smaller targets can be improved by adding the image cropping based on the region detection processing.

Acknowledgements

Parts of this work were funded by the Japan Agency for Medical Research and Development (AMED) (GRANT NUMBER JP21he1602001h0105) and the JSPS KAKENHI (Grant Number 20K20216)

References

- [1] Wang CY, Bochkovskiy A, Liao HYM (2021) Scaled-YOLOv4: Scaling cross stage partial network, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 13029–13038.
- [2] Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation.

Development of whole-leg musculoskeletal segmentation system by integrating multiple individually labeled CT image databases

Y. Masaki^{1,2}, Y. Otake², M. Soufi², K. Uemura³, M. Takao³, T. Miyamoto⁴, Y. Tanaka⁴, N. Sugano³, Y. Sato²

¹PSP Corporation, Minato-ku, Japan ²Nara Institute of Science and Technology, Division of Information Science, Ikoma, Japan ³Osaka University, Graduate School of Medicine, Suita, Japan ⁴Nara Medical University, Department of Orthopaedic Surgery, Kashihara, Japan

Keywords musculoskeletal segmentation, convolutional neural network, uncertainty, label fusion

Purpose

The aim of this study is to develop an automated musculoskeletal segmentation system from a whole-leg CT image by integrating existing labeled CT databases with various limited field-of-views (FOVs), which were created for the purpose of segmentation and analysis of various anatomical structures in multiple institutions. It is important to effectively utilize existing labeled databases for different anatomical structures in different institutions. A previous study has reported an automated musculoskeletal segmentation from whole-leg MR images [1], but as far as we know, automated whole-leg segmentation from CT images has not been reported. In addition, its training database consisted of whole-leg FOV images and it was specially constructed for whole-leg segmentation. That work [1] does not aim to integrate different training databases with various limited FOV. We focus on integration methods of the segmentation results of multiple pre-trained segmentors, each of which was developed using a method proposed in our previous work [2], from individual CT databases. In this study, we conducted experiments to evaluate the accuracy of automated whole-leg musculoskeletal segmentation, and compare two different methods for integrating multiple pre-trained segmentors.

Methods

To construct segmentors for different anatomical structures from CT data of different FOVs, we used 2D Bayesian U-Net [2], which estimates the segmentation label and prediction uncertainty. In this study, we develop automatic whole-leg musculoskeletal segmentation systems using three pre-trained models trained in hip musculoskeletal (Osaka University Hospital (OUH), 40 cases, 19 muscles, and three bones), lower leg bone (Nara Medical University, 35 cases, 28 bones),

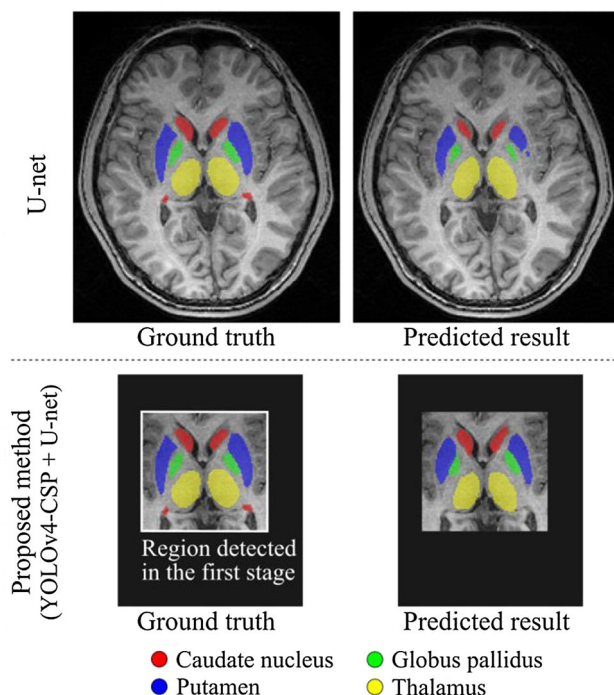


Fig. 1 Visualization results of segmented basal ganglia

and lower leg muscle (Nara Medical University, 10 cases, four muscles) databases. We used two integration methods: One is based on bone bounding box (BB) detection from a 2D projection of the CT image, that is, digitally reconstructed radiograph (DRR). The other is based on the uncertainty captured from Bayesian U-Net. Regarding the former method based on bone BB detection, we used YOLOv5 trained by transfer learning using the DRR of whole-leg CT images of 162 cases from OUH for the bone BB detection. The appropriate FOV in the whole leg CT was calculated for each segmentor by the detected bone BB, and the whole-leg segmentation label was integrated. Regarding the latter uncertainty-based integration, each segmentor generates entire prediction label and uncertainty map from the input whole-leg CT image, regardless of whether it is a training FOV or not. The uncertainty based method is expected to be more scalable because it selects the label with the smallest uncertainty value. In contrast, the YOLO-based method which requires a procedure design depending on available training databases. The automatic whole-leg musculoskeletal segmentation accuracy was evaluated using Dice and ASD metrics.

Results

We evaluated three cases of whole-leg CT from OUH. We created the ground truth labels by manually verifying and correcting the labels automatically generated by the Bayesian U-Net. The Dice scores for muscles and bones for the three methods evaluated in this study are shown in Table 1. Figure 1 shows the quantitative and qualitative results of each method. The Dice scores for the uncertainty-based integration and YOLO-based integration for musculoskeletal structures shown in Table 1 were 0.904 ± 0.059 and 0.940 ± 0.037 on average, respectively. The uncertainty-based integration has the advantage of not requiring additional training data for the object detection but had some segmentation errors where the uncertainty did not correctly indicate segmentation errors outside of the training FOV, resulting in the lowest segmentation accuracy. The YOLO-based integration required additional training data depending on the training FOV, but since the segmentation accuracy is relatively high, the integration of the segmentation labels was successful.

Table 1 Dice scores of all bones and muscles using two integration methods, (1) uncertainty-based integration and (2) YOLO-based integration

Anatomical structures	Uncertainty	YOLO
	Mean±Std	Mean±Std
Pelvis	0.972±0.012	0.973±0.012
Femur	0.985±0.006	0.989±0.005
Adductor muscle	0.950±0.006	0.960±0.002
Biceps femoris muscle	0.772±0.104	0.958±0.007
Gluteus maximus muscle	0.962±0.005	0.963±0.004
Gluteus medius muscle	0.936±0.006	0.936±0.005
Gluteus minimus muscle	0.828±0.011	0.830±0.010
Gracilis muscle	0.938±0.015	0.950±0.011
Iliacus muscle	0.904±0.003	0.905±0.004
Obturator externus muscle	0.889±0.041	0.890±0.042
Obturator internus muscle	0.919±0.009	0.919±0.009
Pectineus muscle	0.892±0.031	0.900±0.038
Piriformis muscle	0.877±0.027	0.876±0.027
Psoas major muscle	0.909±0.017	0.890±0.038
Rectus femoris muscle	0.913±0.015	0.956±0.009
Sartorius muscle	0.944±0.005	0.949±0.004
Semimembranosus muscle	0.822±0.051	0.944±0.017
Semitendinosus muscle	0.837±0.066	0.939±0.009
Tensor fasciae latae muscle	0.920±0.033	0.920±0.032
Vastus lateralis muscle and vastus intermedius muscle	0.951±0.015	0.973±0.003
Vastus medialis muscle	0.910±0.007	0.956±0.005
Sacrum	0.947±0.036	0.947±0.035
Tibia	0.976±0.014	0.987±0.002
Fibula	0.984±0.003	0.984±0.003
Talus	0.863±0.079	0.965±0.028
Calcaneal	0.938±0.031	0.980±0.013
Anterior compartment	0.789±0.034	0.944±0.015
Lateral compartment	0.915±0.021	0.926±0.012
Deep posterior compartment	0.882±0.004	0.914±0.015
Superficial posterior compartment	0.789±0.019	0.981±0.003

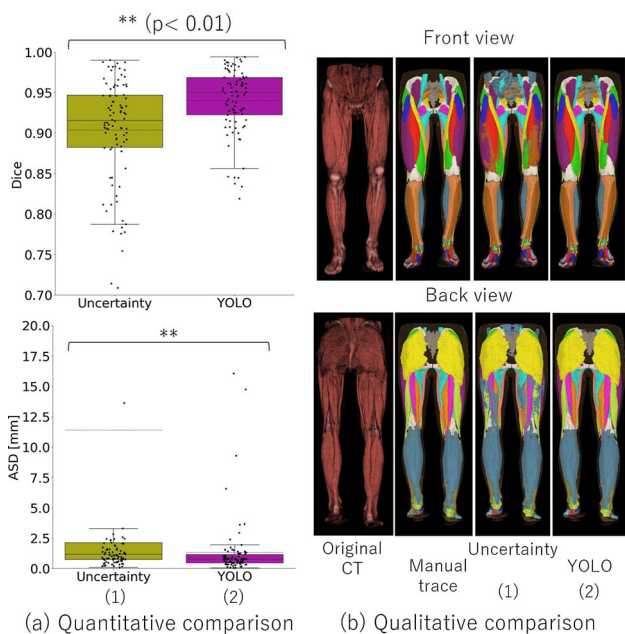


Fig. 1 Quantitative and qualitative comparison of the segmentation results using the two integration methods. The boxplots shows the dice and ASD scores of the muscles and bones shown in Table 1, which were used in this evaluation

Conclusion

We evaluated the accuracy of automatic musculoskeletal segmentation from whole-leg CT images by effectively integrating the multiple pre-trained segmentors from individual training databases in various FOV. The uncertainty-based method has the advantage of not requiring additional training data, but the uncertainty did not correctly indicate segmentation errors outside of the training FOV of each segmentor, resulting in degradation of the integration accuracy. The YOLO-based method was useful in automatic whole-leg musculoskeletal segmentation, although it required additional bone ROI annotations in the DRR of the CT image. In the next step, we will investigate out-of-distribution data detection methods for the uncertainty-based method to separate areas where the uncertainty did not correctly indicate segmentation errors. Furthermore, the YOLO-based

method will be expanded for automated whole-body musculoskeletal segmentation.

References

- [1] R. Ni, C. H. Mayer, S. S. Blemker, J. M. Hart, X. Feng (2019). “Automatic segmentation of all lower limb muscles from high-resolution magnetic resonance imaging using a cascaded three-dimensional deep convolutional neural network.” *Journal of Medical Imaging* 6(4): 1–8.
- [2] Y. Hiasa, Y. Otake, M. Takao, T. Ogawa, N. Sugano and Y. Sato (2019). “Automated Muscle Segmentation from Clinical CT using Bayesian U-Net for Personalized Musculoskeletal Modeling.” *IEEE Transactions on Medical Imaging* 39(4):1030–1040.

Evaluation of automated musculoskeletal segmentation for muscle volume quantification in upright and supine CT imaging

T. Hakotani^{1,2}, M. Hashimoto², Y. Otake^{1,2}, Y. Yamada², M. Yamada², Y. Yokoyama², M. Soufi¹, K. Uemura³, M. Takao³, N. Sugano³, Y. Sato¹, M. Jinzaki²

¹Nara Institute of Science and Technology, Division of Information Science, Ikoma, Japan ²Keio University, School of Medicine, Department of Radiology, Tokyo, Japan ³Osaka University, Graduate School of Medicine, Suita, Japan

Keywords upright CT, muscle segmentation, musculoskeletal deformation, muscle volume

Purpose

Musculoskeletal models that are faithful to medical images are important for analyzing the causes of movement disorders. Since the standing position is the fundamental posture in daily activities, it is desirable to create a musculoskeletal model for the analysis of standing movements based on medical images scanned in the standing position. Our final goal is to make a quantitative comparison of musculoskeletal deformation and muscle volume between supine position, in which muscles are relaxed, and standing position, in which muscles are tense. For this purpose, musculoskeletal segmentation from CT images is needed; however, manual segmentation is labor-intensive and requires expertise. Therefore, automatic segmentation of muscles from clinical CT images is necessary. In this study, we evaluated the accuracy of automatic musculoskeletal segmentation in supine and upright (standing) CT images scanned for the same subject and analyzed its effect on the measurement of volume change with respect to the body position. The upright CT scanner used in this study is a newly developed whole-body imaging scanner [1]. Especially, we addressed the question of how accurately the segmentation tool trained using supine CT images can segment upright CT images in which muscles are largely deformed from supine to upright position.

Methods

Eleven pairs of CT images in supine and upright positions, 22 sets of CT images in total, were used for this study. First, we automatically segmented these CT images using a previously proposed Bayesian U-Net [2]. The model was built for a different purpose, using a different supine CT dataset (added more cases to the dataset in [2]). Second, under the supervision of a medical expert, we manually revised the predicted labels and made ground truth labels of the pelvis, gluteus maximus (Gmax) muscle, gluteus medius (Gmed) muscle, and sacrum. While muscles required substantial manual

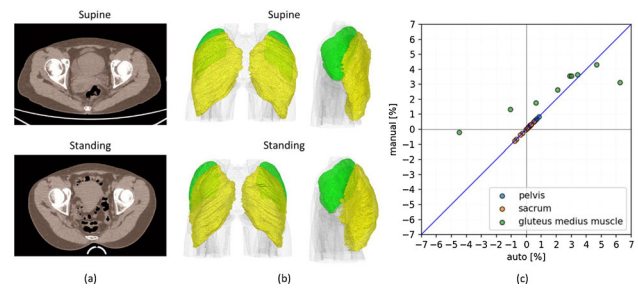


Fig. 1 Results of segmentation for supine and standing CT. (a) Axial CT slices of upright and supine positions of the same subject. (b) Visualization of auto segmentation results. The yellow and green show Gmax and Gmed muscles, respectively. (c) Scatter plots of the volume change ratio from supine to standing position for the pelvis, sacrum, and Gmed muscle. The horizontal axis is for automatic segmentation, and the vertical axis is for manual segmentation

revisions, the pelvis and sacrum bones did not require revisions since the automatic segmentation was highly accurate. Then, we evaluated the accuracy of the automatic segmentation of the Gmax and Gmed muscles. Finally, we evaluated the volume change ratio of each anatomical structure between the supine and standing positions using ground truth labels and auto segmentation labels. For the volume change ratio evaluation, we selected only those cases that covered each anatomical structure entirely in CT images of both the standing and supine positions, resulting in 8, 10, and 9 cases for the pelvis, sacrum, and Gmed, respectively. Regarding Gmax, only 3 cases satisfied the above condition, therefore Gmax was not considered for the volume change analysis due to too small data size.

Results

Figure 1(a) shows axial CT images of the same subject in upright and supine positions, in which large shape differences of muscles are observed while both slices pass through the femoral head center. Figure 1(b) shows the visualizations of the results of auto segmentation of Gmax and Gmed. The distinct shape difference between the supine and standing positions in the Gmax was observed. Table 1 shows the segmentation accuracy of Gmax and Gmed muscles. Compared to the results reported in [2], the segmentation accuracy in this study was higher in both the supine and standing positions. This might be due to the increase in the number of training data from 20 to 29 and the u-net depth from 5 to 6, compared to [2]. Figure 1(c) shows the scatter plots of the volume change ratio from supine to standing position for the pelvis, sacrum, and Gmed muscle. The volume change ratio for bones, that is the pelvis and sacrum, was less than 0.5% on average and was little changed, but for the Gmed muscle, the volume change ratio was around 2.6% and was changed largely compared to the bones. The difference in the volume change ratio for the Gmed muscle between auto and manual segmentations, substantially regarded as an estimation error, was around 0.7% on average, which indicates that automatic muscle volume measurement is potentially possible with an error of less than 1%.

Table 1 Results of evaluation of segmentation accuracy

Position	Gluteus maximus (Gmax) muscle		Gluteus medius (Gmed) muscle	
	Supine	Standing	Supine	Standing
Dice	0.990±0.007	0.984±0.011	0.983±0.007	0.971±0.015
ASD [mm]	0.132±0.102	0.263±0.186	0.169±0.088	0.336±0.206

Conclusion

Muscle segmentation in upright and supine CT images was described and auto segmentation was shown to be promising for accurate quantification of muscles. This work is the first study of 3D muscle shape reconstruction in the standing position. In the standing position, the shape of the muscles is significantly deformed compared to the supine position. Therefore, as shown in Table 1, the segmentation accuracy of the upright CT images was lower than that of the supine CT images with the segmentation tool trained with the supine CT images. Nevertheless, the Gmax and Gmed muscles were automatically segmented with the dice coefficient higher than 0.97 from supine and upright CT images. Although the number of evaluated cases was not large, these results suggest that the model trained with CT images acquired in supine position would be able to be applied for CT images acquired in standing position with reasonable accuracy and promising results were also obtained for the accuracy of the muscle volume measurement. In future work, we will evaluate the segmentation accuracy in more cases and evaluate the volume change and shape deformation from the supine to standing position in a few hundred cases.

References

- [1] Jinzaki M, Yamada Y, Nagura T, Nakahara T, Yokoyama Y, Narita K, Ogihara N, Yamada M (2020) Development of upright computed tomography with area detector for whole-body scans: phantom study, efficacy on workflow, effect of gravity on human body, and potential clinical impact. *Investigative radiology* 55(2):73.
- [2] Hiasa Y, Otake Y, Takao M, Ogawa T, Sugano N, Sato Y (2019) Automated muscle segmentation from clinical CT using Bayesian U-net for personalized musculoskeletal modeling. *IEEE transactions on medical imaging* 39(4):1030–1040.

Slices prioritization for segmentation correction in fetal body MRI scans based on previous corrections data

B. Spektor Fadida¹, D. Link Sourani², L. Ben Sira^{2,3},
D. Ben Bashat^{2,4}, L. Joskowicz¹

¹The Hebrew University of Jerusalem, School of Computer Science and Engineering, Jerusalem, Israel ²Tel Aviv Sourasky Medical Center, Sagol Brain Institute, Tel Aviv, Israel ³Tel Aviv Sourasky Medical Center, Division of Pediatric Radiology, Tel Aviv, Israel ⁴Tel Aviv University, Sackler Faculty of Medicine & Sagol School of Neuroscience, Tel Aviv, Israel

Keywords Deep learning, Annotations bootstrapping, Segmentation correction, Fetal MRI

Purpose

Fetal MRI has the potential to complement US imaging and improve fetal development assessment by providing accurate volumetric information about fetal structures. However, volumetric measurements require the manual delineation, also called segmentation, of the fetal structures, which is a time consuming, annotator-dependent and error-prone task. State-of-the-art automatic segmentation methods for volumetric scans are based on deep neural networks that require a large, high-quality dataset of expert-validated annotations, which is time consuming and very difficult to obtain.

Obtaining expert-level delineations by interactively correcting automatically generated segmentations is usually less laborious and time consuming than performing delineations “from scratch” when

the automatic segmentation is in most cases of acceptable quality and has relatively few errors. Therefore, a variety of deep-learning based segmentation algorithms are being developed in a bootstrapping process. In this process, a network is first developed using few annotated datasets; it is then used on new datasets to generate segmentations that will be manually corrected by an expert, thereby producing additional high quality annotated data for network re-training. This process produces valuable segmentation error correction data which is not currently used and may have potential value.

The goal of our research is to show how to use the segmentation correction data to optimize the radiologist time and effort by correcting first the most significant segmentation errors instead of in random or sequential order as is done in current practice.

Methods

We have developed a novel method to prioritize scan slices for manual examination and correction of a structure segmentation automatically produced by a network. The method is based on a 3D deep learning voxel classification network trained on previous segmentation corrections data. The network inputs the original scan and the segmentation mask and outputs possible segmentation error voxels. The slices selection is performed in three steps: (1) apply errors segmentation network on the input scan and the segmentation mask; (2) assign a segmentation error score for each slice by using the maximum network response in the slice; (3) correct all slices with a segmentation error above a threshold or select a pre-defined percentage of slices to correct based on the highest slices score.

The novel error segmentation network is based on the 3D U-Net described in [1] augmented with an additional input channel for the segmentation mask. We used a patch size of $256 \times 256 \times 48$ to capture a large Field of View without downscaling. The output of the network is the voxel-wise segmentation error prediction for the scan.

We evaluated our method on 101 MRI fetal body scans (gestational age GA of 28–39 weeks) acquired with the TRUFI sequence from Tel-Aviv Sourasky Medical Center of fetuses with. The scans were acquired on the Siemens Skyra 3 T, Prisma 3 T and Auera 1.5 T scanners. Each scan had 55–128 slices, 320–512 pixels/slice and resolution of $0.59\text{--}1.34 \times 0.59\text{--}1.34 \times 2.0\text{--}4.0 \text{ mm}^3$.

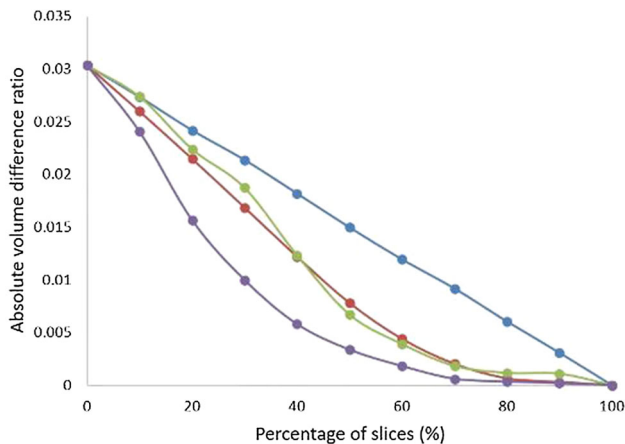
Ground truth annotations were created using a bootstrap approach in which the segmentation masks generated by a pre-trained 3D U-Net [1] were inspected and manually corrected by an expert radiologist as needed. The data pairs consisted of the original and the corrected segmentation masks. Since we used the segmentation masks of the annotation process, the difference between them and the corrected ground truth was only the segmentation error without observer variability.

Results

The slices selection method was evaluated on a 72/5/24 train/validation/test split. The Dice score ($\times 100$, %) and the absolute Volume Difference Ratio (VDR) metrics before and after the correction were used for evaluation. The Dice score was 97.6% and the VDR was 3% with the corrected ground truth on the test set. Table 1 lists the results of the comparison between slices correction using errors segmentation network and slices correction using a random approach with the same number of correction slices (random on all slices and random from non-empty slices). The mean Dice score was 99.3% and the minimum was 97.8% using network predicted slices vs. 98.5% mean and 95.2% minimum using a random approach from non-empty slices. The mean VDR was only 0.8% and the maximum VDR was as low as 3.4% for the corrections based on network prediction vs. a mean VDR of 1.4% and a maximum VDR of 6.3% with random selection from non-empty slices. Thus, correcting slices based on error segmentation network was able to reduce the maximum VDR error by almost half compared to the random approach.

Table 1 Dice score (%) and absolute Volume Difference Ratio (VDR) results before and after correction

		Before correction	Random correction	Random correction from non-empty slices	Network predicted correction
Dice	Mean	97.60%	98.50%	98.90%	99.30%
	Std	2.10%	1.30%	1%	0.70%
	Min	92.80%	94.80%	95.20%	97.80%
VDR	Mean	3.90%	2.00%	1.40%	0.80%
	Std	3.70%	2.30%	1.70%	1%
	Max	0.10%	0.10%	0.10%	0.00%
		13.20%	8.10%	6.30%	3.40%

**Fig. 1** Absolute volume difference of: (1) random prioritization (blue). (2) random prioritization of selection from non-empty slices of the segmentation (red); (3) sequential correction (green); (4) correction based on correction network response (purple)

We also evaluated the VDR after correcting different slices percentages. We compared four segmentation error correction approaches: (1) random slice selection; (2) random slice selection from non-empty slices; (3) sequential selection; (4) selection based on the error segmentation network. Figure 1 shows the results. For each one of the slices percentages, selecting slices based on the error segmentation network resulted in the smallest VDR error. About 30–60% more slices were required for both sequential and random approaches to reach the same VDR error as the network-based approach. For example, to reach an error of 1% VDR, 30% of the slices need to be corrected using network-based approach compared to more than 40% of the slices using either the sequential or the random based approaches.

Conclusion

To the best of our knowledge, this work is the first to utilize segmentation corrections data for segmentation error slices selection. We demonstrated our method on a dataset with minor segmentation errors and showed that selecting slices based on the segmentation errors network can further increase Dice score and reduce VDR error compared to random selection. We also showed that for different target VDR errors our network-based method requires less slices to correct compared to both random and sequential orders.

References

- [1] Dudovitch G, Link-Sourani D, Sira LB, Miller E, Bashat DB, Joskowicz L. Deep Learning Automatic Fetal Structures Segmentation in MRI Scans with Few Annotated Datasets. Proc. Int. Conf. Medical Image Computing and Computer-Assisted Interventions pp. 365–74, 2020.

Synthesis of digitally reconstructed radiograph from x-ray image for estimation of bone mineral density

Y. Gu¹, Y. Otake¹, K. Uemura², S. Mazen¹, M. Takao², N. Sugano², Y. Sato¹

¹Nara Institute of Science and Technology, Division of Information Science, Ikoma, Japan ²Osaka University, Graduate School of Medicine, Suita, Japan

Keywords BMD estimation, radiography, generative adversarial network, vision transformer

Purpose

Quantitative measurement of bone mineral density (BMD) is necessary for the diagnosis of osteoporosis and osteopenia. While dual-energy x-ray absorptiometry (DXA) is the preferred modality for BMD measurement that is recommended in several guidelines, DXA is not always applicable in clinics due to its cost and availability. Thus, a method to accurately measure BMD from conventional X-ray image acquired routinely in the clinic has been deemed necessary. A recent study [1] reported a method to predict the BMD from an x-ray image. However, as the method directly regressed DXA-measured BMD value from an x-ray image, a large data set (> 5000) was used for the training and required pre-processing of data to extract the appropriate region-of-interest from the x-ray image. In this study, we propose leveraging information in CT [2] of the patient in the training dataset to improve the accuracy of BMD estimation from an x-ray image using a limited number of data, which is a common and vital requirement in the medical research field.

Methods

Figure 1 (a) shows the overview of the proposed method. Instead of the direct regression of DXA-measured BMD from the x-ray image, this study first translates the input x-ray image to a digitally reconstructed radiograph (DRR) of the proximal femur region, where its average intensity is known to be highly correlated to the DXA-measured BMD. A model for translating an x-ray image to the proximal femur DRR was trained using the Pix2Pix framework, which uses a conditional discriminator to form the generative adversarial network (GAN), introducing adversarial loss. Unlike the original Pix2Pix that used ResNet as the backbone, this study achieves higher performance by adapting the fashion transformer-based semantic-segmentation model, SegFormer. Our model is pre-trained with the task of bone decomposition from an x-ray image. The final estimation result is obtained by calibrating the average intensity of the proximal femur area of DRR and DXA-measured BMD on training data using the linear least square fitting. We used two datasets in this study: 70-cases (dataset #1) and 130-cases (dataset #2). Details for the datasets are shown in Table 1, where the mean DXA-measured BMD values were almost equal, 0.711 and 0.744 for dataset #1 and dataset #2, respectively. We performed two experiments to compare the effect of increasing the number of training datasets on the prediction accuracy. The first experiment used only dataset #1 in a tenfold cross-validation setup (Experiment #1), i.e., 60 training datasets in each fold, and the second experiment used both dataset #1 and #2 in a fivefold cross-validation setup, i.e., 160 training datasets in each fold (Experiment #2). In experiment #2, we only evaluated the 70 cases in dataset #1 for a fair comparison with experiment #1. The mean absolute error (MAE) in the prediction of the BMD was quantified, and the Pearson Correlation Coefficient (PCC) along with the intraclass correlation coefficient (ICC) between the predicted BMD and the DXA-measured BMD were also quantified. Statistical significance was tested with the single-factor repeated measures ANOVA model.

Table 1 Case characteristics of datasets #1 and datasets #2

Dataset	#1	#2
Number, n	70	130
Female, n (%)	70(100)	96(73.85)
Mean DXA-measured BMD (sd), g/cm^3	0.711(0.162)	0.774(0.147)
Mean age (sd), years	62.29(11)	57.96(13.56)

Results

The experimental results are shown in Fig. 1(b). With SegFormer backbone, MAE showed a statistically significant decrease, whose averages were 0.069 and 0.060 for experiments #1 and #2, respectively, suggesting improved accuracy by increasing the number of training datasets. ICC and PCC also improved from 0.794 to 0.853 and 0.868 to 0.915, respectively. For ResNet, MAE remained the same for experiments #1 and #2. ICC and PCC showed slight improvement, from 0.814 to 0.820, and 0.863 to 0.891, respectively. The average training time was approximately 30 h on an Intel Xeon processor (2.30 GHz, 16 cores) with Nvidia GeForce RTX 3090 (10,496 CUDA cores, 24 GB memory).

Conclusion

This study proposed a method for estimating DXA-measured BMD from a single x-ray image using a DRR of the proximal femur. We observed that the SegFormer in experiment #2 showed the highest PCC, 0.915, with 160 training datasets. The result is close to PCC reported in [1], 0.917, obtained by the training with > 5000 cases, suggesting our proposed framework that efficiently leverages the CT-derived DRR helped improve the accuracy of the BMD prediction from a single x-ray image. Furthermore, our results showed consistently high PCC (higher than 0.86) on two different training datasets,

even with the small number of datasets (70 cases), suggesting the stability of the proposed method with both backbones, ResNet and SegFormer. However, it is worth to be noted that the SegFormer benefits more from a larger training dataset, suggesting a stronger learning ability. Our future work includes improving the accuracy of predicting the DXA-measured BMD and applying the method developed herein to a larger number of cohorts with a larger variation in x-ray acquisition protocol.

References

[1] Hsieh, C.I., Zheng, K., Lin, C. et al. Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. *Nat Commun* 12, 5472 (2021).
 [2] Uemura K, Otake Y, Takao M, Makino H, Soufi M, Iwasa M, Sugano N, Sato Y. Development of an open-source measurement system to assess areal bone mineral density of the proximal femur from clinical CT images. *Archives of Osteoporosis* in press.

Fully convolutional neural network with weighted frequency domain loss for reducing streak artifact in sparse-view CT

T. Okamoto¹, H. Haneishi²

¹Chiba University, Graduate School of Science and Engineering, Chiba, Japan ²Chiba University, Center for Frontier Medical Engineering, Chiba, Japan

Keywords Artifact reduction, Frequency domain, Sinogram, Sparse-view CT

Purpose

Sparse-view computed tomography (CT) is an imaging technique that reduces projections, resulting in lower radiation and faster acquisition time. However, since image reconstruction with insufficient projections is an ill-posed inverse problem, severe streak artifacts are produced on the images reconstructed with the analytical algorithms such as filtered back-projection (FBP). Our goal is to estimate fully sampled sinogram from sparse-view sinogram and reduce artifacts in CT images. In recent years, several researchers have developed artifact reduction methods for CT images or sinograms using a deep learning technique. This paper proposes a deep learning-based new correction method for sinograms using frequency domain information in loss function to reduce artifacts in sparse-view CT.

Methods

We proposed a deep learning model to estimate fully sampled sinogram by correcting for sparse-view sinogram. Since the sinogram is a stacked projections obtained from each angle, the sinogram enlarged vertically as the number of projections increases. In this study, sparse-view sinogram X first enlarges vertically by linear interpolation as preprocessing. With the linearly interpolated sinogram as input, the model F parameterized by θ estimates the corrected sinogram \hat{X} as follows:

$$\hat{X} = F(\text{lerp}(X); \theta). \tag{1}$$

$\text{lerp}(\cdot)$ means the linear interpolation. Training the model F requires optimization of the parameter θ , which is achieved by minimizing the loss function between corrected sinogram \hat{X} and fully sampled sinogram Y corresponding to the ground truth. We implemented a

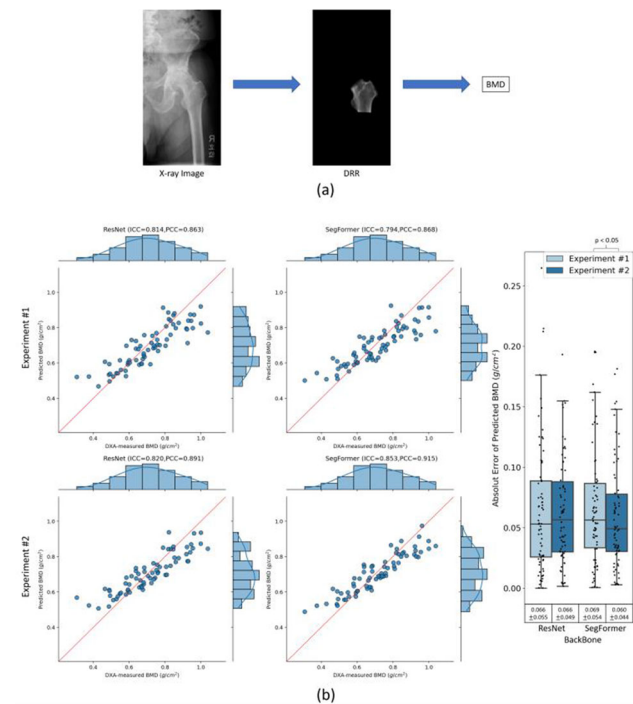


Fig. 1 (a) Overview of proposed method (b) BMD estimation results by the proposed method with ResNet and SegFormer on experiment #1 using dataset #1 and experiment #2 using additional training dataset #2

simple fully convolutional network (FCN) as model F consists of the sequential residual blocks and global skip connection.

The proposed method utilized a frequency domain loss function that calculates the absolute differences of the frequency components between the corrected sinogram and fully sampled sinogram. Attempts to introduce frequency domain loss have been reported in several studies, e.g., Xue et al. [1] proposed a super-resolution model with a frequency feature map of low-resolution images as input. The proposed loss function consists of two terms; the first term calculates the absolute differences in the spatial domain and the second term calculates the absolute differences in the frequency domain. The overall loss function is defined as

$$L(\mathbf{Y}, \hat{\mathbf{X}}) = |\mathbf{Y} - \hat{\mathbf{X}}|_1 + \lambda \cdot |\mathbf{W} \circ (f_{DFT}(\mathbf{Y}) - f_{DFT}(\hat{\mathbf{X}}))|_1, \quad (2)$$

where f_{DFT} means the two-dimensional discrete Fourier transform function, λ means hyper-parameter, and $\mathbf{A} \circ \mathbf{B}$ means the Hadamard product of \mathbf{A} and \mathbf{B} . \mathbf{W} represents the weight matrix for the absolute differences in the frequency domain. As shown in Fig. 1 (b), the frequency pattern of the sinogram exhibits a double-wedge region [2]. Therefore, we designed the weight matrix \mathbf{W} to focus on the high-frequency regions located in the top and bottom ends of the frequency domain. The proposed weight matrix represents the inverted two-dimensional Gaussian function whose value varies vertically (Fig. 1(c)). The element of (i, j) th in $\mathbf{W} \in \mathbb{R}^{M \times N}$ is given as

$$W_j(i) = -a \cdot \exp\left[-(M/2 - i)^2 / 2\sigma^2\right] + a + 1, \quad (3)$$

where a is the height of the curve's bottom peak, and σ is the standard deviation.

Comparison experiments using the publicly available chest CT datasets were conducted to evaluate the effectiveness of the weighted frequency domain loss function. We extracted 12 patients' data from datasets and divided them into eight patients (2696 axial images) for training, two patients (715 axial images) for validation, and two patients (708 axial images) for testing. The sparse-view sinogram with 256 projections and fully sampled sinogram with 1024 projections were generated by applying the 2D fan-beam Radon transform to the axial CT image. We compared the performance among the linear interpolation, the U-net as a benchmark of deep learning approach, FCN without frequency domain loss (FCN, $\lambda = 0$ in Eq. (2)), FCN with frequency domain loss (FCN-FD, $\lambda = 1$), and FCN with weighted frequency domain loss (FCN-WFD, $\lambda = 2$). All models were trained using the Adam optimizer. The initial learning rate was 1.0×10^{-3} for all layers and was decayed by multiplying the value by 0.8 for every 50 epochs. The training was stopped after 200 epochs. The batch size was set to 32. The parameters of the weight matrix were set to $a = 3.5$ and $\sigma = 100$, respectively.

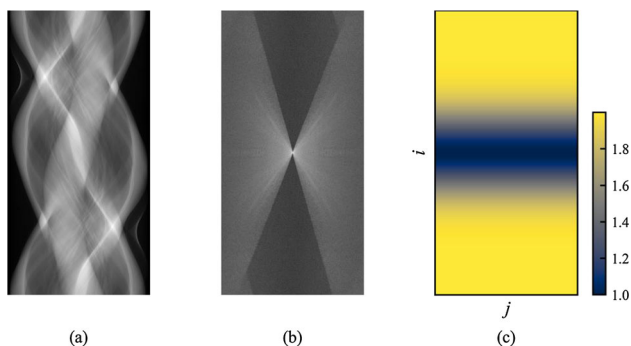


Fig. 1 Frequency pattern of the sinogram. (a) original sinogram, (b) power spectrum of (a), and (c) weight matrix \mathbf{W}

Table 1 Average quantitative results of the sinograms and reconstructed images by FBP

Method	Sinogram	Reconstructed image (CT)
Linear interpolation	0.0058 / 44.89 / 0.9968	0.0267 / 31.70 / 0.9453
U-net	0.0040 / 48.02 / 0.9973	0.0105 / 39.85 / 0.9759
FCN	0.0020 / 54.27 / 0.9992	0.0096 / 40.56 / 0.9802
FCN-FD	0.0020 / 54.51 / 0.9994	0.0096 / 40.61 / 0.9804
FCN-WFD (proposed)	0.0019 / 54.93 / 0.9994	0.0094 / 40.79 / 0.9821

Results

Table 1 summarizes the average quantitative results of the sinograms and reconstructed images by FBP. The proposed method (FCN-WFD) achieved the best performance in terms of root-mean-square error (RMSE), peak signal-to-ratio (PSNR), and structural similarity index (SSIM). In addition, FCN-FD showed the second-best scores. Therefore, we confirmed the effectiveness of the frequency domain loss, and the weight matrix contributes to the improvement of estimation accuracy. Note that we also confirmed that the effectiveness of the proposed method qualitatively.

Conclusion

We proposed a deep learning model to correct sinograms that introduces frequency domain information in loss function to reduce artifacts in sparse-view CT. We introduced the weight matrix is applied to the frequency domain loss function to focus on the high-frequency regions. Experimental results show that the effectiveness of the weighted frequency domain loss function quantitatively and qualitatively.

References

- Xue S., Qiu W., Liu F., Jin X. (2020) Faster image super-resolution by improved frequency-domain neural networks. *Signal, Image and Video Processing* 14:257–265.
- Mazin S. R. Pelc N. J. (2010) Fourier properties of the fan-beam sinogram. *Medical Physics* 37(4):1674–1680.

Blind super resolution of lung CT scans using Wiener deconvolution

S. Bhardwaj¹, K. Inai², H. Kimura², S. Kido³, A. Shimizu¹

¹Tokyo University of Agriculture and Technology, Graduate School of Engineering, Department of Electrical and Electronic Engineering, Koganei, Japan ²University of Fukui, Faculty of Medical Sciences, Fukui, Japan ³Osaka University, Graduate School of Medicine, Department of Artificial Intelligence Diagnostic Radiology, Suita, Osaka, Japan

Keywords Lung CT scan, Deep learning, Super resolution, Wiener deconvolution

Purpose

Computed tomography (CT) is essential for the diagnosis of various diseases, such as pneumonia or lung carcinoma. A higher resolution is required for a more accurate diagnosis. However, the resolution is limited by the geometrical constraints of the CT scanner, such as the X-ray focal spot size or detector element pitch, and the reconstruction algorithm applied. Therefore, machine-learning-based super-

resolution has attracted attention for restoring high-resolution (HR) CT images from low-resolution (LR) CT images.

Deep-learning-based super-resolution techniques have recently achieved remarkable results. These are classified into end-to-end or blur-kernel based approaches, the latter of which are more reliable because its process is more transparent than that of the end-to-end approach, which is a black box. Some of the blur-kernel based approaches assume the underlying blur kernel to be known, such as a bicubic kernel or a Gaussian kernel. However, the blur kernel of a given CT scanner is often unavailable.

This paper focuses on the problem of a blind SR of CT images, that is, the super resolution that occurs when the underlying blur kernel is unknown. A related study was conducted on a deep alternating network (DAN) [1], for which a joint blur kernel estimation and super resolution were proposed but an estimated blur kernel was not explicitly used. By contrast, the deep Wiener deconvolution network (DWDN) [2] applies an explicit deconvolution process in a feature space by integrating a classical Wiener deconvolution framework with learned deep features. However, it was applied to a deblurring problem using an encoder-decoder architecture, but not to the super resolution, for which the encoder-decoder architecture has been shown to provide poorer results.

This paper presents a blind SR method that applies a blind super-resolution by integrating an explicit Wiener deconvolution in a joint blur-kernel estimation and super-resolution framework. We demonstrate an improvement in the super resolution by applying the proposed method to a micro-CT image.

Methods

We employ the architecture proposed in DWDN to explicitly deblur an input LR image in an LR feature space using Wiener deconvolution with an estimated two-dimensional blur kernel. We propose a super-resolution architecture for the feature refinement module by referring to the architecture used in the residual channel attention network (RCAN). Specifically, we pass the deblurred deep features obtained from the Wiener deconvolution to the residual in residual (RIR) module proposed in RCAN. Finally, the output from the RIR module was upscaled through an upscale module.

During the training process, we minimize the sum of the estimator and restorer losses of Eqs. (1)–(3) in an alternating manner, through which the network of the estimator is trained while the network of the restorer is fixed, and vice versa.

$$L_{estimator} = L_1(Est.kernel, GT\ kernel) \quad (1)$$

$$L_{restorer} = L_1(SR, HR) + \alpha * L_{perceptual}(SR, HR) \quad (2)$$

$$Loss = L_{estimator}(Est.kernel, GT\ kernel) + L_{restorer}(SR, HR) \quad (3)$$

Results

Micro CT images with a 70-mm pixel size were used as HR images, and we estimated it from pseudo LR (PLR) images synthesized through blurring and subsampling the HR images. In this study, to realize blind super-resolution, the blurring process employed an isotropic Gaussian kernel with a standard deviation of 0.8–3 pixels for every training iteration. The blurred images were then downsampled by a factor of 4 to generate PLR images whose pixel size was almost similar to that of a clinical CT image. We used HR CT volumes of two cases for training, and an HR CT volume of a case for validation and a different case for testing.

Because a Wiener deconvolution is carried out within the LR feature space (i.e., on extracted features with the same spatial size as the LR image), we need to provide the restorer with a deblur kernel. We assume the standard deviation of the deblur kernel to be one-fourth of the ground truth Gaussian SR blur kernel, i.e., in the range of 0.2–0.75.

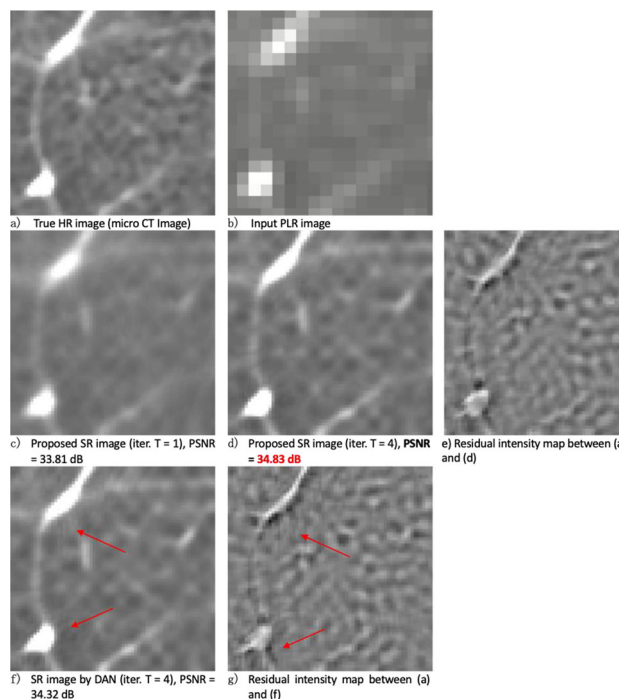


Fig. 1 SR images of test dataset when σ of blur kernel is 2.16 [pixel]

Table 1 Evaluation indices of test dataset

Model	PSNR (dB)	SSIM
Proposed Model	33.49	0,907
DAN	33,37	0,906

The model with the largest peak signal-to-noise ratio on the validation images was applied to the test images. It can be seen from Fig. 1(c) and (d) that the proposed SR image at iteration T = 4 is significantly improved compared to that at T = 1, which shows the superiority of the minimization in an alternating manner and using the estimated blur kernel. In addition, the proposed SR images have fewer artifacts compared with those of DAN, as shown by the red arrows in Fig. 1(f) and (g).

Table 1 shows the PSNR and SSIM of the test dataset, where the proposed model is superior to DAN by 0.12 dB and 0.001 on average, respectively. We suppose that one of the reasons for the improvement is the successful estimation of the blur kernel. For instance, the average L1 loss, shown in Eq. (1), of the estimated kernels over the test dataset during experiment 1 was 0.00047.

Conclusion

We proposed a blind SR method that integrates a Wiener deconvolution in a joint blur kernel estimation and a super-resolution framework. We applied the proposed method to pseudo LR images generated from micro-CT images and found the effectiveness of explicitly using a blur kernel.

Acknowledgements

This work was partly supported by MEXT/JSPS KAKENHI Grant Numbers 20H03908.

References

- [1] Huang Y, Li S, Wang L, Tan T (2020) Unfolding the alternating optimization for blind super resolution. *Advances in Neural Information Processing Systems*, vol. 33.
- [2] Dong J, Roth S, Schiele B (2020) Deep Wiener Deconvolution: Wiener Meets Deep Learning for Image Deblurring. *Advances in Neural Information Processing Systems*, vol. 33.

Compensation of deformable motion for vascular interventional cone-beam CT with a vessel-targeted approach

A. Lu¹, W. Zbijewski¹, M. Unberath², J. Siewerdsen^{1,3}, C. Weiss^{1,3}, A. Sisniega¹

¹Johns Hopkins School of Medicine, Department of Biomedical Engineering, Baltimore, United States ²Johns Hopkins University, Department of Computer Science, Baltimore, United States ³Johns Hopkins University School of Medicine, Russell H. Morgan Department of Radiology, Baltimore, United States

Keywords Motion compensation, Interventional radiology, Cone-beam Computed Tomography, Deep learning

Purpose

Cone-beam CT (CBCT) has seen increased use in the interventional suite, with application to intraprocedural guidance and quality assurance. CBCT has demonstrated improved visualization of small vasculature, critical in selective embolization procedures, compared to 2D imaging, due to reduction in anatomical clutter and ability for 3D localization. However, interventional robotic C-arms suffer from moderately long image acquisition time (> 6 s) that is susceptible to involuntary patient motion from a complex combination of motion sources. The resulting deformable motion is poorly suited to compensation approaches that invoke temporal periodicity or use external tracking devices and physiological signals to infer internal motion. Image-based autofocus has demonstrated potential for soft-tissue deformable motion compensation by optimizing image properties associated to motion-free images (e.g., sharpness). However, conventional autofocus metrics are often agnostic to the imaging task, and act on the complete image, which may result in poorly conditioned problems that challenge the autofocus optimization. Robust autofocus compensation could be obtained with targeted approaches featuring metrics specifically designed to promote image features associated to intricate vascularity depiction, such as presence of sparse, tubular-shaped structures. While pertinent to enforcement of vascularity, such metrics can promote unrealistic appearance in regions not containing vascular structures and must be accurately targeted to a region of interest. This work presents an approach for deformable motion compensation targeted at vascular imaging in the abdomen via a novel vesselness autofocus cost-function and a strategy for accurate identification of target vascularity.

Methods

The method builds on two premises: moving vasculature is sufficiently conspicuous in 2D projection data (prior to CBCT reconstruction); and that vascular structures exhibit a tubular-like appearance that can be exploited for 3D motion-compensated reconstruction. Building from these premises, the approach in Fig. 1 was developed as a two-stage framework: i) automatic segmentation

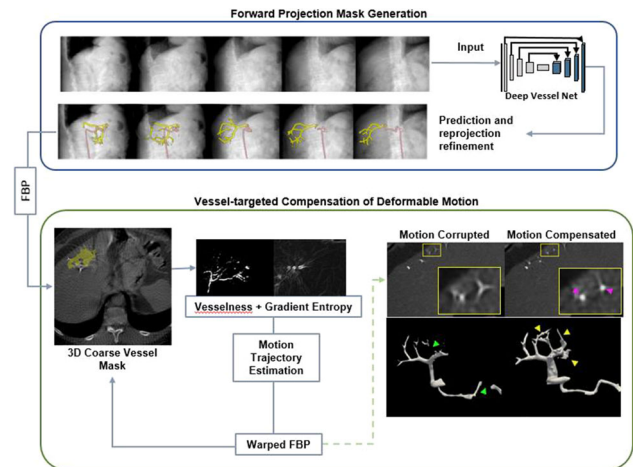


Fig. 1 Vascular motion compensation framework integrates neural network inference and iterative optimization. The U-Net trained on synthetic TACE forward projections produces semantic segmentations of catheter and vasculature in each projection view. These segmentations can be backprojected to form a coarse 3D mask that captures the region containing vessels. This region is targeted for deformable iterative motion compensation, using a vascular autofocus cost function to improve image quality and reduce motion artifacts, particularly within the ROI

of vascular regions-of-interest (ROIs) in projection data to yield an approximate volumetric distribution of the target vasculature; and, ii) a deformable motion compensation approach based on a novel autofocus metric designed to promote the presence of vessel-shaped structures in the target region.

Vascular target generation

Segmentation of 2D distributions of contrast-enhanced vasculature and the associated catheter were obtained with a two-class deep convolutional neural network (CNN) based on the U-Net architecture. The deep-CNN included 4 encoding blocks followed by 4 decoding blocks, each including three 3×3 convolution layers followed by ReLU activation and 2×2 pooling, resulting in a total of 1.73 million parameters. The 2D segmented vasculature was backprojected to generate an approximate volumetric ROI containing the target vascular tree and catheter in the reconstruction domain. Residual inconsistencies between the masks obtained in individual projection views were mitigated by removing regions of the volumetric mask receiving contributions from less than 30% of the total number of projection views.

Targeted motion compensation framework

The autofocus motion compensation algorithm is based on quantification of the image “vesselness” within the target region, where vesselness is described in terms of the 3D Hessian within the ROI, yielding a metric assigning large scores to voxels likely to be members of a tubular structure. This metric is incorporated in the autofocus cost function to promote spatially sparse, vesselness for voxels inside the ROI, as expected for reconstruction of fine vascular structures. The autofocus cost function also includes a gradient entropy term acting outside the ROI to simultaneously drive global image sharpness as described in [1].

The 4D motion vector field is modelled as a spatio-temporal arrangement of 14,256 B-spline knots that is integrated into a warped backprojection method to generate a motion-compensated volume. The motion trajectory minimizing the autofocus cost-function was obtained via stochastic numerical optimization with the covariance matrix adaptation—evolutionary strategy (CMA-ES).

Table 1 Intersection over union (IOU) values

	vasculature	Vascular Mask Generation IOU			Motion Compensation Target IOU Improvement
		refined vasculature	catheter	refined catheter	
test set (motion free)	0.69	0.74	0.82	0.83	25.01%
motion-affected case	0.42	0.50	0.70	0.69	
clinical	0.13	0.34	-	-	

Training and evaluation datasets

The vascular target segmentation network was trained and evaluated on simulated datasets obtained from 54 liver MDCT scans from the CT-ORG dataset. A total of 560 CBCT projections were generated per scan to include synthesized catheters (2.5 mm diameter, 1400 HU contrast) and a vascular tree [2] spanning from the catheter entrance to (15–35) points randomly distributed in the liver. The vascular structures ranged 1.0–3.3 mm diameter with 1400 HU contrast to liver parenchyma. Smooth, deformable, time-varying motion fields with 15 mm maximum motion amplitude were induced to each volume to generate motion-contaminated data. Performance assessment was obtained with a test dataset obtained analogously to the training containing 2800 projection views. Validation studies were performed using two scans from a clinical C-arm CBCT (Artis Zee, Siemens Healthineers) acquired during TACE procedures, in which vascularity was manually segmented.

Results

The vascular target segmentation method yielded accurate depiction of the vascular anatomy and catheter region in the test dataset, as demonstrated by the intersection over union (IOU) values shown in Table 1. The primary detriment to segmentation accuracy was (non-vascular) anatomical clutter arranging into vessel like structures. The projection consistency method used to refine network predictions resulted in near complete removal of such hallucinated, false-positives in segmentation (Table 1), resulting in an average improvement of 7% in IOU in cases with no patient motion and 19% in presence of motion. The segmented vascular trees featured a realistic visual appearance (Fig. 1). Lower IOU values were observed in clinical cases, largely attributable to detection of enlarged vessel regions, compared to the reference. The enlarged estimated vascularity however followed the same spatial distribution depicted by the manual segmentation, resulting in accurate identification of the vessel-containing region of interest for autofocus.

Autofocus motion compensation was performed on a simulated case yielding restoration of vascular structures. Threshold-based segmentation of the vascular anatomy in the motion-compensated CBCT demonstrated an IOU improvement of 25% compared to the motion-affected image, using the motion-free volume as reference. Mitigation of motion-induced distortion and improved visualization of vascularity was also observed in the reconstructed volumes, improving the connectivity of the vascular tree (green arrows in Fig. 1), recovery of fine vascular structures obfuscated by motion corruption (magenta arrows in Fig. 1), and restoration of the tubular-shape in highly distorted vessels (inset, Fig. 1).

Conclusion

Targeted motion compensation for interventional vascular CBCT was developed via combination of a deep learning-based approach for segmentation of vascular structures and catheters in projection data and a novel, vessel-promoting, autofocus cost function. The results in the testing datasets yielded accurate segmentation of the vascular target region in projection data and improved visualization of small vessels affected by complex deformable motion. The proposed targeted concept with identification of target anatomical structures and metrics specifically tailored to vascular imaging tasks poses an step

towards robust motion compensation for reliable delineation of small vascularity in interventional CBCT.

References

- [1] Sisniega A, Lu A, Huang H, Zbijewski W, Unberath M, Siewerdsen JH, Weiss CR (2022) Targeted Deformable Motion Compensation for Vascular Interventional Cone-Beam CT Imaging. SPIE Medical Imaging 2022: Physics of Medical Imaging (abstract ID: MI101-230).
- [2] Wagner MG, Laeseke Paul, Speidel MA (2019) Deep learning based guidewire segmentation in X-ray images. Medical Imaging 2019: Physics of Medical Imaging.

Automatic intraoperative iceball segmentation during MRI-guided focal cryoablation for prostate cancer

I. Gangavaram¹, P. Moreira¹, J. Tokuda¹, K. Tuncali¹, C. Tempany¹
¹Brigham and Women's Hospital, Department of Radiology, Boston, United States

Keywords cryoablation, automatic, segmentation, iceball

Purpose

Percutaneous cryoablation is a cancer treatment intended to destroy malignant cells by inducing cryoinjury in the affected tissue volume. It has been used to treat cancer in different areas, including the kidney, prostate, liver, breast, and bone. Typically, one or more cryoprobes are inserted into or around the tumor to rapidly remove heat from the tissues. An ellipsoidal frozen volume forms as the cryoprobe absorbs heat. This frozen volume, also called iceball, continues to expand until the desired ablation margin is achieved. The iceball is visible in most cross-sectional imaging modalities such as ultrasound, CT, and MRI, potentially allowing real-time confirmation of the tumor coverage by the iceball and the avoidance of nearby critical structures. In particular, MRI offers excellent visualization of both the target tissue and the iceball, which makes MRI a good choice for focal cryoablation of prostate cancer [1]. However, monitoring the leading edge of the iceball in 3 dimensions (3D) is not intuitive and often requires a careful and time-consuming slice-by-slice review of the volumetric image. Ensuring a proper target volume (i.e., visible tumor plus a safety margin) coverage while minimizing damage to surrounding structures is crucial to achieving the expected clinical outcomes and depends on adequate iceball monitoring. The failure to properly monitor the iceball growth can lead to insufficient or excessive ablation causing local recurrence and postprocedural morbidity. This study presents a preliminary evaluation of an artificial intelligence (AI)-based automatic iceball segmentation software to monitor the iceball growth during MRI-guided cryoablation. The software was incorporated into an open-source medical imaging software, 3D Slicer (<https://www.slicer.org/>), for 3D visualization and validated using retrospective clinical data obtained during MRI-guided focal cryoablation procedures.

Methods

This study developed a custom automatic segmentation software implemented in Python with imaging machine-learning frameworks, MONAI, and SimpleITK. Visible iceballs on intraprocedural MRI were automatically segmented using a convolutional neural network (CNN) model based on the 3D U-Net [2]. The model was trained using intraoperative MR images of the prostate obtained

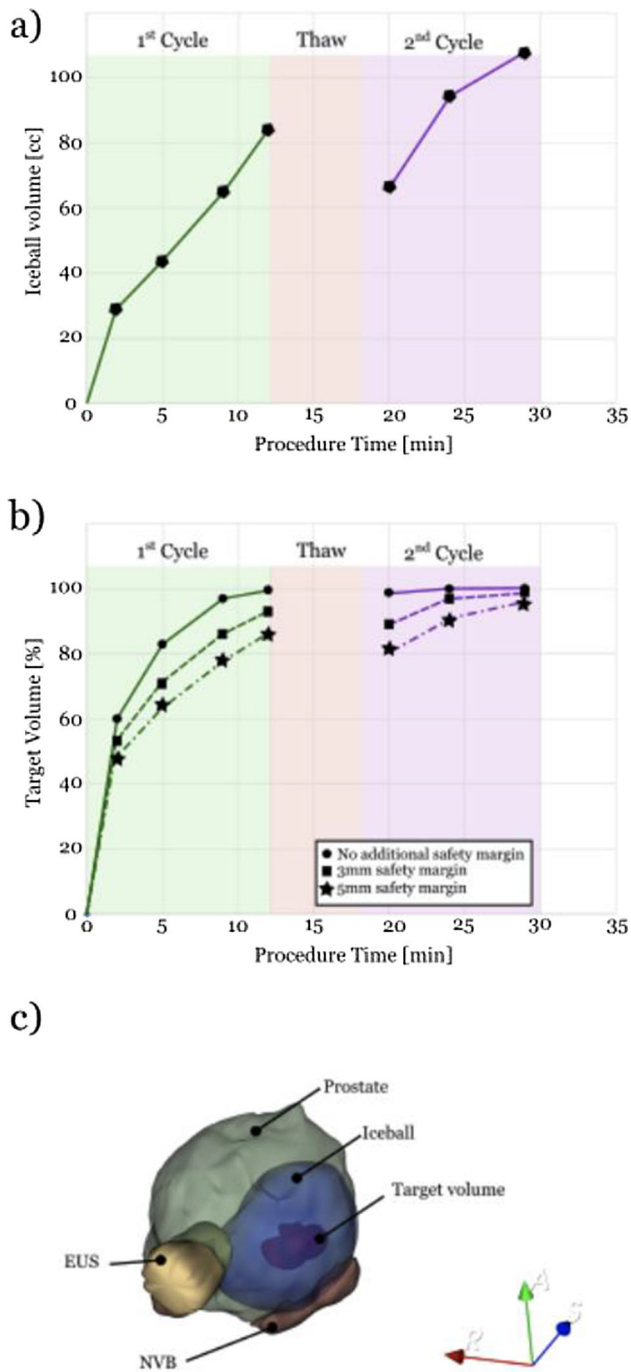


Fig. 1 A representative case of using AI-based segmentation software. Top: The progression of the iceball volume throughout two freezing cycles. Center: The evolution of the tumor coverage over the procedure time considering just the desired target volume and adding safety margins to compensate for possible segmentation and registration errors. Bottom: One advantage of using 3D Slicer is the possibility to visualize the 3D surface model of the patient anatomy, including the segmented iceball, the target volume, the neurovascular bundle (NVB) and the external urethral sphincter (EUS)

during 46 MRI-guided cryoablation procedures. In all cases, 17-gauge cryoprobes (IceSeed, Galil Medical Inc., Yokneam, Israel) were

inserted transperineally using a guiding template and a custom-made planning software. The number of cryoprobes depended on target size and location. Among the 46 cases, we had a median of two cryoprobes and a maximum of 6 probes. The intraoperative monitoring images were acquired during the procedure at different time stamps using a T2-weighted turbo spin-echo sequence (20 slices, 3 mm slice thickness, FOV of 157×180 mm, and 168×192 matrix). As part of a previous study, iceballs were manually segmented on 119 images consisting of 77 training, 21 validation, and 21 test images. The training and validation images were randomly selected and used for training, while the test images were used to compute Dice Similarity Coefficient (DSC) between the manual and automatic segmentation. The model was incorporated into a custom plug-in module for 3D Slicer so that the segmentation result can be converted to a 3D surface model and visualized along with 3D anatomical models on the fly.

Results

The average DSC between the manual and automatic segmentation was 0.843 ± 0.208 . Figure 1 presents the results of a representative case, where an 11 cc target volume located at the peripheral zone was ablated using two cryoprobes. The top two plots show the change of the iceball volume throughout two freezing cycles and the percentage of target volume encompassed by the iceball, respectively. Additional safety margins were added by equally dilating the segmented target volume in all directions to compensate for possible target segmentation and image registration errors during the procedure. Figure 1b shows two examples of safety margins and how additional margins affect the target volume coverage. The use of this type of safety margin will depend on several factors, such as the prostate region where the target is located, proximity to critical structures, and size of the target volume. In order to assist the physician, the software allows for the visualization of the 3D surface model of the target volume, the segmented iceball and models of the prostate gland, neurovascular bundles (NVB), and the external urethral sphincter (EUS), as presented in Fig. 1.

Conclusion

This study presented an AI-based software for intraoperative automatic iceball segmentation. Our preliminary evaluation results suggest the feasibility of using a CNN, in particular a 3D U-net, for iceball segmentation. The software achieved an average DSC value above 0.8, which is suitable for our application. However, it was possible to observe the presence of iceball segmentation artifacts in the rectum and other darker image areas. Those artifacts won't significantly affect the target coverage analysis but might impair the 3D visualization and automatic calculation of ablation margin. We are currently improving the automatic segmentation by adding patient-specific input data, such as the type of cryoprobe and their tip location before the freezing cycles. This additional information will help the algorithm avoid artifact segmentations outside the iceball. Besides, it is well known that MRI image quality may vary, therefore, the algorithm should be robust to variations in the image quality. The software will be integrated with the current 3D Slicer module used for focal cryoablation in our institution and will assist the physician during the intraoperative monitoring. As future work, we intend to extend this work to other organs where the 3D evaluation of the ablation margin is equally challenging.

References

- [1] Bomers, Overduin, Jenniskens, Cornel, van Lin, Sedelaar, Fütterer (2020) Focal Salvage MR Imaging-Guided Cryoablation for Localized Prostate Cancer Recurrence after Radiotherapy: 12-Month Follow-up. *J Vasc Interv Radiol.*(1): 35–41.

- [2] Meyer, Mehrdash, Rak, Bashkanov, Langbein, Ziaei, Kibel, Tempny, Hansen, Tokuda (2021) Domain adaptation for segmentation of critical structures for prostate cancer therapy. *Sci Rep* 11.

Bicuspid aortic valve morphology and aortic hemodynamics: an experimental analysis using an MRI-compatible pulsatile flow circulation system

K. Hattori¹, N. Nakama², J. Takada², E. Kawasaki², K. Hamada², M. Nagao³, Y. Goto⁴, H. Niinami⁵, K. Iwasaki⁶

¹Waseda University, Advanced Biomedical Science, Tokyo, Japan
²Waseda University, Integrative Bioscience and Biomedical Engineering, Tokyo, Japan
³Tokyo Women's Medical University, Diagnostic Imaging and Nuclear Medicine, Tokyo, Japan
⁴Tokyo Women's Medical University Hospital, Radiological Service, Tokyo, Japan
⁵Tokyo Women's Medical University, Cardiovascular surgery, Tokyo, Japan
⁶Waseda University, Modern Mechanical Engineering, Tokyo, Japan

Keywords bicuspid aortic valve, hemodynamics, Magnetic Resonance Imaging, pulsatile flow circulation sys

Purpose

Aortic hemodynamics associated with bicuspid aortic valve (BAV) is affected by the BAV morphology. However, due to the morphological variety of BAV, the relationships between BAV morphology and aortic hemodynamics have not been well clarified. We experimentally investigated influences of BAV morphology on aortic hemodynamics using an MRI-compatible pulsatile flow circulation system.

Methods

Two types of BAV models with cusp angles of 240°–120° (asymmetric BAV) and 180°–180° (symmetric BAV) were prepared using bovine aorta and pericardium. The MRI-compatible pulsatile flow circulation system, which duplicated the thoracic aortic circulation, composed of an elastic left ventricle model, aortic valve model, aortic arch model, aortic compliance chamber, resistive unit, left atrium pressure chamber, and polymeric valve as an alternative to the mitral valve, was developed. The elastic left ventricle model was operated using a pneumatic control system to simulate periodically contracting and relaxing myocardium. The aortic arch model with the geometries based on the literature values of young human aortic arch were devised. The aortic arch model was created in a tunnel shape rather than a tubular shape, to provide a static signal for background phase correction during MRI. The characteristics of aortic valvular outflow jet and circulation values of secondary rotational flow in the aortic arch model were qualitatively and quantitatively evaluated using 4D-flow MRI. Streamlines at peak systole were compared among 4 BAV morphologies including right and left coronary cusps fused BAV (R/L type), right and non-coronary cusps fused BAV (R/N type), left and non-coronary cusps fused BAV (L/N type), and symmetric BAV. Circulation values of secondary flow were calculated at 3 cross-sections (level corresponding to the proximal, middle, and distal parts of the ascending aorta) to quantify the magnitude of helical components of secondary flow.

Results

Markedly eccentric aortic valvular outflow jets directed to the aortic wall faced to the smaller leaflet were present in the 3 asymmetric BAVs. In the R/L type of BAV, an eccentric jet impinging on the outer curvature of the ascending aorta was present. In the R/N type of BAV, a left-posterior directed jet shifting to the outer curvature of the proximal aortic arch was present. In the L/N type of BAV, an eccentric jet impinging on the left-anterior wall of the proximal

ascending aorta was observed. In the symmetric BAV, mildly eccentric outflow jet, which did not impinge on the ascending aortic wall, was present. The asymmetric BAVs induced larger circulation of secondary flow in the ascending aorta, compared to the symmetric BAV.

Conclusion

Our study indicated that the angles and orientations of the BAV impacted on the locations of outflow jets impinging on the aortic wall and circulation values of secondary flow. In the asymmetric BAVs, the direction of jet was influenced by the position of smaller leaflet. Our data suggests that the R/L type of BAV may be a risk factor inducing an asymmetric ascending aortic aneurysm bulged toward the aortic outer curvature. The R/N type of BAV may be a risk factor of an ascending aortic aneurysm involving the transvers arch, whereas the L/N type of BAV may induce an aortic aneurysm involving the proximal part of the ascending aorta.

The CathPilot: a novel approach for accurate interventional device steering and tracking

J. Zhou¹, A. Quadri², A. Sewani¹, Y. Alawneh¹, R. Gilliland-Rocque¹, C. Magnin³, A. Dueck⁴, G. Wright⁴, M. Tavallaei^{1,4}

¹Ryerson University, Biomedical Engineering, Toronto, Canada
²University of Waterloo, Mechanical and Mechatronics Engineering, Waterloo, Canada
³Magellan Biomedical Inc., Richmond Hill, Canada
⁴Sunnybrook Health Sciences Centre, Toronto, Canada

Keywords steerable catheter, cable-driven manipulator, cardiovascular interventions, minimally-invasive

Purpose

Catheter-based procedures, the primary surgical interventions for treating cardiovascular diseases, have high failure (~ 20%) and high complication rates (~ 30%) [1]. This is due to the key challenges of accurate device steering and navigation. The long and flexible devices (e.g., catheters and guidewires) engage with the tortuous anatomy along their length and are manipulated remotely from outside of the patient's body. These limitations constrain the reachable workspace of the device tip and reduce its controllability. Furthermore, these procedures are guided with 2D projection X-ray fluoroscopy that does

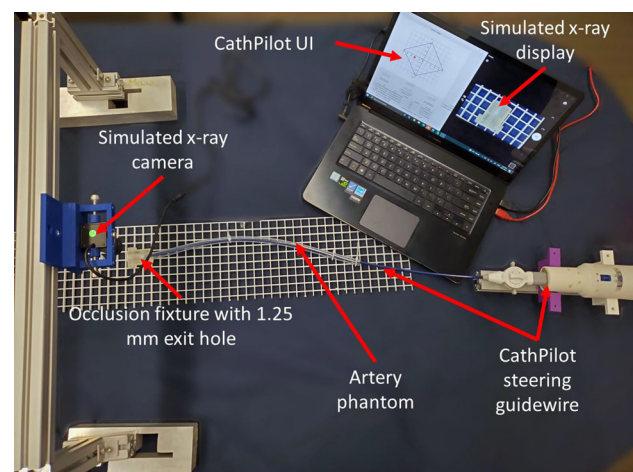


Fig. 1 Experimental setup with, in this case, the CathPilot steering the guidewire to access a 1.25 mm target hole in an artery phantom under simulated x-ray

Table 1 CathPilot performed significantly faster in crossing the occlusion

	Crossing time [s]	Success Rates
Non-Steerable Catheter (n=15)	420 ± 219	53,3%
Steerable Catheter (n=15)	239 ± 182	93,3%
CathPilot (n=15)	81 ± 133	100%

not provide 3D spatial feedback. In angioplasty, these limitations lead to challenges in crossing the occlusion with a guidewire for revascularization and ultimately technical failure of the procedure. In this study, we assess the performance of a novel steerable catheter (CathPilot) in crossing such occlusions in ex vivo phantom models.

Methods

The CathPilot is an expandable cable-driven manipulator that allows for localized device steering and tracking relative to the anatomy [2]. Once the expandable frame is at the location of interest, the user deploys it by retracting its delivery sheath. The frame expands and acts as a mechanical reference for the manipulation and tracking of the device. The steering capability of the device is independent of the local environment and the tortuosity of the path. Sensors coupled to the cables allow for tracking the position of the device to display it in a user interface (UI) for full 3D feedback in conjunction with conventional x-ray. We have proved that this system has similar steering capability regardless of the shape and size of the local anatomy [2]. For assessment of the system, we designed an ex vivo phantom study with an occluded arterial phantom model to compare the CathPilot's performance against a conventional non-steerable catheter and a steerable catheter. The artery model had an inner diameter of 10 mm. Using each method, three users were asked to access a 1.25 mm target hole within the occluded lesion (5 lesions used) using a 0.89 mm guidewire under a simulated fluoroscopy setup (Fig. 1). With a 10-min time limit, we compared crossing times and success rates.

Results

As predicted, the CathPilot performed significantly faster in crossing the occlusion (, two-way ANOVA) (Table 1). It was also always successful at hitting the 1.25 mm target, while the conventional non-steerable and steerable catheters sometimes failed.

Conclusion

The novel CathPilot promises to overcome the limitations of conventional devices; it directly addresses the challenges of catheter control by providing direct manipulation, accurate positioning, and tracking of the device tip. This allowed the CathPilot to perform significantly better in an angioplasty phantom model compared to its conventional counterparts. Future directions include optimization and testing for other procedures.

References

- [1] Bradbury A, Adam D, Bell J, Forbes J, Fowkes G, Gillespie I, Ruckley C, Raab G (2010) Bypass versus Angioplasty in Severe Ischaemia of the Leg (BASIL) trial: A survival prediction model to facilitate clinical decision making. *J. Vasc. Surg.* 51(5):5S–17S.
- [2] Zhou J, Quadri A, Sewani A, Alawneh Y, Gilliland-Rocque R, Magnin C, Dueck A, Wright G, Tavallaei M (2021) The CathPilot: A Novel Approach for Accurate interventional Device Steering and Tracking. *IEEE/ASME Trans Mechatronics*.

Development of a 3D printer-based model to assess contact pressure of cryoballoon ablation

T. Hayashi^{1,2}, E. Kawasaki³, J. Takada³, M. Murakami¹, S. Saito¹, K. Iwasaki^{2,4}

¹Shonan Kamakura General Hospital, Department of Cardiology, Kamakura, Japan ²Tokyo Women's Medical University—Waseda University Joint Graduate School, Cooperative Major in Advanced Biomedical Sciences, Tokyo, Japan ³Waseda University, School of Advanced Science and Engineering, Tokyo, Japan ⁴Waseda University, School of Advanced Science and Engineering, Tokyo, Japan

Keywords catheter ablation, cryoballoon ablation, 3D printer model, simulation model

Purpose

Atrial fibrillation is one of the arrhythmias which can be a predictor of stroke and heart failure. The estimated number of patients in Japan is 1 million in 2020 and it is expected to increase in the future. One of the treatments of atrial fibrillation is catheter ablation. Pulmonary vein isolation with catheter ablation is reported to be an effective method for the treatment of atrial fibrillation, and currently, balloon catheter ablation and radiofrequency catheter ablation are established treatments [1]. Pulmonary vein isolation using a balloon catheter is characterized by shorter operation time and simpler catheter manipulation compared with conventional radiofrequency catheter ablation [2]. However, due to the limitation of balloon size, some cases have been reported to be unsuccessful. Some studies implied that contact force of cryoballoon to pulmonary vein might be a great predictor of success in cryoballoon ablation, but it is still not proven yet. Therefore, in this study, we developed a left atrial and pulmonary vein model that can detect the contact force by the cryoballoon catheter.

Methods

Based on the patient's computed tomography data, a 3D printed left atrial and pulmonary vein model was fabricated. Using the 3D printed model, an elastic left atrial and pulmonary vein model was developed using a silicone. Then, a newly-designed sensing unit using a pressure transducer was placed around pulmonary veins of the silicone model

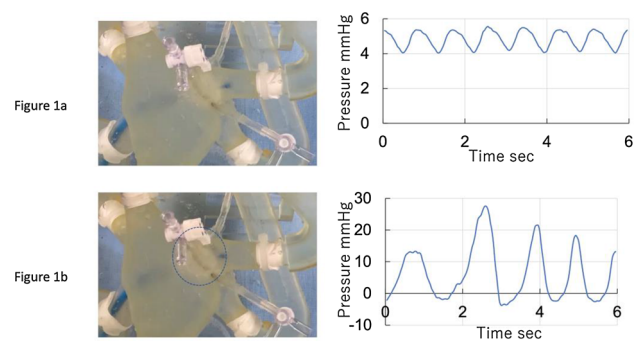


Fig. 1 a Changes in pressure acting on pulmonary vein model without contact with cryoballoon, **b** Changes in pressure acting on pulmonary vein model in contact with cryoballoon. The blue dotted area indicated the region where the cryoballoon was in contact with pulmonary vein model

to detect the contact pressure. To simulate the in vivo environment, pressures of left atrial and pulmonary vein were duplicated in the experiments.

We sought to assess pressures acting on the pulmonary vein model in contact with cryoballoon under physiological pressure environments. First, cryoballoon was inserted into the left atrium–pulmonary vein model by the physician. Then, pressures acting on the pulmonary vein model without contact with cryoballoon and in contact with cryoballoon were measured and compared.

Results

The left atrial internal pressure was duplicated in the physiological ranges between 4.4–7.3 mmHg. The results showed that when the cryoballoon was not in contact with the pulmonary vein model, the maximum pressure acting on the pulmonary vein model was 5.7 mmHg (Fig. 1a). When the cryoballoon was in contact with the pulmonary vein model, the maximum pressure acting on the pulmonary vein model was 28 mmHg (Fig. 1b). We could successfully measure the contact pressure acting on the patient-specific atrial and pulmonary vein model.

Conclusion

We could successfully develop the elastic atrial and pulmonary vein model based on a patient's CT data. Using the patient-specific model, the contact pressure acting on the pulmonary vein model could be measured. The methodology developed in this study will be useful to investigate influences of left atrium anatomical characteristics on success of cryoballoon ablation in terms of the contact pressure.

References

- [1] Haïssaguerre M, Jaïs P, Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N Engl J Med.* 1998 Sep 3;339(10):659–66.
- [2] Kuck KH, Brugada J, Cryoballoon or Radiofrequency Ablation for Paroxysmal Atrial Fibrillation. *N Engl J Med.* 2016 Jun 9;374(23):2235–45.

Development of an in vitro functional mitral regurgitation model for preclinical evaluation of transcatheter mitral valve repair devices

H. Morimura¹, Y. Okamoto¹, J. Takada¹, K. Iwasaki¹

¹Waseda University, Tokyo, Japan

Keywords Mitral valve repair, Mitral regurgitation, in vitro, model

Purpose

Various transcatheter mitral valve repair devices are emerging, however, its preclinical evaluation method has not been established. As they repair and preserve the native valves, valve repair devices should be evaluated on disease models instead of normal valves. However, diseased animal model is not easy to produce in the point of its cost, technique, and reproducibility. Thus, we aimed to develop a repairable in vitro mitral regurgitation model for preclinical evaluation of transcatheter mitral valve repair devices. We focused on functional mitral regurgitation because their surgical results are controversial, therefore, transcatheter devices may play a bigger role compared to degenerative mitral regurgitation.

Methods

Whole swine hearts were obtained from an abattoir (n = 6). The hearts were dissected and the mitral complex including the left atrium were explanted. The annulus was dilated with an original dilator which had been chosen according to each annular size. They were

simultaneously immersed in collagenase for effective dilation and to prevent tissue injury. Left atrium was trimmed and sutured to a silicone sheet which was then inserted between left atrium and left ventricle models of the pulsatile circulation simulator. Papillary muscles had been immersed in glutaraldehyde to prevent tissue injury before they were secured to the left ventricle box. The pulsatile circulation simulator was then driven at a pulse rate of 70 bpm. Mean flow and aortic pressure were adjusted to approximately 4–5 L/minute and 110–140/70–90 mmHg respectively. Regurgitant fraction was measured before and after the valve repair procedure: the edge-to-edge technique. We chose the edge-to-edge technique as a repairing method because this technique became the basis of MitraClip which is the only transcatheter mitral valve repair device available in the United States, Europe, and Japan.

All data were recorded continuously for 6 beats and averaged. Continuous variables are expressed as mean ± standard deviation. Paired *t*-test was employed to evaluate the difference between the regurgitant fraction before and after the procedure.

Results

The annular size of the porcine hearts was 140.3 ± 5.9 mm in average, which were dilated to 153.4 ± 7.1 mm ($p < 0.05$). Mean flow and aortic pressure before and after the procedure was 4.1 ± 0.2 L/minute vs 3.9 ± 0.4 L/minute ($p < 0.05$) and $122/80$ mmHg vs $122/78$ mmHg ($p = 0.95$) respectively. Mean regurgitant fraction was $46.9 \pm 2.6\%$ and $27.8 \pm 3.2\%$ ($p < 0.05$) before and after the procedure.

Conclusion

We developed an in vitro functional mitral regurgitation model. Regurgitant fraction was 46.9% which was compatible with moderate to severe mitral regurgitation. This fulfills the indication of mitral valve repair. This model was also repairable using the edge-to-edge technique which was proved by the regurgitant fraction as low as 27.8%. Moreover, the narrowing of the effective orifice area induced by the edge-to-edge technique was reflected in the decrease of mean flow. Therefore, this model may not only evaluate the positive effects but also detect adverse effects of the procedure.

We are planning to test the model whether it can be repaired with other techniques such as the annuloplasty technique. In vitro degenerative mitral regurgitation model is under development. These models may help the development and evaluation of novel transcatheter mitral valve repair devices.

Respiratory motion detection in X-ray fluoroscopy using CNN

I. Vernikouskaya¹, D. Bertsche¹, W. Rottbauer¹, V. Rasche¹

¹Ulm University Medical Center, Dept. of Internal Medicine II, Ulm, Germany

Keywords motion detection, convolutional neural network, heatmap, image fusion

Purpose

Augmenting X-Ray (XR) fluoroscopy with anatomic overlays (roadmaps) rendered from 3D CT or MRI images is an essential technique to improve the guidance of the catheterization procedures. Current augmentation methods use static roadmaps without adaptation to the actual patient's cardiac and respiratory state. Following initial registration, respiratory motion in head-feet direction is, however, a major cause of introducing mismatch to the superposition. In our previous study [1] we have applied the convolutional neural network (CNN) to fluoroscopic image pair as input with the target values being the displacement of the diaphragm between these frames to extract the

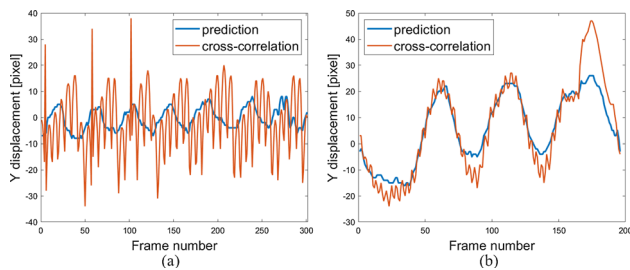


Fig. 1 Superposition of the network prediction (blue) and displacement of the rapid pacing catheter tracked using cross-correlation (orange) for two fluoroscopy runs from the test dataset. **a** MAE = 9.9 pixels. **b** MAE = 4.8 pixels

respiratory motion. Good agreement with the reference method implying template matching over the diaphragm edge using cross-correlation could be achieved and the related motion waveform could be reliably extracted. However, the predicted displacements were highly overestimated when used for the adjustment of the roadmap. In this work, we propose to use instead of the diaphragm a reference catheter which remains static to the heart and thus correlates well with the respiratory induced motion.

Methods

In transcatheter aortic valve implantation (TAVI) procedures rapid right ventricular pacing is used to ensure balloon stability during deployment of the artificial aortic valve. The respective catheter with its single electrode is clearly visible in the X-ray fluoroscopy images and moves synchronously with the aortic valve which is a target structure in these procedures. As such it appears promising to be used as a target for deriving displacements between fluoroscopic frame pairs by a respectively trained CNN.

Training data were derived from TAVI procedures performed at the Ulm University Medical Center. 64 fluoroscopy runs of 512×512 pixels resolution containing the rapid pacing catheter were preprocessed and labeled. The displacements were calculated by defining a rectangular region of interest (ROI) of 15×15 pixels around the catheter tip and tracking its motion through the run with template matching using normalized cross-correlation. All possible permutations of frame pairs were created for each run, which results in a total of 410,298 samples with target displacement values in the range of ± 66 pixels. Samples were split into training-, validation- and test- sets with a ratio of 81%/5%/14%.

The problem was addressed with a rather generic network similar to those for optical flow estimation FlowNetSimple proposed by Dosovitskiy et al. [2], allowing the network to decide itself how to process the image pairs and extract the displacement. The network, consisting of multiple convolutional- and pooling operations before providing the output value from the final fully-connected layer was implemented in Keras and trained on a GeForce GTX 1060 6 GB graphics card for 10 epochs.

Results

Figure 1 visualizes the superposition of the target values derived from cross-correlation and prediction for fluoroscopy runs of the test dataset. Mean absolute error (MAE) averaged over the entire test dataset comprised 8.7 ± 2.5 pixels or 3.0 ± 0.9 mm. The respiratory waveform could be extracted from all runs of the test dataset and could be used for respiratory rate and phase detection.

Figure 2 demonstrates an example of motion compensation using cross-correlation over the rapid pacing catheter (b) and network prediction (c) compared to the non-compensated reference frame (a) for a single fluoroscopic frame of the test run from Figure 2 (a). Whereas cross-correlation highly overestimates the motion amplitude, an excellent match could be achieved when compensated with the

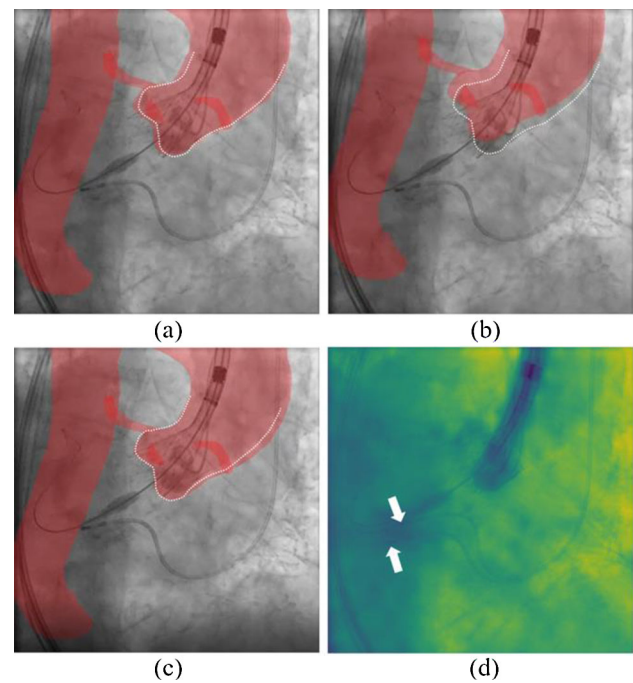


Fig. 2 Motion-compensated roadmaps (a–c) and attention heatmap plot for the respective frame pair (d). **a** Reference overlay without motion compensation. **b** Model overlay compensated using cross-correlation. **c** Model overlay compensated using network prediction. **d** Heatmap plot overlaid on respective image pair with blue regions meaning no network focus and yellow regions meaning strong network focus

predicted displacement. According to attention heatmap plot (d), used to visualize the importance of different regions of the input images for the prediction, the network composes its decision from multiple features in the image and catheter (white arrows) seems not to be important for prediction. These features vary across different fluoroscopy runs and often includes diaphragm and heart edge.

Conclusion

Convolutional neural network has been shown to be capable to extract respiratory motion waveform from fluoroscopic frames with accuracy sufficient to reliably detect respiratory phase and rate. As compared to the diaphragm, rapid pacing catheter represents a uniform structure and as such is much easier to be automatically tracked providing more reliable displacement values for the training. Apparently, smoothed waveforms were obtained as compared to the reference method based on tracking of the rapid pacing catheter with cross-correlation. Large absolute displacements were often underestimated by the network reflecting, however, much better the motion pattern of the target organ as compared to the motion of the rapid pacing catheter, leading to improved motion compensated image fusion. The drawback of the proposed approach is using a single scalar value for describing the motion and can be further addressed by extending the network to predict the 2D displacement vector instead.

Acknowledgements

This work was supported by German Federal Ministry of Education and Research under the funding code BMBF-13GW0372C. Responsibility for the content of this publication lies with the authors.

References

- [1] Baldauf C, Bäuerle A, Ropinski T, Rasche V, Vernikouskaya I (2019) Convolutional neural network (CNN) applied to

respiratory motion detection in fluoroscopic frames. In Proceedings of the International Congress of Computer Assisted Radiology and Surgery (CARS) 14(Suppl 1):S14–S15.

- [2] Dosovitskiy A, Fischer P, Ilg E, Häusser P, Hazirbas C, Golkov V (2015) FlowNet: Learning Optical Flow with Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV): 2758–2766.

A system integration strategy for an autonomously-driven needle in prostate biopsy interventions

R. Lisk¹, F. Tavakkolmoghaddam², G. S. Fischer², J. Tokuda¹, N. Hata¹

¹Brigham and Women's Hospital, Radiology, Boston, United States

²Worcester Polytechnic Institute, Robotics Engineering Department, Worcester, United States.

Keywords Image-guided therapy, Prostate biopsy, MRI-compatible robot, Automated needle insertion

Purpose

MRI-targeted biopsy of the prostate is critical to the diagnosis of prostate cancer, boasting improved diagnostic outcomes and identification of more clinically-significant cancers through the use of intraoperative image feedback. Such biopsies include the in-bore MRI-guided approach, which uses both preoperative and intraoperative MR images in the diagnostic workup and as a needle placement guide. In-bore MRI-guided prostate biopsies can be further divided into transperineal and transrectal approaches. Consistent with findings that transrectal biopsies are strongly associated with sepsis, the transperineal method is quickly becoming widely accepted as the superior method for prostate biopsies with or without MRI guidance.

However, the longer stroke length of a transperineal biopsy makes such operations prone to needle deflection, leading to placement inaccuracy, inadequate sampling, and missed targets. Deflection can be countered via a closed-loop feedback system which integrates real-time image processing with robotic needle control processes to drive needle insertion and guidance; however, no such integrated needle tracking and guidance system exists. Therefore, the unmet need that this research seeks to address is the lack of system integration methods to seamlessly integrate real-time MR imaging with transperineal in-bore robotic needle guidance. If this need remains unmet, consequences will include decreased needle placement accuracy and lower rates of diagnostic success in transperineal prostate biopsy procedures, or continued use of transrectal biopsies and higher postoperative infection rates.

The objective of this study is to address this need through the development of system integration methods to utilize intraoperative images to drive transperineal targeted needle placement with

responsive guidance via real-time MRI imaging, with the hopes of eventually translating our findings to the fusion approach. This paper presents the design, development, and preclinical evaluation of an Open Network Interface for Image-Guided Therapy (OpenIGTLink)-integrated robotic system for targeted transperineal prostate biopsy guided by continuous MRI.

Though some may argue that it is not necessary to perform prostate biopsies with the assistance of real-time MRI guidance, we first emphasize the increased difficulty of achieving accurate needle placement in the transperineal approach and the need for a guidance system. Furthermore, we are using in-bore biopsy as a research platform to validate the robotic system prior to translating the system to the fusion approach. In-bore MRI imaging is an ideal research tool that facilitates data collection in the research process, as image quality is high and we avoid the error that would be caused by image fusion while registering pre-procedural MRI to intra-procedural TRUS. However, once the system is completed and validated, in-bore MRI may no longer be necessary as we hope to translate our system to the fusion approach.

Furthermore, we respond in advance to concerns about the cost and complexity of a robotic system by emphasizing our efforts to simplify the system as much as possible and pointing to past robot-assisted needle intervention technologies that have achieved cost-effectiveness via a number of strategies. Cost to the patient for an in-bore MRI prostate biopsy is comparable to, if not lower than, the cost of the alternative fusion approach due to insurance payment schema.

Methods

System design: The core of the system consists of the robot needle guidance system and its associated robot controller software connected via a fiber optic cable, a bridging software, and a Slicer-based command center module. The robot controller, bridge software, and Slicer module achieve bidirectional communication via OpenIGTLink server and client connections. The custom Slicer module provides (1) IGTLink server control, (2) coordinate system calibration, (3) robot control via target setting and command, (4) robot status monitoring, and (5) incoming MR image reception. An MR marker-based calibration device (Z-frame) is utilized in conjunction with the Slicer module interface to automatically register the MR image to the physical coordinate system and generate a calibration matrix for conversion between the two systems. Finally, images are imported as they are generated from the MRI scanner onto a Linux-based workstation connected to the hospital's private network via the Digital Imaging and Communication in Medicine (DICOM) protocol, where they are processed and utilized in the robot control process initiated by the Slicer module. A diagram describing the system architecture is included in Fig. 1.

Implementation: The surgeon begins by selecting the position of the target lesions on the prostate images via the Slicer user interface. The target position and a registration transform are sent to the robot control software as OpenIGTLink messages, allowing the robot control software to perform inverse position kinematics and trajectory planning to identify the necessary joint positions for accurate insertion [1]. As the robot moves according to these joint position commands, the updated positions are messaged back to the robot control software, bridging software, and the navigation workstation Slicer module. Used in concert with continuous intraoperative MR imaging which monitors and adjusts the position of the prostate target, these tools provide the ability to update the surgical plan in real-time to ensure accurate needle guidance by the robot [2].

Validation: First, frequency of DICOM image transfer from the MRI machine to the Slicer navigation interface was measured to ensure the system can handle a sufficient MR image transmission rate to detect changes in lesion position and needle navigation path due to patient movement or needle deflection and change the robot command sequence accordingly. Second, a gel phantom test was performed to

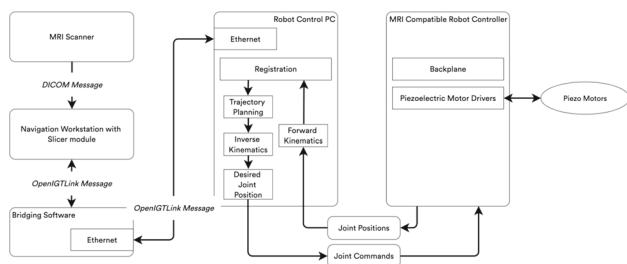


Fig. 1 System control architecture of the prostate biopsy robot [1]

and to quantify global calibration error using a dedicated validation phantom. To evaluate uplink bandwidth limitations, we tuned the Tx video compression rate and were able to send at 100Mbps without loss. On the downlink transmission, we tested several video framerates and observed no packet loss at 20 fps (480Mbps). At 25 fps (580Mbps) some packets started to be missing and at 30 fps (700 Mbps) the image quality was degraded. End to end latency is around 300 ms which is currently not acceptable for clinical practice. This latency is mainly due to application processing and to the final restitution framerate (20 Hz). 5G latency is estimated around 15 ms and is so not impacting significantly image restitution.

Conclusion

5G in the OR can make equipment easier to install, to connect and can help sterilization. This setup allowed to test a complete real time AR solution over 5G and helped assessing 5G possibilities and limitations. 100Mbps were transferred uplink and 480Mbps downlink without loss. Future work should include a validation of X-ray calibration accuracy and a global calibration error quantification. AR application performance needs to be improved before a final demonstration in University Hospital Rennes to gather clinicians' feedbacks. However, this constitutes a promising step towards a 5G connected OR.

This study was partly funded by the European Union's Horizon 2020 research and innovation program under grant agreement n°856,950 (5G-TOURS project). This work also benefited from State aid managed by the National Research Agency under the future investment program bearing the reference ANR-17-RHUS-0005 (FollowKnee project).

References

- [1] El Jaouhari S, Gibaud B, Lemonnier P, Pasquier G, Poiseau E, Guiffard E, Hardy P, Cordonnier E (2019) Introduction to DICOM-RTV: a new standard for real-time video communication in hospitals IEEE International Conference on E-health Networking, Application & Services (HealthCom).
- [2] Zhou Y, Lelu G, Labbé B, Pasquier G, Le Gargasson P, Muriene A, Launay L (FR) [CAR-LE-21-00079]—A Novel 2D Ultrasound Probe Calibration Framework using an RGB-D Camera and a 3D-printed Marker.

Brain tumor surgery using middleware OPeLiNK for integrating intraoperative information

Y. Fujii¹, T. Ogiwara¹, T. Goto¹, K. Hongo¹, Y. Muragaki², T. Horiuchi¹

¹Shinshu University School of Medicine, Neurosurgery, Matsumoto, Japan ²Tokyo Women's Medical University, Faculty of Advanced Techno-Surgery Institute of Advanced Biomedical Engineering & Science, Tokyo, Japan

Keywords OPeLiNK, middleware, Smart Cyber Operating Theater, telemedicine

Purpose

The removal of brain tumors requires not only imaging information such as MRI and navigation systems but also a variety of other information such as neurological function and biological information. However, every medical device in the operating room operates as a stand-alone device and the network have not progressed up to the present date. There was no information and system collaboration with the equipment inside the operating room. Their intraoperative data

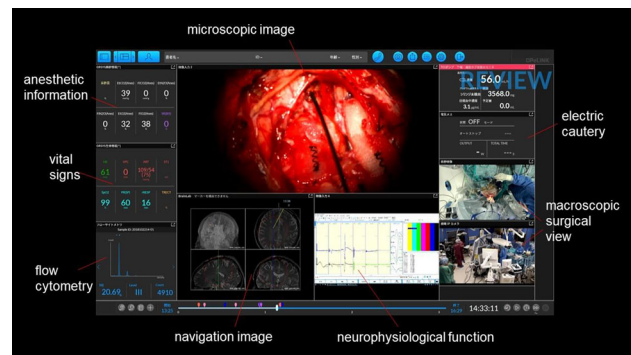


Fig. 1 This picture shows the integrated display of “OPeLiNK”. All of the intraoperative information, such as macroscopic and microscopic surgical view, navigation image, histogram of flow cytometry, neurophysiological monitoring, vital signs, and anesthetic information is displayed on the same screen

were difficult to evaluate retrospectively because the timeline of each medical device was not synchronized. To integrate this intraoperative data, a novel operating room, “Smart Cyber Operating Theater (SCOT)”, which connects the medical devices in the operating room via a network has developed [1]. In this SCOT, the intraoperative information is time-synchronized, recorded, and stored by the middleware “OPeLiNK”. All information of pre-and perioperative data, such as operative video, location and working of surgical tools, updated navigation data by intraoperative MR imaging [2], anesthetic information, intraoperative histopathological data, intraoperative neurophysiological monitoring data, and so on can be displayed in the same screen (Fig. 1). The collected mass information can be reproduced anytime and anywhere in one synchronized timeline. These data are analyzed by a supervising surgeon in the “Strategy desk”. The main surgeon checks the arranged data and receives advice from the Strategy desk during the procedures. Clinical experience of brain tumor surgery using OPeLiNK in our institute is reported.

Methods

Brain tumor surgeries performed at SCOT, which had been started in July 2018, were enrolled. In all surgeries, intraoperative information was integrated by OPeLiNK. Intraoperative MRI with a magnetic field strength of 0.4 T, neuronavigation, serial intraoperative histopathological investigations of the resected tissue by rapid diagnosis and flow cytometry, and comprehensive neurophysiological monitoring were used and their data were visualized in OPeLiNK. The surgical procedure was discussed between the main surgeon and supervising surgeon in the Strategy desk through OPeLiNK intraoperatively, if necessary. Clinical and radiological data from patients who underwent resection at SCOT were analyzed retrospectively.

Results

Sixty-five patients were involved. Histopathological diagnosis was glioma in 29 patients, pituitary adenoma in 29 patients, acoustic tumor, radiation necrosis, and primary central nervous system vasculitis in 1 patient, respectively. Intraoperative discussion with the Strategy desk through OPeLiNK was useful for not only surgeons but also for medical staff in the operation room. Advice for the extent of resection and craniotomy from the Strategy desk was conducted by OPeLiNK using conversation and drawing. Although there was no critical change in surgical procedure after discussion, advice from the Strategy desk using OPeLiNK was more detailed than conventional advice using a cellular phone. Entering the comment intraoperatively was useful for postoperative review. OPeLiNK, which displays multiple intraoperative information, was also used at the postoperative conference, which enabled detailed discussion. With the aid of OPeLiNK, interhospital communication was achieved. Intraoperative

communication with the Strategy desk which was set in another hospital was conducted in a glioma surgery case.

Conclusion

We have reported clinical experience with OPeLiNK for brain tumor surgery in our institute. OPeLiNK was useful for not only sharing intraoperative information with doctors outside the operation room but also postoperative review and education for young doctors. OPeLiNK could be useful for telemedicine in the near future.

References

- [1] Ogiwara T, Goto T, Fujii Y, Nakamura T, Suzuki Y, Hanaoka Y, Ito K, Horiuchi T, Hongo K.: Endoscopic Endonasal Approach in the Smart Cyber Operating Theater (SCOT): Preliminary Clinical Application. *World Neurosurg.* 147: e533–e537. 2021 <https://doi.org/10.1016/j.wneu.2020.12.114>. Epub 2020 Dec 30.
- [2] Fujii Y, Ogiwara T, Watanabe G, Hanaoka Y, Goto T, Hongo K, Horiuchi T.: Intraoperative low-field magnetic resonance imaging-guided tumor resection in glioma surgery: Pros and cons. *J Nippon Med Sch.* 2021 Sep 14. https://doi.org/10.1272/jnms.JNMS.2022_89-301. Online ahead of print.

Clinical experiences of robotic scrub nurse system in neurosurgery

K. Yoshimitsu¹, K. Masamune¹, H. Yamaga², Y. Muragaki¹
¹Tokyo Women's Medical University, Institute of Advanced Biomedical Engineering & Science, Tokyo, Japan ²SHOWA University Northern Yokohama Hospital, Department of Neurosurgery, Kanagawa, Japan

Keywords Robotic scrub nurse, Smart cyber operating theater, Surgical Instrument exchange, Neurosurgery

Purpose

Scrub nurse is one of the significant clinical staff supporting surgical procedure in an operating theater. Smooth instrument exchange with predicting next procedures assists surgeon's operating rhythm efficiency without any distraction, though these performances are generally limited to skilled scrub nurses who were well trained and experienced many surgical cases. However, a chronic shortage of scrub nurses has been severe. To compensate for this shortage, scrub nurse robot (SNR) which aiming to achieve an efficacy at the level of a skilled human scrub nurse during surgery has been proposed [1], and developed by several research teams [2]. Although those have been started developing since 2005, to the best of our knowledge from literature, there is only one robot which has been installed into clinical cases [2]. Because most of all other robots were based on basic research and were focusing to develop novel functions by engineering techniques, there have been few reports describing motivations to install into clinical cases. The aim of our study is fabrication and experiencing clinical cases of such practically yet robotic scrub nurse assistant (RSN) enabled by newly proposed rule as a robotic assistance for a human scrub nurse.

Methods

Specifications for the proposed RSN were designed through discussions with neurosurgeons observing neurosurgical operating cases several times. The RSN was designed with assuming practical installation of the robot from mechanical set up, wearing sterilized plastic covers, installation into operating field, starting manipulations along operating staff's work flows, and wrapping up entire system.

We figured out that when a human scrub nurse was waiting for the moment of exchanging with holding a next instrument for coming procedure, he/she had to keep their eyes on the operating field even though they had other priority tasks which should be treated on the treys. Then, the preliminary concept of the proposed RSN was defined as a robotic assistant repeating exchange task which is the best suited for robotic routine works. The first version of RSN was designed in a CAD software (Inventor professional 2015, AutoDesk) having simple liner mechanism carrying surgical instruments between an operating surgeon and a human scrub nurse. The mechanism we newly proposed in this study employs two liner stages to drive a smart phone sized surgical tray. The size of entire components was designed as 230 × 850 mm with two degrees of freedom. Stroke length of the liner stages is 700 mm. The blue colored treys size 100 × 200 mm for carrying instruments and travel edge-to-edge in 0.8 s. When an operating surgeon asked a next instrument, a human scrub nurse push a foot switch as a trigger for machine movement (Fig. 1A).

The first version of RSN was experienced two clinical cases of neurosurgery and obtained evaluation from operating surgeons and human scrub nurses by asking technical and clinical questioners. Based on the clinical feedback, the robotic hardware design was improved for adjusting to more efficient assist and designed as second version of RSN.

Results

Figure 1A describes the first RSN installed into clinical case. During surgical procedure, a human scrub nurse put a next instrument on a blue colored tray beside him/her and did other tasks till the timing for exchanging (Fig. 1B). When an operating surgeon returns an instrument on a facing empty tray, both of treys start to move to the other side. Then, an operating surgeon grasps a next instrument and a human scrub nurse gets the returned instrument and prepares it for next usage after washing. Thus, the RSN contributes surgical procedure without wasting time for other duty tasks. Table 1 describes the number of instrument exchanging by utilizing the first RSN. In the first clinical case, the RSN was used 34 times and 42 times in the second case. During the cases, the RSN was mainly used under micro surgical procedure in which only micro scissors, dissector, and forceps were exchanged repeatedly among them. However in cases of handing gauze, and when an operating surgeon could not release their eyes from microscope, the robotic exchange did not have chance to work and human scrub nurse served the instrument directly. After experienced two cases of clinical performance, clinical staffs were asked several questioners from the point of view of user experience and robotic assisted surgery. Based on the feedback, a second version of RSN was designed with functional and hardware improvement (Fig. 1). Visible improved points are the shorten length of stroke by

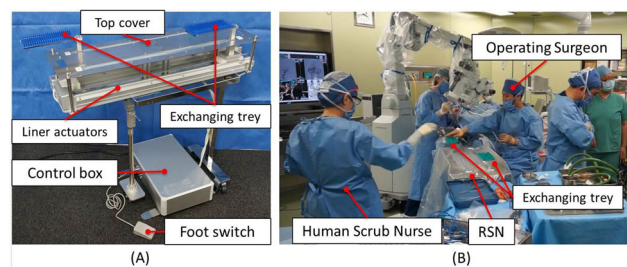


Fig. 1 **A** The first version of robotic scrub nurse assistant. A human scrub nurse pushes foot switch for exchanging instruments on treys. Two liner actuators are mounted under top cover. A sterilized plastic cover is overlaid on the top cover. **B** The first version of robotic scrub nurse assistant in clinical case. An operating surgeon and a human scrub nurse performed surgical instrument exchange with using newly installed robotic scrub nurse assistant

Table 1 Number of instrument exchanges by utilizing the first RSN

	Robotic Exchange	Human Scrub Nurse		
		Gauze	Non-eyes release	Others
First clinical case	34	7	0	8
Second clinical case	43	4	4	15

switching from liner actuators to rotating mechanism and instrument serving point is mechanically exact same position by sensing feedback. The size of whole components was designed as 500 mm diameter with same sized blue colored trays for exchanging on one degree of freedom for table rotating mechanism.

Conclusion

We proposed a robotic scrub nurse assistant (RSN) and designed its feasibility of the robot achieving surgical instruments exchange between an operating surgeon and a human scrub nurse. The first version of RSN experienced two clinical cases in neurosurgery and resulted positive feedback from clinical staffs. Our design with two liner stages is unique approach to perform effective and assists smooth exchanging without distracting current environment beside operating field as a robotic assistant. Ideas in discussion were taken into account for the improvement for the second version of RSN and we conclude this paper with the design on the second version of RSN.

Our approach may potentially overcome the limited utility of these conventional surgical procedures. This study warrants further investigation by fabrication of the real RSN and its mechanical and functional evaluations in clinical experiences. From the results of clinical usages, this presented work shows promising potential for being adapted as part of other surgical procedures and future advanced computer assisted operating theatre.

References

- [1] Miyawaki F, Masamune K, Suzuki S, Yoshimitsu K, Vain J (2005) Scrub nurse robot system—Intraoperative motion analysis of a scrub nurse and timed-automata-based model for surgery, *IEEE Trans. Ind. Electron.*, 52(5):1227–1235.
- [2] Kochan A (2005) Scalpel please, robot: Penelope's debut in the operating room. *Industrial Robot-an International Journal* 32 (6):449–451.

Construction of a 3D organ model using a robot, visual SLAM, and deep learning

T. Ishikawa¹, N. Koizumi¹, Y. Nishiyama¹, Z. Jiayi¹, Y. Watanabe¹, T. Fujibayashi¹, M. Matsuyama¹, M. Yamada¹, R. Tsumura², K. Yoshinaka², H. Tsukihara³, K. Numata⁴

¹The University of Electro-Communications, Graduate School of Informatics and Engineering, Chofu-shi, Tokyo, Japan ²National Institute of Advanced Industrial Science and Technology, Health and Medical Research Institute, Tsukuba, Japan ³The University of Tokyo, Department of Cardiothoracic Surgery, Bunkyo-ku, Tokyo, Japan ⁴Yokohama City University Medical Center, Gastroenterological Center, Yokohamashi, Kanagawa, Japan

Keywords 3D model, RUDS, Visual SLAM for Ultrasound, Segmentation

Purpose

As a diagnosis modality, ultrasound (US) is superior to MRI and CT in terms of radiation safety and flexibility. However, the quality of

ultrasound imaging depends on skill level of the operator. Ultrasonography by 2D imaging relies on the ability of the medical professional, to subjectively manipulate a 3D model of the anatomic and pathologic structures in their head [1]. Obtaining US images using a robot and constructing organ models of individual would reduce the burden on unskilled medical professionals and could be used in treatment planning, measurement of therapeutic response, preoperative simulation, surgical monitoring, and the process of communication between patients and medical professional (Informed consent). The purpose of this study was to construct a simple and convenient phantom kidney model suitable for probe localization estimation, by deep learning using a camera and organ segmentation.

Methods

The Robotics Ultrasound Diagnosis System (RUDS) [2] was applied to implement our method. RUDS comprises a tip component for fine alignment with the organs, a bed component for rough alignment with the patient, and a support component for tip movement (including sliding, fanning, compression and rotation). The tip has a spring escape mechanism because it is difficult to measure probe movement with the RUDS measuring instrument without excessive compression of the patient. For this reason, we used ORB-SLAM2, a Visual SLAM and base ORB feature, for estimation of probe localization, and performed real-scale estimation using a depth camera (Intel Realsense D435). Localization accuracy was improved using a calibration ChArUco board that was placed about 20 cm away from the camera. The accuracy of SLAM was evaluated as MAE (mean absolute error) and RMSE (root mean square error). The ground truth data were obtained by motion capture (OptiTrack Flex3, NaturalPoint, Inc.), during closed loop orbital motion of the probe in the direction of the body axis. This trial was performed 6 times.

Noise in localization data due to SLAM was removed using a moving average filter that averages the before and after data. The obtained US images were segmented by U-Net to extract the target organ. In learning, 1640 images were used in training, 174 images were used in validation, and 100 images were used in the test datasets. Comparisons were made using various encoders (EfficientNet-b0, EfficientNet-b5, ResNet50, and Vgg16) and the most accurate encoder was applied. IoU and Dice Loss evaluation index were used. The loss function was Dice Loss and Adaptive Moment Estimation (Adam) was used as the optimizer. A 3D model was constructed by combining the segmentation images with the corresponding estimated localization data. US images were obtained in the minor axis direction of the kidney along the major axis direction.

Results

The accuracy of probe localization estimation was within 1 mm for MAE and RMSE, and maximum error was 3.23 mm. The farther from the starting point, the greater the position estimation error due to the influence of accumulation error. Maximum error was 2.27 mm after removing noise. The use of a moving average filter improved estimation accuracy by reducing the impact of outliers on the estimates. This technique could identify the approximate location of the lesion and can be applied in diagnosis. In segmentation, IoU of Resnet was 0.9929, Dice Loss was 0.0037; Resnet had the highest precision among the four encoders. Figure 1 shows the constructed 3D kidney model.

Conclusion

Location estimation using only a camera and application of a noise filter was simple, convenient, accurate, and effective. Accuracy could be further improved by changing the features. As the probe was moved only in the body axis direction in this study, the system needs to be improved to incorporate tilting, rocking, and rotating motions and to control contact force between the probe and the patient's body.

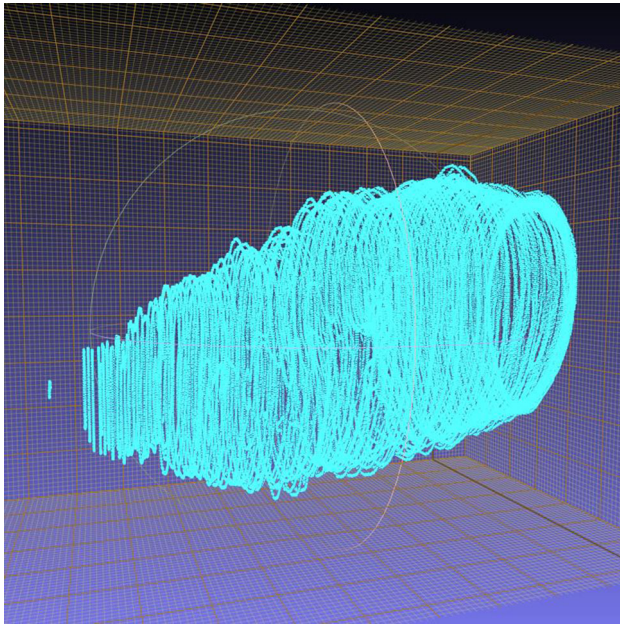


Fig. 1 3D model based on Visual SLAM

Table 1 Results of MAE, RMSE (Mean of 6 trials) and Maximum error by Visual SLAM

Noise Filter	Before	After
Maximum Error [mm]	3,23	2,27
MAE [mm]	0,57	0,53
RMSE [mm]	0,67	0,62

It is difficult to obtain a quality US image using a sliding motion because the body is not flat, and shadows are caused by areas of poor skin contact. Therefore, a system is required that avoids shadowing with tilting, rocking, and rotating motions. And these motions enable imaging of the major axis direction of the kidney along the minor axis direction. This method enables the construction of 3D model with little scatter. This study used a phantom as the target. In future, we aim to test this method on a real body as the target, and to develop for a clinical setting.

References

- [1] Kim SH, Choi BI. Three-dimensional and Four-dimensional Ultrasound: Techniques and Abdominal Applications. *J Med Ultrasound* 2007, Vol.15, No.4
- [2] Matsuyama M, Koizumi N, Otsuka A, Kobayashi K, Yagasaki S, Watanabe Y, Zhou J, Nishiyama Y, Matsumoto N, Tsukihara H, Numata K (2022) A novel complementation method of an acoustic shadow region utilizing a convolutional neural network for ultrasound-guided therapy. *Int J Comput Assist Radiol Surg.* Jan;17(1):107–119.

More accurate cerebral aneurysm clipping using preoperative ultra-high-resolution computed tomography and computational fluid dynamic analysis

H. Kimura¹, K. Hayashi², Y. Fujimoto¹, S. Osaki², A. Tomiyama², T. Sasayama¹

¹Kobe University Graduate School of Medicine, Department of Neurosurgery, Kobe, Japan ²Kobe University Graduate School of Engineering, Department of Mechanical Engineering, Kobe, Japan

Keywords Computational fluid dynamics, Computed tomography venography, Cerebral aneurysm, Simulation

Purpose

The common denominator in all neurosurgical procedures is to perform detailed and accurate preoperative simulations as much as possible before the operation and not to take potluck in the actual operation. Although clipping of cerebral aneurysms seems to be a well-established surgical technique, there have been several uncertainties to accomplish safe surgery. First, in approaching the aneurysm, we need to preserve not only cerebral arteries but also the venous drainage system. The latter could have not been precisely evaluated preoperatively even if a rotational angiography was performed. Second, in exposing the aneurysm, we need to dissect the surrounding brain tissue and adherent arteries from the aneurysm. When the aneurysm wall would consist of a thin wall, the manipulation could cause a premature rupture. Especially, at the blind area or in the deep surgical field, the thin-walled area can be unrecognized even in the microscopic inspection where the surgical procedure would cause a devastating outcome. Furthermore, aneurysms harboring partially thick-walled regions sometimes could cause unintentional failure to occlude the aneurysm lumen and also a thromboembolic complication in clipping. To solve these matters, we introduced ultra-high-resolution three-dimensional computed tomography angiography and venography (UHR- 3D CTA/V) to visualize the three-dimensional venous drainage system. Moreover, we introduced computational fluid dynamic (CFD) analysis to predict the aneurysm wall properties. We had reported that one of the popular parameters, oscillatory shear index (OSI), could detect the aneurysm wall properties [1]. The low OSI areas in the aneurysm wall corresponded to the thin-walled regions. On the other hand, it has been reported that high OSI areas in the aneurysm wall corresponded to thick atherosclerotic walls [2]. We have applied these technologies and findings to simulate preoperative cerebral aneurysm clipping and achieved more accurate surgery. In this presentation, we demonstrate its efficacy and the outcome.

Methods

We performed 35 unruptured cerebral aneurysm surgeries from January 1, 2019, to December 31, 2021, which were included in this study. Contrast-enhanced volume data for UHR- 3D CTA/V were acquired with a 160-detector row UHR-CT scanner (Aquilion Precision) using helical scanning. For CFD analysis, raw data from CTA were transferred to an image processing workstation (Ziostation 2; Ziosoft Inc., Tokyo, Japan) for further image processing and analysis. In the Ziostation, Hemoscope Ver1.5 (EBM Corp.) was used for

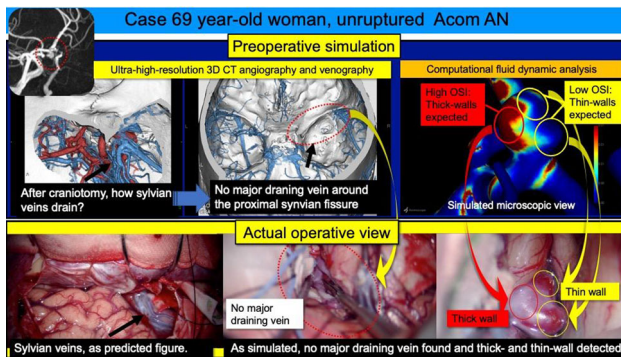


Fig. 1 Preoperative simulation and actual microscopic views in a case underwent clipping for an anterior communicating artery aneurysm. (Acom AN, Anterior communicating artery aneurysm; OSI, oscillatory shear index; 3D-CT, three-dimensional computed tomography)

meshing and a hemodynamic analysis was performed, including visualization of the results including OSI color maps. Using UHR- 3D CTA/V images and OSI color maps, the optimal surgical approach routes and the way to expose the aneurysm dome and apply the aneurysm clips properly were simulated precisely, Fig. 1.

Results

13 were males and 22 were females with an average age of 62.4 years. There were 26 middle cerebral artery aneurysms, 10 internal carotid artery aneurysms, 10 anterior communicating artery aneurysms, and 2 posterior circulation aneurysms. In all cases, a trans-Sylvian approach was selected. When dissecting the estimated thin-walled area of the aneurysm wall, especially at the blind area or in the deep surgical field, the manipulation was performed more carefully, and when the estimated thick-walled areas of the aneurysm walls were going to be clipped, the possibility of incomplete closure and the occurrence of an embolic complication was taken into consideration. We were able to perform a Sylvian fissure dissection at a simulated plane in all cases and close the aneurysm lumen without intraoperative rupture. There were no complications associated with the procedure and no cases with decreased modified Rankin Scale. A representative case is demonstrated in the figure.

Conclusion

UHR- 3D CTA/V and CFD analysis provided more accurate cerebral aneurysm clipping by enabling preoperative simulation for an approach route and prediction of aneurysm wall properties. Further accumulation of surgical cases with this method will demonstrate the efficacy in detail.

References

- [1] Kimura H, Osaki S, Hayashi K, Taniguchi M, Fujita Y, Seta T, Tomiyama A, Sasayama T, Kohmura E (2021) Newly identified hemodynamic parameter to predict thin-walled regions of unruptured cerebral aneurysms using computational fluid dynamics analysis. *World Neurosurgery* 152:e377–e386.
- [2] Furukawa K, Ishida F, Tsuji M, Miura Y, Kashimoto T, Shiba M, Tanemura H, Umeda Y, Sano T, Yasuda R, Shimosaka S, Suzuki H (2018) Hemodynamic characteristics of hyperplastic remodeling lesions in cerebral aneurysms. *PLoS One* 13(1):e0191287.

A deformable image-based registration approach to obtain shape correspondence for statistical shape modeling of finger bones

A. Lange¹, S. Heldmann¹, J. H. Moltz², L. Walczak^{3,4}, C. Engel², M. Detering², A. Rörich², F. Güttler⁵, S. Yasar-Schlickewei⁶, J. Georgii²

¹Fraunhofer Institute for Digital Medicine MEVIS, Lübeck, Germany

²Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany

³Fraunhofer Institute for Digital Medicine MEVIS, Berlin, Germany

⁴Charité—Universitätsmedizin Berlin, Berlin, Germany

⁵Jena University Hospital—Friedrich Schiller University, Institute

of Diagnostic and Interventional Radiology, Jena, Germany

⁶Universitätsklinikum Hamburg-Eppendorf, Klinik und Poliklinik für

Unfallchirurgie und Orthopädie, Hamburg, Germany

Keywords statistical shape model, shape correspondence, deformable image registration, orthopedics

Purpose

Statistical shape models (SSMs) are a state-of-the-art approach to encode complex anatomical shape information. Here we focus on modelling finger bones to support orthopedic applications such as joint replacements and design of implants. A necessary prerequisite for the computation of a SSM is that all shape data have corresponding descriptions and are parameterized on a common reference space. In this work we present a deformable image-registration approach to obtain this shape correspondences.

Methods

Our approach is based on CT images of hands and segmentations of the metacarpals and phalanges. We use 200 CT data sets from clinical routine acquired at two different clinical sites. The metacarpal, proximal, intermediate, and distal bones are segmented with an nnU-Net. The ground truth segmentations for the training of the nnU-Net were created manually and the final segmentation results are reviewed by a radiology technician and corrected if necessary. From the segmentations surface meshes are generated with the Marching Cubes method and stored as Winged Edge Meshes (WEMs). All processing is done in MeVisLab (www.mevislab.de).

First, we determine a reference hand that contains all the phalanges and has a low flexion of the fingers. Since the CT images are not normalized in their orientation and bending, the next step is to perform an initial alignment of each bone. To do this, we calculate a local coordinate system for each bone. The local bone coordinate system is based on principal component analysis (PCA) and the center of gravity of each bone. For consistent axis orientation, the relation of the bones of a single finger is analyzed. We heuristically derive a right-handed, orthonormal local coordinate system including the long axis of the finger (pointing to distal) and the axis of rotation of the finger joint. The third axis is determined via cross product. We then transform all the bones into the local coordinate system of the respective reference bones to achieve a good initial alignment of the bones.

Subsequently, we perform rigid prealignment of bone masks, volumetric deformable image registration and finally non-rigid surface registration separately for each finger bone. The deformable registration is similar to the variational approach presented in [1] but we consider an enhanced objective function directly taking the

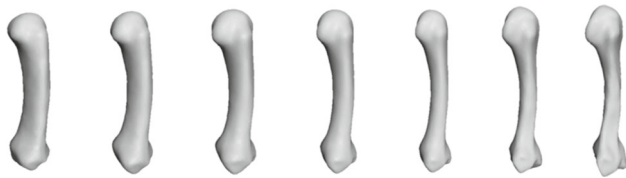


Fig. 1 Variation of SSMs for single bone (Fig. 1) and complete finger (Fig. 2). The average SSM is shown in the middle, next to it the variation of the first mode, followed by the simultaneous variation of the first 5 and 10 (only Fig. 1) modes

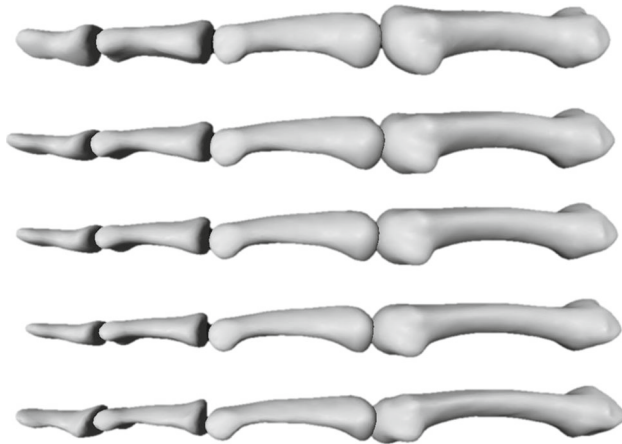


Fig. 2 Complete finger

surface distance between the CT images of the respective bone into account. Therefore we consider additional penalty terms measuring the distance between the segmentation masks and the distance of the deformed reference WEM to the bone segmentation surface. After the deformable registration the deformation field is used to transform the WEM of the reference bone to the template bone. To better fit the propagated WEM to the template surface we run an additional deformable registration between those. We use the Sum-of-squared differences as distance measure and a diffusive regularization for the deformable registrations.

The result of the registration pipeline is the deformed reference WEM, so all individual bones of the same structure now have (approximate) corresponding surface meshes. Now that the correspondence problem has been solved, SSMs for individual bones can be calculated using the familiar steps of Procrustes Analysis and PCA. We do not want the bending of the fingers to be included in the shape variation, so a further step is required to calculate SSMs for complete fingers. After transforming the bones back to their original coordinate system we run a position-based dynamics simulation [2] where the fingers are moved to a normalized position in which they are fully stretched. To achieve this, the simulation engine described by [2] was extended with rigid body dynamics, joint constraints, and a kinematic chain mechanism. For stretching, proximal, intermediate, and distal bones are pulled in the direction of the metacarpal bone long axis (calculated during the initial alignment step described above). The individual fingers are now stored as a single WEM, and the SSMs can be computed as for a single bone.

Results

The Figs. 1 and 2 show the mode variation of the resulting statistical shape model, exemplified for the metacarpal of the index finger and the entire index finger with metacarpal. The SSMs were created using

about 100 different fingers. Fingers with incomplete structures (due to too small a field of view or broken bones) were discarded, so the number of fingers used varies for each SSM. Shape models with significantly more data sets are in preparation, as well as a more thorough analysis and validation of the resulting shape models. The registration pipeline is fundamental for the determination of shape correspondences and thus has a major impact on the quality of SSMs. We will use the local coordinate systems of each bone to calculate cylinder coordinates to evaluate the registration results.

Conclusion

We presented the first results of a deformable image-based registration approach to obtain shape correspondences for SSMs of phalanges and metacarpals. The use of additional objective function terms and the additional registration between deformed WEM and distance map improves the registration results and thus leads to good shape correspondences. With the help of simulation, SSMs can be generated not only from individual bones, but also from larger structures such as entire fingers with metacarpals. These SSMs can now be used to generate new realistic instances.

Acknowledgement: This work was supported by the Fraunhofer Internal Programs under Grant No. PREPARE 840226.

References

- [1] Rhaak, J., Heldmann, S., Kipshagen, T., and Fischer, B. (2013). Highly accurate fast lung ct registration. In SPIE Medical Imaging 2013: Image Processing, Lake Buena Vista, Florida, USA.
- [2] Walczak, L., Georgii, J., Tautz, L., Neugebauer, M., Wamala, I., Sndermann, S., Falk, V., Hennemuth, A., Using Position-Based Dynamics for Simulating Mitral Valve Closure and Repair Procedures, Computer Graphics Forum, 2021, <https://doi.org/10.1111/cgf.14434>

Age prediction from pelvis bone shape using large-scale CT database based on geometric deep learning

B. Zhang¹, Y. Otake¹, S. Mazen¹, K. Uemura², M. Takao², K. Aida³, S. Satoh³, M. Hashimoto⁴, T. Akashi⁵, N. Sugano², Y. Sato¹

¹Nara Institute of Science and Technology, Information Science, Ikoma, Japan ²Osaka University, Osaka, Japan ³National Institute of Informatics, Tokyo, Japan ⁴Keio University, Tokyo, Japan ⁵Juntendo University, Tokyo, Japan

Keywords age estimation, geometric deep learning, Dynamic GCNN, large scale data

Purpose

It is well known that there are some pelvic geometric features characteristic of age, such as the pelvic sagittal inclination. We aim to understand the relationships between the pelvis geometry and age from two aspects, (1) age prediction from the pelvis shape in the areas such as forensic science and criminology, (2) prediction of the degree of aging from the pelvis geometry for health science applications. Although the previous studies using a conventional method like Random Forest (RF) have shown promising results in age prediction from skeletal structures, the rapid development of a geometric deep learning (DL) provides another approach to extract its geometric features. This research compared the performance of the conventional method with recent representative geometric neural network, DGCNN [1] in age prediction tasks. The experiment data contains 32,926 pelvis geometry data, automatically segmented from CT images by a pre-trained neural network. The purpose of this study was

to investigate the possibility of using geometric DL to analyze the pelvis geometric features.

Methods

We addressed the problem of the prediction of age from the pelvis geometry. Two methods includes one conventional method RF and one geometric DL method DGCNN were used. (1) RF (baseline): All vertex coordinates (about 60,000 dimensions, i.e., 20,000 vertices) from pelvis mesh polygon data, which represent the pelvis geometry, were reduced to 100 dimensions using principal component analysis (PCA), then a random forest regression was applied. (2) DGCNN: 2048 vertices in the pelvis mesh were used as input point set. All pelvis meshes in our experiments are generated from auto segmentation from CT images.

Results

All results in our experiments were got from twofold cross validation. First, all 32,926 cases have been used to test the performance of DGCNN compared with conventional RF. As shown in Fig. 1, the mean absolute error (MAE) of DGCNN's results (after 20 h training on RTX-TITAN for each fold) shown a lower MAE (6.01 years) than PCA/RF's (6.94 years). The intraclass correlation coefficient (ICC) of DGCNN's results (0.828) is also higher than PCA/RF's (0.752). In the box plot below, another finding is that the age prediction shows better results in the young group, but worse results in the elder group, which may suggest that pelvis age features are more evident at the young group. An experiment of the age prediction on gender-balanced dataset (i.e., the number of cases aligned with 1-year bin) was also conducted to compare the results of the female pelvis and male pelvis, whose results shows better MAE and ICC in predicting female pelvis age, which indicates pelvis age feature in female is more evident than in male. Furthermore, experiments of reducing training data were included in our research. As shown in Fig. 1 and Table 1, we tested the performance of DGCNN with 10,000 (mae 6.83 years), 3000 (mae 7.84 years), 1000 (mae 8.95 years) cases whose results show the importance of sufficient training data. Also, compared with other age estimation research from bones (vertebral body (best mae 8.2 years with about 700 cases using tenfold cross validation), ischial tuberosity (mae 8.6 years), iliac crest (9.4 years), femur) in [2], our results of small training data experiments is acceptable.

Table 1 Results of different methods and number of cases. DGCNN(Female/Male)-Using only Female/Male data as train set

Method(condition)	MAE(year)	ICC
RF (32,926 cases)	6,94	0,75
DGCNN (32,926 cases)	6,01	0,83
DGCNN (female: 17,851 cases)	5,94	0,85
DGCNN (male: 15,075 cases)	6,12	0,75
DGCNN (10,000 cases)	6,83	0,83
DGCNN (3,000 cases)	7,84	0,69
DGCNN (1,000 cases)	8,95	0,61

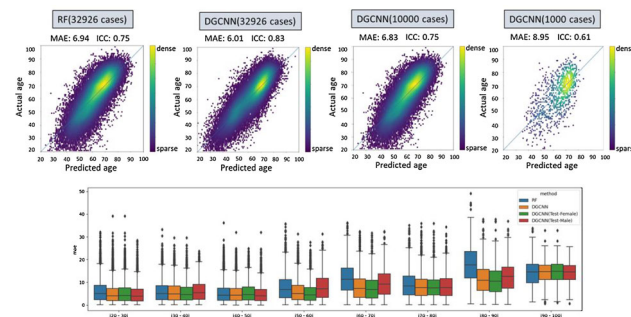


Fig. 1 Scatter plot and box plot of comparison between RF and DGCNN in age prediction task

Conclusion

In our experiments, we evaluated the performance of the geometric neural network in the analysis of geometric features in age prediction from the pelvis shape, which shows great ability of geometric neural network in dealing such tasks. Straightforward future work includes generation of a saliency map to visualize features extracted by the geometric neural network which could be compared to manually extracted features.

References

- [1] Phan, A. V., Le Nguyen, M., Nguyen, Y. L. H., & Bui, L. T. (2018). Dgcn: A convolutional neural network over large-scale labeled graphs. *Neural Networks*, 108, 533–543.
- [2] Imaizumi, K., Usui, S., Taniguchi, K., Ogawa, Y., Nagata, T., Kaga, K., ... & Shiotani, S. (2021). Development of an age estimation method for bones based on machine learning using post-mortem computed tomography images of bones. *Forensic Imaging*, 26, 200477.

Robots in healthcare: now and in the future

D. Wilhelm¹

¹Klinikum rechts der Isar, Technical University Munich, Surgery, München, Germany

Keywords healthcare system, robots, autonomy, personnel shortage

Purpose

Robotic systems are increasingly applied in healthcare (HC) but are confined to heavy load tasks (e.g. within a hybrid OR), are used for undemanding services (e.g. transport and supply) or as master–slave systems aim at increasing the precision of interventional procedures. Only minor they have become substitutes of medical personnel, only minor they have improved the quality of health care delivery and only minor they have truly been integrated in our clinics. The future health system is facing some critical problems, with the shortage of personnel and the maintenance of the quality of care being first in line. Robots offer quite attractive features to cover with these problems but need to be designed accordingly, have to provide autonomous tasks and have to become full team members. The article aims at the identification of weak points of the health care system and how robots can be used to shape its future.

Methods

The results and thoughts presented herein do originate from expert discussions and studies of the available literature, but do also originate from experiences made in course in daily practice. Also, aspects which were elaborated during the work on the patient hub concept [1] and have been debated in panel discussion on the OR of the future and on robots in healthcare are included. Still, the presented theses are speculative and visionary and thus cannot be based on a fully scientific background.

Results

Personnel shortage

The most pressing challenge we are facing for the healthcare system is the shortage of personal, which became even more obvious during the COVID-19 pandemic. As it foreseeable, that we will not be able to replace missing workers by human personnel, care delivery must become less human depending and missing work craft has to be replaced by autonomous systems. Autonomous robotic systems represent a core technology in this respect and can help to take over simple and repetitive tasks, e.g. for the handling of medical goods, for bedding and mobilizing patients and rehabilitation.

Climate change

The HC system will also be affected by the warming of the atmosphere, as it is responsible for almost 5% of CO² emissions. Transport and delivery of medical goods in this regard are the main contributors and could be optimized by reducing the rate of single-use devices and scaling down supply chains. Increasing the in-hospital sterilization capacities and implementing local fabrication facilities of medical devices might offer a solution here, however, would require human resources. Robots again can play a decisive role here and become an enabling technology, e.g. during the reprocessing of sterile goods and for 3D printing based manufacturing lines.

Demographic change

The aging of the population is becoming a relevant burden for society due to the increasing number of disabled people and people in need of care. Since families and the HC system cannot cope with this development, solutions must be found that support the independence and self-subsistence of the elderly. Care robots, mechatronic extra-skeletons and smart assistive technologies for the home are key elements for caring for elderly people in a way that is gentle on staff and can also help to maintain their quality of life.

Specialization and precision medicine

The healthcare system is driven by striving for improved quality of service and precision medicine. Currently available systems, mainly master–slave devices have failed in contributing here as no superiority has been shown for robotic assisted surgeries so far. Nevertheless, robots are the most powerful solution for further reducing the access trauma, for miniaturizing devices and for the realization for autonomous capabilities by coupling with smart imaging solutions. As it was demonstrated with OCT-based micro-robotic solutions for eye surgery, comparable solutions might be a driving technology for example for endovascular surgeries, brain surgery and endoscopic interventions. While surgeons become more and more specialized which makes their individual performance of high value, assistive systems to take over less-demanding tasks (e.g. skin suturing, retraction, suction) could become a meaningful and resource-sparing aid and once again could be realized by robotic solutions.

Multi-drug resistance and isolation

As observed by patients suffering from multi-drug resistance even before the current pandemic an increasing number of patients are requiring isolated care. The isolated care is not only demanding in terms of personnel, but also produces enormous amounts of waste, which have a negative effect on CO² emissions. Robots again offer here a valuable solution as they can remain in an isolated environment, as they can be disinfected which makes additional protective measures unnecessary and as they strictly follow to standard operative procedures thereby reducing the risk of unintended contamination.

Conclusion

Solutions to overcome the pending, or already present challenges in the HC system are urgently required and must provide autonomic functionalities to save personnel, have to reduce the amount of waste and HC related traffic to lower the CO² emission and should enable us to develop smarter and less invasive approaches for the treatment of an increasing number of sick and care-dependent patients. Numerous robotic solutions to cope with this problems have already been introduced [2], but need to be further adapted according to these requirements and fully integrated into a cooperative environment. The alignment between human and robotic tasks and the maintenance of ethical and legal aspects still have to be taken as unsolved problems

for the further involvement of robots, however when solved could open up the basis for a highly efficient patient centred HC system.

References

- [1] Amato C, McCanne L, Yang C, Ostler D, Ratib O, Wilhelm D, Bernhard L: The hospital of the future: rethinking architectural design to enable new patient-centered treatment concepts. *International Journal of Computer Assisted Radiology and Surgery* 2021:1–11.
- [2] Cresswell K, Cunningham-Burley S, Sheikh A: Health care robotics: qualitative exploration of key challenges and future directions. *Journal of medical Internet research* 2018, **20**(7):e10410.

Toward fully automated robotic auscultation platform with LiDAR camera-based registration

R. Tsumura¹, Y. Koseki¹, N. Nitta¹, K. Yoshinaka¹

¹National Institute of Advanced Industrial Science and Technology, Health and Medical Research Institute, Ibaraki, Japan

Keywords Auscultation, Registration, Point cloud, Force control

Purpose

Since most developed countries are facing an increase in the number of patients per healthcare worker due to a declining birth rate and an aging population, relatively simple and safe diagnosis tasks may need to be performed using robotics and automation technologies, without specialists and hospitals. Since the 1800s, auscultation has been an essential component of clinical examination and is a highly cost-effective screening tool to detect abnormal clinical signs [1]. Additionally, recent studies have reported that auscultation is a potential diagnostic tool for COVID-19 patients and can be used as a follow-up tool for noncritical COVID-19 patients [2]. In this study, we aim to develop a robotic auscultation platform that enables estimation of the landing positions and safe placement of the stethoscope at the estimated position. The contribution of this paper is to establish a proof-of-concept of the robotic platform that enables autonomous positioning of the stethoscope based on external body information while satisfying the patient's safety in terms of the contact between the stethoscope and body surface. To the best of our knowledge, this is the first dedicated robotic system designed for autonomous auscultation.

Methods

The developed robotic platform is composed of a 6-degree-of-freedom cooperative robotic arm, the light detection and ranging (LiDAR), and a spring-based mechanism holding an electric stethoscope (Fig. 1). The platform enables autonomous stethoscope positioning based on external body information acquired using LiDAR camera-based multi-way registration. The platform also ensures safe and flexible contact, maintaining the contact force within a certain range through the passive-actuated mechanism. The pipeline for estimating the landing positions to place the stethoscope with the developed robotic auscultation platform is organized into three components: (i) acquisition of the point cloud data for covering the entire chest and registration of the acquired point cloud data to reconstruct the entire chest shape; (ii) estimation of the landing positions based on the reconstructed body shape and the anatomical

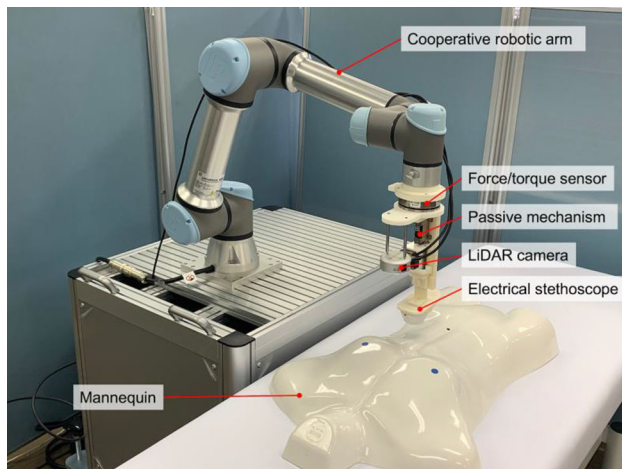


Fig. 1 System overview of robotic auscultation platform

landmarks on the body surface; (iii) placement of the stethoscope at the estimated positions while maintaining a certain safe contact force.

Results

Our preliminary results confirm that the robotic platform enables estimation of the landing positions required for cardiac examinations based on the depth and landmark information of the body surface. The registration error in the 3D space occurred in the range of 5.1 to 7.6 mm on average. It also handles the stethoscope while maintaining the contact force without relying on the push-in displacement by the robotic arm. The generated contact forces were precisely achieved to the targeted forces (5, 10, 15 N). The maximum error was 7.2% of the targeted force.

Conclusion

The developed robotic platform enables the estimation of the landing positions and handling the stethoscope while maintaining the contact force, which promises the potential of automatic remote auscultation. The developed robotic platform has the potential to address the critical issue of the increase in the number of patients per healthcare worker. The use of this technology may further enhance the efficiency of screening for abnormal clinical signs, including COVID-19.

References

- [1] Pereira D, Castro A, Gomes P, Neves Cunha Areias JC, Silveira Nogueira Reis Z, Tavares Coimbra M, Cruz-Correia R (2016) Digital Auscultation: Challenges and Perspectives. *Encycl E-Health Telemed*: 910–927.
- [2] Raj V, Renjini A, Swapna MS, Sreejyothi S, Sankararaman S (2020) Nonlinear time series and principal component analyses: Potential diagnostic tools for COVID-19 auscultation. *Chaos, Solitons and Fractals* 140: 1–8.

Digital transformation of radiology service in the era of artificial intelligence

S. Mun¹, K. Wong¹, S. C. Lo¹

¹Virginia Tech, Arlington Innovation Center—Health Research, Arlington, USA

Keywords Digital Transformatio, Artificial Intelligence, Workflow, Productivity

Purpose

Modern Radiology is a result of digital transformation over the past 40 years. The transformation has several phases to help radiology services evolve into new clinical, operational models. All radiology imaging devices became digital systems, and the picture archiving and communication system (PACS) now manages all images and workflow in and out of the department. The digital imaging community began to experiment with machine learning (ML) and artificial intelligence (AI) tools to improve diagnostic processes in radiology. Computer aided diagnosis (CAD) was the first attempt of AI in radiological imaging. Today there are more than 100 FDA-approved imaging AI products. [Tadavarthi 2020] However, the adoption of the AI products has been low, and some fear that radiology is facing an “artificial intelligence winter”. This paper reviews the adoption of AI technology from radiology’s digital transformation and offers possible trajectories of making AI meaningfully intelligent for radiology.

Methods

Phase I: During the ’70 s and ’80 s, most radiological imaging devices became digital, and this conversion opened the possibility of managing, displaying, and manipulating digital images on high-resolution displays. During the late ’90s, the US military medical community promoted the concept of filmless radiology and teleradiology to reduce the logistical challenges of managing film-based medicine. Several manufactures adopted the concept, and prototype systems were installed at Samsung Medical Center in Korea, Veterans Hospital in Baltimore, and Madigan Army Hospital in Seattle. These various early users developed many different justifications based on their local business environment. As expected, digital imaging and PACS technology faced many obstacles early from potential users, the imaging device industry, and film industries. Phase II. Digital imaging fueled the diffusion of imaging technology to many specialty areas, such as radiation oncology, surgical planning, and robotic surgery. Teleradiology removed the time and distance barriers of diagnosis and became a model for global telemedicine. The PACS network became the digital hub managing the flow of images within the department and the entire hospital. PACS also affected the interaction between radiologists, referring physicians, and teaching residents. As a result, teleradiology became a common practice, especially in the US. The development of the DICOM standard, despite early resistance from many imaging device manufacturers, laid the critically essential and powerful foundation to allow the digital transformation of radiology based on ML/AI of big data. Phase III: The investments that the radiology community is making in ML/AI indicates that there would be three parallel trajectories, Fig. 33, for the meaningful roles of AI as a part of digital transformation toward efficient precision medicine.

Results

Track III-A; The concept of CAD research in medical imaging has evolved into two distinct clinical applications; computer-aided diagnosis (CADx) and computer-aided detection (CADe). During the early development period, the community became aware of the importance of large image data sets, difficulties dealing with variable image quality, scalability, high cost of image labeling, and generalizability caused by bias. The performance of CADx was generally poor, and though some CADe applications improved the reading efficiency but increased false positives. Generally, many current AI tools do not offer significant benefits to the uses worthy of costs. As AI tools and research infrastructure improve, we expect to see more robust CAD products. Much of the work has now evolved into quantitative imaging and radiomics. Track III-B Create New Knowledge and Insights Radiomics (also known as quantitative imaging) attempts to extract additional information from radiology images using powerful analysis techniques and more sophisticated AI tools. Radiomics provides possibilities of identifying predictive and prognostic imaging biomarkers from the images. Such possibilities

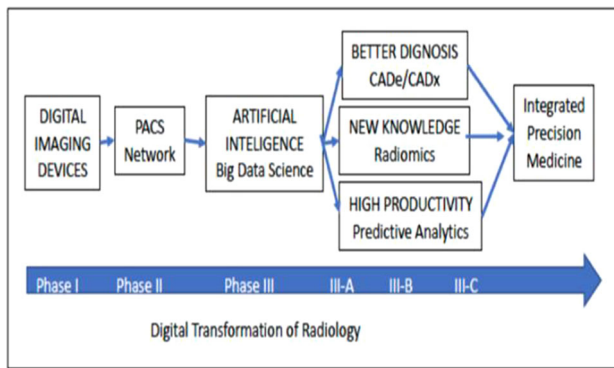


Fig. 1 Three parallel trajectories for ML/AI in radiology

could have a powerful impact on the patient care process within radiology and pathology. The tools from such research will also improve the performance of CADx. Radiomic will have to address many similar challenges that the CAD community faced, such as image quality, reproducibility, and suitable images that represent a realistic clinical environment. Lack of standardization and harmonization in data quality and analytical tools and terminologies will remain challenges for clinical adoption. Track III-C: Improve the Productivity of Radiology Service The PACS-based digital operations have allowed many innovations in radiology. However, much of the radiologist's work remains manual operations. The interpretation (reading) time varies greatly depending on the type of study. Radiological reading times have been steadily increasing as modern imaging systems generate an increasingly larger volume of images per study. The AI can provide powerful tools for predictive analytics [Choy2021]. There is an increasing awareness of the importance of productivity improvement and workflow optimization by the PACS vendors and academics. There are many data points with the PACS network and hospital EHR from which workflow data can be collected for predictive analysis and optimization.

Conclusion

It is conceivable that 20–30 AI tools must be integrated into different parts of the radiology workflow, such as imaging devices, radiology information system (RIS), radiology workstations, and hospital information systems, and PACS. Such integration may require the use of next-generation image management and communication system (IMAC). Instead of technology push, we should focus on offering meaningful solutions for radiology facing increasing financial and efficiency pressure as a core of the digital transformation.

References

- [1] Tadavarthi Y, Vey B, Krupinski E, Prater A, Gichoya J, Safdar N, Trivedi H (2020). The State of Radiology AI: Considerations for Purchase Decisions and Current Market Offerings. *Radiology: Artificial Intelligence* 2020 2(6):e200004. <https://doi.org/10.1148/ryai.2020200004>
- [2] Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, Geis JR, Pandharipande PV, Brink JA., Dreyer KJ (2018) Current applications and future impact of machine learning in radiology. *RADIOLOGY*, 288(2), 318–328. <https://doi.org/10.1148/radiol.2018171820>

How can we build surgery globally? Two very different models with impressive results

R. Andrews¹

¹NASA Ames Research Center, Los Gatos, United States

Keywords global surgery, entrepreneurial neurosurgery, UN & WHO goals for 2030, LMIC healthcare

Purpose

It has been documented by the 2015 Lancet Commission on Global Surgery and other reports that in addition to the humanitarian costs of lack of surgery (one-third of all deaths globally—more than four times the deaths due to HIV/AIDS, malaria, and tuberculosis combined) the economic costs are also staggering (already over US \$500 billion in 2021 and projected to be over US \$1.5 trillion by 2030).

In 2015 the United Nations (UN) issued 17 Sustainable Development Goals (SDGs) for 2030, Goal #3 being “Ensure healthy lives and promote well-being for all at all ages“. Improved global surgery is essential to achieve SDG #3. To that end, the World Health Organization (WHO) and the Harvard Program for Global Surgery and Social Change has developed the National Surgery, Obstetrics, and Anesthesia Plan (NSOAP) to provide goals for surgery by 2030.

The question is, “How do we build surgery worldwide to meet the goals for 2030 set by the UN and the WHO?“.

Methods

The challenges to develop surgery faced by low- and middle-income countries (LMICs) require entrepreneurial solutions that are cost-effective and tailored to the particular circumstances of the country or region involved. Two examples of very different models to provide and improve neurosurgery resources are described. Establishing neurosurgery ensures that the resources are available to address surgery needs far beyond neurosurgery alone: anesthesia, intensive care, radiology, laboratory and blood bank facilities. Less obvious but equally important aspects of full-service neurosurgery include pre-hospital (ambulance and paramedic) care and rehabilitation services.

The two examples described employ different strategies—tailored to their contrasting situations—but both demonstrate the marked benefit for surgery that one neurosurgeon who is dedicated and entrepreneurial can bring about [1, 2]. A common theme between the two examples is that the improvement has taken place in the private sector, remarkable given both examples have markedly advanced the resources for their respective indigent populations.

Results

Peshawar and the Khyber Pakhtunkhwa (KP) province, Pakistan In 1990, following neurosurgical training in Ireland, Tariq Khan returned to his hometown, Peshawar. He became the third neurosurgeon in the Khyber Pakhtunkhwa (KP) province, where the neurosurgeon-to-population ratio in the province was even less than that for Pakistan as a whole: less than 1:5,000,000. In the 1990s several advances were made: (1) with the large number of neuro-trauma patients, he realized prevention was both effective and feasible given the limited infrastructure resources at the time: community education programs for trauma prevention were created; (2) with the International Committee of the Red Cross, he helped establish a rehabilitation unit for spinal cord injury patients in Peshawar; (3) he received approval for a neurosurgery residency program. In the early 2000s he convinced colleagues that building a state-of-the art hospital in Peshawar was feasible. He and his colleagues convinced a bank to provide funds for their first hospital, full-

service Northwest General Hospital, which opened in 2009. Over the next decade he and his colleagues accomplished the following: (1) residency training in all medical/surgical specialties was approved; (2) Schools of Medicine (100 students per year initially, expanded to 150 recently) and Nursing (50 students per year) as well as training programs in allied health professions, including Physical Therapy, were established; (3) a second 300-bed hospital was opened; (4) a fully-equipped ambulance service (including trained paramedics) was begun. In addition, the Peshawar Chapter of the ThinkFirst Head Injury Prevention Program received the International Chapter of the Year award in 2019, celebrated with a community program involving 1000 participants.

Indonesia The situation faced by Eka Wahjoepramono was quite different than that faced by Tariq Khan. The deficit in neurosurgery stretched across an archipelago of 17,000 islands over 5,000 km east to west. His neurosurgery program—based in a Siloam Hospital in suburban Jakarta—began in 1996, when neurosurgery resources were very basic and limited to a few other cities throughout the country. The initial challenge was raising awareness of their neurosurgery resources in the population in as much of Indonesia as possible. This was addressed by various means—seminars, media events, and inviting international neurosurgery experts to assist in developing the local expertise. Additional neurosurgeons were recruited to his department to provide skilled subspecialty service and the Indonesia Brain Foundation was created to help fund surgery for indigent patients. When the Siloam Hospitals group expanded to over 40 hospitals throughout Indonesia in the mid 2000s, the opportunity to expand neurosurgery throughout the country was not lost. Neurosurgeons and neurosurgeons-in-training in Jakarta (who came from all over the country) were recruited to serve in a hospital in their home region—a win–win for both the neurosurgeon and the population in their home region. By 2021 the nationwide count of full-time neurosurgeons reached 28, with senior members traveling wherever needed in the system to mentor junior colleagues faced with challenging cases. The distribution of Siloam Hospitals with neurosurgery in 2021 is provided in Fig. 1. Through an entrepreneurial and inclusive approach to neurosurgery—both among the neurosurgeons and the hospital stakeholders—in a couple of decades dramatic progress in providing neurosurgery for the widely dispersed population of Indonesia has been achieved.

Conclusion

Where healthcare infrastructure is rudimentary (as in Peshawar and KP province), a ground-up approach is appropriate. Where healthcare infrastructure is present (as in the Siloam Hospital system in Indonesia), a cost effective, diplomatic approach can provide rapid expansion of quality care. An innovative, entrepreneurial, and

inclusive approach can expand not only neurosurgery resources but surgery in general (as well as overall healthcare) in LMICs.

References

- [1] Khan TM (2022). Building Both Neurosurgery and Healthcare in Peshawar Pakistan. World Federation of Neurosurgical Societies Newsletter—to appear online February 2022. Available at www.wfns.org.
- [2] Wahjoepramono E (2021). Role of Entrepreneurship in Building Global Neurosurgery: Indonesia. World Federation of Neurosurgical Societies Newsletter—online December 2021. Available at www.wfns.org.

3D bronchus anatomical structure measurement on real bronchoscopic images based on depth images estimated by deep neural network

C. Wang¹, Y. Hayashi¹, M. Oda^{1,2}, T. Kitasaka³, H. Takabatake⁴, M. Mori⁵, H. Honma⁶, H. Natori⁷, K. Mori^{1,8,9}

¹Nagoya University, Graduate School of Informatics, Nagoya, Japan ²Nagoya University, Information and Communications, Information Strategy Office, Nagoya, Japan ³Aichi Institute of Technology, School of Information Science, Toyota, Japan ⁴Sapporo Minami-Sanjo Hospital, Department of Respiratory Medicine, Sapporo, Japan ⁵Sapporo-Kosei General Hospital, Department of Respiratory Medicine, Sapporo, Japan ⁶Seamen's Insurance Hokkaido Healthcare Center, Medical Examination Department, Sapporo, Japan ⁷Keiwakai Nishioka Hospital, Department of Internal Medicine, Sapporo, Japan ⁸National Institute of Informatics, Research Center for Medical Bigdata, Tokyo, Japan ⁹Nagoya University, Information Technology Center, Nagoya, Japan

Keywords 3D reconstruction, structure measurement, depth image, monocular scale decision.

Purpose

This paper proposes a 3D bronchial anatomical structure measurement method based on real bronchoscopic images. During bronchoscopy, it is essential to measure the size of the anatomical structures. Most clinical applications measure the anatomical structure using preoperative CT volumes or using a stereo bronchoscope [1]. However, several pulmonary diseases are challenging to diagnose because it is hard to know the dynamic changes of the lesion region in real-time via CT volume. Stereo bronchoscope is currently unavailable.

We propose to measure the 3D bronchial spatial structure in real-time by using real bronchoscopic images. We use a deep learning-based method to estimate the depth image for 3D shape reconstruction and use the diameter of each branch as a scale factor for measurement. Our original contribution is to propose a measurement method for the anatomical structure in bronchus by using the real bronchoscopic images obtained from monocular bronchoscope. Using real bronchoscopic images, the anatomical structures in bronchus are accurately measured.

Methods

The proposed method uses real bronchoscopic images as input, and the output is the physical measurement result of anatomical structures. There are four steps in the proposed method: (1) depth image generation, (2) scale factor decision, (3) 3D surface reconstruction and (4) 3D measurement of the anatomical structures.

(1) Depth image generation



Fig. 1 Siloam Hospitals in Indonesia with Neurosurgery—2021

Since we cannot obtain depth images from bronchoscope directly, we estimate depth images using an image domain translation technique named CycleGAN [2].

(2) Scale factor decision

It is necessary to recover the scale information to measure the physical size of the anatomical structure because the scale has been lost in images obtained from monocular bronchoscope. Since the diameter in different branches is different, we use the diameter of each branch as the scale factor of the current scene. We count the number of the bronchi after the bronchoscope passed main bronchus (branching level) by using the branching level estimation method in literature [2]. We build a mapping between each branch and its diameter by using segmented bronchus region from CT volume. The diameter of the current branch is used as the scale factor of the current scene.

(3) 3D surface reconstruction

We reconstruct the 3D shape of the bronchus by using depth image and scale factor. The 3D shape reconstruction is achieved by using real bronchoscopic images, depth images and the intrinsic camera parameter. The reconstructed 3D shape is converted to physical space using the scale factor obtained in (2).

(4) 3D measurement of the anatomical structures

We manually select the points located on anatomical structure in reconstructed 3D shape for measurement. In our example, two points are selected to calculate the diameter of the bronchial orifice (BO) approximately (we assume the BO is circle-like and two points are on the line where the diameter lies).

Results

We used eight in-vivo pairs of chest CT volumes and videos during bronchoscopy to validate the proposed method. Six cases were used to train CycleGAN and two cases were used for validation. The bronchoscopic videos were taken by a bronchoscope (BF-260, Olympus, Tokyo, Japan). The chest CT volumes were taken by a CT scanner (XVision, Toshiba Medical Systems, Tokyo, Japan). We used approximately 4500 real bronchoscopic images and 5000 virtual depth images for training. The image size used for training was 256×256 pixels. We set the batch size to 10 and epoch to 200 to train the CycleGAN. We picked four images showing circle-like BO for evaluation. We measured the diameters of the BO from bronchoscopic images using the proposed method. Also, we measured the diameters of the corresponding BO in chest CT volumes as ground truth. For each image, we measured the diameter of BO three times and calculated the average diameter and the error, which were shown in Table 1. The average error was 0.325 mm. We showed an example in Fig. 1. We showed a real bronchoscopic image, measured points in reconstructed 3D shape, corresponded points for measurement in CT volume and corresponded points in the visualized 3D shape. The diameter value measured in the reconstructed 3D shape and CT volume was very close.

Table 1 Calculation of the average diameter and the error

Case-Image index	Diameter of BO in CT (mm)	Diameter of BO in the reconstructed surface (mm)	Error (mm)
1-2570 th	19	18,8	0,2
1-3020 th	14	14,1	0,1
2-3530 th	12	11,2	0,8
2-6470 th	8,7	8,9	0,2

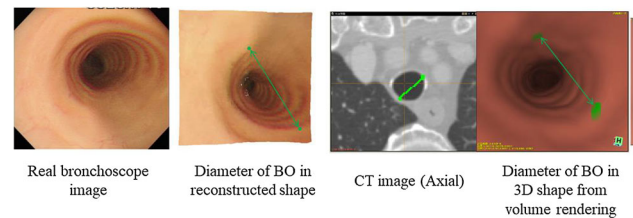


Fig. 1 Validation result of the BO diameter measurement of 2570th image in Case 1. In this real bronchoscopic image, the bronchoscope observes the main bronchus. We chose two points in reconstructed 3D shape and selected two points in approximately the same place in CT volume. Green lines linked the selected points in CT volume and their 3D shapes are visualized as volume rendering images of CT volume. The diameter of BO in the reconstructed shape was 18.8 mm, and the CT volume was 19.0 mm. Note that the viewpoint of real bronchoscopes and virtual bronchoscope is slightly different

Conclusion

We propose an anatomical structure measurement method in bronchus by using real bronchoscopic images. The evaluation result in in-vivo cases showed that the measurements of anatomical structures by the proposed method are close to the measurements in CT volume. Future work includes measuring more anatomical structure in the bronchus and the validation on more datasets.

Acknowledgements

We would like to thank to the member in our lab for useful discussions and comments. Parts of this study were supported by JSPS KAKENHI (Grant Numbers 17H00867, 26108006, 17K20099, and 21K19898), the JSPS Bilateral International Collaboration Granta and the JST CREST Grant Number JPMJCR20D5, Japan

References

- [1] Nobuyama S., Sato T., Handa H., Nishine H., Inoue T., Mine-shita M. and Miyazawa T. (2017). Comparison of Airway Measurements for Tracheobronchial Stenosis Between Stereoscopic Bronchoscope and MD-CT. *J Bronchology Interv Pulmonol* 24(4): 296–302.
- [2] Wang C., Hayashi Y., Oda M., Kitasaka T., Takabatake H., Mori M., Honma H., Natori H. and Mori K. (2021). Depth-based branching level estimation for bronchoscopic navigation. *Int J CARS* 16: 1795–1804.

26th Annual Conference of the International Society for Computer Aided Surgery

Chairs: Kensaku Mori, PhD (JP), Cristian A. Linte, PhD (US)

Development of an intraoperative cancer localization navigation system to support complete resection of prostate cancer

A. Mukasa¹, N. Koizumi¹, Y. Nishiyama¹, Y. Onodera¹, M. Matsuyama¹, T. Fujibayashi¹, S. Shoji²

¹The University of Electro-Communications, Graduate School of Informatics and Engineering, Chofu, Japan ²Tokai University School of Medicine, Departments of Urology, Isehara, Japan

Keywords RARP, YOLACT++, TRUS, Registration

Purpose

Robot-assisted radical prostatectomy (RARP) has been widely used as treatment procedure for localized prostate cancer both in Japan and overseas because of its excellent and delicate operation. On the other hand, since tumor localization cannot be confirmed in real time during the operation due to its characteristics, physician aims for complete resection of the cancer based on the preoperative nuclear magnetic resonance imaging and histopathological findings of the biopsy. Although the usefulness of RARP using intraoperative transrectal ultrasound (TRUS) imaging as navigation in complete cancer resection has been reported, they are challenging to monitor in real time because deformation caused by actions taken during surgery and non-visualization caused by changes in contact pressure with the ultrasound probe. Accordingly, this study investigates the development of an automatic monitoring system for localized of TRUS visible lesions inside prostate on ultrasound images during surgery.

Methods

The flowchart of the proposed method is shown in Fig. 1. First, TRUS images are segmented for the prostate and tumor by pre-trained YOLACT++ [1]. Next, the presence or absence of a tumor in the images is determined after which the tumor is classified as either detected or undetected. If a classification of undetected is assigned

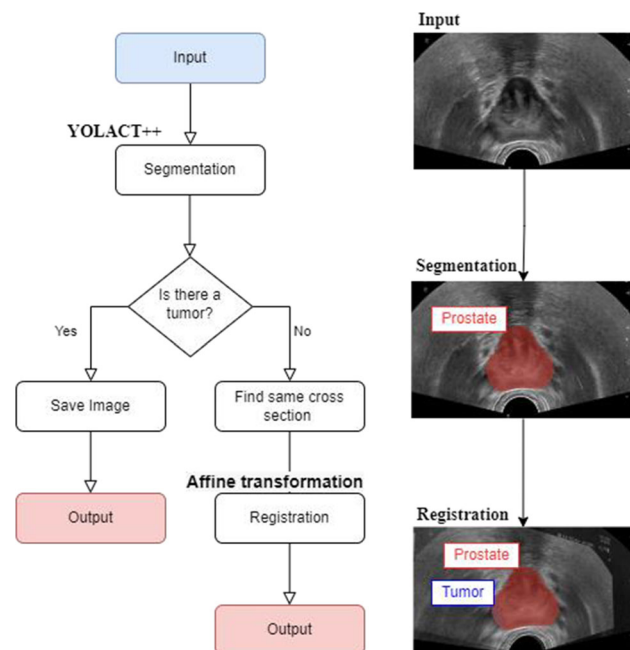


Fig. 1 Flowchart of the proposed method and an example of segmentation and registration

Table 1 Jaccard coefficient of prostate and tumor segmentation

	Prostate	Tumor
U-Net	0,930	0,877
YOLACT	0,966	0,939
YOLACT++	0,966	0,948

despite prior knowledge that a tumor is present, contact pressure has likely weakened, and the tumor position has moved outside the detection range. Based on this, the proposed method searches the detection database for cross-sections with high similarity levels[2] and registration between undetected and detected images to visualize the tumor in real time and stabilize its localization.

Results

We compared the accuracy of YOLACT++ with other segmentation models and found that it had the highest F value and Jaccard coefficient for both prostates and tumors. For prostates, the Jaccard coefficient was 0.966 and for tumors, the Jaccard coefficient was 0.972 (Table 1). Thus indicating that YOLACT++ is capable of very accurate segmentation. Next, the Euclidean distance for the accuracy of the tumor contour near the prostate capsule was 0.419 mm on average. The real-time performance of YOLACT++ was verified to be 30.56 FPS. A registration accuracy comparison between affine and projective transformations showed that the affine transformation had higher accuracy than the projective transformation.

Conclusion

We confirmed the effectiveness of YOLACT++ in prostate and tumor segmentation by obtaining very high accuracy levels, but tumor segmentation accuracy levels were erratic, possibly due to the inability to clarify tumor contours. Therefore it is necessary to perform image processing to enhance the contour. In general, tumors are more likely to have a localization than that confirmed by preoperative MRI. The area may be 3 mm to 5 mm wider than the area of the patient. In the present study, the error in contouring around the prostatic capsule was approximately 1 mm, which was evaluated to be sufficient. Additionally, the real-time processing speed of the transrectal ultrasound was approximately 22 FPS, which is sufficient for use during surgery. In order to deal with changes in prostate contour and severe prostate gland deformation during surgery, it will be necessary to consider a nonlinear registration method such as the B-spline method.

References

- [1] Bolya, Daniel, et al. (2019) Yolact: Real-time instance segmentation. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- [2] Igarashi R, Koizumi N, Nishiyama Y, Tomita K, Shigenari Y, Shoji S (2020) Sagittal alignment in an MR-TRUS fusion biopsy using only the prostate contour in the axial image,” ROBO-MECH Journal, Vol.7, No.1

Optimal quantitative metrics reflecting qualitative evaluation in deep learning-based intraoperative anatomy recognition task

Y. Ishikawa¹, D. Kitaguchi^{1,2}, H. Matsuzaki¹, A. Yamada¹, K. Hayashi¹, N. Kosugi¹, H. Hasegawa^{1,2}, S. Kojima^{1,2}, K. Nakajima^{1,2}, K. Obuchi^{1,2}, N. Takeshita^{1,2}, M. Ito^{1,2}

¹National Cancer Center Hospital East, Surgical Device Innovation Office, Chiba, Japan ²National Cancer Center Hospital East, Department of Colorectal Surgery, Chiba, Japan

Keywords Qualitative Evaluation, Anatomy Recognition, Colorectal Surgery, Deep Learning

Purpose

There have been many efforts to realize image navigation surgery using deep learning. Semantic segmentation techniques can be used to highlight crucial anatomical structures to be recognized in endoscopic surgery, which may lead to avoidance of organ damage and recognition of the correct dissection layer by making the intraoperative surgeon aware of the location of critical organs. Although the evaluation metrics commonly used in semantic segmentation tasks are Dice coefficient and intersection over union (IoU) [1], especially in an intraoperative anatomy recognition task, these metrics often do not match the needs of surgeons, and there are following limitations to using them. First, even a tiny false positive can be visually stressful for the surgeon during surgery. Second, the boundaries of the anatomical structure are covered with tissues, membranes, and the like are originally ambiguous and unimportant for evaluating recognition performance. Therefore, there is a need to develop novel evaluation metrics specific to intraoperative anatomy navigation well reflecting the sense of surgeons to accelerate this research field further. This study aims to explore the optimal metrics that are highly correlated with the qualitative evaluation by surgeons in the anatomy recognition task of endoscopic surgery.

Methods

The target task in this study was semantic segmentation of the ureter, which is an important anatomical structure in laparoscopic colorectal surgery. A total of 125 images capturing ureter were extracted from 55 intraoperative videos. All the images have been manually annotated the regions of the ureter under the supervision of a surgeon, and this manual annotation region was defined as the ground truth. As a deep learning model, PSPNet and DeeplabV3+ were adopted. To vary recognition performance intentionally, we created 20 models by varying the number of data and loss patterns. The patterns consisted of four patterns of positive data (300, 600, 1500, and 15,000 images), the same patterns of negative data, and five patterns of loss functions (Focal Loss, Boundary Loss, Dice Loss, Focal Dice Loss, and Tversky Loss). Finally, a total of 2009 images overlaid predicted ureter regions were created. For quantitative evaluation of the predicted results, not only existing metrics (Dice coefficient, Precision, Recall, and the like) but also a novel proposed metric was used. The proposed metric is based on the Dice coefficient and weight only isolated false positives, which is the predicted mask whose contour is not overlapping with the ground truth. The proposed metric was calculated by varying the weight from 0 to 30 in increments of 5. Subsequently, a total of 20 surgeons performed a qualitative evaluation for the predicted ureter regions by the models in each image using the Visual Analogue Scale (VAS) assuming usage as an intraoperative image navigation system. For the qualitative evaluation by surgeons, we excluded outlier images: surgeon's misrecognition of the ureter, the same qualitative rating despite the different predicted ureter regions, high qualitative rating even though the ureter region

which was not predicted at all, and inadequate annotations. The correlation coefficients between the quantitative and qualitative evaluations were calculated to explore the metrics with the highest correlation.

Results

After excluding outlier images, 1297 images were used to calculate the correlation coefficient between quantitative evaluations with several metrics used in this study and the qualitative evaluation by surgeons. As for the correlation coefficient, the conventional Dice coefficient was 0.61, Precision was 0.17, Recall was 0.56, Hausdorff distance was -0.52 , Average symmetric surface distance was -0.58 , and Surface Dice was 0.53. With regard to the parameter of the weight for the isolated false positives in the proposed novel metrics, 15 was the optimal value, and its correlation coefficient was 0.64. Compared with the existing metrics, the proposed metric showed the highest correlation coefficient.

Conclusion

Since our proposed novel metric had the strongest correlation with qualitative evaluation by surgeons, it best reflected the sense of surgeons in the ureter segmentation task in laparoscopic colorectal surgery. These results indicated that the proposed metric could be the gold standard in intraoperative anatomy recognition tasks. It is considered that existing metrics are not necessarily generic, and the optimal evaluation metric depends on the recognition target and the situation in which they are used. The results of this study can be a significant indicator for future research and development related to image recognition using surgical images.

References

- [1] Csurka G, Larlus D, Perronnin F (2013) What is a good evaluation measure for semantic segmentation? Proceedings of the British Machine Vision Conference:1–11.

Neural network approach to detection and 3D localization of guidewires from 2D intraoperative fluoroscopy

L. Mekki¹, A. Uneri¹, N. Sheth¹, R. Vijayan¹, A. Sisniega Crespo¹, G. Kleinszig², S. Vogt², G. Osgood³, J. Siewerdsen¹

¹Johns Hopkins University, Biomedical Engineering, Baltimore, United States ²Siemens Healthineers, Erlangen, Germany ³Johns Hopkins University, Department of orthopaedic Surgery, Baltimore, United States

Keywords Deep Learning, Guidewires detection, Image-guided interventions, 3D localization

Purpose

Guidewires are used in orthopaedic surgery as guides for insertion of larger instruments, such as cannulated screws for fracture fixation. Accurate positioning of guidewires often involves trial-and-error and long fluoroscopic exposure times. We present an approach that leverages a deep learning approach for object detection (guidewires) in 2D fluoroscopy and model-based 3D-2D image registration to identify corresponding detections in multiple fluoroscopic frames and compute their 3D location, Fig. 1. The method was evaluated in cadaveric specimens in four anatomical sites pertinent to orthopaedic trauma surgery. The study included various levels of other instrumentation (“clutter”) in the fluoroscopic scene to evaluate the robustness of the method in complex, clinically realistic scenes.

Methods

The algorithm extends an early implementation [1] with an improved neural network architecture for keypoint detection (guidewire tip and direction) and a large, realistic training set to accommodate complex fluoroscopic scenes with a high density of surgical instrumentation.

Data-driven object detection and instance segmentation. The detection component of the method consisted of a neural network trained for the instance segmentation of guidewires and 2D tip coordinates from two fluoroscopic images. The architecture incorporated a combination of Mask-RCNN and Keypoint-RCNN models such that both a guidewire segmentation and its tip coordinates were detected. A third order B-spline was fit to the predicted guidewire segmentation along the distal 10 mm to the predicted tip location. The derivative of the fit determined the predicted guidewire direction.

Model-based correspondence and 3D localization. The 3D localization component of the method backprojects the predicted 2D tip location from the fluoroscopic image plane using knowledge of the system geometry (mobile C-arm; Cios Spin, Siemens Healthineers). The backprojected rays were analyzed to determine corresponding detections (i.e., belonging to the same guidewire, since multipole guidewires are often present) in each fluoroscopic view. The 3D location was determined by the nearest point of intersection among corresponding backprojections, and 3D direction was determined from the intersection of the planes formed by the 2D direction vectors in each view and a line from the tip to the x-ray source.

Network training. A dataset of 10,000 training and 1000 validation images was generated from 8 cone-beam CT (CBCT) scans of the pelvis, lumbar spine, thorax, and shoulders in cadavers. Fluoroscopic scenes exhibiting various numbers and orientations of guidewires were simulated by forward-projecting guidewire mesh models derived from 3rd order B-spline curves. A range of 1–4 guidewires with diameter varying 1–4 mm and attenuation $0.15\text{--}0.25\text{ mm}^{-1}$ was simulated in each image. For each image, 1–3 CAD models of other surgical instruments were generated (including clamps, retractors and hemostats) and forward-projected with attenuation varying $0.15\text{--}0.25\text{ mm}^{-1}$. Quantum noise was added to reflect varying tube current (e.g., 1–5 mA) and patient size. The combined Mask-Keypoint-RCNN model was trained for 75 epochs with a binary cross

entropy loss function, the learning rate was set to 10^{-4} , and Adam optimization was performed with a batch size of 20.

Experimental studies in cadaver. The method was tested on 31 CBCT volumes, each containing 2 guidewires—15 without additional instrumentation clutter, and 16 with multiple non-guidewire instrumentation (“clutter”) added to reflect realistic clinical conditions. For each CBCT volume, pairs of views were selected such that: the 2 views were separated by at least 45° ; only one tip of each guidewire was visible; and the guidewire tips were unoccluded by other instruments (to facilitate quantitative analysis and truth definition). The accuracy of 2D detection was evaluated on fluoroscopic image pairs, with true detections defined as those with Euclidean distance $< 10\text{ mm}$ between the predicted and true tip coordinates. All detections (including false-positives) were taken as input to 3D localization to determine 3D tip location for each guidewire. Comparison to a previously reported framework [1] was performed using equivalent fluoroscopic views, including quantum noise, instrumentation, etc. not included in previous work.

Results

The experiments demonstrated performance of the proposed method for the first time in real image data reflecting clinically realistic anatomy and instrumentation. Results are summarized in Table 1. The method was robust to instrumentation clutter, exhibiting a recall of 80% without clutter and 78% in the presence of heavy clutter of various (non-guidewire) instrumentation. Performance was significantly improved in comparison to a previously reported method [1], which exhibited a recall of 60% and 52%, respectively. Directional accuracy was also significantly improved. The enhanced performance is attributed to the superior network structure as well as major augmentation of the training set to include realistic levels of quantum noise (dose levels) and realistic instrumentation clutter. The performance of the detection network could be tuned somewhat with respect to recall and precision, recognizing that the subsequent method for identifying corresponding detections is robust to false-positive detections (rejecting non-correspondent detections between fluoroscopic views). Therefore, as shown in Table 1, precision was lower than recall, but still suitable to subsequent correspondence and 3D localization steps, which was accurate to 1.9–2.7 mm with or without instrumentation clutter. Further improvement in detection accuracy (counting only True-Positive detections in backprojection) could reduce this further to 1.2–1.4 mm, motivating further refinement of the detection algorithm to achieve performance potentially superior to a surgical tracker.

Conclusion

The framework for 2D detection and 3D localization demonstrates major improvements in accuracy compared to previous methodology and, in particular, is robust to the presence of instrument clutter. Future work will include translation to clinical data, comparison to current clinically available 2D detection tools, and investigation of 3D localization accuracy in comparison to (tracker-based) 3D navigation.

Table 1 2D detection and 3D localization of the previous [1] and proposed methods

	Previous Method [1]		Proposed Method	
	No Clutter	Heavy Clutter	No Clutter	Heavy Clutter
2D Detection				
Recall	60%	52%	80%	78%
Precision	59%	24%	69%	41%
Tip error	$(2.5 \pm 1.8)\text{ mm}$	$(2.6 \pm 1.9)\text{ mm}$	$(2.1 \pm 1.6)\text{ mm}$	$(2.7 \pm 1.9)\text{ mm}$
Direction error	$8^\circ \pm 27^\circ$	$15^\circ \pm 36^\circ$	$4.7^\circ \pm 4.5^\circ$	$6^\circ \pm 14^\circ$
3D Localization				
Tip error (All)	$(5.1 \pm 8.8)\text{ mm}$	$(6.6 \pm 9.6)\text{ mm}$	$(1.9 \pm 3.3)\text{ mm}$	$(2.7 \pm 4.6)\text{ mm}$
Tip Error (TP Only)	$(0.9 \pm 1.1)\text{ mm}$	$(1.3 \pm 2.2)\text{ mm}$	$(1.2 \pm 3.0)\text{ mm}$	$(1.4 \pm 3.0)\text{ mm}$

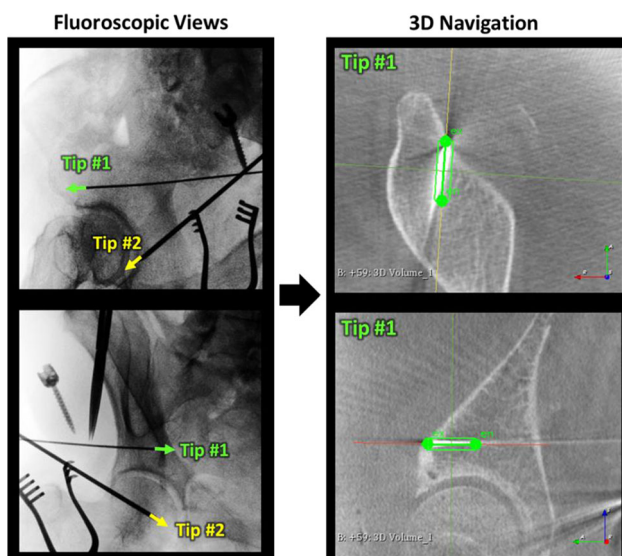


Fig. 1 Example fluoroscopic image pair showing guidewires along with other instrumentation. Predicted 2D detections are shown on the left, and 3D localization (for guidewire #1) is on the right

References

- [1] Bataeva I., Shah K., Vijayan R., Han R., Sheth N.M., Kleinszig G., Vogt S., Osgood G.M., Siewerdsen J.H., Uneri A., (2021) “Intraoperative guidance of orthopaedic instruments using 3D correspondence of 2D object instance segmentations,” Proc. SPIE Medical Imaging Vol. 11,598.

Pathological evaluation of human pancreatic tissue injuries by machine compression for computer-aided safe pancreatic compression devices

K. Inai¹, D. Kim², N. Takano², M. Uno¹, S. Noriki³, H. Naiki¹, E. Kobayashi⁴

¹University of Fukui, Division of Molecular Pathology, Eiheiji, Japan

²Teikyo Heisei University, Faculty of Health and Medical Science, Toshima, Japan ³Fukui Prefectural University, Department of Social Welfare Science, Eiheiji, Japan ⁴The University of Tokyo, School of Engineering, Bunkyo, Japan

Keywords compression device, human pancreas, pathology, wall thickness of pancreas

Purpose

The development of pancreatitis due to leakage of pancreatic juice after pancreatic resection is a notorious complication after pancreatic body and/or tail resection [1]. To establish a computer-aided safe pancreatic compression device, we evaluated the pattern and incidences of human pancreatic tissue damages after mechanical compression.

Methods

The pancreatic tissues were obtained from the deceased in University of Fukui Hospital. The provision and use of autopsy pancreas was subjected to ethical review by the University of Fukui. After isolation of upper abdominal organs, the stomach, duodenum, and pancreas were removed en bloc, followed by extracting pancreas body to tail from there. The compressions were started at 10.9 ± 4.2 h post-mortem. The pancreas compressed at the speed of 1/160 mm/s until the wall thickness became 2 mm, and the reaction force from the pancreas was measured while maintaining the stopped state for 120 s. Followed by the compression, the tissue was soaked in 10% formaldehyde solution for extension fixation. The fixed organ was cut out by scalpels to make the paraffin sections. The paraffin sections were thin sliced and the sliced sections were stained by hematoxylin-eosin dye and Azan stain. The histological analyses were performed by board-certified pathologists [2]. Statistical analyses were performed using χ^2 test and Student's *t*-test.

Results

Pancreatic destructions showing seven different patterns

To investigate the destruction patterns due to compression, slide sections were observed in detail and found seven different patterns emerged in the slide sections. The representative aspects of tissue damage were named as follows: “transmural destruction”, “destruction reaching to the acini”, “intra-acinic destruction”, “disruption among acini”, “dissection between acini and adjacent stroma”, “stromal destruction arising from surface layer”, and “intra-stromal

Table 1 Incidence of pancreatic destruction between pancreatic body and pancreatic tail

Severity of pancreatic destruction	Pancreatic body (number of slides)		Pancreatic tail (number of slides)		P value
	Yes	No	Yes	No	
Transmural destruction	5	56	3	58	
Destruction reaching to the acini	36	25	22	39	$p < 0.02$
Intra-acinic destruction	11	50	17	44	
Disruption among acini	18	43	20	41	
Dissection between acini and adjacent stroma	34	27	33	28	
Stromal destruction arising from surface layer	5	56	17	44	
Intra-stromal destruction	14	47	22	39	
No tissue injury	1	60	2	59	

Statistical calculation was performed using χ^2 analysis.

destruction”, respectively. Of these 7 patterns, the first four patterns were predicted to leak the pancreas juice from destructive acinic exocrine glands, while the remaining three tissue injuries were relatively safe because they did not reach until acini. Therefore, the latter was considered as slight injury patterns.

Significant higher frequency of destruction reaching acini in pancreatic body

Using the above-mentioned destruction patterns, we then compared the incidence of destruction in pancreatic body with those in pancreatic tail. Most destruction patterns represented the same frequency between pancreatic body and pancreatic tail, whereas destruction reaching to the acini developed statistically high frequency of occurrence in pancreatic body than that in pancreatic tail (36/61 vs 22/64 slide sections, $p < 0.02$, Table 1).

The association between pancreatic destruction and thickness of pancreas

Additionally, the association between pancreatic destruction and thickness of pancreas was evaluated using the cases of transmural destruction and patients showing no pancreatic injury and only had slight injuries in the organ. At a glance, it appears that the cases with

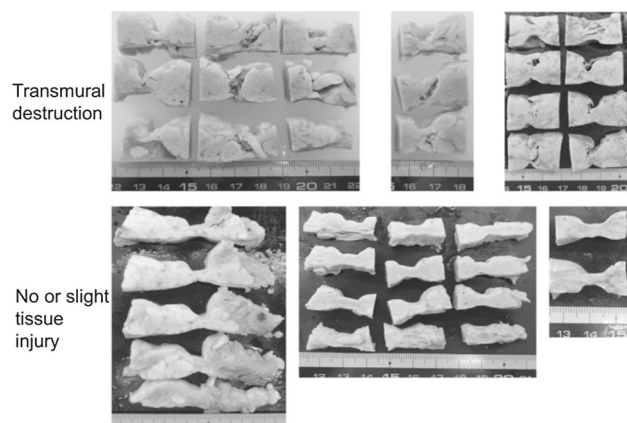


Fig. 1 Cases with transmural destruction and non-injury or slight tissue injury patients

transmural destruction seemed to be thicker than non-injury or slight tissue injury patients (Fig. 1). Indeed, as it expected, the mean wall thickness with tissue destructions represented of significant thick in both pancreatic body (1.9 ± 0.28 cm vs 1.3 ± 0.21 cm, $p < 0.01$) and in pancreatic tail (1.98 ± 0.21 cm vs 1.2 ± 0.1 cm, $p < 0.01$), respectively.

Conclusion

Some relationships may exist between wall thickness of pancreas and the severity of destruction due to machine compression. Our experimental findings using human pancreas would provide us new knowledges in order to establish a computer-aided safe pancreatic compression device.

Acknowledgment

This study was partially supported by JSPS KAKENHI (grant number 26108008, 16K09930, 20H03908, 18K12116).

References

- [1] Kobayashi E, Tsuchiya S, Akagi Y, Tomii N, Nakagawa K, Inai K, Muragaki Y, Asano T, Kim D (2020) A novel reaction force-fluorescence measurement system for evaluating pancreatic juice leakage from an excised swine pancreas during distal pancreatectomy. *J Hepatobiliary Pancreat Sci* 27:877–86.
- [2] Inai K, Noriki S, Kinoshita K, Sakai T, Kimura H, Nishijima A, Iwasaki H, Naiki H (2016) Postmortem CT is more accurate than clinical diagnosis for identifying the immediate cause of death in hospitalized patients: a prospective autopsy-based study. *Virchows Arch* 469:101–9.

Development of a laparoscopic system for laser thermal therapy equipped with an ultra-compact thermographic camera

M. Harada¹, Y. Morimoto², J. Ohya³, K. Masamune⁴, Y. Itazaki¹, T. Sugihara¹, H. Tsujimoto¹, Y. Kishi¹, H. Ueno¹

¹National Defense Medical college, Surgery, Tokorozawa, Japan

²National Defense Medical college, Physiology, Tokorozawa, Japan

³Waseda University, Modern Mechanical Engineering, Shinjyuku, Japan

⁴okyo Women's Medical University, Advanced Techno-Surgery, Shinjyuku, Japan

Keywords laparoscopy, laser thermal therapy, cancer, liver

Purpose

Laser thermotherapy is a therapeutic method to induce cell death of cancer cells by heating cancer tissues with a near-infrared laser. It is known that controlling tissue temperature during heating has a significant role in the therapeutic effect.

We established a laser thermotherapy system that can automatically control the laser power to keep the temperature of the cancer tissue constant by acquiring two-dimensional temperature mapping in real time, and succeeded an eradication of cancer tissue in a subcutaneous tumor mice model using the system.

The purpose of this study was to develop a rigid endoscopic system equipped with an ultra-compact thermal imaging camera in order to apply the laser thermotherapy with automatic temperature control to eradicate cancer tissue located in abdominal organ under laparoscopic surgery. Additionally, we examined the therapeutic effects of the developed endoscopic system on malignant tumors of the abdominal organs using an animal model.

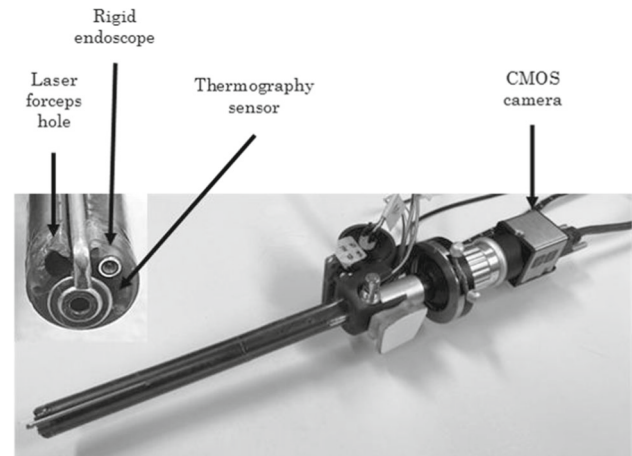


Fig. 1 Optical fiber for laser irradiation

Methods

The constructed thermal endoscope was consisted of a rigid endoscope (the shaft had a maximum diameter of 14 mm and a length of 288 mm), an ultra-compact infrared thermography sensor (HTPA32 × 32d L2.1, Heimann Sensor), and a hole for introducing an optical fiber for laser irradiation (Fig. 1). Bright field images were obtained by a CMOS camera connected to the rigid endoscope. Two dimensional temperature distribution was visualized by the thermography sensor with a frame rate of 8.3 fps and a spatial resolution of 32 × 32 pixels (a temperature range of 20–80 °C corresponds linearly to a pixel value of 0–255).

Laser irradiation (808 nm) through the optical fiber was performed in a non-contact setting. During laser irradiation, by automatically extracting 9 × 9 pixels around the pixel that indicates highest temperature among all of the viewed pixels, and averaged temperature of the 81 pixels was calculated, we defined the averaged temperature as "temperature of irradiated target". Based on the information of the "temperature of irradiated target", the target tissue was heated with keeping the temperature constant by being automatically calculating the appropriate power of the laser irradiation using a PC.

Using the thermal endoscopic system, we examined (1) the performance of the automatic temperature control and (2) the therapeutic effect on hepatocellular carcinoma in an orthotopic liver cancer model rat under laparoscopic condition.

Results

(1) The "temperature of irradiated target" during heating was set at 70 °C, and the "temperature of irradiated target" was recorded for 5 min after the target temperature reached 70 °C. The recorded "temperature of irradiated target" for the 5 min showed that the median was 69.8 °C (min: 67.8 °C, max: 77.4 °C) with a variation of < 68 °C: 0.2%, 68–72 °C: 93.2%, and > 72 °C: 6.6%.

(2) The cancer model rats were randomly divided into two groups (thermal group (n = 6) and control group (n = 7)), and the tumor volume was measured one week after the thermal treatment. Tumor volume was significantly smaller in the thermal group (median of hyperthermia group: 1.0 × 102 mm³, median of control group: 9.4 × 102 mm³, $P = 0.0043$).

Conclusion

We have established a temperature-visualizing laparoscopic system for laser thermal therapy. This system enables visualization of temperature distribution, thus providing to keep the temperature of the target tissue constant during laser irradiation. Using the system, we have succeeded an eradication of tumor in a rat model of hepatocellular carcinoma under laparoscopic condition.

Towards surgical margin assessment with photon-counting spectral microCT imaging

I. Chitra Ragupathy^{1,2}, T. Tramm^{2,3}, J. Albertus Nijkamp^{1,2}

¹Aarhus University Hospital, The Danish Centre for Particle Therapy, Aarhus N, Denmark ²Aarhus University, Department of Clinical Medicine, Aarhus N, Denmark ³Aarhus University Hospital, Department of Pathology, Aarhus N, Denmark

Keywords Spectral micro CT, Image registration, Breast tumor, Histology

Purpose

The main quality indicator for tumor surgery is the surgical resection margin, where tumor cells close to the surface of the tumor specimen are an indication for poor local control. Currently, the surgical resection margin is determined post-operatively at the pathology department, days after the surgical procedure. Ideally, you would like to assess the margin pre-operatively, enabling removal of extra tissue when needed. For breast cancer, microCT imaging has been widely studied as a tool for quick margin analysis, but with mixed results. With high contrast between tumor and surrounding fat, the tumor can be easily recognized based on morphology, but limited contrast between tumor and glandular tissue or fibrosis makes it challenging to determine the full extent of the tumor. With the development of energy resolved photon counting detectors, there is a great opportunity to acquire spectral microCT images with improved signal to noise, and the addition of spectral data could improve the contrast between tumor and normal tissues.

The focus of this work was to show proof of concept of improved tumor-to-normal tissue contrast with spectral microCT imaging of a fresh-frozen mouse-breast tumor, validated with histology imaging. Furthermore, we investigate if spectral imaging of paraffin embedded tissue is comparable to fresh-frozen tissue.

Methods

Our prototype imaging system consist of a self-shielded x-ray imaging cabinet (www.metrixndt.com), a Hamamatsu microfocuss x-ray generator (L9421-02, 20–90 kV, 8 W, focal spot 7 μ m), a spectroscopic 2 \times 2 Medipix3 detector with 512 \times 512 pixels, a pixel pitch of 55 μ m, and 2 photon energy thresholds (www.amscins.com), and a rotating sample-stage. We further built cryo-chamber around the sample-stage to enable scanning of frozen samples.

Our imaging sample was a fresh-frozen, non-irradiated mouse-leg with a transplanted spontaneous mouse-breast tumor of 7 \times 5 mm on the foot, obtained from a proton irradiation study. The sample was imaged with 360 projections over a full rotation with 50kvp and 160uA. Three energy threshold scans were acquired sequentially: 7–14, 14–21, and 21–28 keV, with exposure times of 0.64, 12.8, 27.84 s per projection, respectively. The source-detector and source-object distances were 192.6 and 108.8 mm, respectively.

Once the frozen tissue was imaged, the sample was placed in a formalin (10% formaldehyde) followed by decalcification of the bones using formic acid. Then, the tissue was cut into 5 pieces and embedded in paraffin. These paraffin blocks were imaged using the same imaging parameters as mentioned above. Finally, one histology slide was obtained from each paraffin block, and annotated for ground truth labeling by a pathologist.

The projection data was reconstructed into 3D scans using FDK algorithm provided in TIGRE (<https://github.com/CERN/TIGRE>). Bad pixels of the detector were masked, and replaced by linear interpolation of the surrounding pixels. To create energy bin reconstructions, photon counts of the high threshold were subtracted from the low threshold. Furthermore, the 7 keV threshold data was used to

reconstruct a single energy 3D scan, mimicking a standard microCT scan. Finally, the 28 keV threshold data was also reconstructed into single energy 3D scan, to represent the 28 + keV energy. Ring artefacts removal was done by pre-processing the sinogram data with the `remove_stripe_based_sorting` (<https://github.com/algotom/algotom>).

To enable direct comparison between the fresh-frozen scans, the paraffin block scans, and the histology slides, deformable image registration was used to bring all scanning data to the histology slides. For registration, SimpleElastix was used (<https://simpleelastix.github.io/>), where we used a 2-step affine-b-spline registration. First, the fresh-frozen scans of 14–21, 21–28, and 28 + were registered to the 7 + and 7–14 scans (which are intrinsically aligned), as there was some tissue deformation over time due to de-hydration in the cryo-chamber. In the next step, the 7 + fresh-frozen scan was registered to the 7 + scans of the paraffin blocks. The resulting deformation-vector-field (DVF) was applied to all fresh-frozen scans to also deform them towards the paraffin blocks. In the final part the paraffin blocks were registered with their histology slides. As this is a 3D-to-2D registration, first the best fitting slice from the paraffin block was determined by performing a 2D-to-2D affine registration between each paraffin block slice and the histology slide, and deriving the normalized cross correlation. The best fitting slice was subsequently deformably registered to the histology slide. Again, the resulting DVF was applied to all scanning data to obtain spectral imaging data (7 + , 7–14, 14–21, 21–28, and 28 +) of fresh-frozen tissue, and paraffin-embedded tissue in the reference frame of the histology slides.

Once the registration was performed, the slices of fresh-frozen tissue and the paraffin blocks for all the scans were masked for each annotation presented in the histology slide and the mean spectral intensity was calculated.

Results

In the figure example images from histology, fresh-frozen tissue, and paraffin embedded tissue can be seen. Image registration visually was acceptable for the tumor region, but challenges existed for the area's

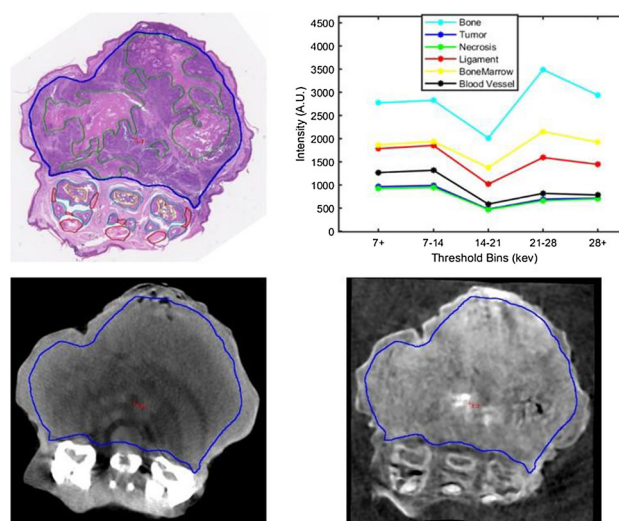


Fig. 1 Top left: histology slide with annotations, showing the tumor area in blue, necrosis in green, ligaments in red, bone in light-blue, and bone marrow in yellow. Top right: mean spectral intensity of the annotated regions from fresh-frozen tissue. Bottom left: 7 + scan data of the fresh-frozen tissue after registration with the histology slide. The blue tumor outline is presented for reference. Bottom right: 7 + scan data of the paraffin block after registration with the histology slide. Note that registration of the area with the bones was challenged due to a tear in the histology slide

where the histology slides were torn or missing some tissue. Spectral data comparison between the different tissue types showed promising results for the fresh-frozen tissue (Fig. 1). There was a clear difference in intensities, and also the change in intensity with increasing spectral bin showed differences between the tissue types. For the paraffin blocks there was hardly any spectral difference between tissue types, probably due to the processing (no water content) and the decalcification of the tissue.

Conclusion

We were able to acquire acceptable spectral microCT scans of a mouse breast tumor with our prototype imaging setup. Based on qualitative data analysis we can say that the spectral information has potential to improve the contrast between tumor and normal tissue. From our data we can also conclude that paraffin embedded tissue, which is easier to image and register with histology, can not be used as a surrogate for fresh tissue. In future we will focus on speeding up image acquisition to minimize tissue changes, and acquisition of additional spectral bins, especially < 7 keV, and at higher spectral resolution to further improve tissue discrimination.

Laparoscopic image classification based on surgical areas in laparoscopic gastrectomy

Y. Hayashi¹, K. Misawa², K. Mori^{1,3}

¹Nagoya University, Graduate School of Informatics, Nagoya, Japan

²Aichi Cancer Center Hospital, Department of Gastroenterological Surgery, Nagoya, Japan ³National Institute of Informatics, Research Center for Medical Bigdata, Tokyo, Japan

Keywords Computer-aided surgery, Laparoscopic surgery, Deep learning, Surgical navigation

Purpose

Laparoscopic surgery has been widely performed as one of minimally invasive surgery. However, this type of surgery is more complicated than conventional open surgery. Therefore, there are many studies about a computer-aided surgery (CAS) system for laparoscopic surgery, such as a surgical robot system and a surgical navigation system. Recent progress of the deep learning technique allows analyzing laparoscopic videos to assist surgery [1]. These studies conducted segmentation of anatomical structures and surgical tools, recognition of surgical phase, assessment of surgical skill, and so on, in several kinds of laparoscopic surgery. In laparoscopic gastrectomy for gastric cancer, surgeons resect the stomach, including the tumor and the associated lymph nodes. Before resecting the stomach, surgeons cut the blood vessels around the stomach in sequence. Therefore, recognizing surgical areas related to the blood vessels from laparoscopic images provides valuable information to the CAS system for assisting laparoscopic gastrectomy. This paper describes a method for classifying laparoscopic images based on the surgical areas in laparoscopic gastrectomy for gastric cancer.

Methods

The proposed method classifies the laparoscopic images into seven classes based on the surgical areas. As described above, surgeons process the blood vessels around the stomach during laparoscopic gastrectomy. Therefore, we define four classes related to the blood vessels processed during surgery. These classes are (Class 1) left gastroepiploic artery and vein, (Class 2) right gastroepiploic artery and vein, (Class 3) right gastric artery, and (Class 4) left gastric artery and vein. The other three classes are defined as (Class 5) abdominal cavity (Class 6) inside trocar (Class 7) outside the body. Example laparoscopic images in each class are shown in Fig. 1. The proposed method consists of two parts, classification of the laparoscopic images

using Bayesian convolutional neural networks and modification of classification results using the uncertainty of the prediction. First, the proposed method classifies the laparoscopic images into the seven predefined classes using a Bayesian densely connected convolutional network (DenseNet). Bayesian DenseNet is implemented by extending DenseNet using Monte Carlo (MC) dropout technique [2]. In the training of this model, we perform fine-tuning from the ImageNet pre-trained model. We use cross-entropy as a loss function and Adam as an optimizer. After the training process, laparoscopic images are classified into seven classes using the trained models. We also measure the uncertainty of the predictions using the predictive entropy [2]. Then, we modify the classification results using the uncertainty and temporal information. We assume that the nearby surgical area is observed in the temporally successive images of laparoscopic videos. The proposed method changes classification results for each laparoscopic image with high uncertainty to classification results for the temporally previous images in the laparoscopic video.

Results

We applied the proposed method to five laparoscopic videos obtained during laparoscopic gastrectomy for gastric cancer. We extracted laparoscopic images from the videos every 10 s. Laparoscopic images per one case were about 1000 images. Leave-one-out cross-validation was performed for the performance evaluation. The average classification accuracy of five cases before and after modification using the uncertainty was 80.8% and 82.4%, respectively. These results showed that the modification process of classification results based on uncertainty and temporal information helped improve the classification accuracy. Examples of correctly classified images for each class based on the surgical areas are shown in Fig. 1. This figure shows that the proposed method could classify laparoscopic images into seven classes based on the surgical areas. Since the proposed method obtains the surgical areas from only laparoscopic images during surgery, the proposed method will provide helpful information to CAS system.

Conclusion

This paper described a method for classifying laparoscopic images based on the surgical areas in laparoscopic gastrectomy for gastric cancer. We classified laparoscopic images into the seven classes based on the surgical areas using Bayesian convolutional neural networks and prediction uncertainty. Experimental results showed that the proposed method could classify laparoscopic images captured during laparoscopic gastrectomy. Future works include application to additional cases and developing a computer-aided surgery system based on the classification results.

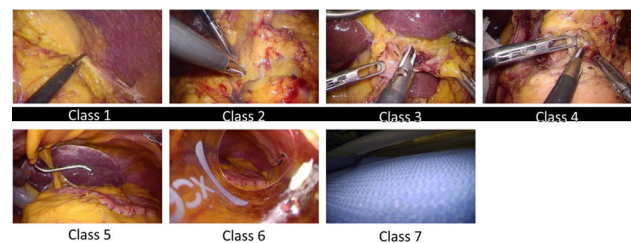


Fig. 1 Examples of correctly classified laparoscopic images in seven classes based on surgical areas. (Class 1) left gastroepiploic artery and vein, (Class 2) right gastroepiploic artery and vein, (Class 3) right gastric artery, (Class 4) left gastric artery and vein, (Class 5) abdominal cavity (Class 6) inside trocar, and (Class 7) outside the body

References

- [1] Mori K, Morimitsu S, Yamamoto S, Ozawa T, Kitasaka T, Hayashi Y, Oda M, Ito M, Takeshita N, Misawa K (2021) Blood vessel regions segmentation from laparoscopic videos using fully convolutional networks with multi field of view input. Int J of CARS 16 (suppl.1): S56-S57.
- [2] Kendall A, Gal Y (2017) What uncertainties do we need in Bayesian deep learning for computer vision? Proc of NIPS‘17: 5580–5590.

Surgical navigation in long-length tomosynthesis images using dynamic focal plane reconstruction

I. Butz¹, X. Zhang¹, Y. Huang¹, A. Johnston¹, T. F. Witham², P. Helm³, J. Siewerdsen^{1,2}, A. Uneri¹

¹Johns Hopkins University, Biomedical Engineering, Baltimore, United States ²Johns Hopkins Medicine, Neurosurgery, Baltimore, United States ³Medtronic, Littleton, United States

Keywords surgical navigation, long-length tomosynthesis, intra-operative imaging, geometric accuracy

Purpose

Surgical navigation for clinical applications in spine involves use of optical trackers that provide real-time positions of surgical instruments within a 3D image of the patient anatomy. Conventional methods that rely on preoperative modalities (CT/MR) are subject to anatomical changes depending on the time of scan or patient pose on the table (prone vs. supine). Intraoperative imaging, for example via cone-beam CT (CBCT) from an O-arm™ imaging system (Medtronic, Littleton MA) can provide an up-to-date view of the anatomy, however, are limited by the detector extent—20–30 cm image length that capture 4–5 vertebrae.

In this work we present a novel navigation approach using a new intraoperative imaging modality that can provide coronal and sagittal tomosynthesis images with extended field-of-view (FOV) [1]. Using a registered surgical tracking system, images are dynamically reconstructed at navigated locations to absolve geometric distortions due to limited depth resolution and provide real-time tracking along the full length of spine that is particularly valuable for guiding placement of long spinal constructs.

Methods

As illustrated in Fig. 1, projection data is acquired through a slot-collimator and linear (z) translation of the O-arm gantry. Limited-angle tomosynthesis reconstruction is performed by weighted back-projection of the data at a given “focal plane,” nominally defined to align with the anatomy of interest. The reconstructed images (Fig. 1) have modest depth resolution and are subject to geometric distortions outside the focal plane, where structures appear distorted and/or shifted [2].

Preliminary approach for localizing instruments in long-length images builds on the existing framework for CBCT-based navigation using a StealthStation™ (Medtronic, Louisville CO)—work underway will avoid the additional 3D scan by directly registering the tracker to the tomosynthesis image. Tracked instrument positions were mapped to the CBCT coordinates frame using fiducial point registration, and the CBCT image was registered to coronal/sagittal long-length images via 3D-2D image registration [1]. Two approaches for navigation were evaluated: (1) “static” focal plane, in which long-length images are reconstructed at the O-arm isocenter; and (2) “dynamic” focal plane reconstruction, which adjusts the focal plane according to the current position of the navigated instrument.

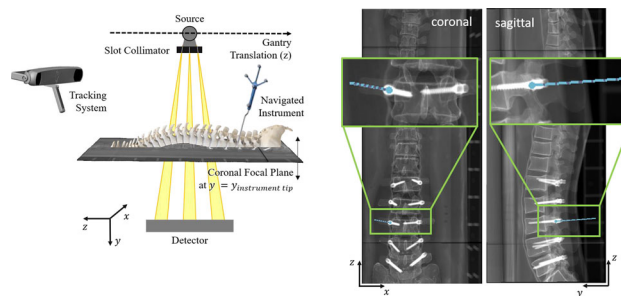


Fig. 1 Illustration of system geometry for long-length tomosynthesis imaging along with example coronal and sagittal image reconstructions. Dynamic focal plane navigation involves the reconstruction of (coronal and sagittal) images at the focal plane defined by the navigated instrument tip position

Geometric accuracy of the proposed navigation approaches was evaluated on a full-length spine phantom (Sawbones, Vashon WA) using CBCT images with (40 × 40 × 16) cm FOV and long-length tomosynthesis images with 51 cm length. Target registration error (TRE) was measured on a bar with 9 target fiducials, placed alongside the spine:

$$TRE_i = \sqrt{|(x_{i,C}, z_{i,C}) - (x_{i,NAV}, z_{i,NAV})|^2 + |(y_{i,S}, z_{i,S}) - (y_{i,NAV}, z_{i,NAV})|^2}$$

where the TRE for each target *i* is measured using the manually segmented positions on coronal ($x_{i,C}, z_{i,C}$) and sagittal ($y_{i,S}, z_{i,S}$) tomosynthesis images and the registered instrument tip position ($x_{i,NAV}, y_{i,NAV}, z_{i,NAV}$). The measurements were repeated 9 times, each time shifting the phantom along the *x* direction.

Results

Table 1 reports the geometry accuracy of static and dynamic focal plane navigation with respect to the standard CBCT navigation. The static approach was observed to suffer from the geometric distortions, which resulted in TRE values ranging 0.8–22.8 mm, subject to the target offset from the focal plane. The dynamic focal plane approach was able to absolve these errors and achieve a median TRE of 1.4 mm across the 51 cm length of the spine.

TRE of the dynamic approach within central region (within CBCT FOV) yielded 1.1 mm median TRE, comparable to the standard CBCT-based navigation (0.8 mm). A slight increase in error was observed in the peripheral (inferior + superior) regions, which resulted in 1.4 mm TRE across the 51 cm length of the spine. This is likely attributed to mechanical effects during image acquisition, such as, gantry sag, which can be accounted for via a direct registration of tracker to segments within long-length images.

Table 1 Geometric accuracy of the navigation methods was measured in terms of TRE (median and CI95) across overall, central, and peripheral regions: overall (all nine fiducials placed along the z-axis); central (fiducials within the FOV of CBCT); and peripheral (fiducials within the long-length FOV and outside that of CBCT)

	CBCT	Static	Dynamic
TRE (overall)	0.8 (1.1) mm	8.8 (17.4) mm	1.4 (3.0) mm
TRE (central)	0.8 (1.1) mm	7.1 (14.9) mm	1.1 (1.5) mm
TRE (peripheral)	N/A	9.2 (18.3) mm	1.9 (3.0) mm

Conclusion

The presented approach demonstrates a novel surgical navigation method on a new long-length radiographic imaging modality that is well-suited for spine surgery. In contrast to navigation on a single, static focal plane, the approach achieved 1.4 mm accuracy, comparable to conventional CBCT navigation, while providing valuable extended spatial context to guide instrumentation of long spine constructs beyond the CBCT FOV. The solution also offers to reduce time and radiation dose by avoiding multiple 3D CBCT scans across the full length of spine. Future work will perform direct registration of the tracker and long-length tomosynthesis image to avoid dependency on an additional CBCT scan and account for mechanical effects during image acquisition to improve accuracy in peripheral regions.

References

- [1] Zhang X, Uneri A, Wu P, Ketcha MD, Jones CK, Huang Y, Lo SL, Helm PA, Siewerdsen JH (2021) Long-length tomosynthesis and 3D-2D registration for intraoperative assessment of spine instrumentation. *Phys Med Biol* 66(5):055008.
- [2] Zhang X, Uneri A, Wu P, Ketcha MD, Doerr SA, Jones CK, Helm PA, Siewerdsen JH (2020) Multi-slot extended view imaging on the O-arm: image quality and application to intraoperative assessment of spinal morphology. *Proc SPIE Medical Imaging* 11315.

Laplacian mesh-based surface deformation recovery using scene flow for robotic minimally invasive surgery

J. Chen¹, K. Hara², I. Sakuma^{1,2,3}, E. Kobayashi³, N. Tomii³

¹The University of Tokyo, Bioengineering, Tokyo, Japan ²The University of Tokyo, Medical Device Development and Regulation Research Center, Tokyo, Japan ³The University of Tokyo, Precision Engineering, Tokyo, Japan

Keywords Surface deformation recovery, Scene flow, Laplacian mesh, Occlusion

Purpose

In robotic minimally invasive surgery (RMIS), analyzing tissue surface deformation is critical for mitigating the risk of tissue damage and quantitatively evaluating surgery performance. The stereo image-based deformation recovery method can be easily applied to RMIS as no extra imaging device other than a binocular camera is required. However, previous works in surface deformation recovery using stereo images for RMIS suffered from the occlusion caused by the surgical instrument. In the occluded areas, the pixels of the target surface cannot be observed, so its 3D structure cannot be directly reconstructed. The occlusion problem can be partially alleviated using surgical instrument segmentation. However, segmentation cannot perfectly label all instrument pixels and the remaining pixels that do not belong to the target surface still cause large deformation recovery errors. To successfully recover the surface 3D structure in the occluded areas, estimation methods using the deformation in the observable areas have been developed. However, many existing solutions rely on the known biomechanical property of the tissue, which is not available in an actual surgical setting. To address these issues, we propose a Laplacian mesh-based method using scene flow for surface deformation recovery. The proposed method is robust to the occlusion caused by the surgical instrument and has been evaluated in a phantom experiment.

Methods

A. Scene flow

We first use stereo matching proposed by Chang et al. [1] to obtain the disparity maps from the input stereo videos. With the disparity map and the calibrated camera parameters, the 3D locations of features are reconstructed using the triangulation algorithm. Second, for each pixel in the current frame, its corresponding pixel in the next frame is found using optical flow proposed by Hui et al. [2]. Scene flow is calculated by combining the 3D reconstruction and optical flow results. It indicates the displacement of each 3D point between frames and is used to drive the deformation of the reconstructed surface.

B. Outlier detection

The scene flow from the previous step contains outliers caused by occlusion, specular highlights, and duplicated textures. Outliers in the scene flow cause large errors in the deformation recovery. Therefore, before using the scene flow to update the mesh surface, outlier detection has to be carried out. The proposed outlier detection method analyzes the gradient of the scene flow. The gradient is used to form a 3 by 3 infinitesimal strain tensor. Singular value decomposition is performed on the tensor to find its maximal principal component (MPC). The corresponding MPC of an outlier is usually larger than an empirical threshold between 1 and 2, while that of the valid scene flow is smaller.

C. Mesh update

The proposed method begins with a dense mesh reconstruction from the 3D reconstructed points without occlusion from the first frame of the input stereo video. In the following frames, the mesh is updated using the scene flow. To maintain surface smoothness while allowing local deformation, we transform the mesh from the absolute Cartesian coordinates into the delta coordinates using a Laplacian matrix derived from 2-ring neighbors. For each vertex with a valid scene flow vector, a new location is calculated by adding up its current location and the scene flow vector. These new locations are used as constraints to form a new linear system combining the initialized Laplacian matrix and delta coordinates. Solving the linear system, we find the new locations of each vertex in a new frame.

D. Error compensation

Over frames, errors between the updated vertex locations and the reconstructed 3D points will accumulate, especially for the vertices once in the occluded areas. To mitigate the error, for a vertex with an invalid scene flow vector in the previous frame, if its scene flow vector becomes valid in the current frame, it will be modified to the projected point in the plane formed by the three closest points in the reconstructed point set.

E. Phantom experiment

A tissue phantom was stretched during the experiment. A calibrated binocular camera was used to record the deformation procedure. The video was used to generate 3D point sets for reference. After that, virtual forceps were added into the video, causing occlusion on the phantom surface. The video with the forceps was

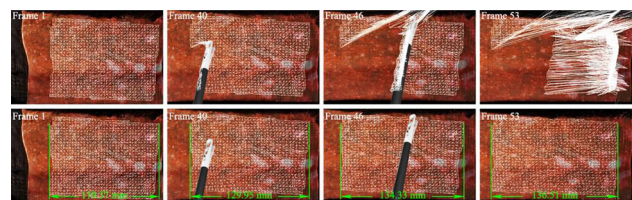


Fig. 1 Mesh surface (with white edges) recovery results, visualizing with low mesh density for clarity. First row: mesh failure when directly using the scene flow. Second row: results of the proposed method

Table 1 Means and standard deviations of errors

Error (mm)	All		x-direction		y-direction		z-direction	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
The proposed method	0.66	1.07	0.15	0.30	0.09	0.21	0.57	1.04
Without error compensation	0.82	1.12	0.17	0.28	0.12	0.22	0.72	1.11

used as the input for the proposed method. We then calculated the errors between the output mesh surfaces and the reference 3D point sets. Error is defined as the Euclidean distance between a vertex and its projected point in the plane formed by the three closest points in the reference point set.

Results

As shown in Fig. 1, the recovered mesh surface from the proposed method overcame the occlusion caused by the forceps. Surfaces were successfully reconstructed in the occluded areas. Table 1 shows the overall errors between the recovered mesh surface and the reference 3D point sets. The errors of 50,302 vertices in 103 frames were measured. The error indicates the fitness between the recovered mesh surface and the reference point set. The overall error of the proposed method was 0.66 mm. Among the *x*, *y*, and *z* directions, the errors were 0.15, 0.09, and 0.57 mm, respectively.

Conclusion

The proposed method successfully recovered dense mesh surfaces from each frame of the input stereo video. Visualization and quantitative results show that the method is robust to occlusion caused by surgical forceps. Surfaces in the occluded areas could still be reconstructed and maintain smooth structures. The overall error of the recovered surfaces is 0.66 mm compared to the reference point sets. The proposed method is promising to handle surface deformation recovery in RMIS with occlusion. This study was partly supported by the JSPS–NNF joint research project (No. 120197410).

References

- [1] Chang JR, Chen YS (2018) Pyramid stereo matching network. CVPR 5410–5418.
- [2] Hui TW, Loy CC (2020) LiteFlowNet3: resolving correspondence ambiguity for more accurate optical flow estimation. ECCV 169–184.

Current issues and future prospects of stereotactic micro-radiosurgery for acoustic tumors

M. Hayashi¹

¹Tokyo Women's Medical University, Neurosurgery, Tokyo, Japan

Keywords Gamma knife, acoustic tumor, microanatomy, NF2

Purpose

Gamma knife surgery (GKS), is the most accurate treatment among the various kinds of stereotactic radiation therapy available Gamma knife surgery (GKS) is seen as the most accurate and, having been widely used for the treatment of acoustic tumors, its results have been well documented with a lot of literatures [1, 2]. According to its relative long term clinical results, tumor progression free rate was 91–97%, hearing preservation rate was 49–55%, and facial nerve preservation rate was 93–100%. These results were not inferior to those of surgical resection. Based on such data GKS has been, and were recognized as a common sense option for general

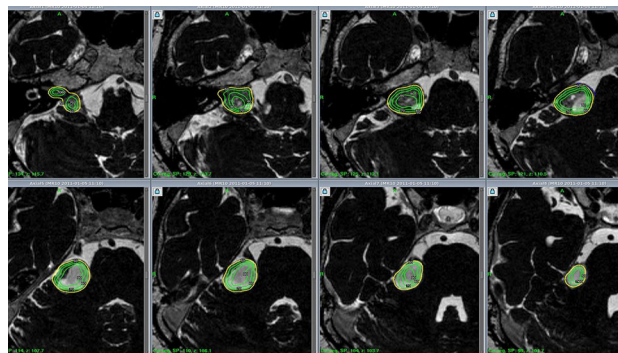


Fig. 1 This figure demonstrated the dose planning for acoustic tumor (Koos stage 4) in Gamma Plan which was dedicated software in Gamma knife surgery. The tumor was covered with 50% isodose line which was separated from the facial nerve and the brain stem

neurosurgeons. However, these detail was not yet mandatory to define critical indications for this data as we have only 25 years follow up with current unique dose planning using the prescribed marginal dose (12–13 Gy). Therefore, we do not have a definite treatment consensus for younger patients who are less below the age of 50 years yet with life reliability.

Methods

Recently, dose planning based on well developed MR imaging preciously taking account precise knowledge of microanatomy has become possible, enabling critical separation of not only the facial nerve, but also the cochlea nerve from the radiation field (50% isodose line) critically, and the origin of tumor could be suggested from superior vestibular, inferior vestibular, or chochlea nerve sheath.

Results

We propose the indications for stereotactic radiosurgery for acoustic tumors as follows; for large size tumors (Koos stage 4), we strongly recommend surgical resection. For small-middle size tumors (Koos stage 1–3), follow-up observation is fundamentally negative, and we recommend surgical resection for the patients with serviceable hearing who are less than 50 years old, and then recommend stereotactic radiosurgery for other patients with serviceable hearing who rejected surgery and are more than 50 years old (Fig. 1).

Conclusion

Especially, for neurofibromatosis type 2 tumors, we have very positive clinical results that serviceable hearing preserved in 91% (15/16) patients with Gardner and Robertson class 1. We strongly recommend stereotactic radiosurgery as early as possible in order to keep retain the auditory function.

References

- [1] Tsao MN, Sahgal A, Xu W, De Salles A, Hayashi M, Levivier M, Ma L, Martinez R, Régis J, Ryu S, Slotman BJ, Paddick I.: Stereotactic radiosurgery for vestibular schwannoma: International Stereotactic Radiosurgery Society (ISRS) Practice Guideline. J Radiosurg SBRT 5(1), 5–24, 2017.
- [2] Horiba A, Hayashi M*, Chernov M, Kawamata T, Okada Y: Hearing Preservation after Low-dose Gamma Knife Radiosurgery of Vestibular Schwannomas. Neurol Med Chir (Tokyo). 56(4),186–192, 2016.

Clinical results and challenges of photodynamic therapy for glioblastoma

M. Nitta¹, T. Kobayashi¹, T. Saito¹, S. Tsuzuki¹, A. Fukui¹, S. Koriyama¹, A. Kuwano¹, T. Kawamata¹, Y. Muragaki¹

¹Tokyo Women's Medical University, Neurosurgery, Shinjuku-ku, Japan

Keywords photodynamic therapy, glioblastoma, immunotherapy, clinical results

Purpose

Photodynamic therapy (PDT) is a local therapy based on cell death and feeding vascular thrombosis, and is one of the important treatment option for lung cancer, esophageal cancer, and primary malignant tumors in Japan. Recently, immunostimulation by PDT has also been suggested. Since January 2014, we have performed more than 200 cases of PDT surgery mainly for glioblastoma.

Methods

The treatment results for initial and recurrent glioblastoma were retrospectively analyzed, and the effect of combined use with immunotherapy was examined.

Results

For primary glioblastoma, 30 cases with PDT and 164 cases without PDT were compared, PFS was 19.6 months vs. 9.0 months ($p = 0.016$), and OS was 27.4 months vs. 22.1 months ($p = 0.03$). A significant prognosis-prolonging effect was observed in the treatment group (Nitta et al. JNS 2019).

Similarly, for recurrent glioblastoma, 70 cases with PDT and 38 cases without PDT were compared. PFS after reoperation was 5.7 months vs. 2.2 months ($p = 0.004$), and OS after reoperation was 16.0 months vs. 12.8 months ($p = 0.031$). A significant prognostic effect was observed in the PDT group. In multivariate analysis, PDT and KPS were prognostic factors. However, there are many cases of recurrence in the early postoperative period, which is a future issue.

Furthermore, when the clinical results of 19 patients with primary glioblastoma who underwent immunotherapy with an autologous tumor vaccine and PDT in combination were examined, the median PFS was 17.4 months and the median OS was 63.1 months, which are extremely good results exceeding 5 years. It is suggested that immune activation by PDT may be a long-term survival factor.

Regarding adverse events, although there were no postoperative complications that could be concluded to be caused by PDT, we have experienced cases of contrasted lesions due to inflammation and edema in the surrounding area after a certain period of time after excision using PDT. The possibility of reaction by PDT cannot be denied.

Conclusion

PDT is an effective treatment for newly diagnosed and recurrent glioblastoma, and is particularly effective in combination with immunotherapy.

Real-time continuous navigation with 3D/4K exoscope for brain tumor surgery

K. Saito¹, N. Sasaki¹, Y. Seiya¹, R. Onoda¹, K. Kobayashi¹, H. Nakatomi¹, Y. Shiokawa¹, M. Nagae¹

¹Kyorin University Faculty of Medicine, Department of Neurosurgery, Tokyo, Japan

Keywords navigation, exoscope, electromagnetic navigation, brain tumor

Purpose

Maximal resection without neurological deficit (maximal safe resection) is required in brain tumor surgery. Microscopic brain tumor resection is the most common while high-definition 3-dimensional exoscope (3D/4K exoscope) is increasingly used. Heads-up surgery with exoscope allows surgeons to set their sights on multi-monitors simultaneously and, together with neuro-navigation and electrophysiological monitoring, is capable of information integrated neurosurgery.

Neuro-navigation supports neurosurgeons to achieve maximal safe resection by showing the location of tumor, eloquent brain fibers, and brain vessels. Electromagnetic (EM) navigation, in contrast to optical tracking navigation, enables continuous guidance despite interruptions in the line of sight. Further, the latest EM navigation has overcome magnetic interference of metal head frame.

Using 3D/4K exoscope and EM navigation, information integrated surgery with real-time continuous navigation can be achieved. Here we present a novel system with 3D/4K exoscope and EM navigation for brain tumor surgery.

Methods

This is a retrospective analysis of 33 patients with brain tumor who underwent tumor removal using exoscope (ORBEYE; OLYMPUS or Hawk Sight; Mitaka Kohki Co) with EM navigation (Stealth Station S8; Medtronic) at Kyorin University Hospital between July 2020 and October 2021. Malleable tracking suction instruments were used for EM continuous navigation. All patients are fixed with metal head frames during the surgery. CT and MR images used for navigation were obtained with a ultra-high-resolution CT scanner (Aquilion Precision; Canon Medical Systems) and a Vantage Galan ZGO; Canon Medical Systems), respectively. A 55-inch 4K 3D monitor was positioned in front of the operator, and a navigation monitor and a monitor of electrophysiological monitoring was placed on either side so that the surgeon could see all the monitors simultaneously.

Results

We included 33 patients (15 men and 18 women; age range, 16–82 years; median age, 49 years; 12 glioblastoma, 13 lower grade glioma, 3 metastatic brain tumor, 1 meningioma, and 4 others). Of these 33 patients, 27 underwent surgery under general anesthesia and 6 underwent awake craniotomy.

Continuous tracking by EM navigation was successful in 33 of 33 cases (100%) despite the use of metal head frames. The improved performance of EM emitter of Stealth Station S8 system seemed to overcome the metal interference. The surgeons could always recognize location data obtained by real-time EM navigation during exoscopic tumor removal. On the other hand, surgeons should look away from the microscope while checking a navigation monitor during microscopic tumor removal.

Further, ultra-high-resolution images could visualize tiny blood vessels even perforator. When integrated to the navigation system, surgeons could have a clear image of tumor and surrounding vessels during surgery and find it easy to preserve important blood vessels.

Supra-total resection or gross total resection was achieved in 9 (75%) of the patients with glioblastoma. Surgical morbidity included hemiparesis in 1 (3.0%) patient, hemianopsia in 1 (3.0%) patient. Postoperative infarction was observed in 2 (8.0%) patients with high grade glioma, which was significantly lower compared to 23 of 77 (29.9%) patients with glioblastoma who underwent microscopic tumor resection ($p < 0.05$).

The limitations of this system are as follows. First, the accuracy of the navigation becomes lower and lower because of the brain shift during the tumor removal. The images integrated to navigation should be updated using intraoperative MRI and so on. Second, the malleable suction instruments are not suited for the fine neurosurgical techniques compared to the normal metal suction tubes. The development of improved EM trackable instruments is warranted.

Conclusion

Using high-resolution exoscope and the state of art EM navigation, information integrated surgery with real-time continuous navigation can be achieved. This novel system is highly useful for maximal tumor resection, avoiding ischemic complications, and preserving brain function in brain tumor surgery.

Clinical utility of intraoperative long-film imaging for thoracolumbar fusion surgery

A. Johnston¹, Y. Huang¹, A. Uneri¹, A. Sacino², R. Planchard², A. Liu², T. Witham², J. Siewerdsen^{1,2}

¹Johns Hopkins University, Department of Biomedical Engineering, Baltimore, United States ²Johns Hopkins University, Department of Neurosurgery, Baltimore, United States

Keywords Long-film, Spinal Deformity, Imaging, Imaging Systems

Purpose

Long-length imaging is an important modality for preoperative planning in spinal deformity surgery. The long field of view (FOV) aids in planning the deformity correction via analysis of global spinal alignment (GSA) parameters. Intraoperative long-length spinal imaging would be similarly beneficial for the assessment of deformity correction while the patient is still on the operating table. Previous work on long-length intraoperative imaging includes stitching multiple fluoroscopic frames but is subject to workflow challenges and parallax error. Mobile cone-beam CT (CBCT) is also widely prevalent but does not normally cover a long FOV and carries a considerable radiation dose.

A novel approach has recently emerged for 2D “long-film” (LF) imaging on a mobile CBCT system (O-arm™ Imaging System, Medtronic) with a multislot collimator and longitudinal translation of the gantry to acquire AP and lateral images with up to 50 cm FOV and dose comparable to a standard radiograph. The long FOV could enable surgeons to assess the deformity correction in the operating room. We evaluate the performance and workflow of LF imaging in thoracolumbar fusion, including integration of LF imaging with deep learning techniques to automatically label vertebrae and analyze GSA.

Methods

LF Imaging System. The O-arm includes a multi-slot collimator that is automatically positioned during LF imaging and translated longitudinally during pulsed x-ray exposure and detector readout to acquire a LF image with up to 50 cm FOV. The radiation dose for LF imaging was measured with an air ionization chamber in terms of dose-area product (DAP).

Clinical Study. The utility of LF imaging was assessed in a clinical study under an IRB-approved protocol including 35 patients undergoing thoracolumbar spine surgery. Outcome measures include analysis of automatic vertebrae labeling, GSA measurement, and workflow (below).

Automatic Spine Labeling. Automatic vertebral labeling was performed (for the first N = 8 subjects from the clinical study) using a region-based deep neural network that combines information from multi-slot projection data to simultaneously label vertebrae in AP and lateral LF images via multi-instance detection [1]. The accuracy of labeling was evaluated in comparison to expert-defined ground truth, and GSA metrics—lumbar lordosis (LL) and medial thoracic kyphosis (MThK)—calculated automatically using an algorithm based on spline fits to the spine labels (SpNorm) [2].

Workflow Analysis. The workflow analysis involved recording the total LF imaging time, defined as from wheels-in to wheels-out of the

O-arm in the operating room, including a breakdown of sub-steps: tableside positioning, definition of LF start and stop positions, AP scan, lateral scan, and removal from tableside. In cases involving both CBCT and LF imaging, the total imaging time (CBCT + LF) was also measured.

Results

Dose measurements showed LF imaging with a 50 cm long FOV to give reference point DAP = 816 mGy.cm², was equivalent to ~ 2.5 s of fluoroscopy. Figure 1 shows example AP and lateral LF images from the clinical study along with automatic vertebral labels computed by the deep neural network and GSA computed via SpNorm. Automatic vertebrae labeling was 95.5% accurate—with 85/89 vertebrae identified as true-positive detection/classifications, 2 false-positive detections, and 2 false-negative (missed) detections owing to obstruction of upper thoracic vertebrae by the shoulders in the lateral view. Two errors were associated with a lack of enforcing the natural anatomical order of the spine in the current algorithm. Automatic evaluation of GSA metrics (LL and/or MThK) was performed using the vertebral labels—for example, in the case of Fig. 1: LL = 38.9° (compared to 38.3° measured manually by a trained reader). Moreover, the intraoperative GSA enables comparison to GSA in preoperative images (in this case, an increase from LL = 24.9°) to assess agreement with the surgical plan.

Table 1 summarizes the LF workflow in terms of imaging time. The average total LF imaging time was 16.6 min (range: 12.1–23.1 min), with 83% of the time owing to tableside positioning and defining the LF start / stop positions. By comparison, the average total time to obtain a CBCT scan (a process for which the radiology technicians were well experienced) was 12.5 min. The LF imaging time was observed to continually decrease over the first 8 cases, reflecting the learning curve of the technologists (e.g., 32.3 min for subject #1, compared to 13.1 min for subject #8). In cases for which the O-arm is already used for CBCT (e.g., for 3D confirmation of instrumentation placement), acquisition of a LF image only added 8.5 min (range 5.4–13.4 min).

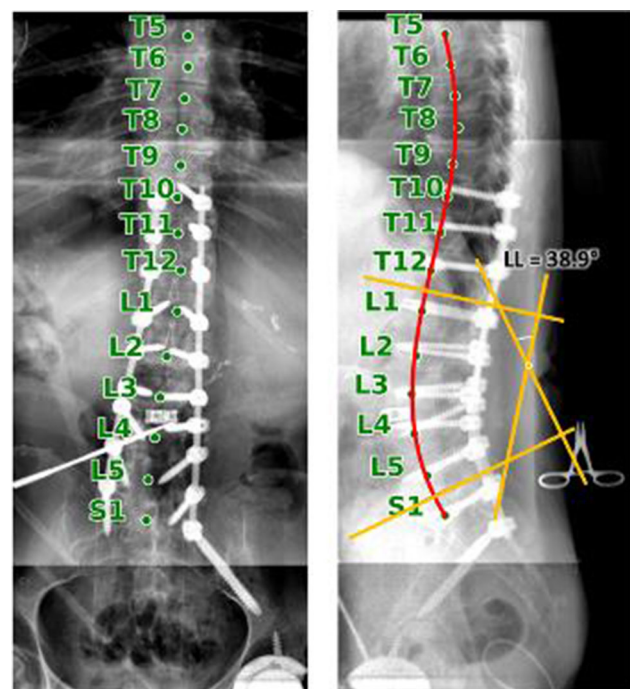


Fig. 1 Example AP and lateral LF images with automatic vertebrae labeling and analysis of GSA

Table 1 Workflow analysis: time required for CBCT and LF imaging, including sub-steps

N= 8 subjects	Total Imaging Time (CBCT+LF)	Total Imaging Time	LF Imaging Process Steps					
			Tableside Positioning	Define LF Start Position	Define LF Stop Position	LF Image Scan (AP)	LF Image Scan (Lateral)	Remove from Tableside
Average (min)	25.0	16.6	4.0	4.0	1.7	1.4	1.4	4.1
Maximum (min)	32.4	23.0	8.0	7.2	2.5	1.7	2.3	5.0
Minimum (min)	20.0	12.1	1.4	2.3	1.3	2.0	1.1	2.1

Conclusion

LF imaging in thoracolumbar fusion surgery provides intraoperative AP and lateral images of the spine covering a long FOV at a low radiation dose (equivalent to a few seconds of fluoroscopy). Automatic vertebrae labeling showed 95.5% accuracy and GSA measurement in 8 cases from an ongoing clinical study.

As technologists became more familiar with LF imaging, workflow time was reduced and comparable to obtaining a CBCT. Patient positioning proved to be a limitation for some GSA measurements, where the cervical spine could not be reached in the FOV without repositioning the arms to a neutral position, which is undesirable for the surgical setup. Ongoing work investigates positioning the arms consistent with surgical requirements and better accommodating LF imaging of the cervical and upper thoracic spine.

Acknowledgements

This study was supported by a research agreement with Medtronic.

References

- [1] Huang, Y., Jones, CK., Zhang, X., Johnston, A., Aygun, N., Witham, TF., Helm, PA., Siewerdsen, JH., Uneri, A. (2022) Automatic labeling of vertebrae in long-length intraoperative imaging with a multi-view, region-based CNN, *SPIE. Med. Imaging 2022 Image-Guided Proceed. Robot. Interv. Model.* Paper No. 12034–22.
- [2] Doerr, S. A., De Silva, T., Vijayan, R., Han, R., Uneri, A., Ketcha, M. D., Zhang, X., Khanna, N., Westbroek, E., Jiang, B., Zygorakis, C., Aygun, N., Theodore, N. and Siewerdsen, J. H. (2020). Automatic analysis of global spinal alignment from simple annotation of vertebral bodies. *J. Med. Imaging.* 7(03):1

Development and assessment of a user-friendly method for non-invasive detection of knee implant loosening using 3D-CT image analysis

M. A. ter Wee¹, J. G. G. Dobbe¹, A. J. Kievit¹, G. J. Streekstra¹

¹Amsterdam UMC, Biomedical engineering & physics, Amsterdam, Netherlands

Keywords Knee implant loosening, 3D-CT image analysis, Software development, Methodological error

Purpose

Current diagnostic methods to detect aseptic loosening of the tibial component after Total Knee Arthroplasty are inconclusive or invasive in nature, causing unnecessary (revision) surgery. Alternatively, advanced 3D-CT image analysis technology can be used to detect tibial implant displacement directly under external valgus and varus loading conditions.

In case of loosening, the relative displacement of the implant with respect to the tibia is often smaller than 1 mm, which requires detection at a submillimeter level. The purpose of this study was to develop methodology and user-friendly software for swift image analysis and to assess its accuracy and precision.

Methods

Software for visualization and automated image analysis was implemented in a stepwise approach to: (1) load CT images of femur and tibia in varus and valgus positions, (2) segment the implant and tibia in the valgus scan, (3) register the segments to the varus scan and (4) measure the relative displacement of the implant with respect to the tibia. Registration and threshold-connected region growing were used as in [1].

Accuracy and precision were assessed by repetitively scanning (N = 10) a cadaver specimen with total knee implant with a Brilliance 64-channel CT scanner (Philips Healthcare, Best, The Netherlands) (isotropic voxel spacing of 0.45 mm), including the partial femur and whole tibia, without external loading. The images were analyzed as described above, where any apparent displacement of the prosthesis with respect to the tibia can be attributed to methodological error. The mean displacement served as measure of accuracy, the standard deviation served as measure of precision. The norm was used to report residual translations along the x-, y- and z-axes, and rotations about the x-, y-, and z-axes.

In the standard approach, we segmented the tibia component and the tibial cortex (excluding the trabecular bone) and registered these models to the varus scan. Three alternative approaches were evaluated with this standard approach with the goal to speed up the analysis procedure while evaluating the corresponding effect on accuracy and precision: (1) segmentation of the tibia including the trabecular bone, (2) registration with a reduced number of mesh points of the tibia and (3) registration using only the proximal 20% of the tibia.

In addition to the accuracy and precision experiments, we evaluated the extent of tibia bending between valgus and varus loading in one patient scan by quantifying the relative displacement of the proximal 20% with respect to the distal 20% of the tibia. In an attempt to minimize the effect of tibia bending in detecting implant loosening we quantified and compared the observed implant displacement for this patient when using either the whole tibia or the proximal 20% in our analysis.

Independent t-tests were performed to assess the level of significance in our experiments.

Results

Comparing the accuracy and precision of the three alternative image analysis approaches to the standard approach is summarized in Table 1. Segmentation including the tibial trabecular bone did not show a significant difference in accuracy, precision or image analysis time compared to the standard approach.

Table 1 Overview of the accuracy, precision and total time of image analysis. $p < 0.05$ indicates a significant difference to the standard approach (first column). Total analysis time and change (%) compared to standard are given in the last row

		Standard approach				
		Excl. trabecular bone 100% resampling 100% tibial length	Incl. trabecular bone 100% resampling 100% tibial length	Excl. trabecular bone 1% resampling 100% tibial length	Excl. trabecular bone 0.1% resampling 100% tibial length	Excl. trabecular bone 100% resampling 20% tibial length
Translation [mm]	Accuracy	0.03	0.03	0.03	1.19	0.03
	Precision	0.01	0.02	0.01	3.17	0.01
	p (2 tail)		0.84	0.96	0.11	0.60
Rotation [°]	Accuracy	0.09	0.09	0.09	0.48	0.10
	Precision	0.05	0.05	0.05	0.92	0.05
	p (2 tail)		0.99	0.91	0.07	0.51
Time [s]	Total analysis time	250	260	140	138	160
			(4% increase)	(44% reduction)	(45% reduction)	(36% reduction)
	p (2 tail)		0.94	<0.01	<0.01	<0.01

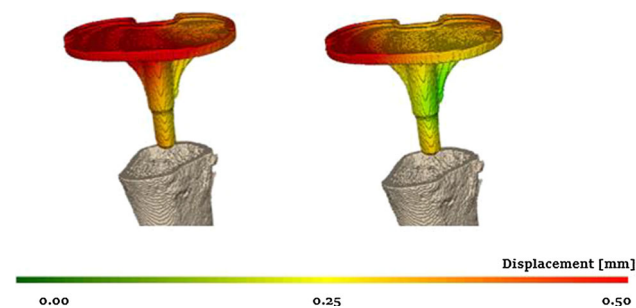


Fig. 1 Heatmaps showing observed implant displacement across the surface when the whole tibia is used for registration (left; larger bending effect included) or the proximal 20% (right; less bending included)

Resampling the tibia contour to 1.0% of approx. 500.000 points did not change the accuracy or precision, but reduced the total image analysis time with 110 s. Further resampling to 0.1% resulted in data outliers which increased the level of error while the image analysis time hardly improved (2 s) compared to 1% of the mesh points.

Registration using the proximal 20% tibia or the entire tibia length showed no difference in accuracy or precision.

Tibial bending showed a relative displacement of 0.4 mm and 0.2 of the proximal tibia with respect to the distal tibia between valgus and varus loading. This relative displacement implies bending of the tibia and affects measurement of implant displacement. During surgery the implant appeared to be loose. Including the entire tibia in the detection of implant loosening showed larger displacements (0.46 mm and 0.68) compared to including the proximal 20% of the tibia (0.29 mm and 0.59) (Fig. 1).

Conclusion

The total image analysis time can be markedly reduced by resampling the tibia mesh to 1% of approx. 500.000 points without showing an effect on the accuracy and precision. No difference between in- or excluding the trabecular bone in the tibia segment was found. Since tibial bending interferes with measuring implant displacement, it is recommended to include only a proximal tibia segment in these evaluations, although future research is needed to confirm whether that approach is beneficial.

References

- [1] Dobbe JGG., de Roo MGA, Visschers JC, Strackee SD, Streekstra GJ (2019) Evaluation of a quantitative method for carpal motion analysis using clinical 3D and 4D CT protocols. *IEEE Trans. Med. Imag.*(38):1048–1057.

Predicting facial changes following bony movement via deep correlation for orthogonal surgical planning

X. Fang¹, D. Kim², T. Kuang², X. Xu¹, H. H. Deng², J. Gateno^{2,3}, M. A. K. Liebschner⁴, J. J. Xia^{2,3}, P. Yan¹

¹Rensselaer Polytechnic Institute, Biomedical Engineering, TROY, United States ²Houston Methodist Research Institute, Department of Oral and Maxillofacial Surgery, Houston, United States ³Weill Medical College of Cornell University, New York, United States ⁴Baylor College of Medicine, Department of Neurosurgery, Houston, United States

Keywords Facial simulation, Point sets, Deep learning, Surgical planning

Purpose

Orthognathic surgery is a bony procedure (osteotomy) designed to correct jaw deformities. While facial soft tissue is not directly operated upon, it changes “automatically” to adapt to the bony movement. While surgeons can successfully plan the bony surgery, predicting facial changes following osteotomy is still a challenging task. Facial change prediction using finite-element method (FEM) simulation is reported to be the most accurate method. However, it is computationally expensive and time-consuming, thus being infeasible for near-real-time prediction following bony surgical planning [1]. To solve the problem, in the past we have developed FC-Net, a deep learning-based approach for facial-change prediction [2]. While the computational speed has been significantly improved, the accuracy of FC-Net also needs to be improved. Therefore, in this work, we propose a deep learning approach with cross point-set attention to enhance the prediction accuracy.

Methods

An attention based deep learning framework, Motion Transformer Network (MotionFormer), is developed to predict the facial changes (the movement of facial points) from the movement of bony segments (points). MotionFormer takes three inputs, pre-operative facial points, bony points, and bony movement vectors, to predict the movement of the facial points. The overview of the developed MotionFormer is shown in Fig. 1.

MotionFormer first extracts the structural features of pre-operative facial and bony points using two PointNet++ networks, respectively. Facial and bony point coordinate values are normalized into

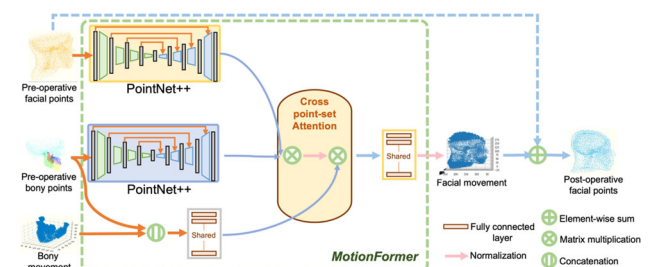


Fig. 1 Overview of the proposed Motion Transformer Network (MotionFormer)

the range of [0,1] before feeding into the networks. Then, shared fully connected (FC) layers are used to transform the bony movement into bony movement features. All these features are fed to a cross point-set attention module. In the cross-set attention module, facial change features are calculated in two steps. In the first step, a dot product similarity is computed between the extracted facial features and bony features, which represents the relationship, i.e., affinity, between bony and facial points. The similarity is further normalized by the number of bony points. In the second step, the normalized similarity matrix is used to transform the bony movement features into the facial movement features. Sequentially, the shared FC layers reduce the dimension of the facial movement feature to 3, which is then normalized into [-1,1] to predict facial movement vector. Finally, the locations of post-operative facial points following bony change are predicted by adding the computed facial movement to the pre-operative facial points. All post-operative facial points are scaled back to the physical coordinate system by using the scale factor computed from the input points. We trained the proposed MotionFormer by adopting the same loss function as FC-Net, which consists of a shape loss, a point density loss, and a local-point-transform loss.

The performance of MotionFormer was evaluated using 40 sets of retrospectively analyzed patient data, 32 for training and 8 for testing. The data was randomly selected from our digital archive of patients who had undergone double-jaw orthognathic surgery (IRB# Pro00008890). Both facial and bony surface of post-operative CT were registered to the pre-operative ones based on surgically unaltered bone volumes. Virtual osteotomies were performed on pre-operative bone surface based on post-operative one. Then, the bony movement of each bony segment, i.e., surgical plan for the actual surgery, were retrospectively established by manual registration. The post-operative face served as the ground-truth for evaluation. For efficient training, 4096 points were down-sampled from each original pre-operative bony and facial segments, respectively.

The prediction accuracy of MotionFormer was evaluated quantitatively using surface deviation between the predicted and post-operative faces. Average surface deviation was calculated for 6 individual facial regions, including the nose, upper lip, lower lip, chin, right cheek, and left cheek, defined by the facial landmarks. In addition, the results of our proposed method were also compared with our previous FC-Net. A one-tailed paired t-test was used for statistical comparison. Finally, computational time for predicting each patient was recorded.

Results

The accuracy results of our proposed method and the comparison to FC-Net are shown in Table 1. The results of paired t test showed that our MotionFormer was statistically significantly outperformed the FC-Net ($p < 0.05$) in the upper lip, right cheek, and entire face. The average computational time was around 70s.

Table 1 Comparison between our MotionFormer and FC-Net

Methods	Surface deviation (Mean ± Std, mm)						
	Nose	Upper Lip	Lower Lip	Chin	Right cheek	Left cheek	Entire face
FC-Net [2]	1.66±0.30	2.09±0.98	1.97±0.94	2.21±1.55	1.79±0.38	1.31±0.25	1.68±0.25
MotionFormer	1.33±0.44	1.57±0.73	1.77±0.53	1.52±0.79	1.23±0.55	1.13±0.30	1.31±0.31
p-value	0.066	0.024	0.220	0.144	0.018	0.095	0.027

Conclusion

The MotionFormer successfully predicted the facial change following bony change by utilizing the cross-set attention. The prediction error of the proposed method significantly improved the prediction accuracy in the facial regions including clinically critical region, i.e., upper lip, compared to the state-of-the-art method, FC-Net, while keeping the computational efficiency. In the future, we will develop an enhanced cross point-set attention module by leveraging the distance information between bony and facial points to further improve the prediction accuracy and perform qualitative analysis for clinical evaluation.

References

- [1] Kim D, Kuang T, Rodrigues YL, Gateno J, Shen SG, Wang X, Stein K, Deng HH, Liebschner MA, Xia JJ (2021) A novel incremental simulation of facial changes following orthognathic surgery using fem with realistic lip sliding effect. *Medical Image Analysis* 72, 102095.
- [2] Ma L, Kim D, Lian C, Xiao D, Kuang T, Liu Q, Lang Y, Deng HH, Gateno J, Wu Y, Yang E (2021) Deep Simulation of Facial Appearance Changes Following Craniomaxillofacial Bony Movements in Orthognathic Surgical Planning. *International Conference on Medical Image Computing and Computer-Assisted Intervention*:459–468.

Analysis of tracking performance with two different HoloLens 2 augmented reality markers for orthopedic oncological surgeries

A. Pose Díez de la Lastra¹, R. Moreta-Martínez¹, F. von Haxthausen¹, M. García-Sevilla¹, L. Hernández-Álvarez², J. A. Calvo-Haro³, R. Pérez-Mañanes³, J. Pascau¹

¹Universidad Carlos III de Madrid, Departamento de Bioingeniería e Ingeniería Aeroespacial, Leganés, Spain ²Instituto de Tecnologías Físicas y de la Información del Consejo Superior de Investigaciones Científicas, Grupo de investigación en Criptografía y Seguridad de la Información (GiCSI), Madrid, Spain ³Hospital General Universitario Gregorio Marañón, Servicio de Cirugía Ortopédica y Traumatología, Madrid, Spain

Keywords Augmented reality, Microsoft HoloLens 2, Orthopedic Oncological surgery, Tracking performance analysis

Purpose

In recent years, augmented reality (AR) has increased its adoption in many areas, including medicine. This technology overlaps three-dimensional virtual models onto physical objects in the real world, improving the information available to the user. An important limitation of AR is the registration between the virtual and the natural environment. Although some studies perform this step manually, the resulting accuracy and user-dependent error are unsuitable for clinical applications. Automatic registration is an alternative when a correct alignment is crucial, such as during surgery. In these cases, AR can overlay patient-specific information to the real patient, such as anatomical models or the preoperative plan, assisting the surgeon during the intervention.

Either smartphones or head-mounted displays are suitable options for the deployment of AR applications. Among all, Microsoft HoloLens is considered the best AR platform in surgical interventions [1]. In this work, we analyze Microsoft HoloLens 2 in orthopedic oncological surgeries. More specifically, we evaluate the accuracy of the AR projection performing automatic registration with two AR

markers based on a visual pattern or retroreflective spheres. Our goal was to determine which one offers the better tracking performance to guide osteotomies.

Methods

We simulated a surgical procedure in the laboratory with a 3D printed phantom based on a patient with a pelvic osteosarcoma. This phantom contained a portion of the patient's tumor and the surrounding bone. We created all the virtual models of the patient structures from the preoperative CT on 3D Slicer. In addition, we used Autodesk MeshMixer to create a surgical guide as a negative of the bone's surface to be adjusted in a specific location of the patient. Finally, we developed two different AR markers to evaluate the HoloLens 2 tracking performance in both cases. The surgical guide contained a socket to fit both AR markers, thus enabling an automatic registration between HoloLens and the real world. The markers, phantom, and surgical guide were 3D printed in polylactic acid (PLA) with fused deposition modeling (FDM) technique using a double extruder desktop 3D printer (Ultimaker 3 Extended).

We also developed two different Microsoft HoloLens 2 applications on the Unity platform. The first one used the HoloLens camera and the Vuforia® development kit to recognize a black and white two-dimensional 3D printed pattern. The second app tracked the position and orientation of a rigid body containing three retroreflective spheres with the HoloLens infrared camera.

During the experiments, we used a Polaris Spectra optical tracking system (OTS) as a gold standard for error measurements, recording positions with a tracked pointer and including a reference rigid body in the phantom. This information was directly streamed to 3D Slicer via PLUS toolkit and an OpenIGTLink protocol. The real phantom and the virtual model were registered with point-based registration calculated from eight small conical holes (\varnothing 4 mm \times 3 mm depth) included in the phantom surface. After this registration, we fitted the corresponding AR marker (pattern or spheres) into the surgical guide socket. Once the marker was identified, the app projected virtual models of the patient's tumor, bone structure, and two cutting planes. Orthopedic oncological surgeons designed these planes to perform an osteotomy of the hip along the iliac crest and the acetabular region.

To evaluate the AR projection error, two different users collected several points in the intersection between the virtual cutting planes and the phantom. They repeated this procedure three times, removing and placing back the surgical guide between repetitions. We measured the AR error as the minimum distance from each collected point to

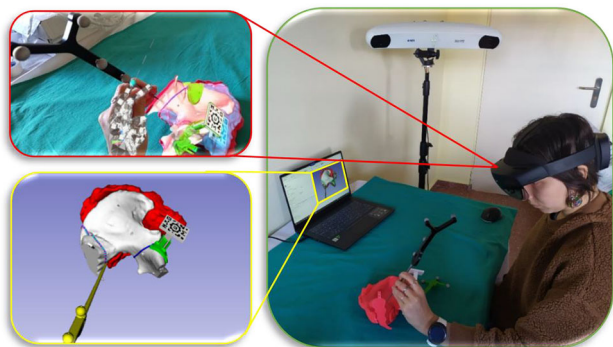


Fig. 1 Measurement of the iliac crest plane projection accuracy using the pattern-based AR marker. The left-hand side of the figure shows the HoloLens 2 app visualization (upper) and the 3D Slicer view during the point acquisition (lower). The Polaris Spectra OTS is in the background of the right-hand side picture

Table 1 Projection errors obtained for each cutting plane relying in either the pattern-based or the spheres-based AR references

	Iliac crest plane [mm]				Acetabular plane [mm]			
	Mean	Std	Median	IQR	Mean	Std	Median	IQR
Pattern-based AR reference	1.225	0.985	0.931	1.346	1.051	0.708	0.964	1.011
Spheres-based AR reference	2.033	1.446	1.821	2.092	1.472	1.077	1.174	1.282

the reference plane. Figure 1 shows the acquisition of the iliac crest plane points using the pattern-based AR reference.

Results

Table 1 summarizes the projection accuracy obtained for each cutting plane using both AR markers. Overall, the pattern-based AR reference offers a more accurate projection of the cutting planes than the spheres-based one: 1.225 ± 0.985 mm versus 2.033 ± 1.446 mm in the iliac crest plane, and 1.051 ± 0.708 mm against 1.472 ± 1.77 mm in the acetabular case. We performed two Mann–Whitney U tests to analyze the data dependence on the AR marker and plane, obtaining a significant difference (p value $<$ 0.05) in both cases.

Conclusion

The results obtained in the laboratory prove that both methods are sufficiently accurate for these surgical procedures, considering the state-of-the-art in osteotomy guidance using AR [2]. In our case, the pattern-based AR marker results are slightly better than the spheres-based. However, the latter would allow combining HoloLens 2 with other surgical navigation frameworks based on optical tracking systems. The next step would be to test our applications in actual surgeries and determine whether the error is still low enough in those uncontrolled scenarios. In this work, we expect to offer a good alternative to current registration issues and an easy solution to apply AR in these scenarios.

Acknowledgements

Supported by projects P118/01625 and PrePlanRT AC20/00120 (Ministerio de Ciencia, Innovation y Universidades Instituto de Salud Carlos III, Association Espanola Contra el Cancer and European Regional Development Fund “Una manera de hacer Europa”).

References

- [1] Azimi E, Qian L, Navab N, Kazanzides P (2017) Alignment of the Virtual Scene to the Tracking Space of a Mixed Reality Head-Mounted Display. arXiv retrieved from <http://arxiv.org/abs/1703.05834>
- [2] Viehöfer AF, Wirth SH, Zimmermann SM, Jaberg L, Dennler C, Föhnstahl P, Farshad M (2020) Augmented reality guided osteotomy in hallux Valgus correction. BMC Musculoskeletal Disord 21:1–6. <https://doi.org/10.1186/s12891-020-03373-4>

Flexible ureteroscope with omnidirectional bending based on frameless-structured tube and crossed control wiring

A. Yamada¹, T. Tani¹

¹Shiga University of Medical Science, Department of Research and Development for Innovative Medical Devices and Systems, Otsu, Japan

Keywords flexible ureteroscope, omnidirectional, frameless-structured tube, crossed control wiring.

Purpose

Flexible ureteroscopes (fURSs) are widely used in minimally invasive diagnostic and therapeutic procedures for upper urinary tract diseases, including tumors, renal stones, and internal obstruction. Geometric parameters and bend shapes vary among fURSs, and comparative studies have been conducted [1]. To expand their range of access, fURSs must be rotated along their long axes because they are only capable of bidirectional bending. This constraint to bend-and-rotate manipulation limits access, especially into the lower calyx, and it is associated with biomechanical stress to operators [2]. Recently, a fURS capable of omnidirectional bending was tested, and its wider access range and shorter access time were assessed using mock procedures [2]. Its right-to-left directional maximum bending angle was only 100°, which represented only 36% of the angle achievable in the up-and-down direction. To overcome these performance and range-of-access limitations, we proposed a fURS with a frameless-structured tube and a unique wire crossing element (WCE). We developed a prototype, evaluated its steering properties, and assessed its feasibility using a kidney phantom.

Methods

The proposed fURS prototype (Fig. 1) comprises a flexible, solid polyamide-12 tube for the body and a stretch-retractable, porous expanded polytetrafluoroethylene (ePTFE) tube with 40% porosity for the distal end. The tubes employ a central lumen and eight wire lumens. The ePTFE tube's distal end was connected to an austenite stainless steel (SUS304) washer with the same lumens. These tubes were connected through a SUS304 WCE.

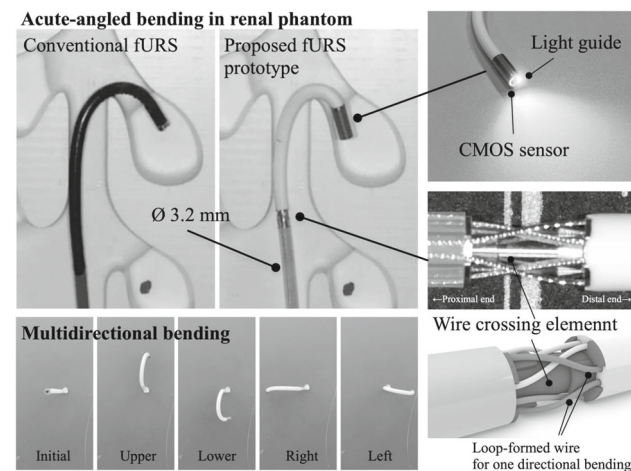


Fig. 1 Flexible ureteroscope prototype (CMOS, complementary metal–oxide–semiconductor; fURS, flexible ureteroscope)

Table 1 Bending results (IST, image sensor type; ABD, active bending direction; UaD, up-and-down; RaL, right-and-left)

Flexible ureteroscope				Tip active bending (degrees)				Passive bending (degrees)	
Name	IST	ABD	Company	1 cm	2 cm	3 cm	4 cm	UaD	RaL
Prototype	Digital	Omnidirectional	Ours	65	175	235	265	275	275
WiScope	Digital	2 (UaD)	OTU Medical	50	110	170	200	185	50
FLEX-X2 (1)	Fiber-Optic	2 (UaD)	Karl Storz	56	119	162	186	No data	No data

A control wire that bends the scope in one direction forms a loop via the insertion of a wire into a wire lumen from the tube's proximal end, bending of the wire into a general U-shape at the washer, and backing of the wire into the adjacent wire lumen. The prototype is equipped with four loop-formed control wires. Each loop-formed wire passes through the wire lumens at different cross-sectional positions on the ePTFE and polyamide-12 tubes because the WCE enables the adjacent loop-formed wires to cross without contacting each other. The polyamide-12 tube's proximal end is connected to a handheld controller with coaxial two angle-adjusting wheels to generate tendon-driven wire motions for bending the distal tube in up-and-down and right-and-left directions.

The controller was fabricated by using a three-dimensional printer. A lens barrel made of SUS304 was connected to the washer, and a complementary metal–oxide–semiconductor (CMOS) image sensor (400 × 400 pixels) and light guides were installed into it. The prototype has no working channel.

We assessed the prototype's bending capabilities. We recorded the maximum bend angle in the active bending of the prototype and comparative fURS (WiScope, OTU Medical Inc., San Jose, CA). The passive maximum bending angles when a payload was applied at their tips were similarly recorded. Then, we measured the maximum tip bend angle for every fURS with the tip extended out from an access sheath mockup at 1-, 2-, 3-, and 4-cm intervals [1]. We investigated the impacts of the wire routing design using the WCE by measuring the WCE's movement distance associated with prototype bending. We performed a feasibility study to assess the prototype's ability to reach a sharp-angled calyx using a flexible ureteroscopy simulator (K-Box, Coloplast, Humlebaek, Denmark).

Results

The prototype's maximum bend angle was 275° in both the up-and-down and right-and-left directions. The prototype accomplished an omnidirectional maximum bend angle of 275° by combining both directional bending planes. Table 1 summarizes the bending results compared with other fURSs. The prototype recorded about a 1.6-fold larger maximal active bending angle with the restricted environment and a 5.5-fold larger passive bending angle. The crossing wire-routing design decreased the WCE's offset distance by about 75% compared with the conventional straight wire routing. This implies that the body tube's shape has less impact on the tip bending performance. This observation suggests that we can steer this fURS without the distal bending performance being degraded even when the fURS body tube has been bent acutely at the fURS insertion point at the patient's body. In the feasibility study using our fURS prototype, we could reach the phantom's acute-angled calyx where only one of the conventional six major CMOS image sensor type fURSs has previously reached [1]. Additionally, the combination of the adaptive shape deformation unique to the ePTFE tube and the balanced omnidirectional bending facilitated access to the upper, middle, and lower calyces omnidirectionally without scope rotation.

Conclusion

Using a frameless-structured flexible tube, we developed a fURS prototype providing balanced omnidirectional bending of the tip and adaptive shape deformation without the need for scope rotation. A WCE installed in the tube for routing the control wires enabled the durable bending performance. This prototype's improved steering

performance, exemplified by its ability to access an acutely angled calyx, could improve treatment outcomes and shorten operation times associated with calyceal stone removal.

References

- [1] Dragos LB, Somani BK, Sener ET, Proietti S, Ploumidis A, Iacoboaie C, Doizi S, Traxer O (2017) Which Flexible Ureteroscopes (Digital vs. Fiber-Optic) Can Easily Reach the Difficult Lower Pole Calices and Have Better End-Tip Deflection: In Vitro Study on K-Box. A PETRA Evaluation. *J Endourol.* 31(7):630–637.
- [2] Tambo M, Inoue T, Miura H, Matsuzaki J, Nutahara K, Hamamoto S, Okada S, Fukuhara H, Fujisawa M, Matsuda T. (2020) A Novel Flexible Ureteroscope with Omnidirectional Bending Tip Using Joystick-Type Control Unit (URF-Y0016): Initial Validation Study in Bench Models. *J Endourol.* 34(6):676–681.

Surgical robot capable of estimating grasping force

K. Tadano^{1,2}, M. Kanazawa¹, T. Nishihara¹, T. Kanno¹, K. Shindo¹, K. Takikawa¹, N. Morita¹, T. Inoue¹, J. Sakata¹, K. Kawashima^{1,3}

¹Riverfield Inc., Tokyo, Japan ²Tokyo Institute of Technology, Yokohama, Japan ³The University of Tokyo, Tokyo, Japan

Keywords Surgical robot, Force feedback, Robotic instrument, Grasping force

Purpose

Robot-assisted surgery has grown in popularity all over the world. Master-follower type robot has advantages such as capability of intuitive operation, applicable to tele-surgery and to microsurgery by changing the motion scale between the master and the slave.

A master-follower surgical robot for minimally invasive surgery named Saroa is developed. The robot can estimate grasping forces with pneumatically driven robotic forceps. The importance of haptic feedback in the surgical robot for MIS is widely recognized [1, 2]. The forces are estimated based on the dynamics and pneumatic pressure changes of the forceps. The accuracy and effectiveness of the estimated forces were evaluated.

Methods

Figure 1 shows the developed surgical robot. The follower robot mainly consists of a holder arm and a forceps manipulator. The six



Fig. 1 Surgical robot with grasping force feedback named Saroa

degree of freedom holder arm is a hybrid drive of pneumatic and electric actuators. The electric motors with reduction gears are used to shoulder and elbow joints of the robot for precise and rigid control. The pneumatic vane motors are used for wrist joints to achieve soft contact to the abdominal wall. The joints are not only controlled by position feedback, but also controlled by torque to compensate for its weight and external forces. The forceps manipulator with wrist joints and a gripper are driven by pneumatic cylinder with high back drivability, and estimates translational and gripping force using a disturbance observer. The estimated force is sent to the master haptic device and displayed directly as a reaction force. The estimated forces by the developed forceps were compared with the forces measured using a force sensor.

We implemented a bilateral control of the surgical robot using a master device. The master device is an electrically-driven 7-DOF link mechanism. It has parallel links for 3-DOF translational motion, a gimbal mechanism for 3-DOF orientation input, and 1-DOF grasping input. All degrees of freedom have electric actuators, where the translational and grasping axes are active in this paper.

A simple force-reflection-type bilateral control was used to control the robot system. The position, orientation, and grasping commands obtained from the master device were sent to the slave side via UDP/IP, and the estimated grasping force were sent back from the follower side. The communication delay was negligible small between the master side and the slave side since they were directly connected in a same room.

Results

We confirmed that the mean absolute error between the measured forces and the estimated grasping forces is 0.05 N or less for any condition. The grasping force of the phantom tissue was less than half that with the haptic display compared to the case without the display.

Conclusion

The performance and accuracy of the estimated grasping force by the surgical robot Saroa was experimentally evaluated. The effectiveness of force feedback of the robot was confirmed using phantom tissue.

References

- [1] Wottawa CR, et al., Evaluating Tactile Feedback in Robotic Surgery for Potential Clinical Application using an Animal Model, *Surg Endoscopy*, 2016, 30(8), pp.3198–3209. <https://doi.org/10.1007/s00464-015-4602-2>.
- [2] Abiri A, et al., Multi-Modal Haptic Feedback for Grip Force Reduction in Robotic Surgery, *Scientific reports*, 2019, 9:5016.

Implementation of a macro-mini robot as a needle insertion guidance system

S. H. Shen¹, N. W. Chang¹, Y. C. Hsieh², H. C. Zuo², F. A. Halim², P. L. Yen²

¹Taipei Veterans General Hospital, Department of Radiology, Taipei, Taiwan ²National Taiwan University, Department of Biomechanics Engineering, Taipei, Taiwan

Keywords Macro-mini robot, Robot Assisted Intervention, Needle Insertion, Navigation Intervention

Purpose

Needle insertion surgery is one of the pioneers in the minimal invasive surgery, for instance cryoablation [1]. Due to the nature of the surgery, it requires the involvement of the technology and give more challenges for the surgeon [2]. One of the challenges is to translate the

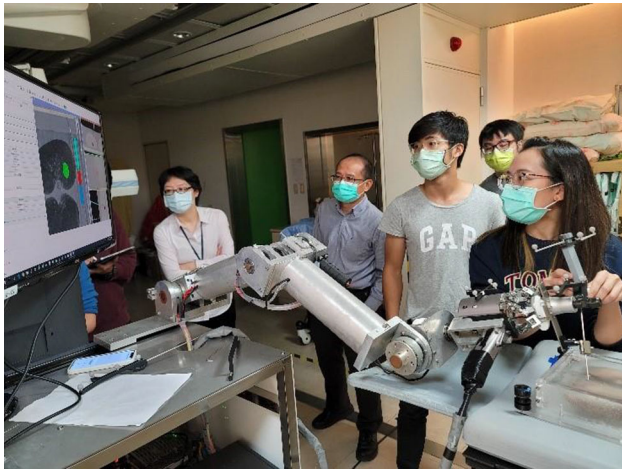


Fig. 1 The Macro-Mini Robot for Needle Insertion Navigation

plan on 2D CT view to the real-world patient. The surgeon capability to guess and get close to the planned path affects the duration of the procedure. Therefore, this study aims to assess the viability of macro-mini robot arm a guidance system for needle insertion to assist the surgeon translating the 2D CT needle plan to the real world and hence get closer to the intended path.

Methods

The system will be based on the CT scan which takes every iteration of needle change. These data will be registered and fused with the previous data sets. The needle entry point and the target point will be assigned on the updated image data. The information regarding their coordinates is delivered to the robot arm and hence the robot can guide the surgeon to the designated path as shown in the CT image plan.

The robot arm is designed adapting the macro mini robot arm which has 11 DOF. It implements the combination of serial and parallel manipulator, respectively. The macro arm is designed to manually moved and can be moved freely. The mini robot module is automated hexapod, which capable to position itself based on the appointed pose. The automated function has an activation trigger upon the macro arm arrived in the designated area. Figure 1 shows the surgeon execute the needle insertion with the guidance of the robot arm.

The User Interface (UI) will utilize the 3DSlicer program, which include the image registration, path planning, until the needle depth tracking visualization. The model is based on the DICOM file from the CT which reconstructed using 3D Slicer function. The registration and tracking process involve NDI to monitor the pose of the phantom and the robot arm. Some sets of markers (passive sensors) are attached to the robot and the phantom base as indicator of their poses.

In the virtual world, the coordinate of intended entry point and target point are selected and the path will be displayed in the model as reference. The Multiplanar view of the needle path allows the surgeon the surgeon to get better view of the safety and eligibility of the planned needle path. When the path is confirmed, the transformation is feed to the controller system to obtain the necessary adjustment of the end effector pose. The required motion-controlled compensation of the mini robot is based on the resolved motion rate. For this specific project, the phantom is a pair of porcine kidneys submerged into wax. The experiment itself is conducted by experienced surgeon, therefore, it would simulate more realistic scenario.

Table 1 Total time consumption on each needle insertion is approximately 15 min

	Trial 1	Trial 2	Trial 3		Average
			1 st needle	2 nd needle	
Base Displacement (mm)	2,4	2,5	2,6	3,9	2,85
Tip Displacement (mm)	4,4	3,1	4,5	7,2	4,8
Insertion Depth (mm)	33,8	30,6	35	32,5	33
Clean procedure time consumption (min)	15	19	8	13	13,75

Results

From 4 needle insertions of 3 different trials, the accuracy of the of the needle base is within 3 mm and the needle tip have less than 5 mm displacement, in which the depth of insertion is on average 33 mm with up to 2° needle deflection. The UI is similar to the traditional approach. The total time consumption on each needle insertion is approximately 15 min. The detailed result shown in the Table 1. The performance of the first 3 needle insertions are quite stable with certain amount of stability and accuracy. However, an outlier of accuracy data shown in the trial 3 2nd needle, which reflect the situation of overextend of the arm during the experiment.

Conclusion

The macro-mini robot arm has potential in guiding the needle insertion process. This preliminary design able to guide the first needle insertion closer to the target path with good procedure time. The current project is limited with rigid phantom and the arm have limited accessibility, therefore, the future project will focus on increase the stability and accuracy while adapting the non-rigid body issues.

References

- [1] Chu, K. F., Dupuy, D. E. (2014). Thermal ablation of tumours: Biological mechanisms and advances in therapy. *Nature Reviews Cancer*, 14(3), 199–208. <https://doi.org/10.1038/nrc3672>
- [2] Siepel, F. J., Maris, B., Welleweerd, M. K., Groenhuis, V., Fiorini, P., Stramigioli, S. (2021). Needle and Biopsy Robots: a Review. *Current Robotics Reports*, 2(1), 73–84. <https://doi.org/10.1007/s43154-020-00042-1>

Compliant Multi-DOF endoluminal forceps with 2.5 mm in diameter

K. Osawa¹, D. S. V. Bandara¹, R. Nakadate², Y. Nagao³, T. Akahoshi⁴, M. Eto⁵, J. Arata¹

¹Kyushu University, Mechanical Engineering, Fukuoka, Japan

²Kyushu University, Center for Advanced Medical Innovation,

Fukuoka, Japan ³Kyushu University Hospital, Advanced Medicine

and Innovative Technology, Fukuoka, Japan ⁴Kyushu University,

Advanced Medical Initiatives, Fukuoka, Japan ⁵Kyushu University, Urology, Fukuoka, Japan

Keywords Endoluminal Forceps, Compliant Mechanism, Endoscopic Submucosal Dissecti, Medical Robotics

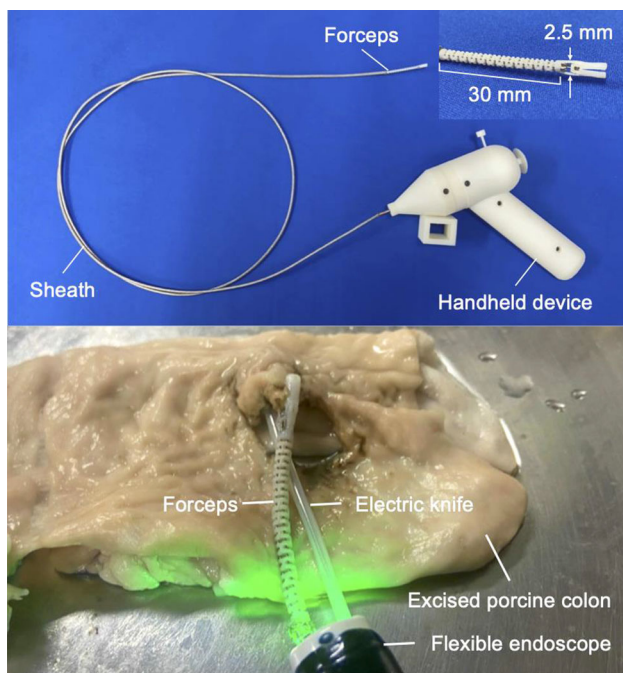


Fig. 1 Prototype developed based on the proposed structure (upper) showed a positive feasibility in an evaluation experiment on animal tissues (lower)

Purpose

Gastrointestinal cancer can be treated by inserting a flexible endoscope through natural orifice. However, treatment instruments with limited-DOF (Degree-Of-Freedom) require the operator to have a high level of skill. In previous studies, articulated devices have been developed. These articulated devices are useful for endoluminal procedure such as ESD (Endoscopic Submucosal Dissection) and NOTES (Natural Orifice Transluminal Endoscopic Surgery) to be able to operate dexterously in narrow lumen, however, still suffering from the limitation such as size and cost [1]. To overcome these limitations, we propose an articulated forceps insertable into a standard endoscope's channel. The forceps are made with a compliant mechanism, which allows the device to be compact and affordable thanks to its monolithic structure. The proposed structure showed a positive feasibility throughout a series of evaluations.

Methods

In this study, we developed a 2.5 mm multi-DOF endoluminal forceps by using a compliant mechanism. The forceps designed based on the proposed structure can be inserted into a 2.8 mm working channel of a prevailing standard flexible endoscope. The proposed mechanism consists of 2 segments: 1-DOF Grasping and 2-DOF bending. These motions are generated by a tendon-sheath mechanism. When a tension is applied to a wire passing through the center of the structure, the tip opens and closes along with the part of structure elastically deformed. Its structure was designed based on the results of FEA (Finite Element Analysis) considering the maximum opening width as an evaluation index. The 2-DOF bending can be performed by an antagonistic wire mechanism for each DOF, providing the tip motion in up, down, left, and right motions by four wires. As a notable feature of the bending segment, the thickness of elastic hinge was designed to become thinner toward the tip to prevent stress concentration on the proximal few hinges when bending largely [2]. The proposed structure can be

produced within a single piece of mechanical element, allowing to be downsized. In addition, the proposed mechanism can be operated without mechanical backlash.

Results

The developed prototype was comprised of forceps, sheath, and handheld parts. The proposed forceps was fabricated by a 3D printer (Objet30, Stratasys, USA) using acrylic resin (Verowhite). A notable feature of the proposed forceps is that the two components of motion (grasping and bending) are monolithically structured by compliant mechanism. The sheath is composed of five stainless wires with 0.3 mm in outer diameter and five stainless flexible tubes with 0.6 mm in outer diameter. Wires and tubes have been designed in consideration of transmission efficiency. The wire surface was coated by PTFE, and the tube was consisted of a coil with a flat cross section of strand. The five wires are mechanically connected to the joystick, providing intuitive control of the tip in 3-DOF (grasping and bendings) by the structure without electric components such as sensors and motors. The motorless structure enables low-cost manufacturing and disposable use.

A series of mechanical performance verification tests on the prototype revealed the successful achievement of the following specifications:

- Outer diameter: 2.5 mm.
- Maximum opening width of grasping: 2.25 mm.
- Length of bending segment: 30 mm.
- DOF: 1-DOF Grasping + 2-DOF Bending.
- Range of motion: -90 to $+90$.

To confirm the effectiveness of developed forceps, we carried out a feasibility test on an excised porcine colon. As an experimental setup, developed forceps and an electric knife (KD-650Q, Olympus, Japan) were inserted through the two working channels of a flexible endoscope (GIF-2TQ260M, Olympus, Japan). An evaluation experiment using handheld device was performed as shown in Fig. 1. The mucosa of excised porcine colon was successfully resected by the electric knife while being lifted by the developed forceps. Therefore, the experimental result showed a positive feasibility of the prototype.

Conclusion

In this study, we developed a novel multi-DOF endoluminal forceps comprised of a compliant mechanism. The proposed structure significantly reduced the number of parts compared to the conventional endoscopic tools within the size of 2.5 mm in diameter. Feasibility and effectiveness of the developed forceps have been positively shown in ex- vivo test. The material of the tip structure will be replaced to a biocompatible material from the current 3D printable acrylic resin. Also, a further investigation is on-going to improve the usability.

References

- [1] Nakadate R, Nakamura S, Moriyama T, Kenmotsu H, Oguri S, Arata J, Uemura M, Ohuchida K, Akahoshi T, Ikeda T, Hashizume M (2015) Gastric endoscopic submucosal dissection using novel 2.6-mm articulating devices: an ex vivo comparative and in vivo feasibility study. *Endoscopy*, 47: 820–824.
- [2] Osawa K, Bandara DSV, Nakadate R, Nagao Y, Akahoshi T, Eto M, Arata J (2021) Development of Multi-DOF Forceps Comprised of a Monolithic Compliant Structure for Endoluminal Surgery. *Proceedings of the 7th International Conference on Advanced Mechatronics*, GS3-5: 40–41.

A flexible endoscope manipulator with diagonal rolling friction mechanism

H. Kim¹, S. Shim², J. Hong¹

¹Daegu Gyeongbuk Institute of Science and Technology (DGIST), Department of Robotics Engineering, Daejeu, South Korea ²Korea Institute of Machinery and Materials (KIMM), Department of Medical Assistant Robot, Daegu, South Korea

Keywords Endoscope motorize, Rolling friction, Slack, Manipulator

Purpose

For conventional endoscopic surgeries, at least two medical staff must participate in the surgery. The surgeon manipulates the surgical instrument, and the endoscopist manipulates the endoscope to provide the view to the surgeon. Research of endoscopic robots has focused on replacing the surgeon's manual operation with accurate and safe robotic control. However, manual manipulation of the endoscopist is remained as an issue particularly for a flexible endoscope. A fine and high level of communication between the surgeon and the endoscopist is also required for safe and accurate surgery, which is not always possible. When manipulating a flexible endoscope by a robotic device, due to gravity and the length between the body and the tip of the endoscope, the flexible cable hangs down, and a slack happened. If the slack of the endoscope is not handled properly, the varying tension of the endoscope causes a problem to control the endoscope. A previous study tried to minimize the error from the slack [1]. However, the system had a large footprint, and was difficult to transmit the rotation motion to the endoscope tip due to the friction. In this study, we propose a tip manipulator with an adapted diagonal rolling friction mechanism to achieve required rotation and translation motion at the endoscope tip, as well as a separate body manipulator to follow the tip manipulator.

Methods

The total system consists of two parts. The first part is a tip manipulator that has four rollers to implement the diagonal rolling friction mechanism. The second part is a separate body manipulator, which enables three degrees of freedom motions. (Fig. 1) 1. Endoscope body manipulator A commercial endoscope (Olympus) is motorized with a belt and pulley. Three degrees of motion (bending, rotation, and translation) are available. The body manipulator acquires the slack value from the tip manipulator. The slack angle is measured by a pin encoder. It is controlled to maintain the slack angle in a predefined range which assures required stability. 2. Endoscope tip manipulator Four rollers are in pairs to make two roller sets. The coupled rollers are arranged in 90 degrees rotated with respect to the axis direction of the endoscope to facilitate rotation and translation motions while canceling the escaping motion perpendicular to the endoscope longitudinal axis. The two roller sets are positioned in the front and rear

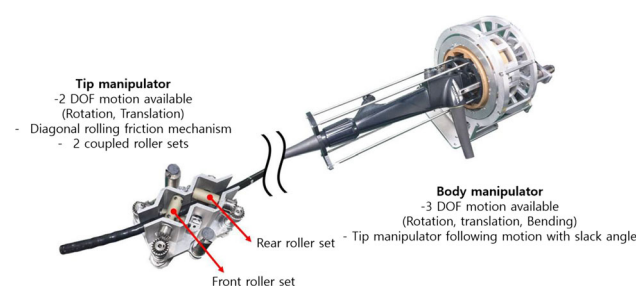


Fig. 1 Endoscope manipulating system with diagonal rolling friction mechanism

Table 1 Rolling direction matrix showing roller rotation direction as the input and the resulting endoscope motions as output

Rear roller set \ Forward roller set	+	0	-
+	Forward translation CW rotation	Forward translation CCW rotation	CCW rotation
0	Forward translation CW rotation	No motion	Backward translation CCW rotation
-	CW rotation	Backward translation CW rotation	Backward translation

of the endoscope path. Rotation and translation motions are achieved by the combination of two roller sets. (Fig. 1) When the front roller set rotates in + direction, the endoscope is engaged to make forward translation and clockwise rotation motion. The rotation of the rear roller set rotates in + direction makes the endoscope have the forward translation and counterclockwise rotation motion. By summation of these two roller set motions, the endoscope can be controlled with 2 degrees of freedom. Table 1 shows the possible combinations of the roller driving directions for rotating and translating the endoscope. The tip manipulator was designed with a size of 93(W) X 123(H) X 138(D) mm. This is similar to the parallel roller system that can only perform the translation. [2] Furthermore, it is much smaller than the two roller system that enables translation and rotation motions.

Results

Repeatability tests were performed for the translation and rotation movement. The manipulators were mounted on an optical table, and the tip manipulator was controlled in velocity control mode. The translation distance and rotation angle were measured. The standard deviation was calculated from the measured values. The experiment showed a repeatability error of about 0.4 mm for translation. The translation error is smaller compared to previous research about 60%. [2]. It means the error that occurred by the slack issue was compensated well by the proposed method.

Conclusion

The proposed diagonal rolling friction mechanism provides a novel and compact way to control the endoscope tip. The slack error was compensated by controlling the endoscope at the tip. And the compact size was achieved by driving rotation and translation motion in two roller sets. Repeatability proved sufficient for endoscope manipulation which can be assured minimize the error of slack. It is expected that this system will be applied as a more accurate and usable way to manipulate the endoscope.

References

- [1] Huang, Y., Lai, W., Cao, L., Liu, J., Li, X., Burdet, E., Phee, S. J. (2021). A three-limb teleoperated robotic system with foot control for flexible endoscopic surgery. *Annals of Biomedical Engineering*, 1–15.
- [2] Thakur, Y., Holdsworth, D. W., Drangova, M. (2009). Characterization of catheter dynamics during percutaneous transluminal catheter procedures. *IEEE Transactions on Biomedical Engineering*, 56(8), 2140–2143.

Stiffness evaluation by a sensorized grasping forceps with modularized gap sensor focusing on the object thicknesses

T. Komaki¹, T. Dohi², K. Kuwana^{1,3}

¹Tokyo Denki University, Graduate School of Engineering, Tokyo, Japan ²Tokyo Denki University, Research Institute for Science and Technology, Center for Research and Collaboration, Tokyo, Japan ³Tokyo Denki University, School of Engineering, Tokyo, Japan

Keywords Sensorized grasping forceps, Stiffness measurement, Thickness measurement, MEMS sensors

Purpose

Recently, VATS (Video-Assisted Thoracic Surgery) is widely used for lung cancer treatment. In VATS, it is difficult to palpate the affected area because the incision area of VATS is smaller than that of open surgery. There have been several studies on detecting hard objects based on the contact reaction to the objects of device such as the methods using pressure sensor or acoustic reflection. In these methods, hard objects were detected by comparing the values between the affected area and around the affected area. In contrast, we have proposed a method of estimating the stiffness by measuring the absolute value related to Young's modulus using a MEMS (Micro Electro Mechanical Systems) 3-axis force sensor and a potentiometer. In our previous study, we have developed a prototype of a sensorized grasping forceps with modularized gap sensor (Fig. 1) for the stiffness measurement method. In this study, we evaluated the stiffness measurement characteristics of the prototype focusing on the thickness of the grasping objects by comparing to the stiffness measured by a force gauge.

Methods

The sensorized grasping forceps consists of a grasping forceps with two MEMS 3-axis force sensors attached to the each tip and a modularized gap sensor. The grasping forceps was designed so that the sensor surfaces are parallel when the forceps are closed. We used gelatin as the grasping object. We prepared 9 types of object. The thicknesses of the objects were 5 mm, 10 mm, and 15 mm at 10%, 20%, and 30% gelatin by weight. We measured the thickness and stiffness 5 times to each object by both the sensorized grasping forceps and the force gauge. The gap of the forceps when the measured value of the Z-axis of the 3-axis force sensor was over 0.1 N is defined as the thickness of the grasped object. The stiffness was calculated using the Z-axis force and the displacement of the thickness.

Results

The average thickness and the average stiffness of the 5 mm, 10 mm and 15 mm object at 10%, 20% and 30% gelatins were shown in Table 1. The thickness was measured almost within 1 mm error. The stiffness increased with the increase of the gelatin concentration, linearly. The slope of the approximate line was 0.5939 when the horizontal axis is the stiffness of a force gauge and the vertical axis is the stiffness of a grasping forceps. We observed that the error become larger with the increase of the thickness and the concentration. One of the causes may be the tip angle of the grasping forceps. It may improve by compensating the stiffness by the tip angle measured by the prototype.



Fig. 1 Photograph of a Sensorized grasping forceps

Table 1 Average thickness and the average stiffness of the 5 mm, 10 mm and 15 mm object at 10%, 20% and 30% gelatins

Density	Thickness [mm]	Measurement thickness [mm]	Force gauge stiffness [MPa]	Grasping forceps stiffness [MPa]
10%	5	4,46	0,66	0,1
	10	11,09	0,13	0,16
	15	15,79	0,16	0,16
20%	5	4,55	0,36	0,29
	10	10,37	0,45	0,44
	15	15,47	0,56	0,43
30%	5	4,99	0,88	0,46
	10	10,69	1,11	0,68
	15	15,59	1,33	0,68

Conclusion

In this study, we evaluated the stiffness measurement characteristics of the prototype focusing on the thickness of the grasping objects by comparing to the stiffness measured by a force gauge. The thickness of the grasping object was measured almost within 1 mm error. Though the error increased with the increase of the thickness, the results show that the forceps may evaluate the stiffness same as the measurement by the force gauge.

General concept of self-assembling surgical tools for minimally invasive surgery

T. H. Fass¹, G. Hao^{1,2}, P. Cantillon-Murphy¹

¹University College Cork, Tyndall National Institute, Cork, Ireland

²University College Cork, Electrical and Electronic Engineering, Cork, Ireland

Keywords self-assembly, origami robot, minimally invasive surgery, magnetic surgery

Purpose

Minimally invasive surgery has led to significantly reduced infection rates and reduced recovery times for surgical procedures, by minimizing the damage to the patient's body. This imposes a restriction on any tool used to be deployable during minimally invasive surgery and, in turn, limits or complicates procedures as a trade-off between incision size and number of tools or tool size that has to be achieved. Self-assembly is the process of a structure to assemble itself and can be found in countless examples in nature. The principle of self-assembly applied to surgery tools could potentially lift the size limitation a port imposes on the used tool. This work seeks to work toward the development of a general concept of self-assembly to be used in minimally invasive surgery.

Methods

Self-reconfiguring behaviour and origami robots were investigated for their possible application in surgical instruments. A general concept of self-assembly has been modeled after the natural self-reconfiguration of proteins. The proposed concept features a chain of magnetic components, linked together by rotational joints to allow for the deployment through a long flexible tube and therefore promote the application in a range of surgical procedures. Simulation tools have been implemented [1] in order to investigate folding behaviour. The

folding dynamics of simple structures have been simulated. More complex structures were analysed for folding reliability by computing potential energy [2] during the folding process. Prototypes have been constructed as proof of concept and to show the scalability of the proposed tools. The folding behavior of the prototypes have been investigated and compared against the simulation. Simulations suggest that the folding is achievable and reliable. While the prototype confirms the simulation in part, it also shows that highly precise manufacturing is necessary for more complex self-folding shapes such as the highlighted anastomosis ring.

Results

The proposed concept consists of a chain of magnetic components, linked together by rotational joints. Once deployed and free of the confinement of the deployment tube, the magnetic attraction between the chain components in combination with the unique restrictions of the joints causes the chain to fold in on itself to a functional device, exemplary shown in Fig. 1. The individual orientation of each magnetic component is unique, which results in a matrix of magnetic components that displays a strong magnetic potential capable of interacting with external magnets. This led to the chosen target applications of magnetic anastomosis creation, to serve as proof of concept. The concept allows for the formation of a magnetic voxel and other secondary structures that in turn can assemble to larger structures. Using the concept, a self-reconfiguring magnetic anastomosis device has been constructed and tested. The prototype proves that independent self-assembly of a magnetic surgery tool is possible. Finally, limits in manufacturability of the prototype have been investigated, showing that the device is potentially scalable in size.

Conclusion

The developed concept is shown to be deployable through small tubes and to fold into volumetric structures upon release. The proof of concept as a self-reconfiguring anastomosis ring displays its potential in minimally invasive surgery. While manufacturability is a limit in the current prototype, supported folding applications of the reconfiguring anastomosis ring prove that application as a surgery tool is possible. The developed concept of self-folding tools is proven to be capable of mimicking a range of possible shapes that can be used. The

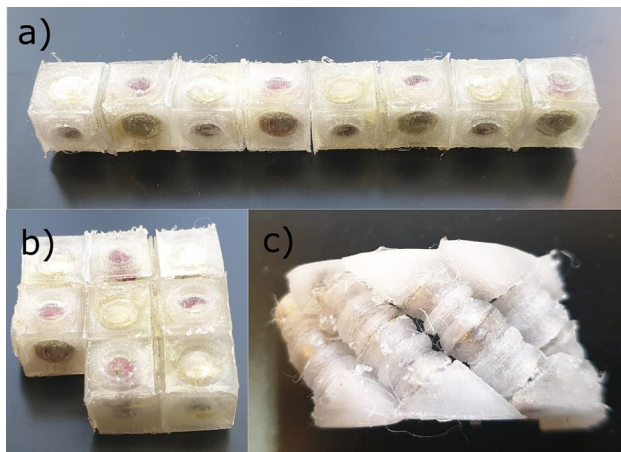


Fig. 1 Chain of magnetic components a) folds into a sheet b) or a complex structure c)

concept allows for potentially multiple stable states of folding that could be exploited to create moving parts in the created tool. If the production methods can be refined to allow for reliable production in the desired size, this could lead to a new class of self-folding surgical tools for minimally invasive surgery.

References

- [1] Fass TH, Hao G, Cantillon-Murphy P (2020) On planar self-folding magnetic chains: Comparison of Newton–Euler dynamics and internal energy optimization. *Robotics Autonomous Syst.* 132:103601
- [2] Fass TH, Hao G, Cantillon-Murphy P (2021) Vectorised Formulation of Newton–Euler Dynamics for Efficiently Computing 3D Multibody Chains. *IEEE/IFTOMM Int Conf Reconfigurable Mech Robots* 5:645–655

MRI-based 3D silicone model for left atrial appendage occluder device size determination

D. Bertsche¹, A. Reinelt¹, P. Metze¹, V. Rasche¹, T. Dahme¹, L. Schneider¹, I. Vernikouskaya¹

¹Ulm University Medical Center, Ulm, Germany

Keywords 3D modelling, 3D printing, non-contrast-enhanced 3D MRI, LAA occlusion

Purpose

Closure of the left atrial appendage (LAA) prevents embolic stroke in patients with atrial fibrillation. The correct selection of the closure device and size is challenging due to the high variability of LAA anatomy. In general, transesophageal echocardiography (TEE) is used for sizing the patient-specific LAA and subsequent device selection. However, TEE-based measurements, as well as x-ray fluoroscopy measurements, tend to underestimate the size of the LAA [1]. Due to the accurate measurements enabled by the three-dimensionality, the closure device selection based on computer tomographic (CT) image volumes and CT-based 3D models is appropriate but increases radiation exposure and contrast agent injection to the patient. Here, we propose generating an MRI-based 3D silicone model for LAA closure device selection, eliminating invasive imaging, radiation exposure, and contrast agent. The silicone model was created by filling a 3D-printed cast model generated with MRI-based segmentation with silicone compound. A Watchman FLXTM (Boston Scientific Corporation, Marlborough, MA, USA) demo device could be inserted into the silicone model to verify device selection.

Methods

Preprocedural, a 3D MRI data set was acquired with a respiratory navigated 3D isotropic six-point mDixon method at 3 T (Achieva 3.0 T, dStream, R5.6, Philips Medical Systems B.V., Best, The Netherlands) at 1.33 mm³ resolution and with a non-contrast-enhanced protocol. Based on the 3D MRI, the part of the left atrium involving the left superior pulmonary vein (LSPV) approach and LAA was segmented using 3DSlicer (Fig. 1a). Similar to previously described work [2], a casting model was created from the resulting segmentation with the inner shell corresponding to the original segmentation size, and the outer shell with an offset of 2 mm (Fig. 1b). The cast model was printed using the PLA filament NX2 1.75 mm (extruder, Lauterach, Austria) and a layer thickness of 0.2 mm. Printing time resulted to 3:36 h on a RAISE3D Pro2 Plus and the associated software ideaMaker 4.0.1 (Raise3D, Irvine, California, USA). The resulting printout was filled with a silicone compound composed of 90% DragonSkinTM 10 MEDIUM (Smooth On,

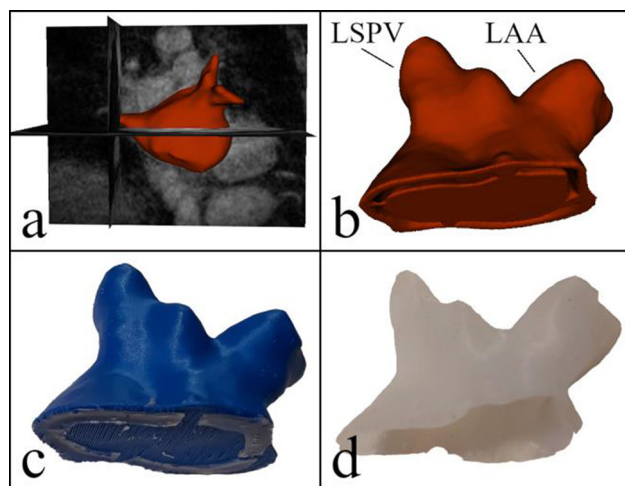


Fig. 1 **a** segmentation based on patient MRI, **b** cast model, **c** printed cast model filled with silicone compound, **d** silicone model

Macungie, Pennsylvania, USA) and 10% Silicone Thinner™ (Smooth On, Macungie, Pennsylvania, USA) (Fig. 1c). After overnight drying, the printout was removed from the silicone (Fig. 1d). A fitting Watchman FLX™ demo device was placed in the LAA of the silicone model. Subsequently, an MRI of the silicone model with the implanted device within a water bath was generated to validate the required compression of the device. Further, MRI-, TEE-, and XR-based measurements of the patient's LAA were performed to compare the respective device size selections.

Results

The LAA could be identified in the non-contrast-enhanced patient MRI (Fig. 2a). In the model MRI, the intensity of the MR signal of the silicone compound enabled the distinction between water, silicone model, and device (Fig. 2b). A Watchman FLX™ of size 24 mm fit in the silicone model (Fig. 2c) and was also implanted in the LAA of the patient. Based on the similar results of the measurements of the LAA using 3D TEE (21 mm), 2D XR (21.66 mm), and 3D MRI (21.3 mm), the size of the optimal Watchman FLX™ of 24 mm

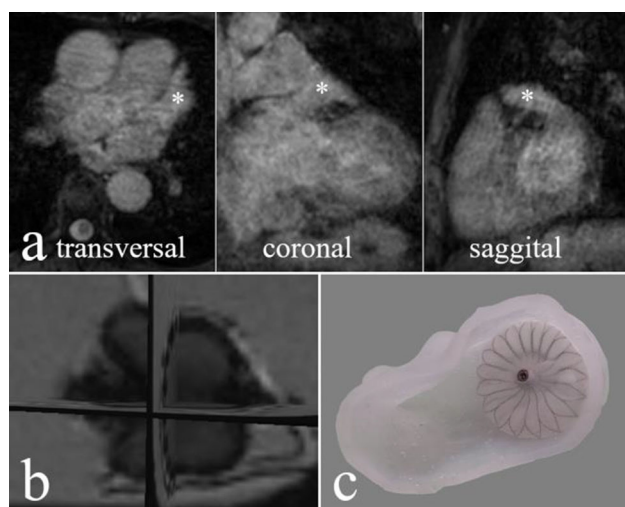


Fig. 2 **a** LAA (*) identifiable in non-contrast-enhanced patient MRI, **b** 3D MRI of (c) within a water bath, **c** silicone model with implanted LAA occluder (Watchman FLX™)

diameter was determined respectively. Similar to tissue, the silicone deformed as the device was implanted into the model. The device was nevertheless compressed. The compression of the device by about 13%, as determined from the MRI of the silicone model in the water bath, was within the required range of 10–30%. The compression of the device implanted in the patient's LAA was determined at 12.5% by 2D TEE measurement. The silicone model comprised the left superior pulmonary vein in addition to the LAA, which supported orientation and handling of the model. Assuming a drying time of 5 h according to the manufacturer's specifications of the silicone, the production of a patient-specific silicone model required 10:30 h.

Conclusion

By using the MRI-based silicone model, the same device size as based on TEE and XR but with contrast agent-free, radiation-free, and noninvasive imaging could be determined. The measurement directly on MRI also yielded the same device size, though the proper selection of the slice to be measured in the volume is essential. Based on the silicone model, in addition to the measurement of the LAA, the presumable compression of the device could be determined, which could facilitate the decision between two possible device sizes in case of ambiguous measurements. Since in this work only one LAA occlusion case was evaluated, the priority of further work is to verify the device size selection based on an MRI-based silicone model using a larger data set.

Acknowledgements

The project on which this report is based was funded by the Federal Ministry of Education and Research under the funding code 13GW0372C. Responsibility for the content of this publication lies with the authors.

References

- [1] Fernández-Vegas C, Millán-Núñez V, Fuentes-Cañamero ME, Nogales-Asensio JM, Doncel-Vecino J, Domínguez MY, Serrano LG, Quintana DS (2014) Comparison of imaging techniques to assess appendage anatomy and measurements for left atrial appendage closure device selection. *J Invasive Cardiol.* 26(9):462–7.
- [2] Bertsche D, Reinelt A, Speidel T, Metze P, Stumpf K, Rasche V, Vernikouskaya I (2021) Multi-modality phantom for accuracy verification of image-based navigation tools. *Int J CARS* 16 (Suppl1):S108–S110.

Effective surgical education and development of an AI operation system designed for safety improvement and optimal cancer surgery by digitizing

Y. Bamba¹, S. Ogawa¹, F. Tokitou¹, H. Kondo¹, M. Itabashi¹, K. Suzuki², S. Morita², K. Iwado³, T. Yamaguchi⁴, K. Masamune⁴, S. Yamaguchi¹

¹Tokyo Women's Medical University, Department of Digestive and General Surgery, Tokyo, Japan ²Tokyo Women's Medical University, Diagnostic Imaging and Nuclear Medicine, Tokyo, Japan ³Tokyo Women's Medical University, Blood Purification Therapy Department, Tokyo, Japan ⁴Tokyo Women's Medical University, Institute of Advanced BioMedical Engineering and Science, Tokyo, Japan

Keywords AI navigation, surgical education, surgical difficulty, Digitization

Purpose

A surgical brain is the ultimate useful model for diagnosis, treatments, surgery, and postoperative management. Theorizing and structuring a surgical brain’s decision-making and surgical skills are important for promoting precision medicine. During a laparotomy with no recording equipment many skills were taken over on-site. Recently, a digitized surgery video can be replayed repeatedly by diffusion during laparoscopic surgery and robot-assisted surgery. These surgical videos and perioperative data in electronic charts are now easily available, with their implementation being less-invasive in AI-navigated image-guided surgery. This effectively avoids risks during the perioperative period and can be achieved by analyzing data in detail using AI [1].

Methods

We have created an image recognition model for AI using IBM Visual Insight. Annotated (tagged) images are trained to create and verify area detection models and judgment models. We also created diagnosis models using AI to analyze preoperative images and surgical videos. Using 570 preoperative rectal cancer images (CT and MRI) of patients treated in our hospital, a diagnosis model of tumor depth was created. For T3-T4 cases with a large tumor diameter, the percentage of correct answers of the MRI images was 88%. In order to create a real-time object recognition model for surgery, 5000 still images were created from surgical videos, and 18,580 locations of 18 types such as forceps, ports, bleeding, intestinal tract, gauze, and blood vessels were annotated and trained. When a still image was created from the test surgery video and the diagnostic results were calculated, the overall diagnostic results had a sensitivity for recall of 82.7% and specificity for precision of 84.1% (Fig. 1) (Table 1). The ovaries, ureters, and suction forceps were easily mistaken for others and the diagnostic results were poor.

Table 1 Overall diagnostic results had a sensitivity for recall of 82.7% and specificity for precision of 84.1%

	Number of objec	Recall (%)	Precision (%)
		in images	
GI tract	160	82,5	92,3
Grasping F.	189	91,5	85,2
Harmonic F.	58	100	75,3
Maryland F.	23	100	88,5
Sphere F.	13	92,3	80
Suction F.	26	50	61,9
Clip F.	32	100	91,4
Clip	59	89,8	86,9
Metal Clip	33	90,9	90,9
Gauze	105	84,8	97,8
Artery	25	84	100
Vein	20	55	55
Blood	86	94,2	100
Liver	38	92,1	85,4
Gb	23	82,6	82,6
Ovary	28	82,1	85,2

Objective prediction of surgical difficulty before surgery is important for safe and efficient surgery. We measured the level of correlation between operation time and amount of bleeding using machine learning, and examined whether it was possible to predict these two parameters. In 216 cases of rectal (Rb) cancer, we created a predictive model for the length of surgery and the amount of bleeding using a machine learning model. In the operation time model of

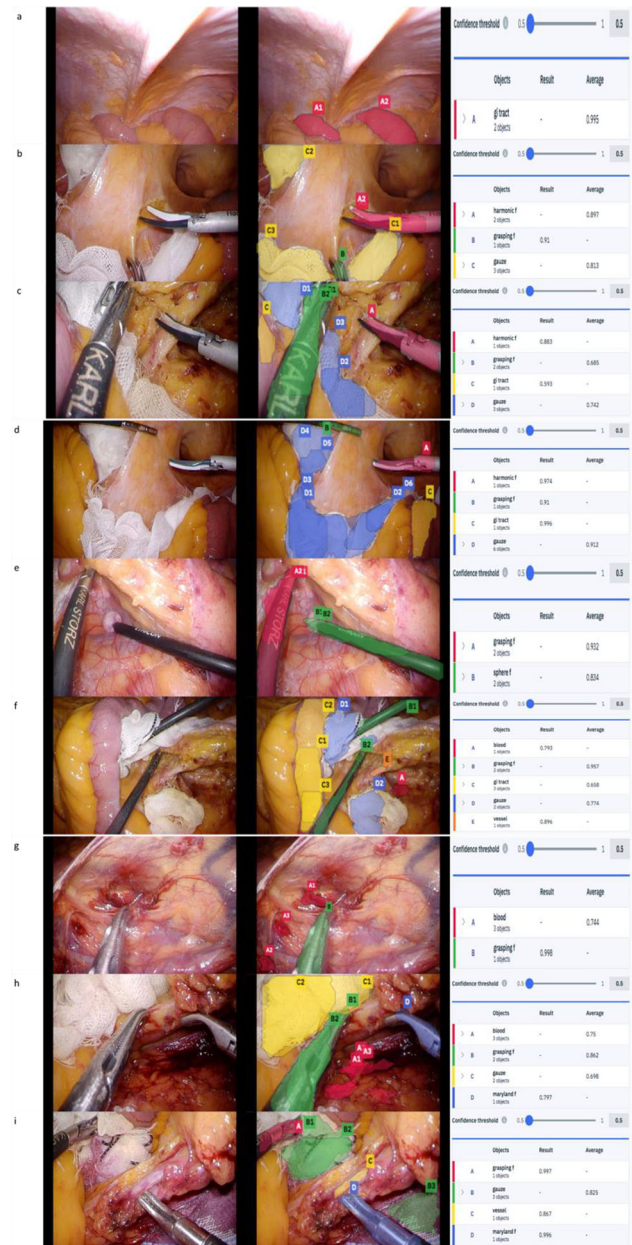


Fig. 1 Examples of object detection in surgical images. The objects were detected accurately

symbolic regression, stoma construction, surgeon, operation method, bleeding volume, preoperative radio-chemotherapy, and lateral dissection were important factors, with an accuracy, precision, and recall of 68.75, 72.41, and 67.74, respectively and an AUC of 0.825. In the bleeding volume model, surgical method, male, and preoperative radio-chemotherapy were important factors, with an accuracy, precision, and recall of 71.88, 72.41, and 67.74, respectively and an AUC of 0.763.

Results

Digitization of surgical videos is important for establishing surgical education and navigation surgery. Creation of an algorithm using surgical big data for a lasting navigation system is also of great advantage to perform less-invasive and safe surgery. For example, scene development, bleeding detection, and use of forceps can be

evaluated using a timeline to understand how surgery has progressed. Surgical risk management is therefore possible, such as raising an alarm when bleeding exceeds a certain level or when the estimated time for each process in the surgical procedure is exceeded in the timeline. Colorectal cancer surgery is relatively easy to standardize and it is considered that AI will improve efficiency and optimization. As progress assistance develops, the possibilities of automated surgery will increase. By using this approach, surgeons can expect to improve surgical results as well as surgical accuracy.

Conclusion

It would be very helpful for surgeons to reflect on the system-obtained data in clinical practice, in other words, to obtain information on time. For instance, risk avoidance is recommended by raising alarms during high-risk surgery, resulting in an improvement in the skills of surgeons using objective assessment of surgical skills, with less experienced surgeons performing surgery with navigation. We have introduced the operation system under development and plan to validate future issues.

References

- [1] Bamba Y, Ogawa S, Itabashi M, Shindo H, Kameoka S, Okamoto T, Yamamoto M (2021) Object and anatomical feature recognition in surgical video images based on a convolutional neural network. *International Journal of Computer Assisted Radiology and Surgery* volume 16, pages2045–2054.

Predicting surgery using normalized brain-function database and the application of tele-surgery support in 5G network

M. Tamura^{1,2}, I. Sato³, T. Yamaguchi¹, K. Yoshimitsu¹, Y. Horise¹, T. Maruyama², M. Nitta^{1,2}, T. Saito^{1,2}, Y. Fujino³, K. Masamune¹, T. Kawamata², H. Iseki¹, Y. Muragaki^{1,2}

¹Tokyo Women's Medical University, Faculty of Advanced Techno-Surgery, Institute of Advanced Biomedical Engineering and Science, Tokyo, Japan ²Tokyo Women's Medical University, Department of Neurosurgery, Tokyo, Japan ³Future University Hakodate, Faculty of System Information Science Engineering, Hakodate, Japan

Keywords tele-surgery support, surgical data coordination, predicting surgery, 5G.

Purpose

We have clinically operated an intraoperative MRI-guided navigation system since 2000 in more than 2000 cases, and since 2004, we have performed intraoperative examination monitoring in awake mapping, that provided navigation images, electrical stimulation, and patient response on simultaneous video in more than 475 cases. Recently, mapping data was added to intraoperative MRI as digital brain function information, and the reaction points on intraoperative MRI were converted to preoperative MRI and standard brain, and the analysis results of 20 cases (reaction point 51) were summarized (JAMIT, 2020). The image conversion accuracy of the reaction points was 2.6 ± 1.5 mm for preoperative MRI and 1.7 ± 0.8 mm for standard brain, and we simulated the projection of the normalized brain data to the individual pre- and intra-operative MR image. In the future, we are planning predicting glioma surgery that includes AI analysis database intraoperatively by adding the tumor removal rate, the incidence of motor / language function complications, the degree of home / social rehabilitation, etc. to the mapping information [1].

The hyper smart cyber operating theater (Hyper SCOT), which was initiated as clinical research from 2019, manages an on-site limited database that integrates surgical information in a time-synchronized manner. For the purpose of provide equal distribution of medical resources in Japan, we started to demonstrate in 2020 to remotely link surgical data between Tokyo and Hakodate to support and realize remote surgery (research grant by the National Institute of Information and Communications).

The purpose is to achieve a high level of uniformity of medical technology including AI analysis, share a wide variety of intraoperative multi-modal information obtained from Hyper SCOT at two bases, and analyze and confirm these data using a high-speed and large capacity 5G network.

Methods

We have already established a system for anonymizing all surgical information, storing and sharing it, and constructed remote confirmation using a closed 5G network (Fig. 1). This 5G base Intraoperative information sharing system can store and sharing of intraoperative information obtained from SCOT as big data. Additionally, this closed 5G network uses the cloud direct of docomo open innovation cloud, which allows the system to be constructed with only 5G router and servers / PC, and the security is guaranteed by the carrier. It also enables medical AI using this big data. In addition, data from multiple medical devices on the server can be integrated into 4 K images, and intraoperative information can be checked at once from a remote location. Therefore, the integrated 4 K images can be used for remote advice by expert doctors. Furthermore, this system also enables medical AI of predicting glioma surgery using normalized brain-function database and medical image analysis AI.

Results

We evaluated network response speed, TCP throughput and the quality of data transmission using 5G. First experiment, the time from the server-to-server response speed and TCP throughput results was calculated. The average delay between servers using ping was 90.6 ms. The throughput measured by iPerf was 108 Mbps. In the second experiment, the latency from the server-to-server of the RTP streaming video transmission using an OBS and VLC was evaluated. In this experiment, the 4 K microscopy video and system time were displayed on the monitor of the transmission side, and the monitor was then captured by 30fps and transferred. The delay was evaluated by comparing the system time of the receiving side and the system time of the video acquired at the receiving side. This results, the delay was about 1.36 s when the bit rate was 70 Mbps. Therefore, 4 K images of 70 Mbps quality could be streamed between the servers,

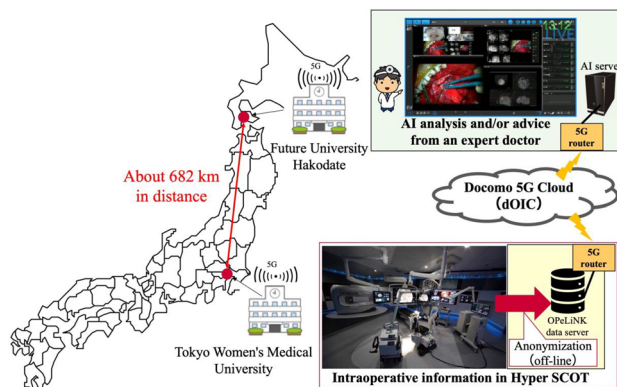


Fig. 1 Overview of 5G base Intraoperative Information Sharing System between Tokyo Women's Medical University and Future University Hakodate

and the bandwidth and speed were sufficient for intraoperative information distribution.

Conclusion

We calculate the time and also obtained results that depend on the remote communication and 5G base station performance, and we constructed a communicative infrastructure for the transfer of surgical images including 4 K images. In the future, real-time AI analysis of intraoperative information will make it possible to easily understand the surgical process and situation during surgery, and skilled doctors in remote areas can easily check the surgical status and provide the tele-surgery support from anywhere through high-speed and large-capacity data transfer technology.

Acknowledgments

These research results were obtained from the commissioned research by National Institute of Information and Communications Technology (NICT), Japan.

References

- [1] Tamura M, Sato I, Muragaki Y (2022) Normalized Brain Datasets with Functional Information Predict the Glioma Surgery. In: Hashizume M (ed) Multidisciplinary Computational Anatomy: Toward Integration of Artificial Intelligence with MCA-based Medicine. Springer Singapore, Singapore, pp 173–180. https://doi.org/10.1007/978-981-16-4325-5_22

Development of an advanced simulator for technical assessment and training of hepaticojejunostomy

K. Yamada¹, S. Onishi¹, S. Ieiri¹

¹Kagoshima University, Pediatric Surgery, Kagoshima, Japan

Keywords laparoscopy, hepaticojejunostomy, simulator, motion analysis

Purpose

Recently, high difficulty laparoscopic surgery has become feasible and widespread. However, the number of cases performed per institution is small at present, making it difficult to standardize and improve the technique through the accumulation of cases. Therefore, the development of a simulator reproducing a disease-specific surgical procedure is essential. We developed a high-fidelity simulator of laparoscopic hepaticojejunostomy, and we studied to clarify the characteristics of the forceps movement during a surgical process using this simulator.

Methods

We developed a high-fidelity laparoscopic hepaticojejunostomy simulator. It is made by combining 1-year-old infant pneumoperitoneum body trunk model (body weight: 10 kg) based on computed tomography data and organ models such as liver, jejunum, and hepatic duct. TrackSTAR (Northern Digital Inc., Ontario, Canada) an electromagnetic tracking system were used to trace the tips of the forceps during a simulated procedure of hepaticojejunostomy. A total of 35 surgeons participated in this study. 19 of them were pediatric surgeons (PS) and the other 16 were gastrointestinal surgeons (GIS). The participants had to perform hepaticojejunostomy using the simulator. The port layout was the right para-axial position. The assessment points were time required to complete the task, total path length of each forceps and average velocity of each forceps tip. All data were expressed as the mean \pm standard deviation. Two-tailed paired and unpaired Student's *t*-test and analyses of variance were

Table 1 Results of the study

	PS (n = 19)	GIS (n = 16)	<i>p</i> value
Required time (s)	939.57 \pm 302.37	755.17 \pm 442.27	0.15
Rt. total path length (m)	55.13 \pm 51.21	143.53 \pm 98.55	< 0.01*
Lt. total path length (m)	188.02 \pm 111.91	44.35 \pm 40.75	< 0.01*
Rt. average velocity (mm/s)	56.18 \pm 46.59	201.78 \pm 112.36	< 0.01*
Lt. average velocity (mm/s)	188.05 \pm 65.29	57.72 \pm 39.70	< 0.01*

*Significant difference

conducted for comparisons using EZR version 1.38 (Saitama Medical Center, Jichi Medical University, Saitama, Japan). All data were defined as being statistically significant at *p* values < 0.05.

Results

Table 1 shows the results. The time required to complete the tasks was 939.57 \pm 302.37 s for the PS group and 755.17 \pm 442.27 s for the GIS group (*p* = 0.15). There were no significant differences in the time required to complete the task. The total path length of the right forceps was 55.13 \pm 51.21 m for the PS group and 143.53 \pm 98.55 m for the GIS group (*p* < 0.01). The total path length of the left forceps was 188.02 \pm 111.91 m for the PS group and 44.35 \pm 40.75 m for the GIS group (*p* < 0.01). The total path length of the left forceps in the PS group was significantly longer than that in the GIS group, and the right forceps in the GIS group was significantly longer than that in the PS group. The average velocity of the right forceps tip was 56.18 \pm 46.59 mm/s for the PS group and 201.78 \pm 112.36 mm/s for the GIS group. The average velocity of the right forceps tip in the PS group was significantly slower than that in the GIS group (*p* < 0.01). The average velocity of the left forceps tip was 188.05 \pm 65.29 mm/s for the PS group and 57.72 \pm 39.70 mm/s for the GIS group. The average velocity of the left forceps tip in the PS group was significantly faster than that in the GIS group (*p* < 0.01).

Conclusion

The right hand movement of the PS group was shorter and slower than that of the GIS group, but the left hand movement was the opposite. The opposite characteristics of PS forceps manipulation compared with GIS may have been due to the differences in clinical procedures. Ieiri [1] reported that a shorter path length and slow manipulation increased the quality of endoscopic procedures. We found here that the movement of the left hand in the GIS group was shorter and slower than in the PS group, suggest that GIS can use their left hand more efficiently than PS. GIS perform laparoscopic surgery primarily in a para-axial setting, such as gastrectomy, colectomy, and pancreatectomy. They are accustomed to performed surgery in an expansive space, so they usually use their left hand more frequently than PS.

References

- [1] Ieiri S, Nakatsuji T (2010) Effectiveness of basic endoscopic surgical skill training for pediatric surgeons. *Pediatr Surg Int* 26 (10): 947–954.

Development of automated skill assessment system for laparoscopic colorectal surgery using automated surgical workflow recognition model

K. Nakajima^{1,2}, S. Takenaka¹, D. Kitaguchi¹, H. Hasegawa¹, S. Kojima¹, N. Takeshita¹, Y. Kinugasa², M. Ito¹

¹National Cancer Center Hospital East, Surgical Device Innovation Office, Chiba, Japan ²Tokyo Medical and Dental University, Department of Gastrointestinal Surgery, Tokyo, Japan

Keywords surgical skill assessment, surgical workflow recognition, Laparoscopic colorectal surgery, artificial intelligence

Purpose

Currently, there are surgical technique assessment tools such as OSATS and GOALS. They evaluate the overall surgical workflow and the surgeon's performance, and then divide the surgical skills into steps and evaluate each item. In Japan, the Japan Society for Endoscopic Surgery (JSES) has established a system of certified endoscopic surgeons to improve the skill level of surgeons and to train surgical supervisors. In JSES, the laparoscopic surgery videos applied every year are evaluated and scored by expert surgeons according to the criteria to determine whether they pass or fail. The overall score is 100 points, and each item is set in detail. However, it cannot be said that these skill assessments are free from subjectivity of the examiner, and it also places a heavy burden on surgeons in that they must watch surgical videos. In recent years, image recognition model using artificial intelligence has been developed, and it has become possible to automatically recognize the current surgical workflow from surgical images. Development of an automated skill assessment system using this artificial intelligence model may lead to the solution of problems in manual skill assessment. The purpose of this study is to construct a skill assessment system using surgical videos of laparoscopic colorectal surgery taken in daily practice at multiple institutions. In the future, the system will be used by endoscopic surgeons to educate surgeons and to reduce the burden on expert surgeons who have been scoring surgical videos.

Methods

A single surgery was divided into 11 phases, and 182 surgical videos of laparoscopic sigmoid colon resection were used as training data for the AI to build an automatic surgical phase recognition model. For 622 videos of laparoscopic sigmoid colon resection submitted for technical certification review by the Japanese Society of Endoscopic Surgeons in 2016 and 2017, we used this automatic phase recognition model to calculate the operative time for each phase and the number of intraoperative transfers between each phase, which is an index of planning and efficiency of the operation, as a parameter for automatic skill assessment. In addition, as an indicator of the skill of the surgical field exposure, we calculated the average of the confidence of the classification for each surgical image for each case when the AI performed the phase classification (named AI confidence). We compared each parameter calculated by the phase classification model in the high score group (mean score plus 2SD or higher) and the low score group (mean score minus 2SD or lower). Each parameter was converted to a standard score and added together to form the AI score, and the correlation between these scores and the JSES total score was examined. In addition, the cutoff value of AI score to pick up the low score group was calculated using ROC curve. The cutoff value of the AI score was calculated using the ROC curve.

Results

The AI model was able to automatically recognize the surgical process with an average F1 value: 73.4% accuracy. As a result of comparison using the analysis results of this model, the surgery time

was significantly longer in the low score group (median 125.7 min) than in the high score group (median 109.0 min) ($p = 0.011$), the number of process transfers was significantly higher in the low score group (median 35 times) than in the high score group (median 16 times) (p).

A multi-stage TCN-LSTM network for recognition of surgical gestures based on kinematic data

A. Goldbraikh¹, C. M. Pugh², S. Laufer³

¹Technion—Israel Institute of Technology, Applied mathematics, Haifa, Israel ²Stanford University, School of Medicine, Stanford, United States, ³Technion—Israel Institute of Technology, Faculty of Industrial Engineering and Management, Haifa, Israel

Keywords Surgical simulation, Motion sensor, Kinematic data, gesture recognition

Purpose

In open surgery, motion sensors are becoming increasingly popular for measuring surgical performance. These sensors are commonly used for calculating performance metrics and assessing surgical skills [1].

In this study, we explore the use of these sensors for identifying surgical gestures in offline fashion, which can be used later for the development of new gesture-based assessment metrics. Toward this end, we present our Multi-stage-TCN-LSTM network. Our network preserves the multi-stage framework of MS-TCN++ [2] while introducing several new components. First, intra-stage regularization was added to the prediction generator, this effected the loss function and improved the results. Second, in MS-TCN++ all stages are based on TCNs. We studied the impact of replacing the TCNs with LSTMs in some of the stages.

Methods

Data were collected using a variable tissue simulator [1]. To simulate different types of tissue we used tissue paper to simulate friable tissue and rubber balloons which are similar to arteries. During the suturing task, three interrupted instrument-tied sutures were placed on two opposing pieces of the material.

The study involved eleven medical students, one resident, and 13 attending surgeons. Participants performed twice on the friable tissue simulator and twice on the artery simulator. Thus, the data-set contains in total 100 procedures. In addition, video data were captured using two cameras, providing top and side views. In this study, the top view data were used for data labeling. Electromagnetic motion sensors were used to record data (Ascension, trakSTAR). The participants' hands were equipped with three sensors attached with medical tape under their surgical gloves.

The task was divided into six gestures: pass the needle through the material, pull the suture, perform an instrumental tie, lay the knot, cut the suture, and the background gesture.

The raw data contained for each electromagnetic sensor three coordinates indicating the location of the sensor in space, and three Euler angles indicating the sensor's orientation. The original measurement rate was 179.695 Hz and was downsampled as preprocessing step to 30 Hz. The input of all networks were the calculated linear and angular velocities, in total 36 channels. The input was normalized by using Z-score where μ and σ were calculated in advance based on the train set.

TCN prediction generator with Intra-Stage Regularization (ISR):

Since the input and the output of each single or dual dilated residual layer have exactly the same dimension, after each layer the output

feature maps fit the dimension of the last layer output feature maps. Hence, after each layer, theoretically, the output feature maps could have been fed to the prediction head. We found that the use of several intra-stage prediction heads in the prediction generator improves the model performance (see Fig. 1). In loss calculation, each prediction head in the entire model contributes equally.

LSTM based refinement stages:

We use a standard LSTM followed by a linear layer for the prediction head. During training, we apply a dropout on the raw output of the LSTM unit. The input to the LSTM units is downsampled to 5 Hz, and the output of the linear layer is upsampled back into 30 Hz.

Implementation details:

All models were evaluated using k-fold cross-validation (K = 5). Cross-validation groups were constructed so that all the data from a specific participant was in the same validation group (leave-k-users-out approach). The training and evaluation were performed on a NVIDIA Tesla V100 Volta GPU Accelerator 32 GB Graphics Card. All the networks were trained using an Adam optimizer where, all hyper-parameter, except learning rate, are default.

All multi-stage networks had at least one TCN component, where each TCN component contained exactly 13 (dual) dilated residual layers. The number of feature maps was 64. Each (dual) dilated residual layer had a dropout probability of 0.5. The multi-stage networks included 3 refinement stages. If the refinement units are LSTMs then the hidden dimension was 256, and each stage had exactly two LSTM layers. We applied dropout with a probability of 0.9 on the output of the last LSTM layer. The loss parameters were $\tau = 4$, $\lambda = 0.5$. The learning rate was 0.005 and the network was trained on a batch size of 1.

Results

The average performances of the algorithms and their standard deviations are summarized in Table 1.

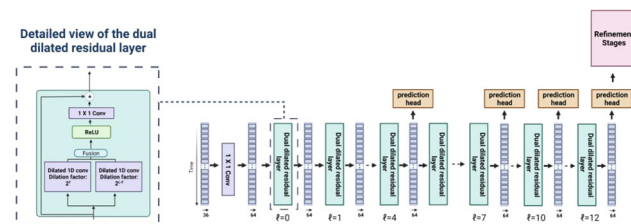


Fig. 1 TCN prediction generator with Intra-Stage Regularization (ISR)

Table 1 Networks comparison on a gesture recognition task

	Acc	F1-macro	Edit	F1@10	F1@25	F1@50
LSTM	81.91 ± 2.40	78.23 ± 2.55	84.04 ± 4.10	86.59 ± 3.52	82.62 ± 4.63	68.94 ± 7.08
MS-TCN++	84.49 ± 2.22	81.19 ± 3.34	88.51 ± 2.94	90.19 ± 3.68	86.78 ± 5.39	73.25 ± 8.36
MS-TCN++ ISR	84.84 ± 2.3	81.4 ± 3.07	89.28 ± 3.76	91.04 ± 3.81	88.22 ± 4.93	75.01 ± 7.55
MS-TCN-LSTM	85.15 ± 2.15	82.1 ± 3.75	89.27 ± 3.75	91.58 ± 3.17	88.95 ± 4.29	76.11 ± 7.26

Conclusion

MS-TCN++ has performed very well on video data. In this study, we adapted it to motion sensor data. This network outperformed the reference networks in gesture recognition. MS-TCN++ is a multi-stage network, based solely on TCN networks. In this study, we claim that MS-TCN++ is only one example of a more general multi-stage framework. This framework is composed of a prediction generator stage which is followed by multiple refinement stages. Each one of these stages contributes to the loss function independently. We expanded this approach by adding our intra-stage regularization. We then showed that different temporal networks may be used for each stage. We replaced the prediction generator with LSTM. The performance was similar to that of MS-TCN++ yet was outperformed by the TCN with intra-stage regularization. We then replaced the TCN based refinement stages with a LSTM. Since previous studies have shown that LSTM-based networks perform best on human kinematic data when the sampling frequency is 5 Hz, we downsampled the outputs of each step before feeding them to the LSTM refinement stage. Then we upsample the final prediction back to 30 Hz. We believe the fact that the MS-TCN++ prediction generation stage and the LSTM refinement stages feed data at different frequencies forces the system to identify patterns across a wide spectrum, thereby enhancing the entire network performance. Ultimately, this results in our MS-TCN-LSTM network.

References

[1] D’Angelo ALD, Rutherford DN, Ray RD, Laufer S, Kwan C, Cohen ER, Mason A, Pugh CM (2015). Idle time: an underdeveloped performance metric for assessing surgical skill. *The American journal of surgery*, 209(4), 645–651.
 [2] Li SJ, AbuFarha Y, Liu Y, Cheng MM, Gall J (2020) Mstcn++: Multi-stage temporal convolutional network for action segmentation. *IEEE transactions on pattern analysis and machine intelligence*.

Automatic purse-string suture skill assessment in transanal total mesorectal excision (TaTME) using 3D CNN-based video analysis

D. Kitaguchi^{1,2}, K. Teramura², H. Matsuzaki¹, H. Hasegawa^{1,2}, N. Takeshita¹, M. Ito^{1,2}

¹National Cancer Center Hospital East, Surgical Device Innovation Office, Kashiwa, Japan ²National Cancer Center Hospital East, Department of Colorectal Surgery, Kashiwa, Japan

Keywords TaTME, purse-string suture, surgical skill assessment, 3D CNN

Purpose

Transanal total mesorectal excision (TaTME) is a relatively new surgical approach for rectal cancer with multiple advantages in terms of oncological margins and pelvic functional outcomes. The skill of purse-string suture for rectal stump closure in TaTME is highly important because purse-string suture failure due to inadequate skill is hypothesized to directly affect the local recurrence of rectal cancer after TaTME [1, 2]. The objectives of this study are to develop an automatic skill assessment system for purse-string suture in TaTME using deep learning and to evaluate the reliability of the score output from the proposed system.

Methods

In order to assess the surgical skill, it is necessary for the deep learning model to recognise actions by analysing videos instead of static images. Three-dimensional convolutional neural network (3D CNN)-based deep learning models enable the analysis of information including both spatial and temporal dimensions and are nowadays utilised for various types of action recognition tasks. Therefore, Inception-v1 I3D two-stream (RGB + Optical-flow), a 3D CNN-based deep learning model, was also applied in this study.

Purse-string suture scenes extracted from 45 TaTME videos were divided into five video fragments. Subsequently, each video fragment was split into consecutive static images and input into the 3D CNN. In the final layer of the 3D CNN architecture, five groups of consecutive static images were aggregated again as a video clip, and image regression analysis was performed for each video clip. As pre-processing, every image was down-sampled from a resolution of 1280×720 pixels to a resolution of 224×224 pixels, and 30 frames per second (fps) to 10 fps. The maximum length of the video clip was 1 min owing to the GPU memory.

From all 45 cases, five video clips were randomly extracted, i.e. a total of 225 video clips were utilised in this study. The dataset was divided into training and test datasets and validated using the leave-one-supertrial-out (LOSO) scheme. In the LOSO validation scheme, one of five video clips in each case was extracted and included in the test dataset, i.e. the numbers of video clips in the training and test datasets were 180 and 45, respectively, and the video clips included in the training dataset were not presented in the test dataset.

Every video in the training dataset and their manual scores were input as the training data, and deep learning-based image regression analysis was performed for every video clip in the test dataset. The scores of purse-string suture skill in TaTME predicted by the trained deep learning model, which is called “AI score”, were output as continuous variables. The absolute error between the manual scores and AI scores was utilised as the evaluation metric for the model performance, and the correlations between the manual scores and AI scores, purse-string suture time and total AI score, and surgeons’ experience and total AI score were also evaluated.

Results

In the video dataset, purse-string suture was performed by an expert, surgeons with intermediate experience, and novice in 27 (60%), 8 (18%), and 10 (22%) cases, respectively, and the mean purse-string suture time was 3.2 (1.1–9.0) min. The mean total manual score was 9.2 (\pm 2.7) points, and the mean total AI score was 10 (\pm 3.9) points, i.e., the AI scores tended to be slightly higher than the manual scores. The mean absolute error between the AI scores and manual scores was 0.42 (\pm 0.39), and a strong correlation was observed between the AI scores and manual scores with statistical significance. There was a strong negative correlation between the AI score and the purse-string suture time with statistical significance, i.e. purse-string suture tended to close quickly due to the procedure with a high AI score. The AI score had no significant correlation between novices and surgeons with intermediate experience; however, the AI score of experts was significantly higher than those of novices and surgeons with intermediate experience.

Conclusion

We succeeded in developing an automatic purse-string suture skill assessment system using 3D CNN-based video analysis. The proposed AI score strongly correlated not only with the manual score but also with the surgeon’s experience and the time required, which indicated that the AI score was highly reliable and robust. To the best of our knowledge, this is the first report on the endeavour of automatic skill assessment for purse-string suture in TaTME. The proposed approach of automatic surgical skill assessment applies not only to TaTME but also to any endoscopic surgery.

References

- [1] Wasmuth HH, Faerden AE, Myklebust TÅ, Pfeffer F, Norderval S, Riis R, Olsen OC, Lambrecht JR, Kørner H, Larsen SG (2020) Transanal total mesorectal excision for rectal cancer has been suspended in Norway. *Br J Surg*;107(1):121–30.
- [2] Van Oostendorp SE, Belgers HJ, Bootsma BT, Hol JC, Belt EJTH, Bleeker W, Den Boer FC, Demirkiran A, Dunker MS, Fabry HFJ, Graaf EJR, Knol JJ, Oosterling SJ, Slooter GD, Sonneveld DJA, Talsma AK, Van Westreenen HL, Kusters M, Hompes R, Bonjer HJ, Sietses C, Tuynman JB (2020) Locoregional recurrences after transanal total mesorectal excision of rectal cancer during implementation. *Br J Surg*;107(9):1211.

Development of a manufacturer independent method for real-time data collection of surgical energy devices

P. van Esch¹, K. Hara¹, N. Takeshita², S. Takenaka², N. Takeshita², I. Masaaki², I. Sakuma¹, E. Kobayashi¹

¹The University of Tokyo, School of Engineering Department of Precision Engineering, Tokyo, Japan ²National Cancer Center Hospital East, NEXT Medical Device Innovation Center, Kashiwa, Chiba, Japan

Keywords OPeLiNK, surgical energy device, image recognition, endoscopic images

Purpose

Research about surgical analysis using intraoperative image is popular now. It is expected that more accurate analysis will be possible by adding the usage status of surgical energy device. [1] However, in an operating room, various types of energy devices are used. Therefore, the aim of this study is to develop an image recognition system that can acquire the usage status of various types of energy devices and integrate them with endoscopic images.

Methods

To be able to read the status of any energy device, a common communication protocol is required. We set the following requirements: the program must have an accuracy of at least 90% and can read and communicate the state of the device in real time, which is defined as less than 0.2 s.

To make sure the measuring system can measure any Energy device, a common state analysis was done, based on which a state tree was developed describing all possible states of any energy device. From this communication protocol, a reading strategy was developed, where each state-option is determined by its parent-state. This reduces the amount of state options, and thereby increases the accuracy and reduces the estimation time. Based on this communication protocol and reading strategy, the measuring system was developed.

The developed measuring system uses image recognition in the form of template matching to determine the state of the device. It consists of two programs, a setup program for the preoperative stage, and a main program for the intraoperative stage. During the preoperative stage, we can register any surgical energy device and make a classification tree for it. Based on this, we then classify, preprocess and store template images in the template database. The data from this database is then used during the intraoperative stage. During this stage the main program gathers visual data from one or multiple cameras filming the display of each energy device. These data are then compared to the preprocessed data from the database, and the states of the energy device are estimated. The main program uses two types of template-matching to estimate the state of the machine. For Light indicators, the luminance values are evaluated. And for numbers, text,

and symbols, we use Normalized cross correlation to estimate the state. The states of the energy device are integrated with endoscopic images using OPeLiNK system which were used in the real-time decision-making system, SCOT [2].

Results

Experiments were performed to evaluate the accuracy in a real operating environment. The experimental setup consists of a camera filming the display of an energy device. The camera is connected to the OPeLiNK system, which then creates multiple video files, each with a 31 s duration. The recording is done during the intraoperative stage, while the state evaluation is done postoperatively through a MATLAB program. The measuring system has been tested for both the Conmed System 2450, and the Harmonic & ENSEAL. Data acquisition is conducted with the approval of the ethic committee of NCC and the University of Tokyo. For the Conmed System, 1441 videos of four colon surgeries performed at NCC were evaluated. The preliminary results show a 97.3% accuracy under the condition where the camera does not move. For the Harmonic device, 316 videos of one colon surgery performed at NCC were evaluated. The preliminary results showed a 99.4% accuracy. The average reading time per frame is 0.129 s.

Conclusion

The preliminary results show a high accuracy and fast reading time that fits within our requirements. This indicates that our system is well suited for real time state estimation of energy devices in a surgical environment.

Acknowledgement: This study is supported by AMED (grant No. 21he2102001h1003).

References

- [1] Meeuwse C. F, Gudeon C. P. A, Arkenbout A. E, van der Elst M, Dankelman J, van den Dobbelsteen JJ (2017) The Art of Electrosurgery: Trainees and Experts. *Surgical innovation* 24(4): 373–378.
- [2] Okamoto J, Masamune K, Iseki H, Muragaki Y (2018) Development concepts of a smart cyber operating theater (SCOT) using ORiN technology. *Biomedizinische Technik* 63(1): 31–37.

A system for streamlined quality assurance of surgical trackers

I. Butz¹, M. Fernandez¹, A. Uneri¹, N. Theodore^{1,2}, W. S. Anderson^{1,2}, J. Siewerdsen^{1,2}

¹Johns Hopkins University, Biomedical Engineering, Baltimore, United States ²Johns Hopkins Medicine, Neurosurgery, Baltimore, United States

Keywords surgical navigation, quality assurance, tracking accuracy, streamlined QA

Purpose

Surgical tracking systems are prevalent components of minimally invasive surgery. Although a variety of quality assurance (QA) systems have been proposed for assessment of accuracy at the manufacturer site or in engineering labs [1], few are suitable to routine clinical use for systematic QA. We propose a streamlined QA system for use by medical physicists and/or service engineers in the

clinical environment. Inspired by analogous work in radiotherapy [2], we develop a streamlined form suitable to the operating theater.

Methods

The system incorporates a simple phantom manufactured without specialized materials or machining—a PVC cylinder with fiducial markers formed by conical divots in the surface. High-resolution CT (or cone-beam CT) of the phantom is acquired as a reference to define fiducial locations. A computer interface was developed for QA measurements, implemented as a module in 3D Slicer using Slicer-IGT and the open-source network communication protocols, OpenIGTLink and PlusProcess. The module supports data acquisition, computation of geometric accuracy, and generation of a structured QA report.

The system was evaluated in longitudinal (weekly) QA measurements using six tracking systems (2 electromagnetic and 4 infrared): 2 Aurora (NDI, Waterloo, Canada); 2 Polaris Vicra (NDI); 1 Polaris Spectra (NDI); and 1 Stealth Station S7 (Medtronic, Minneapolis USA). The time required for QA measurement was recorded from the time of tracker connection to output of the structured report.

For each tracker, fiducial positions $c_{i,j}$, $i \in \{1, \dots, N\}$, $j \in \{1, \dots, M\}$, were measured in world coordinates for each ($N = 10$) divot fiducial on the phantom, collecting $M = 30$ successive samples for each marker. True fiducial locations $f_{i,j}$, $i \in \{1, \dots, N\}$, $j \in \{1, \dots, L\}$, were defined in image coordinates by manual localization in the reference CT, repeated times (once offline—not with each QA test).

Geometric accuracy was assessed in terms of three outcome measures automatically computed and compiled in a structured QA report:

(1) **Tracker jitter (J_j)**. For each marker, the mean measured position c'_i was determined by averaging over M samples, and the corresponding mean-subtracted position measurement $c^{*}_{i,j} = c_{i,j} - c'_i$ was computed for each sample j . The jitter (J_j) was determined for each marker in terms of the standard deviation of the distribution of $c^{*}_{i,j}$, and the median and 95% confidence interval (CI95) over all J_j were computed.

(2) **Absolute distance error (Δd)**. For each marker, the mean reference position f'_i was determined by averaging over repeated localizations L . For all 2-element combinations (i,j) among N markers, the absolute distance error $\Delta d_{i,j} = |d_{i,j} - d'_{i,j}|$ was determined, where $d_{i,j}$ and $d'_{i,j}$ are the distance between measured world coordinates / reference coordinates of marker i and j , respectively, and the median and CI95 over all combinations, $\Delta d_{i,j}$, was determined.

(3) **Target registration error (TRE)**. The 6 DoF point registration between a set of tracked and reference divot positions was computed in a leave-one-out analysis ($N-1$ markers for registration and 1 marker, k , for analysis) via Horn's quaternion method. For the k th marker, the target registration error (TRE_k) was determined as the distance between registered and true marker locations in the CT domain, reporting the median and CI95.

Results

Figure 1 shows measurements of TRE from structured reports for the 6 trackers over the course of 10 weeks. The report additionally includes important metadata, such as tracker name and time stamp, and the verbose report includes raw measurements for retrospective analysis.

A variety of performance and quality findings are evident from the study. For the Aurora trackers, a source of interference (metal) in the original setup was discovered in the third week of trials (marked event (1)), and a reduction in TRE was realized by modifying the setup. Baseline tolerance values in Jitter, Δd , and TRE (e.g., for Aurora 1: median TRE = 1.6 mm, CI95 = 5.6 mm) were established for future reference. The Vicra systems exhibited performance comparable TRE to the Spectra but with increased Jitter (not shown). Here also, baseline and tolerance levels were established (e.g., Vicra 1: median TRE = 1.1 mm, CI95 = 1.8 mm) as a basis for QA action levels. The

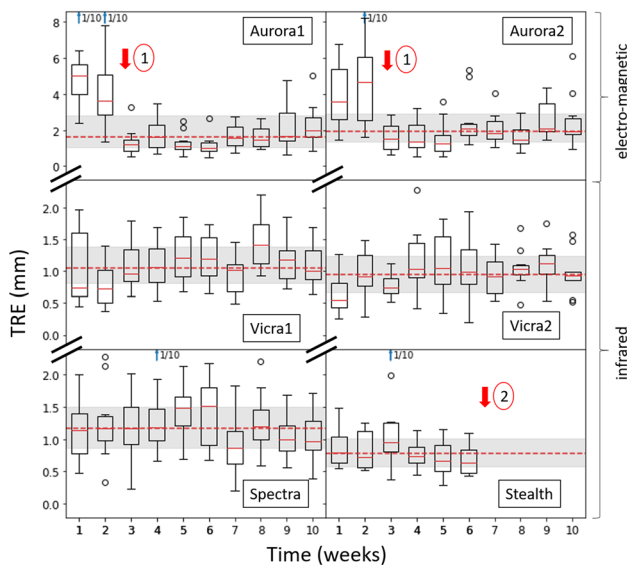


Fig. 1 Longitudinal TRE for six tracking systems over 10 weeks of the ongoing study. Median TRE (red dashed) and IQR (gray). Arrows mark events observed during the trials ((1) change in setup to reduce magnetic interference; (2) system failure). The IQR for each system are valuable in defining tolerance / action levels in routine QA

Table 1 Geometric accuracy and time required for the QA test for six tracking systems drawn from a particular week of the trial. Values shown are median (and CI95)

	Electromagnetic Trackers		Optical Trackers			
	Aurora1	Aurora2	Vicra1	Vicra2	Spectra	Stealth
Jitter (mm)	0.28 (0.55)	0.18 (0.29)	0.08 (0.13)	0.07 (0.12)	0.08 (0.10)	0.05 (0.06)
Distance error (mm)	1.8 (4.1)	1.6 (3.3)	1.3 (2.7)	1.1 (2.5)	1.1 (2.3)	0.8 (1.6)
TRE (mm)	1.6 (3.3)	1.4 (3.2)	1.2 (2.2)	1.0 (2.0)	1.1 (1.6)	0.7 (1.1)
Time (min)	1,3	1,5	1,4	1,2	1,6	not measured

StealthStation exhibited a system failure in the seventh week of the trial (marked event (2)), and retrospective analysis is underway to determine if the failure could be predicted from previous trials. The relative overall constancy is a positive finding, and such data are anticipated to be valuable to routine QA and troubleshooting in dynamic clinical environments where trackers are mobile and subject to collision.

Table 1 summarizes the Jitter, Δd , and TRE for each tracker. The time required for QA measurement in each case was ~ 1.5 min (sample size $M = 30$ and $N = 10$ fiducials), enabling routine checks. The open-source interface is suitable to operation by a technician, medical physics assistant, or service engineer.

Conclusion

The system facilitates routine QA for various modalities of surgical tracker with a short measurement time and streamlined interface. The system will help to translate common laboratory methodology to practical, clinical use. Longitudinal studies suggest tolerance / action levels appropriate to QA in a dynamic clinical setting. For trackers subject to proprietary software, the system is likely of greatest utility to field service engineers (cf., hospital technicians).

References

[1] Franz AM, Haidegger T, Birkfellner W, Cleary K, Peters TM, Maier-Hein L (2014) Electromagnetic tracking in medicine -A review of technology, validation, and applications, *IEEE Trans. Med. Imaging*, 33(8):1702–1725.
 [2] Willoughby T, Lehmann J, Bencomo JA, Jani SK, Santanam L, Sethi A, Solberg TD, Tome WA, Waldron TJ (2012) Quality assurance for nonradiographic radiotherapy localization and positioning systems: report of Task Group 147. *Med Phys*. 39(4):1728–47.

Construction of a surgical process identification system via a 5G network in awake, brain tumor surgery

A. Nakamura¹, I. Sato¹, M. Tamura², T. Yamaguchi², K. Kusuda², K. Yoshimitsu², Y. Muragaki², K. Masamune², Y. Fujino¹
¹Future University Hakodate, Faculty of System Information Science Engineering, Hakodate, Japan ²Institute of Advanced Biomedical Engineering & Science, Tokyo Women’s Medical University, Faculty of Advanced Techno-Surgery, Tokyo, Japan

Keywords Surgical process identification, 5G, Brain tumor, Awake surgery

Purpose

To maximize tumor removal and minimize the risk of postoperative complications during awake brain tumor surgery, the primary surgeon must grasp the patient’s brain structure and determine intraoperatively the resected tumor area. The intraoperative decision-making process adopted by the primary surgeon cannot be easily reproduced by junior doctors.

We have been developing a surgical process identification system using machine learning to automatically identify surgical processes in awake brain tumor removal surgery. This system defines 13 processes and three layers during awake, brain tumor surgery, and uses information from the surgical navigation system log, magnetic resonance images (MRI), and microscopy video. It can identify tumors with an accuracy > 90%. However, the introduction of this system to the field is costly owing to the type of equipment and operation and maintenance. Therefore, if the system can operate remotely via a communications network, and processing and analysis can be performed in real time, it can function as an advanced surgical support system that can be used from remote locations. For this purpose, we need a high-security and secure information sharing system. Therefore, in this study, we use the 5G network of the DOCOMO open innovation cloud (dOIC) [1]. In this study, we report on the construction of a brain tumor removal support system by analyzing the surgical process.

Methods

Figure 1 shows an overview of the proposed system. In this system, the user selects a case to be analyzed on a browser-based interface. Subsequently, the system analyzes the surgical process using information based on the medical devices obtained from the operating room, and displays the results. In this study, we constructed a system that can analyze remotely the surgical process by interconnecting the Tokyo Women’s Medical University and Future University Hakodate via a 5G network. The analysis of the surgical process analysis [2]

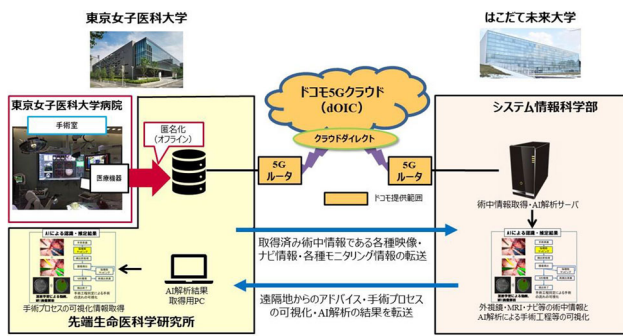


Fig. 1 System overview

Table 1 Surgical process model

1st Layer	2nd Layer	3rd Layer
Pre-tumor removal process	Preparation	Device setting
	Cortical mapping and pathology diagnosis	Verification of the tip position of the surgical instruments
		Examined on the cortex by electric stimulation probe
		Dialing with convulsive wave and marking of brain functional position during cortical mapping
		Cogulation of cortex
Tumor removal process	Tumor approach	Meninges incision and sampling lesion of glioma
		Venous and surface detachment
		Incision of cortex and clipping of arterial
		White matter resection, dissection tumor and removal tumor
	White mapping and pathology diagnosis	White mapping to the cavity of resection
		Dialing with convulsive wave and marking of brain functional position during cortical mapping
		Sampling lesion to the cavity of resection

consists of the following four steps: (1) acquisition of the type of surgical tool and location of the intraoperative procedure using preoperative and intraoperative MRI and surgical navigation system logs, (2) acquisition of the type of surgical tool using surgical microscopy images, and (3) identification of the surgical process using the acquired information. (4) The analyzed results were displayed on a browser interface.

In the first step, the types of surgical tools used by the primary surgeon and the locations of the procedures were obtained based on image processing using preoperative and intraoperative MRI and the surgical navigation log. In this process, the surgical tools are bipolar electric stimulation probes, and the four treatment positions are defined as “brain surface inside of tumor, and near normal tissue” in the brain region and the surgical field.

In Process 2, we used the images of the operating microscope to determine the types of surgical tools used by the surgeon during the operation using deep learning. This process supplements the information that cannot be obtained in Process 1. The surgical tools identified in this process include bipolar tools, electric stimulation probes, and shear blades. YOLO was used for deep learning. The YOLO used in this study uses a single video frame as the input and outputs the name of the tool in the region if a surgical tool is detected in the input.

In Process 3, the information obtained in Processes 1 and 2 is integrated and saved in a time series. Subsequently, using the stored

data as input, we identified the surgical process using a hierarchical hidden Markov model. The model used in this study has three levels: the first level classifies whether the patient is undergoing tumor removal or not, the second level classifies the purpose of the procedure, and the third level classifies the process from the end of MRI to the end of tumor removal. The second level classifies the purpose of the procedure, and the third level defines the surgical process from the end of MRI to the end of tumor removal. Table 1 presents an overview of the models used in this study.

Finally, Process 4 temporarily stores the analyzed results of the process at regular intervals, and outputs the current surgical process at the top of the interface, the transition of the surgical process on the right side, and the current surgical microscopic image on the left side.

Results

In this study, we evaluated the quality of data transmission for the analysis of the assisted surgical process. In this experiment, we evaluated the latency and packet loss of the RTP streaming video transmission using an OBS and VLC. The microscopy video and system time were displayed on the monitor of the transmission side. The delay was evaluated by comparing the system time of the receiving side and the system time of the video acquired at the receiving side. Accordingly, the packet loss data were analyzed by capturing the packets during video delivery. The video delivered had a 4 k quality, and the bit rate was reduced every 10 Mbps from 150 to 70 Mbps. When the bit rate was between 150 and 110 Mbps, packets were received, but the video was not displayed on the VLC, and the delay could not be measured. It was found that the video data used for surgical processing and analysis can be transferred with high quality, and surgical analysis could be performed via the constructed 5G network.

Conclusion

In this study, we constructed and demonstrated a surgical process identification system via a 5G network for awake, brain tumor surgery. In the demonstration experiment on information sharing, the delay was 1.329 s and the packet loss was 3% when the bit rate was 80 Mbps. The experimental results showed that surgical analysis was possible using the constructed 5G network. In the future, we will evaluate the delay and packet loss when the analysis function is used, aiming to make the analysis more detailed and improve accuracy.

References

[1] <https://www.nttdocomo.co.jp/biz/service/doioc/>, accessed January 2022.
 [2] Nagai T, Sato I, Fujino Y, Horise Y, Kusuda K, Tamura M, Muragaki Y, Masamune K (2022) Surgical Process Identification System in Awake Surgery for Glioma. Journal of Japan Society of Computer Aided Surgery 22(2):87–101.

24th International Workshop on Computer-Aided Diagnosis and Artificial Intelligence

Chair: Hiroyuki Yoshida, PhD (US)

Performance evaluation of self-supervised 3D GAN for electronic cleansing in CT colonography

R. Tachibana^{1,2}, J. Näppi², T. Hironaka², H. Yoshida²

¹National Institute of Technology, Oshima College, Suo-Oshima, Japan ²Massachusetts General Hospital and Harvard Medical School, Radiology, Boston, United States

Keywords CT colonography, electronic cleansing, generative adversarial network, self-supervised learning

Purpose

Early detection and removal of the benign precursor polyps of colorectal cancer would prevent the development of the cancer. CT colonography (CTC) provides a safe and accurate method for examining the complete region of the colon, and it has been recommended by the U.S. Preventive Services Task Force and the American Cancer Society as an option for colon cancer screening. Electronic cleansing (EC) performs virtual subtraction of residual materials from CTC images to enable radiologists and computer-aided detection (CADe) systems to detect polyps that could be submerged in the residual materials. However, current EC methods tend to produce EC image artifacts, and radiologists find even the smallest artifacts distracting [1].

Previously, we developed an EC scheme where some of the training samples were self-generated by use of a self-supervised 3D generative adversarial network (GAN) [2]. Unlike previous EC methods, the 3D-GAN EC could be trained to transform an uncleansed CTC volume directly into the corresponding virtually cleansed image volume.

In this study, we evaluated the image quality of EC by the self-supervised 3D-GAN EC scheme for CTC with both quantitative and visual assessments based on a phantom study and clinical CTC cases.

Methods

An anthropomorphic phantom (Phantom Laboratory, Salem, NY) designed to imitate the human colon in CT scans was filled with 300 cc of simulated fecal material, which was tagged by use of a non-ionic iodinated contrast agent (Omnipaque iohexol, GE Healthcare) at three different contrast agent concentrations (20, 40, and 60 mg/ml). The native (empty) and the three different partially filled versions of the phantom were scanned by use of a CT scanner (SOMATOM Definition Flash, Siemens Healthcare, Erlangen, Germany) with 120 kVp, 68 mA, 0.6-mm collimation and 0.6-mm reconstruction interval. We also acquired eighteen clinical CTC patient cases with reduced cathartic preparation, where the CT image volumes were scanned by use of a CT scanner (SOMATOM Definition Flash) and reconstructed at 120 kVp, 1.0-mm slice thickness, 0.59–0.76 mm pixel spacing, and 0.7-mm reconstruction interval.

The architecture of the 3D GAN of the EC scheme consists of a 3D generator (3D U-Net) and a 3D discriminator network. Given an uncleansed CTC image volume, the generator network is trained to generate the corresponding EC image volume. The 3D GAN was pre-trained with a supervised-training dataset, where we used 200 paired volumes of interest (VOIs) extracted from precisely matching lumen locations of the CTC datasets of the colon phantom acquired without and with 20 mg/ml and 60 mg/ml contrast concentrations. The 3D GAN was subsequently trained iteratively with a self-training dataset, where 100 paired VOIs extracted from each input volume were paired with VOIs cleansed by the 3D GAN itself.

For an objective assessment of the image quality of EC, we calculated the peak signal-to-noise ratio (PSNR) between the EC VOIs of the CT datasets of the fecal-tagged phantom acquired with 40 mg/ml contrast agent concentration and those of the corresponding native

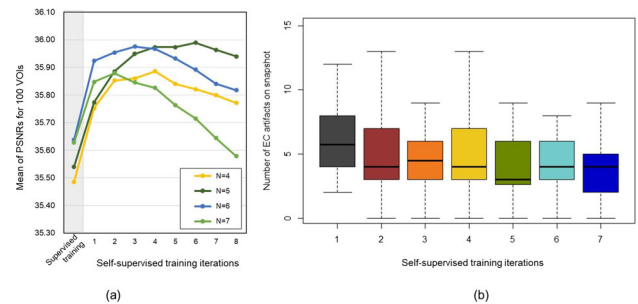


Fig. 1 **a** The mean value of the PSNRs for 100 VOIs of the phantom cleansed by use of the self-supervised 3D-GAN scheme, where N indicates the number of convolution/deconvolution layers. **b** Comparison of the numbers of EC artifacts observed over the self-supervised training iterations on the virtual endoluminal examinations of the eighteen clinical CTC cases

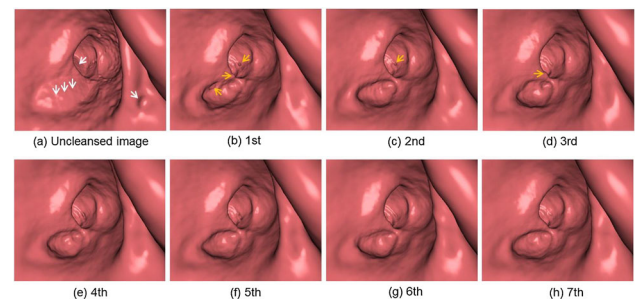


Fig. 2 An example of the progress of self-supervised training on a virtual endoluminal view of a clinical CTC case. In **(a)**, the white arrows indicate tagged fecal materials. In **(b)–(h)**, the orange arrows indicate EC artifacts

phantom. The image quality was assessed for several generators containing 4–7 convolution layers and over eight self-supervised training iterations of the 3D-GAN EC scheme. The statistical significances of the differences of the PSNRs were tested by use of the paired t-test with Bonferroni correction.

For visual assessment of the image quality in clinical cases, we also calculated the numbers of EC image artifacts that were observed on the virtual endoluminal view examinations of the eighteen clinical CTC cases.

Results

In the phantom study, the best image quality was obtained when the generator was trained with five convolution/deconvolution layers at six self-supervised training iterations (Fig. 1a). The PSNR values obtained after the self-supervised training iterations were statistically significantly higher ($p < 10^{-5}$) than those obtained from supervised training.

Visual assessment on the clinical cases indicated that the number of EC artifacts was reduced gradually as the number of self-supervised training iterations increased (Fig. 1b).

Figure 2 illustrates the differences in the virtual endoluminal views generated by the self-supervised 3D-GAN EC scheme with five convolution/deconvolution layers over the self-supervised training iterations.

Conclusion

We developed a self-supervised 3D-GAN EC scheme that can be used to convert uncleansed input CTC image volumes directly into their corresponding output EC volumes. A pilot study was performed to

evaluate the image quality of the scheme. Our preliminary results suggest that the image quality is highest when the generator is constructed with five convolution/deconvolution layers and trained for six self-supervised training iterations.

Acknowledgements

This study was supported partly by JSPS KAKENHI Grant 21K07578 (PI: Tachibana) and the NIH/NIBIB/NCI grants of R21EBO24025 (PI: Näppi), R01CA212382 (PI: Yoshida) and R01EB023942 (PI: Yōshida).

References

- [1] Pickhardt PJ (2017) Imaging and screening for colorectal cancer with CT colonography. *Radiol Clin N Am* 55:1183–1196.
- [2] Tachibana R, Näppi JJ, Hironaka T, Yoshida H (2019) Generative adversarial learning-based electronic cleansing system for CT colonography. *Computer Assisted Radiology and Surgery (CARS): 33th International Congress and Exhibition*, S87–S88.

Automatic detection of bladder tumors in narrow-band imaging cystoscopic images by tiny-YOLO

Mutaguchi¹, M. Oda^{2,3}, E. Kashiwagi¹, J. Inokuchi¹, K. Mori^{2,3,4}, M. Eto^{1,5}

¹Kyushu University Hospital, Urology, Fukuoka, Japan ²Nagoya University, Intelligent system, Graduate School of Informatics, Nagoya, Japan ³Nagoya University, Information Strategy Office, Information and Communications, Nagoya, Japan ⁴Research Center for Medical Bigdata, National Institute of Informatics, Tokyo, Japan ⁵Kyushu University, Advanced Medical Initiatives Faculty of Medical Sciences, Fukuoka, Japan

Keywords Object detection, Artificial intelligence, Cystoscopic images, Narrow-band imaging

Purpose

Bladder tumors have high intravesical recurrence rate after transurethral resection of bladder tumor (TURBT). Especially, overlooking tumors during TURBT occasionally causes early intravesical recurrences. White light imaging (WLI) is the gold standard method, however it sometimes overlooked small and flat tumors. Narrow-band imaging (NBI) has been developed to detect more bladder tumors, and improved the detection sensitivities comparing to the conventional WLI. However, the accuracy of NBI are less objective and poor reproducible because this strategy are operator-dependent method. Recently, artificial intelligence (AI) has been applying in diagnostic imaging. Object detection is the useful AI method to detect the lesions with objective and reproducible, and has possibility to reduce overlooking as same as expert doctor's level. This technique has potential to improve tumor detection rate in cystoscopy examination, however a few studies reported the feasibility of AI in cystoscopy. Furthermore, the accuracy of AI in those prior studies were evaluated by only WLI images, and there are no reports about AI system evaluated by NBI cystoscopic images.

In this study, we aimed to develop an AI system with deep learning to detect bladder tumors in NBI cystoscopic images, and evaluate its detection accuracies.

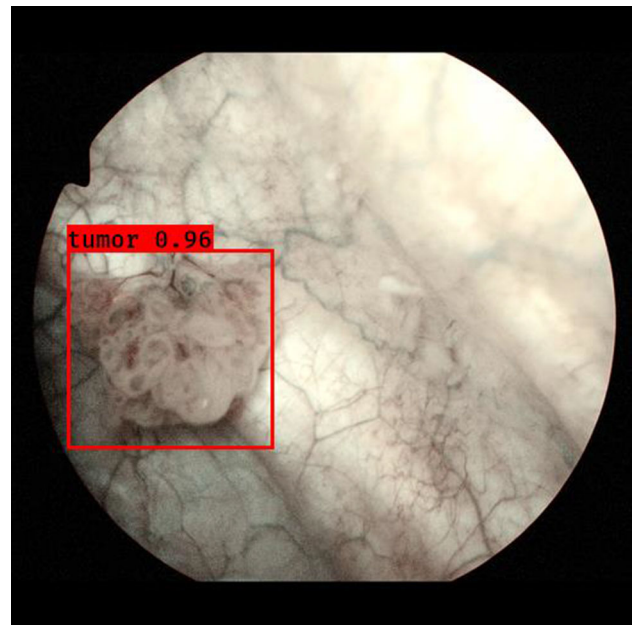


Fig. 1 The detection example by tiny-YOLO in a NBI cystoscopic image

Methods

We constructed an object detection system based on AI architecture termed tiny-You Only Look Once (tiny-YOLO) that is deep neural networks for fast and accurate object detection, Fig. 1. To obtain cystoscopic images and create annotation images, we conducted following process. First, we prospectively observed the tumors during TURBT from December 2019 to April 2021 at Kyushu University Hospital. Second, we converted the TURBT videos into frame images, and selected clear and some blurred images. All selected images were cropped and resized 512 × 512 pixels for suitable to AI training and test. Finally, all tumors in the selected cystoscopic images were annotated as rectangle by using annotation software.

We divided those images into training and testing data. Training data were consisted of the cystoscopic images from December 2019 to January 2021. To assess the detecting accuracy of AI, we used an independent test set that consisted of cystoscopic images obtained from February 2021 to April 2021. As the evaluating metrics, we calculated sensitivity, specificity, positive predictive value (PPV) and F-value. We also assessed AI performance by calculating the area under the receiver operating characteristics curve (AUC).

Results

We obtained 1,410 cystoscopic images containing bladder tumors from 105 patients. 1,191 images from 91 patients were used for the AI training. The accuracy of AI system was assessed by 219 tumor images from 14 patients and 108 non-tumor images. The trained AI required 54 s to evaluate 327 test images. The sensitivity, specificity, PPV and F-value of AI system were 82.2%, 81.4%, 90.0%, and 85.9%, respectively. The AUC for the detection of bladder tumors in test cystoscopic images was 0.905.

Conclusion

We proposed the objective and reproducible system implemented with tiny-YOLO to detect bladder tumors in NBI cystoscopic images. Our system has the possibility to improve detection rates of the bladder tumors during cystoscopy examination, and might be beneficial to reduce recurrence rate of bladder tumor after TURBT in the future.

Audio based analysis of dysphagia using cervical auscultation

J. Fuchtmann¹, D. Ostler¹, D. Bani-Harouni¹, L. Richter², B. Schilling², H. Feußner¹, D. Wilhelm¹, S. Graf²

¹University Hospital rechts der Isar, MITI Research Group / Department of Surgery, Munich, Germany ²University Hospital rechts der Isar, Department of Otorhinolaryngology, Munich, Germany

Keywords Dysphagia, Cervical auscultation, Audio analysis, Deep learning

Purpose

Swallowing is a complex procedure that requires the seamless coordination of multiple muscles and neural signals. Structural and/or neurological lesions can cause different forms of swallowing problems, called dysphagia. If substances cannot be safely swallowed into the stomach, they pass into the lungs and may cause pneumonia, which is associated with high mortality, high morbidity and high cost for the healthcare system.

However, today's diagnostic for dysphagia relies on experienced, highly qualified medical staff, that is only available at specialized centers, in conjunction with complex instrumental techniques. Commonly a Videofluoroscopic Swallowing Study—swallowing of contrast agents under x-ray control—is performed or the act of swallowing is visually assessed using a flexible endoscope, that gets inserted via the nasal orifice.

As dysphagia is quite common, especially amongst elderly people, there exists a high number of unidentified cases. Hence, development of a less invasive and more simple diagnostic procedure for a routine screening of potential patients would be desirable. Recent research primarily focused on the recognition of individual swallows. Khalifa et al. [1] reported a swallow detection accuracy of at least 95% by combining audio and accelerometric signals in a data set with more than 3000 samples. By using high resolution cervical auscultation Yu et al. were even able to proof an existing association between the medical scoring in the "Penetration Aspiration Scale" and features in their recorded signals, classified with techniques such as SVMs, K-means and Naive Bayes. From a medical perspective however, available literature treats dysphagia quite undifferentiated and mixes different sub-types of disorders when collecting and labeling the ground truth data. To yield a better performance, our approach focuses on a differentiated labeling in alignment with international established scoring standards in order to develop a deep learning based low cost solution for diagnostic purposes.

Methods

In our approach, we aim to develop a hardware solution that can be used to record cervical sounds for gathering training data, but also for the instant diagnostic application through analysis. To fit the individual anatomy of the patient we rely on a flexible collar that can be adjusted in its length. Using two medical approved stethoscope heads in conjunction with miniature microphones, ensures the volunteers safety during the data gathering process as well as precise sound capturing of both sides of the neck. Signals are subsequently recorded using a low cost single-board computer, which is also meant to classify the incoming audio once training is completed. The setup and the further workflow is depicted in Fig. 1. Sample generation is currently ongoing and includes healthy volunteers, as well as patients with dysphagia who come to the outpatient clinic at the department of

otorhinolaryngology at the university hospital rechts der Isar, Munich, for dysphagia examinations. While swallow signals are recorded, patients simultaneously undergo a video endoscopic swallowing examination. Video recordings are then rated by medical experts based on internationally well established scales for the classification of dysphagia (EAT 10-Score, Penetration-Aspiration-Scale, Yale Pharyngeal Residue Severity Rating Scale, Murray Secretion Assessment Scale). Videos are stored and synchronized with the audio data. Upon these ratings, audio recordings are labeled by the physicians regarding the underlying disorder and severity.

For training of the classifier, audio recordings were cut into individual samples, containing one swallow each. Average sample length was set to one second with a sample rate of 44,100 Hz, as initial tests have shown that most swallows are thereby completely covered. As we chose a spectrogram feature based approach for the classification, we applied a rectangular sliding window function to the individual sound clips in the data set and computed log-mel-spectrograms based on a Short-Time Fourier Transformation and a Mel filter bank with 64 filters. The window length was thereby set to 1024 samples with an overlap of 50%.

Results

The proposed hardware setup delivers high quality audio of the act of swallowing. Hence, in an initial step, incoming swallow sounds have to be reliably recognized before medical classification can take place. Therefore 566 samples from 25 healthy volunteers with three types of swallows (dry/only saliva, water, food) were recorded and processed as described in the Method section. 411 additional samples of 5 healthy volunteers were extracted for the creation of an idle class, including sounds of silence, breathing, vocalization of the patient or talking in the background.

Using the CNN-14 model by Kong et al. [2], we optimized a deep learning classifier with and without pretrained weights based on the large-scale AudioSet data-set. Comparing our yielded 95% accuracy with the results of e.g. Khalifa et al. [1], who used nearly six times the sample size while achieving 95% accuracy as well, stresses the importance of transfer learning for the application of deep learning based spectrogram classification. While the project is still work in progress, and pathological swallow samples are currently gathered, we additionally trained the very same classifier for the separation of

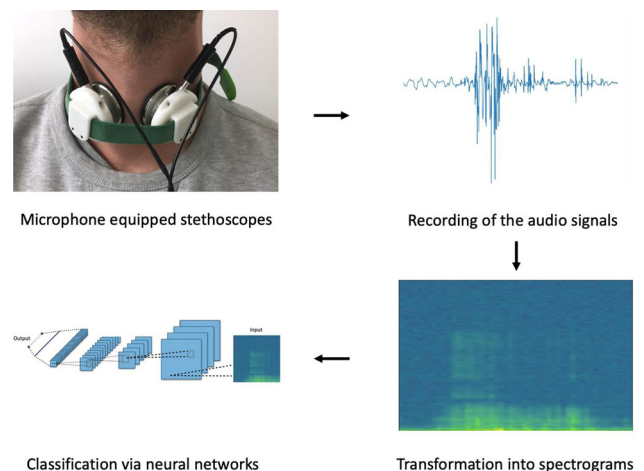


Fig. 1 Workflow of the training process and future application. Audio signals of the swallow-sounds are captured by microphone equipped stethoscopes, recorded and transformed from a waveform into a log-mel-spectrogram. Subsequently processed spectrograms present the input for a neural network, either for training or—later on—for prediction of potential dysphagia

Table 1 Test results with STD for the used CNN-14 network [2] with and without pretraining

	CNN-14		CNN-14 pretrained	
	Accuracy	±0.106	0.95	±0.038
Swallow Recognition	Av. Precision	0.83 ±0.094	0.98	±0.013
	F1 Score	0.66 ±0.156	0.96	±0.032
Swallow Type Recognition	Accuracy	0.35 ±0.026	0.56	±0.026
	Av. Precision	0.39 ±0.037	0.59	±0.033
	F1 Score	0.35 ±0.026	0.56	±0.026

the three types of swallows. However, here current results indicate the need for further improvements. All results, as listed in Table 1, are based on a fivefold cross validation with one fold each for validation and testing with micro averages for multiclass metrics.

Conclusion

Current diagnostic for dysphagia not only requires expert knowledge, but also an expensive technical setup. Goal of this project is to facilitate the diagnostic process by capturing the sounds of swallowing using cervical auscultation and classifying the audio signals based on a deep learning pipeline. While the project is still work in progress, we were able to reliably identify relevant audio events, that can be further analyzed once sufficient pathological samples are recorded. First results for the separation of dry, liquid and solid based swallows indicate the feasibility of classifying audio signals by cervical auscultation. Combining an audio recording functionality and local signal classification within a handheld device would ultimately improve future dysphagia screening not only for patients, but also for physicians.

References

- [1] Khalifa, Y., Coyle, J.L., Sejdic, E.: Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings. *Scientific Reports* 10(1), 1–13 (2020).
- [2] Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., Plumley, M.D.: Panns: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28, 2880–2894 (2020).

Prognosis prediction and its visualization of the patient in coronary care unit using electrocardiogram

T. Kondo¹, A. Teramoto¹, E. Watanabe², Y. Sobue², H. Izawa³, K. Saito¹, H. Fujita⁴

¹Fujita Health University, Graduate School of Health Sciences, Toyoake, Japan ²Fujita Health University Bantane Hospital, Department of Internal Medicine, Nagoya, Japan ³Fujita Health University, Department of Cardiology, School of Medicine, Toyoake, Japan ⁴Gifu University, Faculty of Engineering, Gifu, Germany

Keywords electrocardiogram, prognosis prediction, deep learning, activation map

Purpose

Electrocardiogram (ECG) is a useful test for patients with various diseases including cardiovascular diseases because it is minimally invasive and allows real-time confirmation of waveforms. In particular, the ECG reading of patients admitted to the coronary care unit (CCU) is required to rapid and accurate because of the high risk of further serious illness. However, the ECG waveform of patients in the CCU is affected by the patient's vital signs and complications of cardiac diseases, so the ECG reading varies greatly depending on the experience and skill of the cardiologist. Therefore, we focused on

ECG analysis using machine learning. In a previous study, Raghunath et al. used DNN to predict 1-year survival in 12-lead ECGs [1]. The area under the curve (AUC) was 0.88, indicating that it could predict patient prognosis with high accuracy. However, this study was not conducted for the patients in CCU. Furthermore, since numerical data was input, it was not clear whether detailed analysis of the shape of the ECG waveform could be performed. Therefore, we aimed to develop prognosis prediction and its visualization methods using two dimensional convolutional neural network (2D-CNN) and gradient weighted class activation mapping (Grad-CAM) for ECG images at admission in CCU.

Methods

Of the 10,023 12-lead ECGs of patients admitted to the CCU of the Fujita Health University Hospital, we used ECG data from a total of 618 patients, 309 of whom died while in the CCU and 309 of whom survived. We used II, V3, V5, and aVR inductions, which have been shown to have high prognostic accuracy among 12-lead ECGs¹. In addition, these numerical data were converted into three images per case. From these transformed images, 548 cases were fine-tuned with 7 types of 2D-CNNs (ResNet50, InceptionV3, VGG16, VGG19, DenseNet121, DenseNet169, and DenseNet201) pre-trained using Image Net database. The classification performance of each CNN was compared using unseen 70 cases. The average of three continuous values for three image outputs per patient was used as the evaluation for the patient. In addition, a color map showing the CNN's basis for judgments was obtained by Grad-CAM [2].

Results

To verify the proposed method, we conducted a performance evaluation using hold-out Validation. The training of 7 types of 2D-CNNs were conducted using a computer with an XEON CPU and dual NVIDIA Quadro RTX8000 GPU. We evaluated accuracy, sensitivity, specificity, and AUC as the performance metrics of CNNs. VGG16 showed the best results in all of them; accuracy, sensitivity, specificity, and AUC were 81.4%, 80.6%, 82.4%, and 0.862, respectively. The upper part of Fig. 1 shows the ROC curve with VGG16. These results showed that the prediction was highly accurate for CCU patients, whose ECG waveforms are often complex. The bottom part

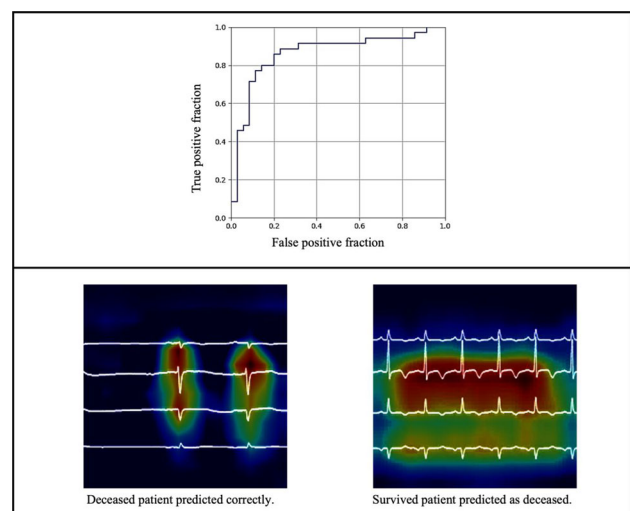


Fig. 1 ROC curve with VGG16 (upper), and outputs of the Grad-CAM and Input images (bottom)

of the Fig. 1 shows a superimposed image of the input image and the color map output by Grad-CAM. (a) In the case of correctly predicting deceased patients, it was found that the CNN focused on the shape of the R-wave, such as its size and direction. (b) In the case where a surviving patient was incorrectly predicted to be deceased, the CNN focused on tachycardia and predicted death. These results suggested that the prediction was based on the regularity of the waveform.

Conclusion

In this study, we conducted mortality prediction by 2D-CNN was performed using ECG waveforms of CCU patients. The predicted correct rate, sensitivity, specificity and AUC were 81.4%, 80.6%, 82.4% and 0.862, respectively. And the visualization by Grad-CAM showed that CNN tended to focus on the regularity and shape of the ECG waveforms, suggesting that this method may be effective in predicting the mortality of CCU patients.

References

- [1] S. Raghunath, A. Cerna, L. Jing, P. vanMaanen, J. Stough, D. Hartzel, J. Leader, H. Kirchner, M. Stumpe, A. Hafez, A. Nemani, T. Carbonati, K. Johnson, K. Young, C. Good, J. Pfeifer, A. Patel, B. Delisle, A. Alsaïd, D. Beer, C. Haggerty, B. Fornwalt, “Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network”, *Nature Medicine*, Vol.26, pp.886–891, 2020.
- [2] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, “Grad-CAM: visual explanations from deep networks via gradient-based localization”, *IEEE*, pp. 618–626, 2017.

3-D split attention CNN-based tumor detection for automated breast ultrasound image

F. Chang¹, Y. S. Huang¹, C. W. Tu¹

¹National Taiwan University, Computer Science and Information Engineering, Taipei, Taiwan

Keywords automated breast ultrasound, one-stage object detection, attention mechanism, YOLOv4

Purpose

The automated breast ultrasound (ABUS) has been widely used in breast cancer screening for early detection. It could scan the whole breast and provide the complete breast three-dimensional (3-D) volume. Despite the time efficiency of ABUS, viewing hundreds of 2-D slices is still time-consuming. Therefore, a computer-aided detection (CADe) system is needed to assist radiologists. Recently, convolutional neural networks (CNN) have been widely used in medical images, and the CNN-based CADe could achieve outstanding performance. Hence, a 3-D split attention CNN-based CADe system is proposed for tumor detection in ABUS in this study.

Methods

The proposed CADe system consists of image resizing, tumor detection, and post-processing. In the image resizing, the input image is resampled to consistent spacing and resized to fit the CNN model. Then, the resized image is fed into the proposed detection model, 3-D split attention YOLOv4 (SA-YOLOv4), for tumor detection. The proposed model is modified from the YOLOv4 (You Only Look Once V4) for its high efficiency and excellent performance. To further improve feature extraction, the split attention (SA) block of ResNeSt is employed as the backbone. Furthermore, the multi-stage training strategy is proposed for false positive reduction. Finally, the non-maximum suppression (NMS) removes overlapped bounding boxes.

Results

This study collected the used dataset from collected by InveniaTM automated breast ultrasound system (Invenia ABUS, GE Healthcare). An automatic linear broadband transducer obtains all ABUS volumes with a covering area of 15.4'17'5 cm. Each ABUS volume is made of 330 serial 2-D images, and each 2-D image consists of 831'422 pixels, and the distance between each slice is 0.5 mm. A total of 348 ABUS images containing 523 tumors are used in our experiment. The results show that the proposed model achieves a sensitivity of 94.2% at false positives per pass of 4.0, and the model has a robust performance for tumors of different sizes. Compared to the original YOLOv4, the proposed model is more accurate and the proposed modifications can remarkably improve detection models.

Conclusion

In this study, a 3-D CNN-based CADe system, SA-YOLOv4, was proposed for tumor detection in 3-D ABUS images. The proposed model employed the split attention mechanism of ResNeSt to enhance feature extraction. In addition, the proposed multi-stage training strategy remarkably reduced the number of FPs generated by the model to half. The proposed model used high-resolution feature maps and low-level shortcuts to improve the sensitivity of small tumors. The results showed that the proposed method remarkably improved the accuracy of YOLOv4 for tumor detection, and the proposed system was robust to tumors of different sizes.

Deep Convolutional Neural Networks (CNN) for lesions localization of in mammograms

X. Lessage¹, H. Benzaouia^{1,2}, S. Murgo³, S. Mahmoudi¹, S. A. Mahmoudi¹, Z. Guessoum²

¹UMONS, Computer Science, MONS, Belgium ²URCA, Sciences Exactes et Naturelles, Reims, Germany ³CHR Mons, Radiology, Mons, Belgium

Keywords X-ray analysis, localization, convolutional neural networks, breast cancer

Purpose

The latest advances in machine learning and in particular with convolutional neurons (CNN) have proven more than once their great accuracy in the detection of anomalies. Deep learning algorithms, in particular convolutional neural networks, have rapidly become a methodology of choice for analyzing medical images. In this paper, we present a new approach for mass detection from mammogram X-ray images using Deep Learning algorithms. An efficient process based on Yolo v5 is proposed in this paper. Experiments have been conducted using an anonymized database from a Belgium hospital

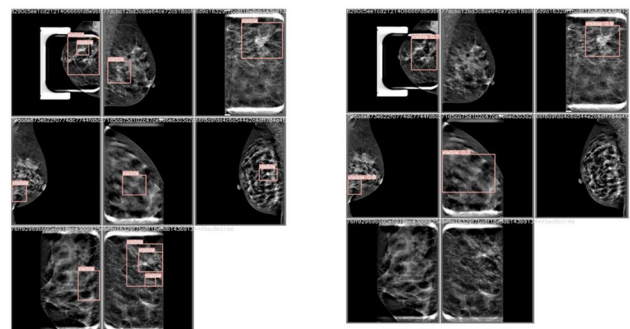


Fig. 1 Comparison between ground truth and prediction

thanks to a retrospective study. It is composed of two classes (benign, malignant).

Methods

YOLO is one of the most successful object detection algorithms used in recent times. We chose the final version of YOLO because of its speed, performance and better ratio of execution time and accuracy.

On the other hand, the annotation is a very important step in the training process. To this aim we have developed and used a QT/Python application to label the images and export them. This solution was used to frame the anomalies as well as the nipples and to store the coordinates in an Excel file with various important information such as (density, types of anomalies, etc.) in order to finally create a file in the desired YOLOV5 format.

The experiments with the YOLOV5 model were executed on a Linux cluster node with 32 CPU cores using a single NVIDIA GeForce GTX 980 with 4 GB memory. Keras 2 with Tensorflow 1.8 backend was used as a deep learning framework.

Results

We have used a database composed of 792 images and we achieved a precision of 84%, a recall and mAP of 50%. These results are very promising considering the number of images used. Ideally, the number of images used to train our model should be around 1500 images and not 792 images, but we did not have additional images with our private database from the Belgian hospital.

In Fig. 1 below, shown a comparison between the ground truth and the prediction of our model on 8 images. On the left what is annotated by the radiologist and on the right what is detected by the YOLO localization model. This model was able to identify anomalies in 4 images with good accuracy.

Conclusion

Our approach based on in-depth learning is very promising for detecting pathologies, based on chest X-ray images. We can improve this result by increasing the number of images per class using data augmentation.

References

- [1] Charan S, Khan MJ, Khurshid K (2018) Breast cancer detection in mammograms using convolutional neural networks. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCo- MET) 2018, IEEE, pp. 1–5.
- [2] Ramachandran S, George J, Skaria S, Varun V (2018) Using yolo based deep learning network for real time detection and localization of lung nodules from low dose CT scans. In Medical Imaging, Computer-Aided Diagnosis (2018), vol. 10575, International Society for Optics and Photonics, p. 105751I.

A computer-aided diagnosis system-3-D SGE-SANet for lung nodule classification on low-dose computed tomography

Y. S. Huang¹, T. C. Lin¹, R. F. Chang¹

¹National Changhua University of Education, Department of Computer Science and Information Engineering, Changhua, Taiwan

Keywords Low-dose computer tomography, Computer-aided diagnosis, 3-D Split attention, 3-D Spatial group-wise enhance

Purpose

Lung cancer has been pointed out as a leading cause of cancer death in the world. Low-dose computed tomography (LDCT) is an essential tool for lung cancer detection and diagnosis because it can provide a complete three-dimensional (3-D) chest image with high resolution. However, the nodules with ambiguous shape and texture, and radiologist's experience would result in different diagnosis results. Therefore, the computer-aided diagnosis (CADx) system was developed for assisting radiologist. Recently, designing a CADx system based on the convolution neural network (CNN) has flourished in the field of medical images because of the automatic feature extraction and powerful performance in diagnosis. Many researches have shown the evidence that a CNN-based CADx system could assist radiologists to make a preliminary decision. Therefore, a 3-D CADx system, 3-D SGE-SANet, was proposed for lung nodule diagnosis.

Methods

In this research, the proposed 3-D SGE-SANet for lung nodule diagnosis was composed of the volumes of interest (VOIs) extraction and a 3-D SGE-SANet nodule classification model. The VOIs were defined and normalized into the range from -1 to 1 first in the VOI extraction. In the nodule classification model, the 3-D SGE-SANet which integrated 3-D spatial group-wise enhance block with 3-D split attention mechanism as the backbone was built. The defined VOIs are then delivered to the 3-D SGE-SANet to determine nodule as malignant or benign.

Results

In this research, the dataset was collected from the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI). In this dataset, the nodules whose diameter ranging from 3 to 30 mm were annotated by four experienced thoracic radiologists, and the malignancy level was evaluated in a 5-point scale where 1 represented a benign nodule and 5 represented a malignant nodule. Because of the collation of scans from different institutions and scan vendors, the pixel spacing in each CT scan were different and the slice thicknesses of the CT scan ranged from 0.45 mm to 5 mm with a median of 2 mm. The resolution of each slice was 512×512 pixels. As recommended by the American College of Radiology, scans with a slice thickness greater than 3 mm, inconsistent slice spacing or missing slices should be discarded. Therefore, there were totally 716 nodules consisting of 302 benign nodules and 414 malignant nodules. For system validation, three performance indices, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area under the ROC curve (AUC), and fivefold cross validation, are used to validate our CADx system. In experiments, the proposed system's accuracy, sensitivity, specificity, PPV, NPV, and AUC are 90.9%, 92.0%, 91.7%, 93.8%, 89.4%, and 0.9618, respectively. In conclusion, it was confirmed that the proposed system had a good diagnostic capability for discriminating malignant nodules from benign ones.

Conclusion

In this study, a CNN-based CADx system made of the VOI extraction and the 3-D SGE-SANet classification model was proposed for lung nodule classification in LDCT images. According to the results, the proposed CADx system in this study could achieve a high performance. In future, the system would be further improved by using other CNN architectures or learning strategies.

Improvement of lung nodule classification performance using progressive growing channel attentive non-local networks

M. Al-Shabi¹, M. Tan¹

¹Monash University Malaysia, School of Engineering, Subang Jaya, Selangor, Malaysia

Keywords Self-Attention, Non-local network, Nodule classification, Curriculum learning

Purpose

Lung cancer is the most frequently diagnosed cancer (i.e., 11.6% of all cancer cases) [1]. Although we can reduce lung cancer mortality by examining with Low Dosage Computed Tomography (LDCT) in early stages, however, manual diagnosis by the radiologist is time-consuming as every CT scan consists of multiple slices.

A big challenge in training deep learning networks is getting the networks to extract low-level features and high-level features at the same time. Therefore, we need to study new methods to gradually grow and train networks to learn low-level features before discovering the higher-level features.

Another challenge is to extract useful/relevant features of deep learning networks for the task at hand. It is extremely challenging to differentiate between benign/malignant lung nodules, due to their heterogeneity. Therefore, to extract useful features, in this work, we extend the Non-Local Network first introduced by [2] for video classification to perform spatial and channel attention. The Non-Local operator is closely related to the Transformer, which is based on self-attention operations, but applied to two-dimensional (2D) images.

Methods

The dataset used in this study is the public LIDC-IDRI dataset, which is the largest public lung nodule dataset and consists of 1,018 CT scans collated from 1010 patients altogether. We also analyzed our method on the LungX Challenge dataset, which was first introduced at the 2015 SPIE Medical Imaging Conference. The dataset consists of 70 CT scans, of which 10 for calibration and 60 for testing. The test set of the LungX dataset contains 60 nodules where 37 are benign nodules and are 36 malignant nodules. In this study, we only

examined our method directly on the testing set and not on the calibration set.

We proposed a new **Progressive Growing Channel Attentive Non-Local Network (ProCAN)** network with a number of novel features. First, we added a channel-wise attention capability by extending the spatial-wise attention capability of the popular Non-Local network. The new **Channel Attentive Non-Local (CAN)** Network design block gives our ProCAN network the ability to detect size-invariant nodules with both spatial and channel-wise attention.

The second novelty of our method is that we developed a Curriculum Learning method based on the nodule diameter/size and radiologists' ratings. Through our experiments, we found that we can improve network performance if we train it on easy examples (i.e., small or big nodules in the nodule size category and clearly benign or clearly malignant in the radiologist ratings' category) before the difficult/hard examples (i.e., mid-sized nodules in the nodule size category and probably benign or probably malignant in the radiologist ratings' category).

The third novelty of our method is: We found that progressively growing the network during training is crucial and improves the results compared with using fixed depth/number of layers in the overall network architecture. We proposed a new strategy to blend the new blocks gradually with a matrix sampled from a 2D *Bernoulli* distribution and show that this improved the overall accuracy and the area under a receiver operating characteristic (ROC) curve (AUC).

The effectiveness of all these new components were analyzed in extensive ablation studies and in comparison, with other state-of-the-art methods in the literature. The experimental results show that our proposed ProCAN model achieves state-of-the-art results in the literature.

Results

From Table 1, we observe that both our ProCAN and Ensemble ProCAN models outperform the state-of-the-art models in the literature on all evaluation criteria. Also, our Ensemble ProCAN model outperforms ProCAN on all evaluated performance metrics, which is consistent with the results in the literature that show that ensemble methods generally outperform non-ensemble ones. Our non-ensemble ProCAN model marginally outperforms all the non-ensemble and ensemble methods in Table 1, excluding Ensemble ProCAN, whereas Ensemble ProCAN considerably outperforms all other methods in Table 1.

Table 1 Comparison of our ProCAN and Ensemble ProCAN methods with other state-of-the-art methods in the literature on the LIDC-IDRI database

Model	Ensemble?	AUC	Accuracy	Precision	Sensitivity	F1-Score
Multi-Crop	No	93.0	87.14	--	77.0	--
HSCNN	No	85.6	84.2	--	70.5	--
Local–Global	No	95.62	88.46	87.38	88.66	--
Gated-Dilated	No	95.14	92.57	91.85	92.21	--
ProCAN	No	97.13	94.11	94.54	93.12	93.81
3D DPN(Ensemble)	Yes	--	90.24	--	92.04	90.45
MV-KBC	Yes	95.70	91.60	87.75	86.52	87.13
MSCS-DeepLN	Yes	94.00	92.65	90.39	85.58	87.91
MK-SSAC	Yes	95.81	92.53	--	84.94	--
Ensemble ProCAN	Yes	98.05	95.28	95.75	94.33	95.04

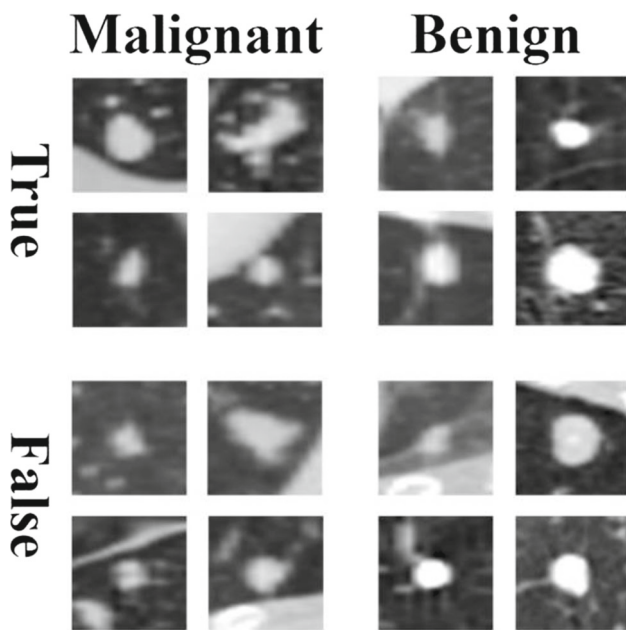


Fig. 1 Samples of correctly and incorrectly classified nodules in the LIDC-IDRI dataset

We provide some examples of benign and malignant nodules that were correctly and incorrectly classified by the ProCAN model. We randomly selected these samples from the LIDC-IDRI dataset and display them in Fig. 1.

In Fig. 1, the first two columns depict malignant nodules in the dataset, whereas the last two columns depict benign. The first two rows depict the correctly classified nodules, whereas the last two rows depict the misclassified nodules. To predict a nodule as benign or malignant, we employed a threshold of 0.5 on the probability scores. In this way, we identified the correctly and incorrectly predicted nodules in Fig. 1.

Conclusion

In this study, we proposed a new ProCAN network for lung nodule classification with many unique features. The experimental results show that our ProCAN model achieves state-of-the-art results compared to other existing methods in the literature.

References

- [1] WHO, “WHO report on cancer: setting priorities, investing wisely and providing care for all,” World Health Organization, 2020.
- [2] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local Neural Networks,” Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7794–7803, 2018.

Computer aided diagnosis based homology method - Identification of the degree of invasion of lung adenocarcinoma

K. Takarabe¹, K. Doi¹, K. Nakane¹, M. Yanagawa¹, H. Yamamoto¹, N. Tomiyama¹, H. Numasaki¹

¹Osaka University Graduate School of Medicine, Department of Medical Physics and Engineering, Suita City, Osaka Prefecture, Japan

Keywords CAD, homology, lung adenocarcinoma, deep learning

Purpose

In recent years, the number of patients with lung cancer has been increasing with the aging of the population. In particular, adenocarcinoma of the lung, a type of Non-Small Cell Lung Cancer, accounts for a large proportion of cases, and it is a noteworthy disease. The degree of lung adenocarcinoma invasion makes a big difference in the prognosis of lung cancer. Minimally invasive adenocarcinoma with minimal invasion and adenocarcinoma in situ with no invasion have a 5-year survival rate of approximately 100%. While the survival rate is 36–80% for invasive adenocarcinoma with large invasion, depending on the stage of the disease. Thus, the degree of invasion in lung adenocarcinoma is an important prognostic factor in lung adenocarcinoma, but percutaneous needle biopsy and surgical lung biopsy to determine the degree of invasion are burdensome to the patient and may cause complications. As a non-invasive method, machine learning is used to classify the degree of invasion. However, there is a concern that the process is a black box. One example of the solutions to the black box problem is the application of the homology method to histopathological images of lung tissue, which allows for automatic classification [1]. In this study, we attempted to analyze chest computed tomography (CT) images using the homology method to determine the degree of lung adenocarcinoma invasion and examined the possibility of a non-invasive test that mathematically discriminates [1].

Methods

Homology is a field called topology, and that is a mathematical concept to quantify contact. Here, we used the indices b_0 and b_1 that present the number of connected clusters and the number of enclosed regions, respectively. We can extract the features of the image based on these indices. We utilized the homology method to the analysis of CT images and evaluated the relationship between the extracted features and the degree of invasion. We classified CT images of 83 patients according to the degree of invasion into 43 cases of invasive adenocarcinoma, 19 cases of minimally invasive adenocarcinoma, and 21 cases of adenocarcinoma in situ and used them. The degree of invasion was confirmed by histological diagnosis. We cropped according to the shape of the tumor and binarized by varying their threshold values in the range of 1 to 225, and applied the homology method (Fig. 1). Here, the threshold value refers to the luminance value in 256 shades. b_0 and b_1 are calculated yielding b_1/b_0 , so we created histograms indicating b_1/b_0 and analyzed the correlation between b_1/b_0 and the degree of invasion.

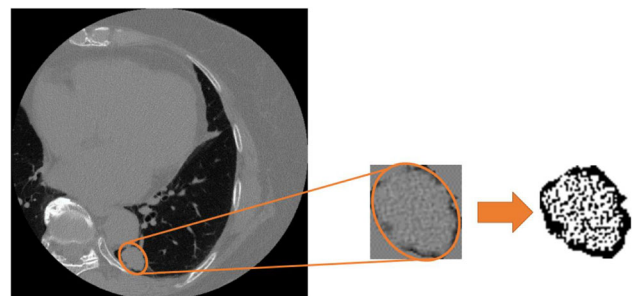


Fig. 1 The image after cropping from the CT image and the binarized image at a threshold of 127

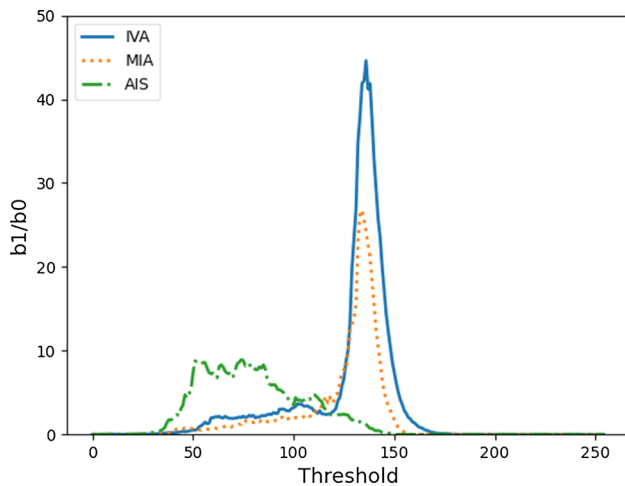


Fig. 2 Comparison of results (b_1/b_0) by class of the degree of invasion of lung adenocarcinoma

Results

The results showed that the height of the peaks and the presence of peaks in the histograms of the solid and part solid types differed depending on the degree of invasion (Fig. 2). In the threshold value of 120–170, there was a significant difference in the mean values of b_1/b_0 between invasive adenocarcinoma and others with a large difference in prognosis ($p < 0.05$). From the above results, it is highly possible that the degree of invasion can be discriminated by the peak value of b_1/b_0 and the position of the peak using the features obtained by the homology method. In addition, the proposed method is a new CT image analysis method using the concept of homology, and the results are mathematically based and visually comprehensible. For this reason, it is expected to be an effective method to solve the black box problem in machine learning. We also performed deep learning using the obtained b_1/b_0 values. The results suggested the possibility of classifying the patients into those with poor prognosis invasive adenocarcinoma and others. Therefore, by implementing machine learning using the features extracted by this method, the accuracy of lung cancer diagnosis by computer is expected to be improved.

Conclusion

The findings that the homology method applies to the binarized chest CT image by varying the brightness value from 1 to 255 as the threshold value, indicated that it is possible to determine the degree of lung adenocarcinoma invasion with a noninvasive method.

References

- [1] Nishio M, Nishio M, Jimbo N, Nakane K. Homology-Based Image Processing for Automatic Classification of Histopathological Images of Lung Tissue. *Cancers (Basel)*. 2021 Mar 10;13(6):1192. <https://doi.org/10.3390/cancers13061192>. PMID: 33801859; PMCID: PMC8001245.

Classifying pulmonary interstitial opacities in chest X-ray images using convolutional neural network and transfer learning

L. L. Lima^{1,2}, M. Koenigkam-Santos², L. Peron¹, K. L. Fontes Lopes¹, M. Repolês Lourenço¹, M. Koenigkam-Santos¹, P. de Azevedo-Marques¹

¹Ribeirão Preto Medical School at University of São Paulo, Medical Imaging, Hematology and Oncology, Ribeirão Preto—SP, Brazil
²University of São Paulo, Interunit graduate program in Bioengineering, São Carlos—SP, Brazil

Keywords Interstitial Lung Diseases, Covid-19, Tuberculosis, Convolutional Neural Network

Purpose

Interstitial lung disease (ILD) includes a group of more than 200 chronic lung disarrangements, which can reduce the ability of the air sacs to capture and carry oxygen into the bloodstream and consequently lead to permanent loss of the ability to breathe [1]. Another lung disease is tuberculosis. According to the World Health Organization, it is among the most infectious diseases globally, and approximately 1.8 million people died in 2015 from this disease [2]. Another significant lung disease is Coronavirus pneumonia (Covid-19), highly contagious that has spread rapidly throughout the world since January of 2020 and remains to infect the whole world [1]. These disorders have a similar presentation on imaging exams of the chest, predominantly represented by pulmonary interstitial opacities.

Chest simple radiography (CXR) is one of the initial imaging exams after clinical suspicion of lung disease, being the most straightforward, cheapest, and most popular radiologic diagnosis method. To help specialists improve the diagnostic accuracy of ILD, tuberculosis, and Covid-19 on CXR images by acting as a second opinion through a computer-supplied suggestion, computer-aided diagnosis (CAD) systems have been developed.

One machine learning technique that has emerged to improve CAD systems is the convolutional neural network (CNN) based on deep learning [1, 2]. Modeling the best CNN architecture for any specific problem by hand can be exhausting, time-consuming, and an expensive task. Another difficulty is to get a database with a large number of images. However, an alternative option is the Transfer Learning technique, which uses a network pre-trained on a large dataset.

In this work, we improved our previous work [1] and performed a new CNN analysis, including tuberculosis disease. So, in this work, we applied the VGG-19 CNN to classify ILD, tuberculosis, and Covid-19 diseases.

Methods

Our institutional review board approved this retrospective study with a waiver of patients' informed consent. The experiments were performed on a graphics processing unit (GPU) NVidia Tesla T4 with 16 Gigabytes of RAM. In our experiments, the framework Keras was used with the Tensorflow backend.

A total of 1044 images were used, separated into 382 healthy, 308 with ILD, 165 with tuberculosis, and 189 with Covid-19 pneumonia. The images were resized for 224×224 with three channels (RGB) instead of 1 channel (grayscale). For each image, another four were produced using transformations such as shear range, horizontal flip, and zoom to augment the database and improve CNN's performance in the experiments.

The CNN architecture used in this work was the VGG-19. A specific Keras pre-processing method for the VGG-19 was used. This method converts the images from RGB to BGR, centering at zero each color channel concerning the ImageNet dataset, unscaled.

For the use of VGG19, a global average pooling layer was added to the output of the last max-pooling layer in the last block. The original three fully connected layers were removed and replaced by only one fully connected layer with four output neurons using the activation function Softmax. The weights of the five blocks were frozen, so only the fully connected layer to the prediction was trained. The number of epochs used was 200, but the callback function

Table 1 Results obtained

Class	Precision	Recall	F1-score
Healthy	0.94	0.95	0.95
ILD	0.91	0.91	0.91
Tuberculosis	0.89	0.86	0.88
Covid-19	0.93	0.94	0.94

EarlyStopping, implemented in the Keras library, was used to interrupt the training if there was an increase in the error rate evaluated in the validation dataset during three epochs in a row. The batch size used was 32, and the optimizer function used was Adam to minimize the categorical cross-entropy.

To evaluate the model's performance, the samples were shuffled, and the method of cross-validation (tenfold) was applied. The statistical metrics: precision, recall, f1-score, and accuracy were used for performance evaluation. These metrics results were averaged over the test sets during the cross-validation process.

Results

The results obtained are presented in Table 1, with an average general accuracy of 92%. The best F1-score result of 0.95 was obtained to classify images of healthy lungs, and the worst F1-score result of 0.88 was obtained for the images of the tuberculosis class exams.

Conclusion

This study performed a CNN modeling integrated with the VGG-19 architecture to predict CXR images as belonging to healthy-lung individuals, patients with ILD, or patients with tuberculosis, or patients with Covid-19 pneumonia. Initial results have shown that the proposed approach presented an excellent potential for classifying CXR images associated with those diseases. Moreover, this method does not require image segmentation, handcrafted features extractors, or clinical features as prerequisites, improving its potential as a CAD tool for classifying pulmonary interstitial opacities in chest X-ray images.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)—Finance Code 001 and by The São Paulo Research Foundation (FAPESP)—Grants #2020/07200-9, #16/17078-0 and #14/50889-7.

References

- [1] Lima, L. L.; Nather-Jr J.C.; Koenigkam-Santos, M.; Peron, L.; Lopes, K. L. F.; Machado, C.; Sarmiento-Neto, F.; Wada, Danilo; Koenigkam-Santos, M.; Azevedo-Marques, P. M. Predicting interstitial and covid-19 diseases on chest x-ray using a convolutional neural network with transfer learning. In: CARS 2021 Computer Assisted Radiology and Surgery, 2021, Munique. Int J CARS—Suppl 1, 2021. v. 1. p. S68–S69.
- [2] Lopes U. K.; Valiati J. F. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Comput Biol Med.* 2017;89:135–143. <https://doi.org/10.1016/j.compbiomed.2017.08.001>.

Unsupervised 3D prediction of the pulmonary progression in COVID-19 based on chest CT

H. Yoshida¹, C. Watari¹, T. Uemura¹, T. Hironaka¹, J. Näppi¹
¹Massachusetts General Hospital / Harvard Medical School, Radiology, Boston, United States

Keywords generative adversarial network, survival analysis, deep learning, COVID-19

Purpose

The severity and rapid progression of the cases with COVID-19 has placed great pressures on healthcare services around the world. Recently, United States reported a new single-day record number of over 1 million new daily COVID-19 cases. Therefore, fast and accurate prediction of the progression of patients with COVID-19 is needed for the logistical planning of the healthcare resources. The purpose of this study was to develop a 3D prediction model of the pulmonary progression in COVID-19 based on the volumetric chest CT images of the patients.

Methods

One-hundred and forty-one patients with a chest CT examination and diagnosed as COVID-19 positive at our institution based on a positive result for SARS-COV-2 by the reverse transcriptase-polymerase chain reaction (RT-PCR) test were included retrospectively in this study. The CT scans were performed with a single-phase, low-dose acquisition protocol by use of multi-channel CT scanners with slice thickness of 0.625–1.5 mm, pitch of 0.3–1.6, tube voltage of 80–140 kVp, and automatic tube current modulation.

Our unsupervised 3D image-based prediction model, vox2pred, is based on an adversarial time-to-event model [1]. Figure 1 shows the architecture of vox2pred. A time generator G is used to generate a “survival-time volume” from the input CT image volume of a patient. The survival-time volume has a single survival time value at each voxel. The discriminator D attempts to differentiate the “estimated pairs” of input CT image volumes and their generated survival-time volumes, from the “observed pairs” of the input CT image volumes and their observed true survival-time volumes of the patients. The training of vox2pred involves the optimization of G and D through a modified min–max objective function so that G can learn to generate a survival time (i.e., the predicted pulmonary progression) that is close to the observed true survival time (i.e., the observed pulmonary progression). In this study, we defined the survival time of a patient as the number of days from the patient's chest CT scan to either ICU admission or death.

To evaluate the performance of the prognostic prediction, we used the metrics of concordance index (C-index) and the relative absolute error (RAE). The RAE is defined as $\sum_i |t_i^{\text{pred}} - t_i^{\text{obs}}| / t_i^{\text{obs}}$, where t_i^{pred} and t_i^{obs} are the predicted and observed progression times for patient i , respectively. A bootstrap method with 100 replications was used to estimate the C-index and RAE. The prognostic prediction

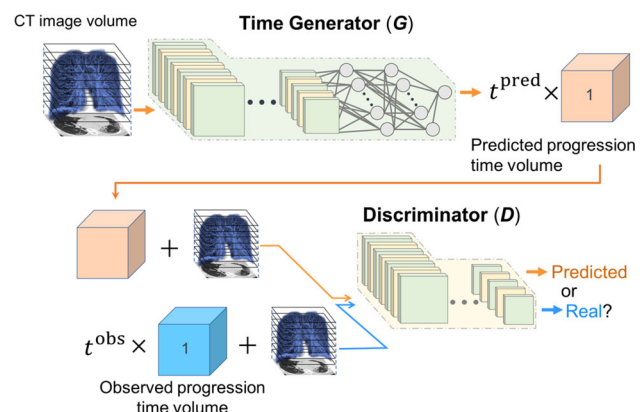


Fig. 1 Schematic architecture of the proposed vox2pred model for the prediction of the COVID-19 progression

Table 1 The C-index and RAE values estimated by the bootstrap evaluation of the %S-WAL, blood tests, and vox2pred model. 95% CI = 95% bootstrap confidence interval. *Two-tailed t-test

Clinical/Radiomic Features	C-index (%) [95% CI]	RAE (%) [95% CI]	P-value*
Percentage of well-aerated lung parenchyma (%S-WAL)	60.7% [59.9, 61.5]	55.9% [55.2, 56.5]	0.0000 >
Total severity score (TSS)	62.4 [61.6, 63.1]	51.8 [51.2, 52.4]	
vox2pred	87.4 [86.6, 88.1]	18.5 [17.4, 19.6]	

performance of vox2pred was compared with those of the percentage of well-aerated lung parenchyma (%S-WAL) [2] (evaluated on 135 cases), and the total severity score (TSS) by use of a two-sided unpaired *t*-test.

Results

Table 1 [HY1] shows the median C-index and RAE values calculated from the bootstrap evaluation. The C-index values (larger is better) obtained from the %S-WAL, TSS, and vox2pred were 60.7% [95% confidence interval (CI): 59.9, 61.5], 62.4% [95% CI: 61.6, 63.1] and 87.4% [95% CI: 86.6, 88.1], respectively. The RAE values (smaller is better) obtained from the %S-WAL, TSS, and vox2surv were 55.9% [95% CI: 55.2, 56.5], 51.8% [95% CI: 51.2, 52.4], and 18.5% [95% CI: 17.4, 19.6], respectively. The prediction performance of vox2pred was significantly higher than that of %S-WAL ($P < 0.0001$), and its prediction error was significantly lower than that of %S-WAL ($P < 0.0001$).

Conclusion

We developed a novel unsupervised 3D prediction model, vox2pred, which can directly predict the pulmonary progression of COVID-19 from the patients' chest CT image volumes. We showed that vox2pred outperforms the current standards of %S-WAL and blood test in predicting the pulmonary progression in patients with COVID-19, indicating that vox2pred can be an effective predictor of the COVID-19 progression.

References

- Chapfuwa P, Tao C, Li C, Page C, Goldstein B, Carin L, Henao R. Adversarial time-to-event modeling. Proceedings of the 35th International Conference on Machine Learning 2018; 734–743.
- Colombi D, Bodini FC, Petrini M, Maffi G, Morelli N, Milanese G, Silva M, Sverzellati N, Michieletti E. Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia. *Radiology* 2020; 296:E86–E96.

Computer-aided prognostic system for lung cancer using chest CT images

Y. W. Lee¹, S. Y. Chang¹, R. F. Chang¹

¹National Taiwan University, Department of Computer Science and Information Engineering, Taipei City, Taiwan

Keywords Lung cancer, Prognosis, Computed tomography, Convolution neural network

Purpose

Lung cancer is the leading cause of cancer death worldwide, with an estimated 1.8 million deaths (18%) in 2020. The overall 5-year survival rate of lung cancer (10–20%) is much lower than other leading cancer. Accurate prognosis of lung cancer patients can provide

essential information for planning personalized treatment. Computed tomography (CT) imaging has been widely used to diagnose lung cancer, the benefits include: fast, painless, non-invasive, and accurate, that providing detailed, cross-sectional views of all types of lung tissue. Although early diagnosis and treatments can effectively increase the survival of lung cancer patients, the heterogeneity of tumors makes it difficult to plan the optimal treatment for each patient. The previous studies indicated that the patients with the same tumor node metastasis (TNM) stage usually have various survival times and tumor behaviors. Therefore, further discovering the additional markers for prognosis becomes imperative, such as clinical-pathologic data and imaging characteristics; it could facilitate the cancer prognosis and help clinicians make wise decisions for personalized medicine. The survival prediction is mainly based on clinical data, including patient characteristics and lung cancer-related pathology data, and the function of imaging focuses on estimating the quantitative features of the tumor region. In this study, we proposed a prognosis system to predict the 5-years survival rate of patients which aggregates the CNN features from CT images, radiomics features, and clinical data from patients.

Methods

In this study, we proposed a prognosis system composed of a 3-D tumor segmentation model (3D HarDNet-MSEG) to automatically acquire the tumor volume and a classification model combining the clinical data and CT imaging to effectively predict lung cancer patients' survival outcomes. The clinical data included patient characteristics and lung cancer-related pathology data, and the CT scans were used to estimate the image characteristics of the tumor region, including radiomics features and CNN-based features. To prevent overfitting and eliminate irrelative data, we also employed a feature selection strategy based on several machine learning tools to preserve the most informative attributes and further improve the predictive ability of the prognosis classifier.

Results

In this study, the dataset was collected from 691 patients in Changhua Christian Hospital between November 2011 to February 2020, including CT scans and associated clinical information from medical records. The former consisted of 612 standard CT scans and 81 low-dose CT scans. The data ratio for the training set, validation set, and test set were 7:1:2, respectively. The results showed that the proposed method achieved the best performance; the accuracy, sensitivity, specificity, and AUC are 81.79%, 86.67%, 74.59%, and 0.8373 for the 5-year prognosis (Fig. 1).

Conclusion

We proposed a prognosis system to effectively predict the survival outcome of lung cancer patients using the chest CT images, including

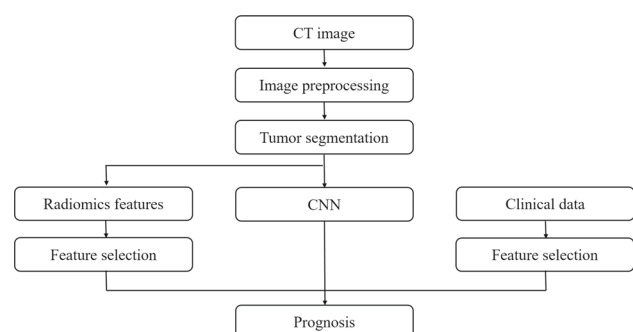


Fig. 1 The overall flow chart of the proposed method

a tumor segmentation model, image feature extractors, and a fully connected neural network classifier. The experiment results demonstrated that the combination of both imaging signatures and clinical data with the CNN model is superior to using each type of feature individually and also obtains better performance than several machine learning classifiers, which further evidence that our method has the potential to develop the survival analysis for the lung cancer patients.

Machine learning algorithm for quantitative CT in patients with COVID-19 pneumonia: utility for favipiravir treatment effect prediction

Y. Ohno^{1,2}, K. Aoyagi³, K. Arakita⁴, Y. Doi^{5,6}, M. Kodo^{7,8}, S. Banno⁸, Y. Fujisawa³, A. Taniguchi⁹, H. Ikeda¹, H. Hattori¹, K. Murayama², H. Toyama¹

¹Fujita Health University School of Medicine, Radiology, Toyoake, Japan ²Fujita Health University School of Medicine, Joint Research Laboratory of Advanced Medical Imaging, Toyoake, Japan ³Canon Medical Systems Corporation, Clinical Application Research Department, Research and Development Center, Otawara, Japan ⁴Canon Medical Systems Corporation, Healthcare IT Software Development Department, Healthcare IT Development Center, Otawara, Japan ⁵Fujita Health University School of Medicine, Microbiology and Infectious Diseases, Toyoake, Japan ⁶University of Pittsburgh School of Medicine, Division of Infectious Diseases, Pittsburgh, United States ⁷Fujita Health University School of Medicine, Respiratory Medicine, Toyoake, Japan ⁸Fujita Health University School of Medicine, Center for Clinical Trial and Research Support, Toyoake, Japan ⁹Canon Medical Systems Corporation, CT-MR Solution Planning Department, Otawara, Japan

Keywords Computer-aided diagnosis, Machine Learning, CT, Texture Analysis

Purpose

The new coronavirus disease 2019 (COVID-19) has been spreading worldwide since late 2019 and become a global pandemic involving over 200 countries or regions and more than 180 million individuals. About 10–20% of COVID-19 patients deteriorate into severe or critical illnesses within 7–14 days after symptom onset. In the overall management of COVID-19, a few specific anti-coronavirus treatments for severe patients are proposed at present, although their significant clinical benefits for severe COVID-19 still requires further confirmation. Favipiravir has demonstrated in vitro activity against SARS-CoV-2, and several randomized studies of COVID-19 conducted in China, Russia and India have indicated the potential clinical benefit of favipiravir. A randomized trial of patients with asymptomatic to mildly symptomatic COVID-19 was also conducted in Japan and suggested it has potential for modest clinical benefits, although radiological severity was not considered in this study. For the current study, we developed a new machine-learning (ML-) based CT texture analysis software for COVID-19, which evaluates radiological findings in lieu of expert chest radiologists and also functions as a second reader of CT images for various pulmonary diseases. We hypothesized that ML-based algorithm for evaluating thin-section CT has equal to or more useful than CT-determined disease severity score or time since disease onset for determination of better candidates for favipiravir treatment in COVID-19 patients. Using CT findings from a prospective, randomized, open-label multicenter trial of favipiravir treatment of COVID-19 patients, the purpose of this study was to compare the utility of ML-based algorithm with that of CT-

determined disease severity score and time since disease onset in this setting.

Methods

From March to May 2020, 32 COVID-19 patients underwent initial chest CT before enrollment were evaluated in this study. Eighteen patients were randomized to start favipiravir on day 1 (early treatment group), and 14 patients on day 6 of study participation (late treatment group). Viral clearance and duration of fever after enrollment were used as patient outcomes. In this study, percentages of ground-glass opacity (GGO), reticulation, consolidation, emphysema, honeycomb and nodular lesion volumes were calculated by means of the software, while CT-determined disease severity was also visually scored based on past literatures. Next, univariate and stepwise regression analyses were performed to determine relationships between quantitative indexes and time from disease onset to CT. With applying quantitative and qualitative indexes for differentiation of patients who started therapy within 4, 5 or 6 days after clinical onset from those who started later, patient outcomes were compared between each two groups by means of the Kaplan–Meier method followed by Wilcoxon's signed rank test.

Results

Time until CT examination had significant correlations with % GGO and % consolidation ($p < 0.05$), and stepwise regression analyses also identified %GGO and % consolidation as significant descriptors ($p < 0.05$). When radiologically divided all patients between 4 and 5 days from clinical onset or later, viral clearance and duration of fever after enrollment for each of the two groups divided according to % GGO, % consolidation, combined descriptors method and disease severity CT score showed significant differences ($p < 0.05$). When radiologically divided all patients between 6 days from clinical onset or later, viral clearance and duration of fever after enrollment for each of the two groups divided according to % consolidation and combined descriptors method showed significant differences ($p < 0.05$). Moreover, CT disease severity score had significant difference of viral clearance between two groups ($p < 0.05$).

Conclusion

ML-based CT texture analysis is equally or more useful than CT disease severity score for predicting the effect of favipiravir treatment on COVID-19 patients.

Predicting the necessity of supplemental oxygen in COVID-19 infected patients with the temporal variation of radiomics features in chest CT

Y. Sekiguchi¹, T. Okamoto¹, Y. Iwao^{2,3}, N. Kawata⁴, Y. Takiguchi⁵, Y. Matsuura⁵, T. Suzuki⁴, H. Haneishi²

¹Chiba University, Graduate School of Science and Engineering, Chiba, Japan ²Chiba University, Center for Frontier Medical Engineering, Chiba, Japan ³National Institutes for Quantum and Radiological Science and Technology, National Institute of Radiological Sciences, Chiba, Japan ⁴Chiba University, Graduate School of Medicine, Chiba, Japan ⁵Chiba Municipal Aoba Hospital, Respiratory Medicine, Chiba, Japan

Keywords COVID-19, predicting the necessity of su, radiomics features, temporal variation

Purpose

The rapid spread of coronavirus disease 2019 (COVID-19) with severe acute respiratory syndrome causes a shortage of medical supplies and hospital beds for COVID-19 patients. With prognosis

prediction at an early stage, these problems could be solved and appropriate treatment of patients could be provided, which leads to ease of burdens on the healthcare system. We aim to predict the prognosis of patients hospitalized with COVID-19 using chest CT scans. In this study, we focused on the disease transition-oriented temporal variation in CT images obtained at the admission to and the discharge from the hospital. Our hypothesis is that if we could quantify temporal variation in those CT images and select features that take into account the disease transition, we could predict accurately the progression of the disease, more specifically, the necessity of supplemental oxygen using those features in the CT image at the admission to the hospital. In this paper, we developed a prediction model for necessity of supplemental oxygen by focusing on the changes in radiomics features in CT images at admission and discharge.

Methods

We extracted radiomics features from three-dimensional (3D) lung fields and used them to predict the necessity of supplemental oxygen. First, we segmented the lung fields using the deep learning method to eliminate the irrelevant information in the feature analysis. Next, we extracted various radiomics features from segmented 3D lung fields in CT images at admission and discharge, respectively. The feature types we selected include basic statistics, gray level co-occurrence matrix (GLCM), gray level size zone matrix (GLSZM), gray level run length matrix (GLRLM), neighboring gray tone difference matrix (NGTDM), and gray level dependence matrix (GLDM) and the features are calculated using inherent evaluation formulas each of feature types. To extract various features, we applied no filter, logarithm filter, square-root filter, square filter, exponential filter, Laplacian of Gaussian (LoG) filter, and wavelet filter to them before extracting features, respectively. These features and filters refer to the previous study [1]. Then, we calculated the difference of features between admission and discharge. The features to be used in the prediction were selected in the following two steps. In the first step, The Mann–Whitney U test was used to select the features that were significantly different between required supplemental oxygen and not required supplemental oxygen for the difference in the extracted features. In the second step, the features selected in the first step were extracted

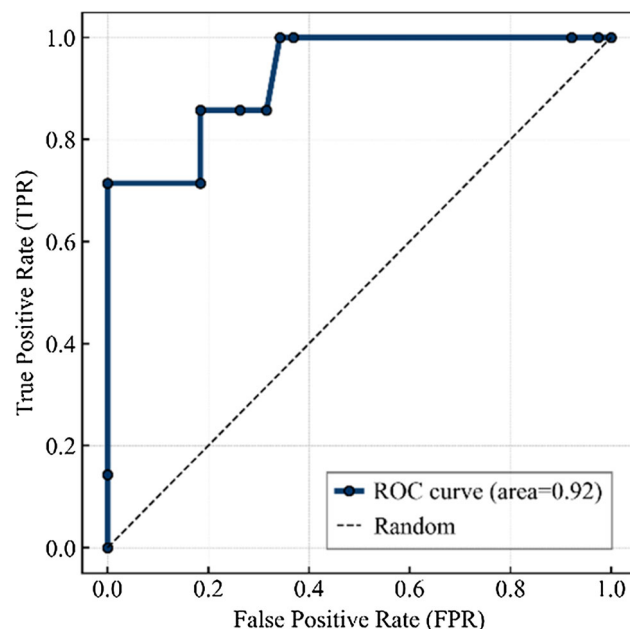


Fig. 1 ROC curve

Table 1 Accuracy evaluation

accuracy	precision	recall
0,93	1	0,57

from the CT images at the admission and narrowed down by using minimum redundancy maximum relevance (mRMR) [2]. For the prediction of the necessity of supplemental oxygen, we trained random forest regression by inputting features selected by mRMR.

We used the CT images obtained from 228 patients in Chiba Municipal Aoba Hospital and divided them into 183 data (required supplemental oxygen: 30 patients, not required supplemental oxygen: 153 patients) for training and 45 data (required supplemental oxygen: 7 patients, not required supplemental oxygen: 38 patients) for testing. We extracted 1581 features from each CT image and trained a random forest regression model using 10 features selected by mRMR. We used the receiver operating characteristic (ROC) curves to evaluate the performance of the regression model. As evaluation indices, in addition to the area under the curve (AUC) of the ROC, accuracy, precision and recall were calculated using the threshold at which the Youden's index in the ROC curve was maximized.

Results

The top three features selected by mRMR were the following three combinations: logarithm filtering and GLCM, square-root filtering and GLRLM, and Square filtering and GLSZM. We predicted the necessity of supplemental oxygen by using the regression model. 10 features extracted from the admission CT images of 45 test data were used as input data to the model. The ROC curve and AUC are shown in Fig. 1 and the results of accuracy evaluation are shown in Table 1. Namely, AUC was 0.92, accuracy was 0.93, precision was 1.0 and recall was 0.57. AUC, accuracy and precision were high, but recall was low compared to other indices. We consider that recall could be improved by increasing the data of patients who were required supplemental oxygen. As a whole, the regression model has the potential to predict the necessity of supplemental oxygen.

Conclusion

We developed a prediction model for the necessity of supplemental oxygen by focusing on the changes in features of CT scans at admission and discharge. We extracted radiomics features from 3D lung fields and used them to predict the necessity of supplemental oxygen. High accuracy suggested that the prediction model has the potential to predict the necessity of supplemental oxygen.

References

- [1] Chao H, Fang X, Zhang J, Homayounieh F, Arru C D, Digmurthy S R, Babaei R, Mobin H K, Mohseni I, Saba L, Carriero A, Falaschi Z, Pasche A, Wang G, Kalra M K, Yan P (2021) Integrative analysis for COVID-19 patient outcome prediction. *Medical Image Analysis* 67:101844.
- [2] Ding C, Peng H (2005) Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology* 3(2):185–205.

27th Computed Maxillofacial Imaging Congress

Chairman: Christos Angelopoulos, DDS (US)

Co-Chair: Yoshihiko Hayakawa, PhD (JP)

Prediction of dysphagia risk by hyoid bone classification using panoramic radiographs and deep learning technology

Y. Matsuda¹, Y. Hanada², Y. Hayakawa², K. Araki¹

¹Showa University School of Dentistry, Division of Radiology, Department of Oral Diagnostic Sciences, Tokyo, Japan ²Kitami Institute of Technology, Dept. of Engineering on Intelligent Machines & Biomechanics, Hokkaido, Japan

Keywords dysphagia, deep learning, panoramic radiography, image classification

Purpose

Dysphagia can lead to swallowing dysfunction and choking, and is a feature of physical weakening and frailty in old age. Early detection and prediction could permit clinical intervention to improve geriatric wellness. The downward displacement of the hyoid bone has been observed in individuals having swallowing dysfunction. This can lead to a widening of the distance between the dorsum of the tongue and the palate and impact appropriate timing of movement of a food bolus from the mouth to the throat.

Our previous study [1] found that individuals with dysphagia have the downward displacement of the hyoid bone, and that this can be viewed using panoramic radiographs. The purpose of the new study is to examine classification of hyoid bone positioning in panoramic radiographs using deep learning technology to predict dysphagia. (IRB approved. No.SUDH0034.)

Methods

A six-grade (Type 0–5) evaluation was applied to the vertical location of the hyoid bone on panoramic radiographs. Location types are shown in the diagram (Fig. 1). This classification was based on our previous paper [1]. Standardization was established using a virtual horizontal line crossing the left and right lower angles of the mandible, this mandibular-angle line being shifted horizontally to contact the lower mandibular border in the midline. Type 0 was set as the

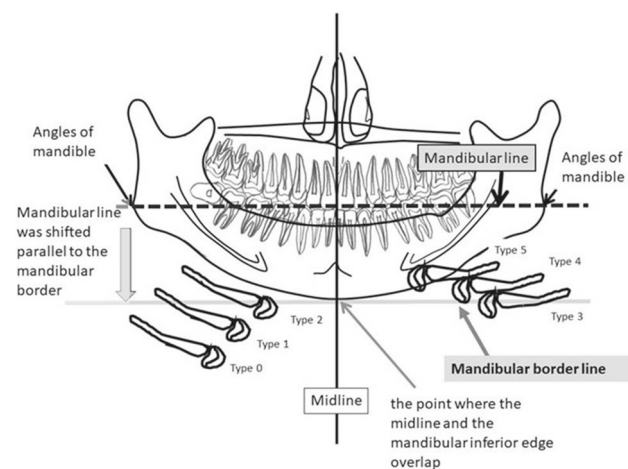


Fig. 1 The classification of vertical hyoid bone position: The position of the hyoid bone on the panoramic radiograph is categorized by "Mandibular border line". This line is shifted line from "Mandibular line" which passes through the mandibular inferior corners on both sides to the lower edge of the mandible. The types were classified according to their position relative to the lower edge of the mandible

lowest hyoid bone height and type 5 as the highest. Panoramic radiographs were examined and exclusion was made for motion artefact, misalignment, jaw tumors and jaw deformities. The object detection tool, YOLO v.5 was used [2]. The data used for learning was 467 images having 1976×976 pixels. Since downsizing causes a risk of missing image features, image cropping to 988×488 pixels was performed at the left and right lower portions of individual images. Hyoid bone detection was performed in these areas. Then 934 images were used as the data set for learning. The test data set was 129 images.

Results

For Type 1–5 hyoid locations, precision (P) values were 0.186, 0.612, 0.711, 0.838, and 0.659 respectively, while recall (R) values were 0.571, 0.950, 0.758, 0.471, and 0.789 respectively. Hence, P and R values were both very low for the Type 1 location. Moreover, the R value was low for the Type 4 location. Other P and R values were higher. Misclassification when it occurred was usually between neighbouring locations. Figure 2 is a panoramic radiograph for which the hyoid bone images on both sides were classed as Type 5, but where the right image could be classified as borderline Type 4. Classification as Type 0 means the hyoid bone was too low to be clearly depicted in the panoramic image and constitutes the true negative fraction.

Using the designated location classifications for the hyoid bone, both Precision and Recall values showed variations based on location designation. There might be need to adjust the classification numbers to reduce neighbouring overlap decisions to improve labeling precision labeling. Differentiation between Types 0 and 1 was difficult making Type 1 a bottle neck for achieving a high-precision classification. Since Type 1 is usually depicted in a small area and often incompletely, the object depiction tool did not work well.

Conclusion

The image classification of hyoid bone location was attempted using panoramic radiographs using deep learning with a goal of predicting the risk of dysphagia. The use of a six level classification led to some confusion based upon neighbouring overlap in decisions by AI. The next step will be to compare Type 0 locations against Type 1–5 combined in the hope of achieving less confusion. We will also increase epic learning repetition. The aim is to achieve a simple means for dentists to be aware of situations where their patients are susceptible to dysphagia and choking, and this could well also have implications for detection of risk of airway restriction and sleep apnea.



Fig. 2 An example of a classification of hyoid bone status. Both are classified in a category Type 5

References

- [1] Ito E, Matsuda Y, Kuroda M, Araki K (2021) A novel dysphagia screening method using panoramic radiography. *Showa Univ J Med Sci* 33(3),74–81.
- [2] https://pytorch.org/hub/ultralytics_volov/5/

Clinical feasibility of automatic CBCT segmentation for orthognathic surgical planning

H. Deng¹, J. Barber¹, Q. Liu², T. Kuang¹, D. Kim¹, J. Gateno¹, P. T. Yap², D. Shen², J. Xia¹

¹Houston Methodist Hospital, Oral and Maxillofacial Surgery, Houston, United States ²University of North Carolina, Radiology and Biomedical Research Imaging Center, Chapel Hill, United States

Keywords Learning-based segmentation, CBCT, Craniomaxillofacial, Orthognathic surgical planning

Purpose

Accurately segmenting bony structures from head cone-beam computed tomography (CBCT) is the first step in computer-aided surgical simulation for orthognathic surgery. The current clinical process is to manually segment the four bony structures from CBCT, i.e., midface, mandible, and upper teeth and lower teeth [1]. CBCT has lower radiation but more severe artifacts than spiral computed tomography. Thus, manual segmentation of CBCT is an extremely time-consuming and labor-intensive process. While recent research has been focused on developing deep-learning-based methods for automatic segmentation, they are only in laboratory settings and yet have been incorporated into daily clinical practice. To this end, we have developed a multi-stage convolutional neural network (CNN) framework for CBCT segmentation [2]. The purpose of this study was to evaluate the performance of our method in clinical setting.

Methods

Our automatic segmentation method is a coarse-to-fine CNN-based framework [2]. In the coarse stage, a scalable joint segmentation and landmark detection model is developed for coarse segmentation and global landmark detection. In the fine stage, both global and local volumes are first cropped out from the original image based on the landmarks detected in the coarse stage. The global volume contains the entire skull model region, and the local volume are the selected special regions that contain the thin bones or regions that are difficult to be segmented. The coarse segmentation is then refined on both global and local cropped volumes. After that, the refined segmentations are fused together to generate the final segmentation result (a.k.a. segmentation mask). Our framework for orthognathic surgery requires 4 labels: midface, mandible, upper teeth, and lower teeth. Our framework is trained using 125 sets of CBCTs and corresponding labels on a standard Linux workstation equipped with Intel dual-Xeon CPU and four 12-GB NVidia GPUs.

Our method was evaluated using randomly selected 15 sets of CBCT data (0.4mm³ isotropically) of patients with jaw deformities since 2020 (IRB: Pro00013802). These testing CBCTs had never been used for training. The evaluation started with feeding the testing CBCTs into our framework to automatically generate four required segmentation masks. They served as the experimental group. The auto-segmented masks were then imported into AnatomicAligner software. A single experienced surgical planner reviewed the auto-generated mask slice-by-slice and improved them as needed. The final segmentation masks that would be used for surgical planning were

exported after they were inspected by a second investigator. They served as the control group for quantitative evaluation.

The computational time for automatic segmentation and the time spent on manually improving the auto-segmentation results were recorded. The evaluation was completed quantitatively and qualitatively. To quantitatively evaluate the results, dice similarity coefficient (DSC) was calculated to assess the difference between the automatic and manually improved segmentation results of each structure. To qualitatively evaluate the results, a three-dimensional (3D) model of each anatomical structure was generated in AnatomicAligner using the auto-segmented masks (without manual improvement). After that, based on his careful inspection and expert intuition, an experienced oral surgeon visually evaluated each 3D model and its corresponding mask to determine whether it was suitable for surgical planning: 1) use directly without improvement; 2) use after minor manual improvement; and 3) do not use unless re-segmenting it from scratch. Any findings beyond the above category were also recorded.

Results

All CBCT scans were successfully segmented into 4 required anatomical structures. The median computational time was 3.5 min (range: 3.3–3.7 min) on a regular Windows computer equipped with dual-Xeon CPU and a single 24-GB NVidia GPU. The median time spent on manually improving the automatic segmentation results was 1 h (range: 0–4 h).

The results of the quantitative evaluation showed that the DSCs for midface, mandible, upper and lower teeth were $94.3\% \pm 1.6\%$, $97.3\% \pm 1.5\%$, $98.7\% \pm 2.7\%$, and $98.4\% \pm 2.8\%$, respectively. The results of the qualitative evaluation showed that the auto-segmentation results of 9 midfaces (60.0%), 10 mandibles (66.7%), and all upper and lower teeth (both 100%) could be directly used for surgical planning without the need for manual improvement. In addition, 6 midfaces (40.0%) and 5 mandibles (33.3%) required a minor manual improvement. Furthermore, there were few free-floating small artifacts unattached to anatomy, and few small holes located on correctly-shaped anatomy that were not along any osteotomy line. They were deemed not to affect the surgical planning and were cat-

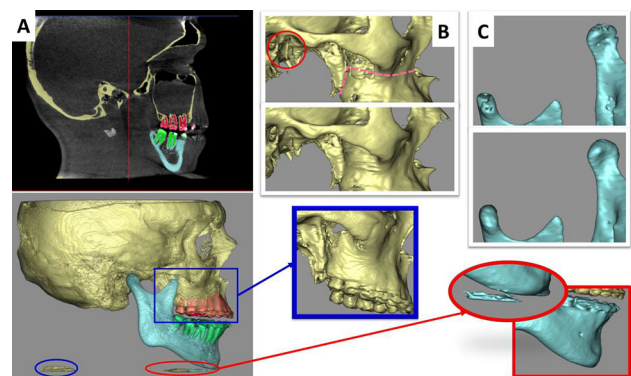


Fig. 1 Examples of the qualitative evaluation. **A:** Top shows the auto-segmented masks of midface (yellow), mandible (blue), upper teeth (red) and lower teeth (green). Bottom shows the corresponding 3D models. Blue Circle: The free-floating artifacts and the holes on pterygoid are not a concern. Red Circle: The artifacts touch the chin and are removed (Red Box). **B:** The holes are along the pink osteotomy line (Top). A part of condyle is misclassified as the midface (red circle). They are corrected (Bottom). **C:** A part of both mandibular condylar heads is missing (Top) and is repaired (Bottom)

egorized as “use directly without improvement”. Figure 1 shows 3 examples reflecting the above results.

Conclusion

The results of this study demonstrate the feasibility and effectiveness of using our deep-learning-based segmentation framework in daily clinical practice. We used to spend at least a day and a half manually segmenting the 4 anatomical structures from head CBCT for surgical planning. Now with our automatic segmentation workflow, we can lightly improve the segmentation masks based on automatic results, if ever needed, for surgical planning. The improvement was mainly in the pterygoid, the condyles and the chin.

Our DSCs are unsurprisingly better than the ones during the algorithm development (88.5% for the midface and 93.5% for the mandible) [2]. This is because in our workflow, instead of manually segmenting the CBCT image from scratch, we directly improve the auto-segmentation masks. It significantly reduces the error in certain fuzzy regions that are uncritical to orthognathic surgery, e.g., intraorbital, where there are large inter- and intra-observer“s discrepancies even when the manual segmentations are performed by the most experienced clinicians.

A technical advantage of our proposed method is that the model is scalable and able to refine the critical regions which are difficult to segment but clinically important, e.g., the pterygoid. Future work includes to train the models using a larger training dataset and refine the algorithm to improve the critical areas.

References

- [1] Xia JJ, Gateno J, Teichgraeber JF, Yuan P, Chen KC, Li J, Zhang X, Tang Z, Alfi DM (2015). Algorithm for planning a double-jaw orthognathic surgery using a computer-aided surgical simulation (CASS) protocol. Part 1: planning sequence. *Int J Oral Maxillofac Surg.* 44(12):1431–40.
- [2] Liu Q, Deng H, Lian C, Chen X, Xiao D, Ma L, Chen X, Kuang T, Gateno J, Yap P-T, Xia JJ (2021). SkullEngine: A Multi-stage CNN Framework for Collaborative CBCT Image Segmentation and Landmark Detection. In C. Lian, X. Cao, I. Rekić, X. Xu, P. Yan (Eds.), *Machine Learning in Medical Imaging* (pp. 606–614). Springer International Publishing.

Poster Session

Adaptive noise reduction scheme for dedicated breast positron emission tomography using multiple convolutional neural networks

M. Tsukijima^{1,2}, A. Teramoto¹, K. Saito¹, H. Fujita³

¹Fujita Health University, Graduate School of Health Sciences, toyoake, Japan ²East Nagoya Image Diagnosis Center, Nagoya, Japan ³Gifu University, Faculty of Engineering, Gifu, Japan

Keywords db-PET, denoising, image processing, deep learning

Purpose

In recent years, dedicated breast positron emission tomography (db-PET), a PET system that provides high-definition imaging of the breast, has become capable of detecting small lesions because it has higher resolution than whole-body PET systems. However, db-PET, which acquires data from detectors arranged in a circular pattern, has a problem that image quality on the chest wall side, the edge of the detector, is degraded due to the decrease in PET counts, resulting in reduced imaging performance. [1] In order to improve the image quality of db-PET images, we propose a denoising method using multiple convolutional neural network filters with residual U-Net.

Methods

A total of 191 cases of 94 breast cancer patients and 97 healthy subjects collected between 11/1/2016 and 6/31/2017 were used for this study. One-minute and seven-minute list mode data were extracted from the collected data, and reconstructed to the input image with high noise level and the ideal image with low noise level were created. The outline of the proposed denoising method is shown in Fig. 1, where the input gives an image taken for one-minute and the output gives a denoised image equivalent to a seven-minute scan. Since the noise level of db-PET varies depending on the slice position, the volume data was divided into four areas, and filters with different characteristics were designed for each area. In this study, residual U-Net was introduced as an individual denoising filter, and training was performed on one-minute and seven-minute images at different slice positions. When the unseen image is given to the filter system after training, different residual U-Net is applied depending on the slice position, and the denoised image is provided. For comparison, we also created a filter that trains all slice images with a single residual U-Net. For these processes, we used a computer with an intel corei9-10900K CPU and a graphics processing unit (NVIDIA GeForce RTX 2080Ti) and used Tensorflow and Keras as the deep learning APIs.

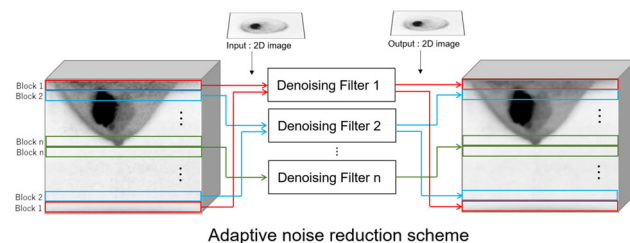


Fig. 1 The outline of the proposed denoising method is shown in the figure, the input gives an image taken for high noise image and the output gives a denoised image with adaptive noise reduction where the noise level of db-PET varies depending on the slice position

Table 1 Average values of PSNR and SSIM of each filter of four imaging areas

	PSNR					SSIM				
	Original	Gaussian	NLM	S-CNN	M-CNN	Original	Gaussian	NLM	S-CNN	M-CNN
Area1	30,6	31,1	30,8	31,5	32,0	0,820	0,834	0,832	0,826	0,878
Area2	37,0	37,1	37,1	37,3	37,9	0,920	0,934	0,928	0,948	0,966
Area3	32,4	33,1	32,5	33,2	33,8	0,803	0,836	0,823	0,881	0,896
Area4	32,2	32,3	32,3	33,2	34,4	0,758	0,802	0,784	0,868	0,967

Results of PSNR and SSIM in each area. Gaussian: Gaussian filter, NLM: Non local mean filter, S-CNN: Filter using single convolutional neural network, M-CNN: Proposed method using multiple convolutional neural networks

Results

To evaluate the effectiveness of the proposed method, we conducted the filter processing using db-PET images of five cases that were not used for training; the PSNR and SSIM of the obtained images were calculated. Table 1 shows the average values of PSNR and SSIM of each filter of four imaging areas. The PSNR and SSIM of the proposed method are the highest in all areas.

Conclusion

In this study, our proposed method for noise reduction of db-PET images using multiple convolutional neural network filters was the best method compared to the other filters, with the highest PSNR and SSIM values for all test data. Figure shows the proposed method, which was trained separately for each transverse body axis with different noise characteristics, and thus was able to optimize the noise reduction process for each region. Although the filter using a single residual U-Net tended to outperform the existing noise reduction methods such as Gaussian filter and non-local mean filter, the PSNR and SSIM values were lower than those of the proposed method. This may be caused by the different noise characteristics of db-PET images, which were trained by the same network. These results indicate that our method will be effective to improve the image quality of db-PET images.

References

- [1] Satoh Y, Motosugi U, Imai M, Omiya Y, Onishi H (2021) Evaluation of image quality at the detector's edge of dedicated breast positron emission tomography *EJNMMI Phys.* 8(1).

Higher resolution of CT images by deep learning using GAN

Y. Taguchi¹, R. Takara¹, S. Ito¹, S. Kuga¹, M. Takeda¹, K. Yoshida¹, M. Nakajima¹, K. Inoue², T. Mitsumoto³

¹Chuoh College of medical Technology, Department of Medical Radiology, Tokyo, Japan ²Tokyo Metropolitan University, Department of Radiological Sciences, Tokyo, Japan ³Teikyo university, Department of Radioisotope Research Center, Tokyo, Japan

Keywords Deep Learning, GAN, CT image, high resolution

Purpose

Currently, research on image processing technology using deep learning is being actively conducted. Focused on high resolution in deep learning image processing technology. There are several high resolution deep learnings, but I used TecoGAN. TecoGAN is deep

learning using GAN (Generative Adversarial Network) [1]. To date, there have been no announcements that TecoGAN will be applied to medical images. TecoGAN is a deep learning program announced by Mengyu C, which can create high resolution videos. [2] CT images are still images, but I came up with the idea of using TecoGAN for CT images in the same way as movies because I collect continuous slice images to create multiple images. In this study, TecoGAN was used to verify the high resolution of CT images.

Methods

In this study, we made a visual and physical assessment. CT images of the chest phantom were used for visual evaluation. Image data acquisition was performed with diameters of FOV 350 mm and FOV 430 mm. To use TecoGAN, the two images were taken with a matrix size of 512×512 and the matrix size was changed to 256×256 . Both have an effective field of view of 350 mm. And with TecoGAN, L-sized images have a high resolution with a matrix size of 1024×1024 . Visually evaluated high resolution images and M and L size images. For physical evaluation, the MTF (Modulation Transfer Function) was calculated using a square wave chart image. The MTF compared three types of images: FOV 350 mm and FOV 430 mm images with an effective field of view of 350 mm, and high-resolution images with FOV 430 mm. Due to the use of TecoGAN, all images are processed in PNG instead of DICOM.

Results

Two types of data, FOV350 mm and FOV430 mm, and a high-resolution image of FOV430 mm created by Teko GAN were evaluated. Visual and physical evaluations were performed using three types of images. In visual evaluation, data collection has significantly improved the quality of high resolution images. L size diameter of 3 types of images. As a result of MTF, the difference between the three types of images was small at frequencies of 1 cycle / mm and 2 cycles / mm. At 3 cycles / mm and 4 cycles / mm, MTF was improved on high resolution L size images (Fig. 1).

Conclusion

Higher resolution CT images using TecoGAN are expected to improve image quality as a result of visual and physical evaluations. In the future, it will be necessary to consider improving the image quality using clinical images and directly improving the image quality of DICOM images.

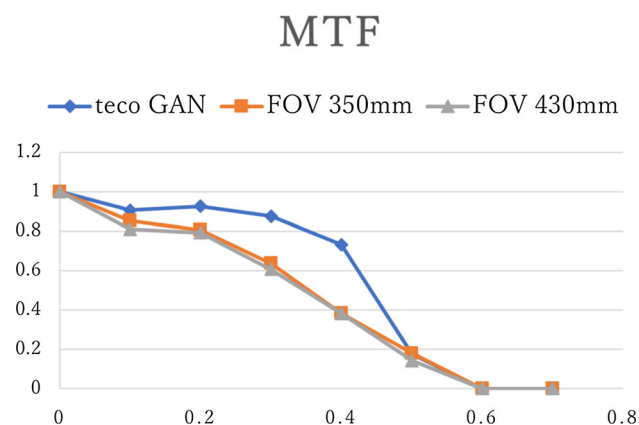


Fig. 1 This graph is MTF for three types images. The vertical line shows the number of Modulation Transfer Function and the horizontal one shows spatial frequency

References

- [1] Ian G, Jean P, Mehdi M, Bing X, David W, Sherjil O, Aaron C, Yoshua B (2014) Generative Adversarial Networks. In *Advances in neural information processing systems*. 2672–2680.
- [2] Mengyu C, You X, Jonas M, Laura L, Nils T (2020) Learning temporal coherence via self-supervision for GAN-based video generation. *ACM Transactions on Graphics*. 39(75):1–75.

Artificial chest X-ray image creation with simulated lung nodules by Glow algorithm

S. Hanaoka¹, Y. Nomura^{1,2}, N. Hayashi³, H. Shibata³, T. Nakao³, T. Takenaga³, O. Abe¹

¹The University of Tokyo Hospital, Department of Radiology, Tokyo, Japan ²Chiba University, Center for Frontier Medical Engineering, Chiba, Japan ³The University of Tokyo Hospital, Department of Computational Diagnostic Radiology and Preventive Medicine, Tokyo, Japan

Keywords Lung cancer, Chest X-ray, Imaginary image creation, Glow

Purpose

We propose a method with which fictional chest X-ray images with lung nodules can be created. With the created images only, we tried to train a detector of lung cancer in each X-ray image.

Methods

We utilized Glow algorithm [1], one of flow-based generative models, for creating a chest X-ray image. Using 45,808 normal cases, the Glow algorithm was trained as unsupervised manner. The image resolution was 512×512 . Furthermore, we developed automatic

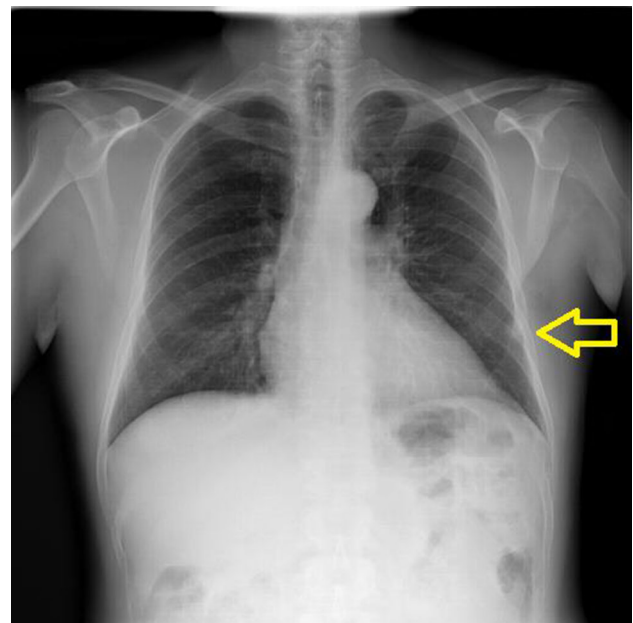


Fig. 1 A created imaginary chest X-ray with a lung nodule

methods for extraction of the lung field area, creation of fictional lung cancer nodules and embedding of them. These three methods were implemented with rule-based methods (i.e., not with deep-learning-based method). The lung field area was extracted by using morphology filter-based method. The nodule creation was performed with 3-d volume simulation. And final embedding was performed by 2D-3D projection method. Simultaneously, the gold standard answer mask (i.e., the position information of the embedded fictional lung cancer) was created. Using all these implementations, total 131,072 fictional X-ray images with nodules were created. As an evaluation, we trained a U-net-based method for lung nodule detection using these 131,072 imaginary datasets. We tested the trained U-net with JSRT chest lung nodule image database [2].

Results

The Glow-based chest X-ray creation method successfully created normal chest X-rays. Additionally, Our lung nodule creator and embedder successfully embedded nodule shadows into these X-rays. Figure 1 illustrates a sample image. As described above, we tried to train a U-net to detect real nodules in X-ray images. We tested our U-net with JSRT chest lung nodule image database and found sensitivity of 0.65 with the false positive detection rate of 0.2 per case.

Conclusion

Our fictional lung nodule case creation method was presented. With the method, 131,072 chest images were successfully created. Using only these created images, our U-net was successfully trained and showed clinically meaningful sensitivity. In our future work, we will try to improve the sensitivity so that it can surpass radiologist's sensitivity. Furthermore, because our dataset is totally artificial, we can safely share it on the internet. We believe that our fictional database is useful for training and evaluation of various computer-assisted tasks, for example, federated learnings.

References

- [1] Kingma DP, Dhariwal P (2018) "Glow: Generative flow with invertible 1×1 convolutions." arXiv preprint arXiv:1807.03039 (2018).
- [2] Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K, Matsui M, Fujita H, Kodera Y, and Doi K.: Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. AJR 174; 71–74, 2000.

2D + 3D registration in deformation-adaptive super-resolution for medical images

T. Zheng¹, H. Oda², T. Hu¹, Y. Hayashi¹, S. Nakamura², M. Mori³, H. Takabatake⁴, H. Natori⁵, M. Oda^{1,6}, K. Mori^{1,7,8}

¹Nagoya University, Graduate School of Informatics, Nagoya, Japan

²Nagoya University, Graduate School of Medicine, Nagoya, Japan

³Sapporo-Kosei General Hospital, Respiratory Medicine, Sapporo, Japan

⁴Sapporo Minami-sanjo Hospital, Respiratory medicine, Sapporo, Japan

⁵Keiwakai Nishioka Hospital, Respiratory medicine, Sapporo, Japan

⁶Information and Communications, Nagoya University, Information Strategy Office, Nagoya, Japan

⁷Nagoya University, Information Technology Center, Nagoya, Japan

⁸National Institute of Informatics, Research Center for Medical Big Data, Tokyo, Japan

Keywords Super-resolution, Registration, 3D + 2D framework, MRI

Purpose

This paper proposes a 3D + 2D registration method for medical images, and additionally applied the registration method in a novel deformation-adaptive super-resolution method. Acquiring high-resolution (HR) medical images such as HR-MRI comes at problems such as too strong magnetic field, and are not ideal for real-time imaging of moving organs such as cardiac imaging. Therefore, obtaining HR images by super-resolution (SR) of low-resolution (LR) images utilizing fully convolutional network (FCN) is a feasible approach. However, commonly the FCN needs to be trained with LR-HR image pairs. Image registration (aligning LR and HR images to obtain LR-HR image pairs) needs to be conducted. Therefore, we propose a 3D + 2D registration method for obtaining precisely-aligned LR-HR medical images. For evaluating effectiveness of our registration method, we also propose a deformation-adaptive FCN for SR. The FCN for SR is trained on LR-HR medical images aligned by our registration method.

The contributions of our paper are (1) a precise 3D + 2D medical image registration method and (2) a deformation-adaptive medical image SR method trained on aligned LR-HR images.

Methods

Overview

Given a 3D medical volume M and another 3D medical volume F , we perform 2D + 3D registration to align M to F precisely. Additionally, for evaluating effectiveness of our registration method, we train an FCN g_0 using aligned LR images $m_{AA,1}$ and corresponding HR images f . For inference, the trained FCN performs SR of a given LR image x_{LR} to x_{SR} .

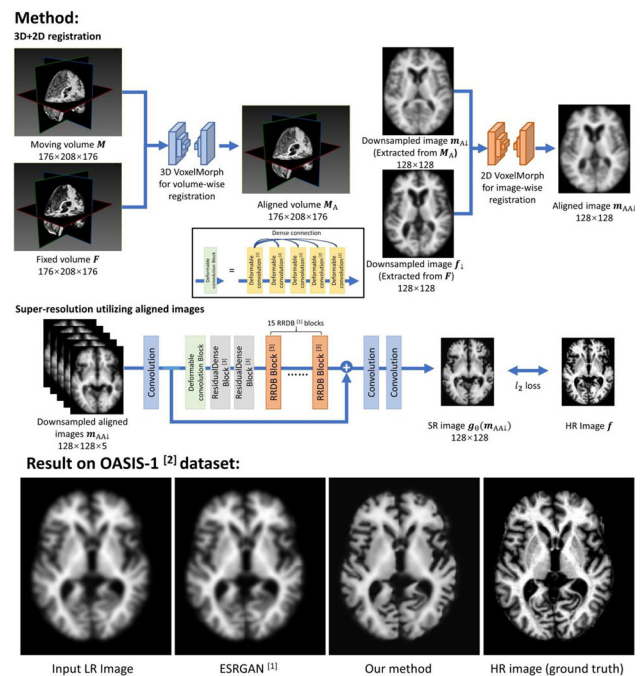


Fig. 1 Our method and results. We utilize 3D and 2D VoxelMorph for precise registration of low-resolution (LR) and high-resolution (HR) images. Then we utilize aligned LR images $m_{AA,1}$ and HR image f to train an FCN for SR. The lower part illustrates our method's result together with linear interpolation and ESRGAN [1]. Our method successfully reconstructed important tissues such as the cortex of the brain, while linear interpolation and ESRGAN's results are blurred

Precise 2D + 3D registration for aligning LR-HR images

Typically, paired LR-HR images are needed for training an FCN for SR. However, since LR and HR images are obtained from different imaging devices or different imaging conditions, it is hard to obtain paired LR-HR images in medical imaging. Therefore, we need to perform registration to obtain paired LR and HR medical images. Existing registration methods always align a whole moving 3D medical volume M to another fixed 3D medical volume F to obtain an aligned 3D medical volume M_A , causing the precision of registration (commonly MSE) to be low. To obtain a higher precision of registration, we first train a 3D VoxelMorph to align 3D medical volumes M and F . The inputs of the 3D VoxelMorph are M and F . The output of the 3D VoxelMorph is an aligned 3D medical volume M_A . Additionally, we extract downsampled 2D images $m_{A\downarrow}$ and f_{\downarrow} from volumes M_A and F to train a 2D VoxelMorph for aligning $m_{A\downarrow}$ to f_{\downarrow} . We name the aligned 2D images as $m_{AA\downarrow}$. Paired $m_{AA\downarrow}$ (the LR image) and f (the HR image) are used for training an FCN for SR.

Training SR FCN using aligned LR-HR images

We train an SR FCN g_0 using paired 2D images $m_{AA\downarrow}$ and f . We utilize an FCN with RRDB block [1], deformable convolution, and multi-slice input as shown in Fig. 1. RRDB block could boost performance of single-image SR; deformable convolution enlarges the receptive field of convolution kernels and model geometric transformations between $m_{AA\downarrow}$ and f ; multi-slice input feeds additional spatial information into the FCN to enhance the accuracy of SR result. We utilize three kinds of loss terms to optimize the FCN. The first loss term is pixel-wise l_2 loss $l_2(g_0(m_{AA\downarrow}), f)$; the second is identity loss $l_2(g_0(f), f)$ to stabilize training; the third is adversarial loss [1] to ensure SR image $g_0(m_{AA\downarrow})$ has a higher perceptual quality. For inference, the input of FCN g_0 is an LR image x_{LR} ; the output is an SR image x_{SR} .

Results

We utilized the OASIS-1 [2] dataset to evaluate our method. The OASIS-1 dataset is a T1-weighted brain MRI dataset consists of 416 cases. We utilized 250 cases for training. We performed $2 \times$ SR (LR image was downsampled $2 \times$ by width and height from HR image). For registration, the 3D VoxelMorph for registration was trained with 250 patches with size $176 \times 208 \times 176$ voxels for 1500 epochs. The 2D VoxelMorph for registration was trained with 25,000 patches with size 128×128 pixels for 1,500 epochs. For SR, the FCN for SR was trained with 37,500 patches of size $128 \times 128 \times 5$ voxels for 100 epochs.

For evaluation, we utilized 50 cases in the OASIS-1 dataset. We compared our method with ESRGAN [1] and linear interpolation. Our method outperformed ESRGAN and linear interpolation quantitatively as shown in Table 1. Qualitative results are shown in Fig. 1. Our method successfully reconstructed important tissues such as the cortex of the brain, while linear interpolation and ESRGAN's results are blurred.

Table 1 Quantitative evaluation of linear interpolation, baseline [1] and our method

	Linear interpolation	Baseline [1]	Our method
PSNR	22.2857	22.9336	24.2553
SSIM	0.8394	0.8616	0.8992
GMSD	0.1825	0.1825	0.1418

Conclusion

We proposed a 2D + 3D registration method for aligning LR and HR medical images precisely. For evaluating effectiveness of our registration method, we also proposed an SR method trained on LR and HR medial images aligned by our registration method. Experimental results showed our method outperformed the recent baseline [1] and linear interpolation qualitatively and quantitatively. Future work includes ablation studies, and evaluating our method on more datasets.

Acknowledgements Parts of this research was supported by MEXT/JSPS KAKENHI (21K19898, 26108006, 17K20099), the JSPS Bilateral Joint Research Project, the AMED (JP201k1010036(2020)) and the Hori Sciences & Arts Foundation.

References

- [1] Wang, Xintao, et al. (2018) "ESRGAN: Enhanced super-resolution generative adversarial networks." Proceedings of the European conference on computer vision (ECCV) workshops.
- [2] Marcus, Daniel S., et al. (2020) "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults." Journal of cognitive neuroscience 19.9: 1498–1507.

Effect of a deep learning-based reconstruction method for pulmonary nodule volumetry analysis in low-dose computed tomography

S. Watanabe^{1,2}, K. Sakaguchi², S. Kitaguchi²

¹Kindai University Hospital, Division of Positron Emission Tomography, Institute of Advanced Clinical Medicine, Osakasayama, Osaka, Japan ²Kindai University Hospital, Radiology Center, Osakasayama, Osaka, Japan

Keywords computed tomography, deep learning-based reconstruction, pulmonary nodule volumetry, low-dose

Purpose

A newly developed deep learning-based reconstruction algorithm (DLR) potentially can reduce image noise while image noise texture change was more negligible than the iterative reconstruction technique [1]. This feature was useful for repeated low-dose lung computed tomography (CT) screening because of the concerning radiation dose. The accuracy of volumetry was important for the nodule growth assessment because the volumetric assessment was used for the lung nodule management protocol in lung CT screening trial such as US National Lung Screening Trial (NLST) [2]. However, the lung nodule volumetry accuracy using DLR has not been investigated. The aim of the present study was to evaluate the influence of a DLR, Advanced Intelligent Clear-IQ Engine (AiCE), for pulmonary nodule volumetry accuracy in low-dose CT.

Methods

A thoracic phantom contained eight artificial pulmonary nodules which have two types of Hounsfield unit (HU): -630 HU and $+100$ HU, and four diameters: 5, 8, 10, and 12 mm was scanned with a 320-detector row CT. The standard dose of the present study was determined in accordance with the average effective dose of 1.5 mSv in NLST. The phantom was scanned at 1/2 and 1/6 standard dose in addition to the standard dose and repeated 20 times. All images were reconstructed with filtered back projection (FBP), a hybrid type iterative reconstruction (adaptive iterative dose reduction: AIDR 3D),

Table 1 Resultant APEs of FBP, AIDR 3D, and AiCE at standard dose, 1/2 dose, and 1/6 dose

		standard dose (3.8mGy)			1/2 dose (1.9mGy)			1/6 dose (0.6mGy)		
		FBP	AIDR 3D	AiCE	FBP	AIDR 3D	AiCE	FBP	AIDR 3D	AiCE
-630HU	5mm	33%	34%	30%	32%	32%	31%	38%	35%	32%
	8mm	18%	16%	11%	20%	18%	13%	26%	18%	16%
	10mm	14%	12%	5%	17%	18%	18%	21%	14%	12%
	12mm	5%	2%	3%	10%	3%	3%	16%	6%	2%
+100HU	5mm	4%	3%	5%	7%	5%	6%	15%	5%	5%
	8mm	2%	1%	4%	4%	1%	4%	7%	4%	3%
	10mm	5%	7%	8%	4%	7%	9%	2%	9%	11%
	12mm	5%	5%	6%	3%	5%	6%	1%	6%	8%

and AiCE. For the evaluation of image noise, standard deviation (SD) of pixel value was measured by region of interest which located at the lung field and the air outside the thoracic phantom. For more accurate comparison of SD, statistical analysis was performed with one-way analysis of variance and Tukey honest significant difference post hoc tests. The differences with a P values less than 0.05 were considered to indicate statistically significant. Moreover, to assess the reproducibility of pulmonary nodule volume, the volume was measured by a commercially available software and calculated the absolute percentage error (APE) from theoretical value for FBP, AIDR 3D, and AiCE. The $> 25\%$ of APE was used to define a clinically relevant nodule volume difference in this study following Response Evaluation Criteria in Solid Tumors.

Results

The SD value reduction rates from FBP in AiCE and AIDR 3D were 79–85% and 62–77%, respectively. Statistical differences were found for SD among all reconstruction methods ($P < 0.01$). The results of APEs are listed in Table 1. For the pulmonary nodule with $- 630$ HU, the APEs of AiCE were lower than those of FBP and AIDR 3D at almost all dose strengths and nodule diameters. However, the nodule with a diameter of 5 mm showed $> 25\%$ of APE in all reconstruction methods and dose strengths. In the diameter of 8 mm at 1/6 standard dose, APE of only FBP represented $> 25\%$. Although AIDR 3D and AiCE showed less or equivocal APEs than FBP in the diameters of 5 and 8 mm for the pulmonary nodule with $+ 100$ HU, AIDR 3D and AiCE represented greater APEs than FBP in the diameters of 10 and 12 mm. However, all images for the pulmonary nodule with $+ 100$ HU resulted in $< 25\%$ of APEs.

Conclusion

AiCE achieved greater noise reduction than AIDR 3D without compromising lung nodule volumetry accuracy based on clinical requirement.

References

- [1] Higaki Toru, Yuko Nakamura, Jian Zhou, Zhou Yu, Takuya Nemoto, Fuminari Tatsugami, Kazuo Awai (2020) Deep Learning Reconstruction at CT: Phantom Study of the Image Characteristics. Academic Radiology 27(1):82–87.
- [2] Aberle DR, Adams AM BC, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM SJ (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. New England Journal of Medicine 365:395–409.

A benchmark for the evaluation of computational methods for bronchoscopic navigation

J. Borrego-Carazo¹, C. Sanchez¹, D. Castells-Rufas², J. Carrabina², D. Gil¹

¹Computer Vision Center, Computer Science, Cerdanyola del Vallès, Spain ²Universitat Autònoma de Barcelona, Microelectronics and electronic systems, Cerdanyola del Vallès, Spain

Keywords Bronchoscopic Navigation, Pose estimation, Evaluation metrics, Synthetic datasets

Purpose

Vision-based bronchoscopy (VB) models require registration of the virtual lung model with video bronchoscopy frames to provide effective guidance during interventions. The registration can be achieved by either tracking the position and orientation of the bronchoscopy camera or by calibrating its deviation from the pose (position and orientation) simulated in the virtual lung model.

Methods based on hand-crafted features and similarity measures have tracking errors and large execution times. Recently, data-intensive learning methods, like neural networks, have provided new state-of-the-art results.

In any case, results are affected by a lack of fair comparability due to the absence of public datasets and the usage of inappropriate metrics. Additionally, learning methods depend on high data availability, often hindering their application in data-scarce environments.

We contribute by addressing both topics. First, a bronchoscopy navigation synthetic dataset for comparability among methods and address data requirements. Second, a comparison of pose estimation metrics (including a novel one) to establish better grounds for training and evaluation.

Methods

Synthetic dataset

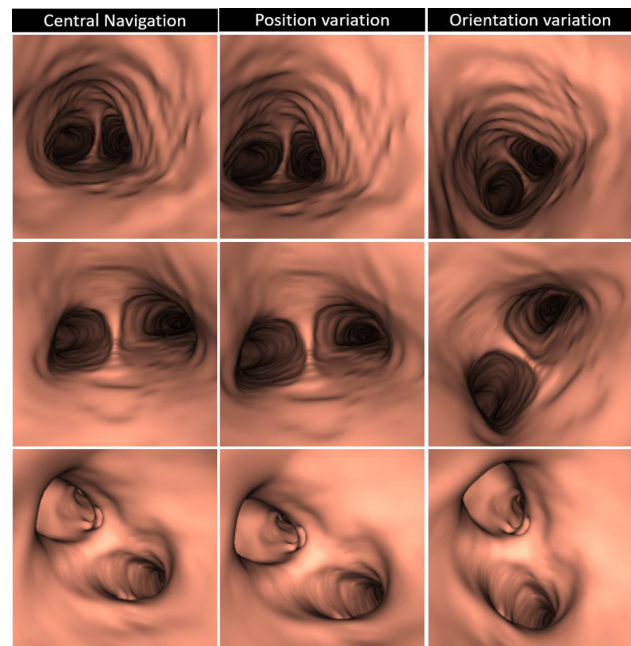


Fig. 1 Example of synthetic frames from a trajectory from patient P18 lower left lobe

Virtual models come from 6 computed tomography scans using [1] to segment the airways. Virtual airways models are simulated using a platform developed in C++ and VTK.

Bronchoscope trajectories are simulated from the trachea to the 4–6 level covering upper-right, lower-right, upper-left, and lower-left lobes. Trajectories are generated from a path through the luminal central line traversed using the arch-length parameter. Different increments in this parameter allow the simulation of varying velocities across the path.

For each central path, variations in position ([- 2: 1: 2] voxels in each axis) and camera orientation ([- 45: 15: 45] degrees of rotations around the navigation vector) are generated. The variation in camera position implicitly also modifies the camera point of view, given its position and a point in the central path at a distance Δd from the current point. The rotation around this navigation vector introduces a variation in the orientation of the image plane. This way, we simulate a full change in the camera central pose.

In order to simulate realistic trajectories, paths with neighboring variations are randomly combined along with the navigation arc-length parameter. In total, our dataset has 876 trajectories per patient and lobe, amounting to a total of 842,712 frames.

The dataset has as input values the synthetic frames from the camera view during the trajectory, and the associated pose inside the VTK airway model coordinate system as ground truth source values. The position is in voxel units and camera view angles are presented in Euler angles.

Figure 1 shows some dataset examples. Rows are different carinas and each column represents variations in position and orientation from the central navigation. Dataset will be made publicly available upon abstract acceptance.

Losses and Metrics for Pose Estimation

A system, f , for camera pose estimation predicts the difference in pose between two images:

$$f : \mathbf{I} \rightarrow \Delta\mathbf{P}, \mathbf{I} \in \mathbb{R}^{2 \times C \times H \times W}, \Delta\mathbf{P} \in \mathbb{R}^6 \tag{1}$$

where \mathbf{I} are the two images, $C = 3$ the RGB channels, height H and width W , and $\Delta\mathbf{P} = (\Delta x, \Delta y, \Delta z, \Delta \alpha, \Delta \beta, \Delta \gamma)$ is the difference in pose, given by (x, y, z) the position coordinates and (α, β, γ) the Euler angles.

In order to train and validate f , we need metrics for assessing the error of the predicted rotation and position. As those are two separate components, we can have different metrics (or losses) for each of them. Let us note L_p and L_r the metrics for, respectively, position and rotation.

The common choice for, both, L_p and L_r , is either the mean squared error (MSE) or, its equivalent, the euclidean norm (L_2). Although they naturally model the euclidean space of positions, they are not adequate for the space of rotations. An alternative is the direction error (DE) [2]

$$L_{rDE} = \cos^{-1}(\mathbf{v}_o \cdot \mathbf{v}_g) \tag{2}$$

where \mathbf{v}_o and \mathbf{v}_g are the predicted and ground truth direction vectors.

The main issue with DE is that the choice of the unitary vector, \mathbf{u} , for building the direction vector affects the perceived rotations since DE is oblivious to errors in rotations around \mathbf{u} . To remedy such a problem we present an alternative metric, the *cosinus error* (CE):

$$L_{rCE} = 1/N \sum_i^N (1 - \cos(\Delta\mathbf{P}_r[i]_o - \Delta\mathbf{P}_r[i]_g)) \tag{3}$$

where N refers to the total number of angular components, i is an index for them, o indicates prediction and g ground truth.

Table 1 Results for the different loss combinations evaluated with the different metrics. Values show mean and standard deviation

Loss	Position Error		Rotation Error	
	L_2	L_2	DE	CE
$L_p + L_r$	0,549 ± 0,573	1.236 ± 6.762	0,724 ± 0,498	0,173 ± 0,194
$L_p + L_r + DE$	0,368 ± 0,191	2.441 ± 8.491	0,713 ± 0,533	0,434 ± 0,385
$L_p + L_r + CE$	0,264 ± 0,169	1.554 ± 8.572	0,542 ± 0,503	0,116 ± 0,181

Results

As experimental setup, we select 30 trajectories per lobe and patient for training, from which we reserve 6 for testing and the rest for training. As network, we build a model, f , composed of two EfficientNetV1-B0 backbones, a convolution fusion layer, a ShuffleNet block, and a fully connected layer for predicting ΔP. Networks are trained until convergence with Adam optimizer, learning rate of 10^{-4} and batch size of 2048.

The network loss is given by the addition of the position and rotation metrics:

$$L = L_p + L_r \tag{4}$$

L_p is given by the MSE error, while for L_r we used the three metrics described in the previous section. For each loss, we evaluated the position and rotation errors in the test set using the L_2 norm for position (in voxels) and L_2 (in degrees), DE (in degrees), and CE for the rotation.

Table 1 shows mean ± standard deviation for the different loss combinations evaluated with the different metrics. Best performers are highlighted in bold. For all metrics, the network trained with the proposed CE achieves the best results. Models trained with metrics adapted to the rotational domain improve also position errors. A reduction in their standard deviation indicates a more stable behavior and, thus, higher generalization and transfer capability.

Conclusion

We have presented a synthetic dataset for bronchoscopy pose estimation to enable fair comparability among methods and improve learning-based VB training. Additionally, we have argued that current rotation metrics and losses are not appropriate and provided a more Future work could consider transfer learning from synthetic to real data, and improving the network architecture.

References

[1] Debora Gil, Carles Sanchez, Agnes Borrás, Marta Diez-Ferrer, and Antoni Rosell. (2019). *Segmentation of distal airways using structural analysis*. Plos one, 14(12):e0226006.
 [2] Scott A. Merritt, Rahul Khare, Rebecca Bascom, and William E. Higgins (2013) *Interactive CT-Video Registration for the Continuous Guidance of Bronchoscopy*. IEEE Transactions on Medical Imaging, 32(8):1376–1396.

Image-based algorithm for automated insertion of CT-guided biopsy needle: proof of concept

L. Kini¹, P. Hadar², B. Haas¹
¹University of California, San Francisco, Radiology, San Francisco, United States ²Hospital of the University of Pennsylvania, Neurology, Philadelphia, United States

Keywords Algorithm, Image-guidance, Biopsy, CT

Purpose

Each year, millions of Americans are diagnosed with cancer, and a significant number undergo biopsy for diagnosis and staging. These biopsies, which are commonly done through CT-guided needle insertions, are essential in guiding further care for our patients. However, these CT-guided needle biopsies are at times technically difficult, particularly if the needle path is not in plane with the axial images from the CT scanner. These out of plane needle paths are extremely difficult for the operator to plan out and execute.

Needle positioning and insertion during CT guided procedures is operator-dependent and solely based on the physician's visuo-spatial estimation of angle and depth from a scan. Therefore, when an out of plane path is required, the operator often struggles to biopsy the lesion, and many times may be unsuccessful. Even when successful, the procedure often takes longer, has higher radiation dose, or a higher likelihood of a non-diagnostic or false negative pathology result.

Instead of relying on human estimation, we propose the employment of an imaging-based algorithm to provide the angle of entry and depth of lesion prior to procedure. Therefore, a physician can have a more accurate estimate of the path for needle insertion, potentially shortening procedure time, limiting radiation exposure, and minimizing under-sampling. We provide a proof of concept of this approach in a single subject, described below, that minimizes the distance between lesion and skin surface while avoiding bone and blood vessels.

Methods

A CT Chest scan was acquired from a single subject from the public access Cancer Imaging Archive (TCIA) non-small cell lung cancer (NSCLC) dataset [1]. Using the Advanced Normalization Tools (ANTs) suite for automated imaging analysis [2], this 3D subject image were automatically registered to a control template, after which the lung soft tissues, blood vessels and bones could be automatically segmented using the labeled atlas template.

Using the segmentation and imaging data, ϕ (psi, yaw around the Z-axis) and θ (theta, pitch around the y-axis) could be derived. Python scripting was used to generate these angles and depths of the lesions relative to a skin insertion site. The goal of the algorithm was to chart a needle biopsy path that was the shortest distance from the surface to the lesion while avoiding major bones or vessels such that a needle cannula can be targeted directly towards the lesion. A short distance between skin and lesion is desirable because it provides for greater tolerance in less-than-ideal needle angles. Paths traversing blood vessels need to be avoided to minimize bleeding complications. The iterative algorithm was as follows:

Algorithm 1: Interventional Radiology Optimization

```

x represents the space of voxels

Require: Image I(x) is a CT CHEST OR CT ABDOMEN/PELVIS for space x
Ensure: Image I is a 3D volume image
TISSUE_NL(x) <- TISSUE_ALL - SEGMENT(I(x))
PATH_NEEDLE(x) <- INITIAL - ROI(x)

while PATH_NEEDLE does not extend outside image limited do
  if AIR is in TISSUE_NL(PATH_NEEDLE(x)) AND BONE and VESSEL not in line with AIR
  then
     $\phi$ ,  $\theta$  <- polar angles to air voxel end while
  else
    PATH_NEEDLE(x) <- GROW - PATH_NEEDLE : Grow INITIAL by 1 voxel radius
  end if
end while

```

In simpler terms, x represents the space of voxels. First the algorithm confirms that the image is the appropriate one (that it is a CT of the Chest or Abdomen and Pelvis), as well as that it is in the correct 3D volume format. $TISSUE_NL(x)$ effectively creates a negative image, where the lesion (SEGMENT) is removed from the total tissue image ($TISSUE_ALL$) and only the normal tissue is present; this is important, because the needle has to traverse normal tissue. $PATH_NEEDLE$ effectively describes the total distance (or depth) from a random point (INITIAL) to the lesion (ROI, represented as a point). When the $PATH_NEEDLE$ is complete, INITIAL will represent the skin surface. BONE and VESSEL are derived from the segmentations, while AIR outside the body is determined by creating the air-skin boundary, computed using the convex hull of all soft tissue segmentations. Assuming that the $PATH_NEEDLE$ is inside the actual image (no computer errors), that anticipated path does not cross any bone or vessels, and the path includes air (reaches just outside the skin), then the angles will be calculated. If these criteria are not met, then a new $PATH_NEEDLE$ needs to still be created, so a new INITIAL point will be created by moving it 1 voxel radius away from the ROI. This will iterate until an accurate $PATH_NEEDLE$ is completed.

Results

The CT scan image overlaid with segmentation is shown below in Fig. 1. Two lesions are segmented in the anterior right lobe of the lung, seen as a lighter blue compared to the green normal tissue. Using the iterative algorithm described previously, we were able to determine the optimal angle trajectory for both lesions that demonstrates the shortest distance to the skin surface without hitting bone or vessels, with angles of ϕ 5 degrees and θ 35 degrees for the largest lesion noted on the figure.

Conclusion

In summary, our proof of concept demonstrates that, in this single subject who underwent a CT of the lung, we were able to employ an image-based iterative algorithm to calculate an optimal trajectory that minimized the distance between the skin surface to the lesion while

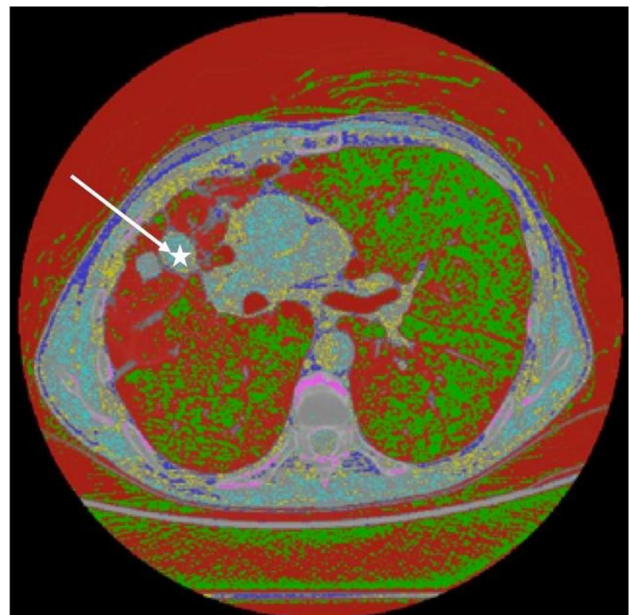


Fig. 1 Segmented CT Lung section with lesion and optimal trajectory for the largest lesion noted, with angles of ϕ 5 degrees and θ 35 degrees

avoiding bone and vessels. Board certified chest radiologists who perform these procedures will prospectively confirm the validity of the path discovered. Future directions will include running this algorithm on several more subjects with CT lung scans, and expanding to subjects with CT abdomen/pelvis scans. Prospective recording of physician-performed needle insertion angles on patients who are undergoing lung biopsy can then be compared to the ideal path calculated by our algorithm.

References

- [1] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging*. 2013 Dec 1;26(6):1045–57.
- [2] Tustison NJ, Cook PA, Holbrook AJ, Johnson HJ, Muschelli J, Devenyi GA, Duda JT, Das SR, Cullen NC, Gillen DL, Yassa MA. The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports*. 2021 Apr 27;11(1):1–3.

Prediction of deep myometrial invasion of uterine endometrial cancer on MRI using vision transformer

Y. Kurata¹, M. Nishio¹, Y. Matsumoto¹, S. Otani¹, Y. Moribata², Y. Himoto¹, A. Kido¹, M. Mandai³, Y. Nakamoto¹

¹Kyoto University Graduate School of Medicine, Diagnostic Imaging and Nuclear Medicine, Kyoto, Japan ²Kyoto University, Preemptive Medicine and Lifestyle-Related Disease Research Center, Kyoto, Japan ³Kyoto University Hospital, Gynecology and Obstetrics, Kyoto, Japan

Keywords Endometrial cancer, MRI, Vision transformer, deep learning

Purpose

Endometrial cancer (EC) is the most common gynecological malignant tumor in developed countries. The presence or absence of deep myometrial invasion is an important factor in determining the treatment strategy of EC and it is usually evaluated using preoperative MRI. There have been several reports to predict the degree of muscle invasion of EC in MRI using machine learning methods. Most of these reports were based on texture analysis, and some of them used convolutional neural networks. In this study, we investigated whether we could predict the deep myometrial invasion of EC using Vision Transformer (ViT), a deep learning method that has recently attracted considerable attention [1].

Methods

This research included 200 patients surgically diagnosed with EC who underwent preoperative MRI. From the 200 patients, 159 patients were randomly selected as training datasets and the remaining 41 patients as test datasets. The presence of deep myometrial invasion was determined by the pathological diagnosis. Two board-certified gynecologic radiologists manually segmented the tumor on each slice of the sagittal T2-weighted image (T2WI) using a 3D Slicer (<https://www.slicer.org/>) by referring to all the available images and pathological records in consensus. The MR images were reformatted to 512*512, and rectangular areas containing the tumor and 50 pixels around it were extracted. Three consecutive slices of sagittal T2WI

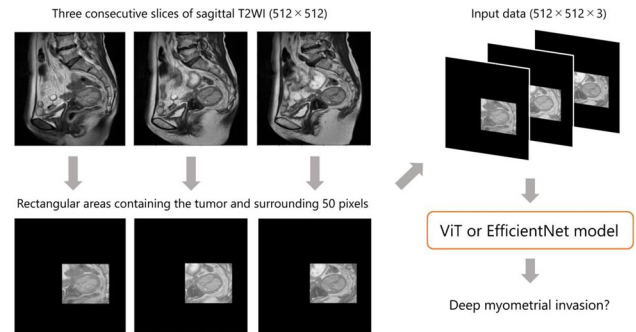


Fig. 1 The workflow of the prediction of deep myometrial invasion using Vision Transformer (ViT) or EfficientNet

including the largest area of EC were used as input data. In addition to ViT, EfficientNet B5-based model was adopted for comparison with the convolutional neural network-based model [2]. Hyperparameter tuning and the model training were performed using the training datasets with five-fold cross-validation, and the final model with the highest diagnostic performance was determined. The ensemble results of the five cross-validation models were adopted as the prediction of the final model. The diagnostic accuracy of the final model was calculated using the test dataset. Two board-certified gynecologic radiologists who were unaware of the clinical information evaluated the test dataset concerning the presence of deep myometrial invasion referring to all the available images. The accuracy of deep myometrial invasion and inter-rater reliability were calculated.

Our ViT model and EfficientNet model were built using PyTorch (version 1.9.0) and TensorFlow (version 2.3.0), respectively. The models were trained on a Linux workstation (Ubuntu version 18.04) with NVIDIA Quadro RTX8000 GPU with 48 GB memory (NVIDIA, Santa Clara, CA, USA).

Results

The average diagnostic accuracy [\pm standard deviation] of five-fold cross-validation of ViT model and EfficientNet model was 77.7% [\pm 1.2%] and 75.5% [\pm 4.2%], respectively. The diagnostic accuracy of ViT model for the test dataset was 78.0%. The diagnostic accuracy of the two radiologists for the test dataset was 87.8% and 80.5% ($\kappa = 0.71$) (Fig. 1).

Conclusion

ViT outperformed EfficientNet-based model for the prediction of deep myometrial invasion of EC on MRI. Our results showed the feasibility of ViT in gynecological MR imaging.

Acknowledgments: This work was supported by JSPS KAKENHI Grant Number JP20K16780.

References

- [1] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. 2020.
- [2] Tan M, Le Q, editors. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*; 2019: PMLR.

Skeletal muscle segmentation by simultaneous learning of particular superficial back muscles using 2D U-Net in torso CT images

M. Kawamoto¹, N. Kamiya¹, X. Zhou², H. Kato³, T. Hara^{2,4}, H. Fujita²

¹Aichi Prefectural University, School of Information Science and Technology, Nagakute, Japan ²Gifu University, Faculty of Engineering, Gifu, Japan ³Gifu University, Dept. of Radiology, Graduate School of Medicine, Gifu, Japan ⁴Tokai National Higher Education and Research System, Center for Healthcare Information Technology, Nagoya, Japan

Keywords trapezius muscle, erector spinae muscle, supraspinatus muscle, adjacent muscle

Purpose

Although the effectiveness of deep learning in segmenting skeletal muscles in torso CT images has been demonstrated, certain challenges remain; for example, cross-sections such as the L3 cross-section and region-specific 3D segmentation have not been adequately investigated. This is partially owing to the high cost of annotating skeletal muscles. Studies concerning the amount of annotation and segmentation accuracy [1] and the 3D segmentation of muscles via localization using skeletal muscle attachment bones [2] are being conducted. However, the simultaneous identification of multiple muscles has been limited to muscles in the lower extremities and has not yet been investigated for the torso, wherein many organs exist.

In this study, we evaluate the accuracy of skeletal muscle segmentation via the simultaneous learning of the trapezius (T) muscle and the adjacent erector spinae (E) and supraspinatus (S) muscles that are located on the superficial layer of the back.

Methods

Herein, transverse images of the T muscle and the adjacent E and S muscles were simultaneously learned using a two-dimensional (2D) U-Net, and the learned model was subsequently used for simultaneous segmentation of the T muscle and the adjacent muscles. We used 30 non-contrast torso CT images obtained using a LightSpeed Ultra 16 (GE Healthcare, Chicago, IL, USA) CT scanner. The image size was $512 \times 512 \times 802\text{--}1104$ [voxel], and the spatial resolution was $0.625 \times 0.625 \times 0.625$ mm. Data augmentation was performed during the training phase. We used shear transformation ranging between $-\pi/8$ to $+\pi/8$, rotation by random angles ranging from -10° to $+10^\circ$, random scaling between -35% and $+35\%$, translation of random distances up to 25% of the image width, and horizontal flipping to generate additional images for data augmentation. The hyperparameters of the 2D U-Net included the number of epochs (50), learning rate (1×10^{-4}), batch size (4), and optimization function (Adam). A combination of cross entropy and Dice losses was used as the loss function.

For accuracy verification, three-fold cross-validation was used, and the accuracy was evaluated based on the percentage of agreement with a ground-truth image manually created by the graph-cutting tool ported to PLUTO, which is a shared platform for medical image diagnosis support. The Dice coefficient as well as recall and precision were used as evaluation indices.

Results

Table 1 summarizes the recognition results of the three-fold cross-validation of the T muscle and the adjacent muscles; the average Dice (DC) coefficient, recall (RC), and precision (PC) values for each muscle type are also listed therein. Table 1 reveals that the segmentation accuracies for the T, E, and S muscles when trained independently are 76.2%, 93.1%, and 0%, respectively. In contrast,

Table 1 Segmentation results on simultaneous learning of the trapezius and adjacent muscles

Training	Trapezius: T			Erector spinae: E			Supraspinatus: S		
	DC	RC	PC	DC	RC	PC	DC	RC	PC
Trapezius	76.2	74.9	79.8	–	–	–	–	–	–
Erector spinae	–	–	–	93.1	93.1	93.2	–	–	–
Supraspinatus	–	–	–	–	–	–	0.0	0.0	–
T& E	89.6	88.6	91.7	92.8	93.4	92.4	–	–	–
T& S	74.3	84.2	72.2	–	–	–	42.5	34.2	62.6
E& S	–	–	–	92.5	93.3	91.9	87.8	87.4	89.4
T& E& S	78.4	74.2	90.1	90.4	88.8	92.7	71.8	69.6	89.8

when these muscles are trained simultaneously (T & E & S), segmentation accuracies were 78.4%, 90.4%, and 71.8%, respectively.

Subsequently, we focused on the T muscle and the adjacent E and S muscles. When we focused on the T muscle, simultaneous learning of the T and E muscles yielded the highest segmentation accuracy (89.6%), while simultaneous learning of the T and S muscles yielded a marginally lower accuracy (74.3%). By contrast, when focusing on the S muscle alone, the segmentation accuracy was 0%; however, it improved on simultaneous learning with the T or E muscles. The highest accuracy for the S muscle was achieved on learning the E and S muscles simultaneously (87.8%). The E muscle offered a high segmentation accuracy (93.1%) when learned independently. Simultaneous learning of adjacent muscles did not improve the segmentation accuracy for the E muscle; notably, the accuracy decreased to 90.4%.

Conclusion

This study investigated the accuracy of segmentation using a U-Net in torso CT images by simultaneous learning of the T muscle and adjacent muscles located in the superficial layer of the back. The results indicate that for the simultaneous segmentation of multiple adjacent skeletal muscles, it is more effective to learn a single skeletal muscle with high segmentation accuracy, such as the E muscle learned herein, than to learn multiple regions simultaneously.

References

- [1] Wakamatsu Y, Kamiya N, Zhou X, Hara T, Fujita H (2021) Relationship between number of annotations and accuracy in segmentation of the erector spinae muscle using Bayesian U-Net in torso CT images Proc. of International forum on medical imaging in Asia 2021, vol 11792. 1179207-1-4.
- [2] Wakamatsu Y, Kamiya N, Zhou X, Kato H, Hara T, Fujita H (2021) Automatic segmentation of supraspinatus muscle via bone-based localization in torso computed tomography images using U-Net. IEEE Access 9:155555–155563. <https://doi.org/10.1109/ACCESS.2021.3127565>.

AI-assisted RoI extraction and DVH calculation tool for Geant4-based radiotherapy simulation

A. Kimura¹

¹Ashikaga University, Faculty of Engineering, Ashikaga, Japan

Keywords RoI segmentation, visualization, radiotherapy simulation, NVIDIA Clara AI

Purpose

Geant4 toolkit has been used for a Monte Carlo simulation of the passage of particles through matter. It takes charge of the important role in high energy physics in order to verify a theory, explore an unknown particle, and so on. In radiotherapy, it has been used in order to precisely evaluate dose distributions in an irradiated human body constructed from medical data or optimize devices of a beam delivery system.

We had developed and released gMocren since 2006 [1]. It is volume visualization software for a radiotherapy simulation. It had been designed according to the requirements of medical users and for the visualization system of Geant4. gMocren can only visualize RoIs, and it has no capability to edit RoIs and use to calculate a DVH.

In this research, a software tool that is named gRovai has been developed for RoI extraction using NVIDIA Clara [2] and DVH calculation for Geant4-based radiotherapy simulation. It supports the extraction of tumors using artificial intelligence from patient image data such as a DICOM dataset, RoIs editing, a DVH calculation, and volume visualization for radiotherapy simulation.

Methods

RoI extraction using NVIDIA Clara

NVIDIA Clara is provided as a docker container for a Linux PC with an NVIDIA GPU. It works on a Linux PC as a standalone AI server and provides an HTML interface to communicate with the other software. NVIDIA Clara server allows access through an HTML-based API. gRovai that runs on a PC sends CT data in NIfTI (the Neuroimaging Informatics Technology Initiative) format files to the NVIDIA Clara server and then receives extracted tumor or tissue region data from the NVIDIA Clara server. It is used to extract an initial RoI as a GTV. gRovai has a function to edit RoI shapes from the initial RoIs. The tumor region of edited RoI will be able to return to and update the trained AI model for NVIDIA Clara.

RoI editing and DVH

A DVH is calculated with a DICOM image dataset and dose distribution calculated by a radiotherapy simulation. NVIDIA Clara extracts tumor or organ regions in the DICOM image. The extracted regions are used as initial RoI data that are defined as a GTV. gRovai uses the dilate function of morphological transformations in OpenCV in order to estimate CTV and PTV automatically based on the extracted GTV. Using the dilation function is a temporary implementation for rapid prototyping of gRovai. It is going to replace a method based on theoretical definitions of CTV or PTV in order to support efficient RoI editing.

Visualization

gRovai can be also visualized medical image data in 3D volume rendering with VTK (Visualization Toolkit). It supports 1D and 2D transfer functions. It has also the function of drawing a DVH.

Results

A spleen region in a patient image dataset can be extracted adequately by using NVIDIA Clara. For instance, a spleen region is extracted from the chest region of a patient image dataset within 10 s. The size of the dataset is $512 \times 512 \times 34$ voxels. Users cannot edit only the GTV, but also CTV and PTV on a 2D image by using a mouse in gRovai as shown in Fig. 1. Finally, a DVH is calculated with the edited RoIs and a dose distribution calculated by Geant4-based radiotherapy simulation.

The patient image and ROIs are visualized as shown in Fig. 1. The contrast of the patient imaged is controlled with window level and window width in the 1D transfer function. It provides that two GUI sliders control window level and window width that are the parameters of the 1D transfer function. The 1D transfer function is drawn on the histogram of CT values in the transfer function pane. The 2D transfer function is provided in spatial gradient values or spatial curvature values of CT values. The 2D histogram is drawn in the CT values versus spatial gradient values in the transfer function pane. The

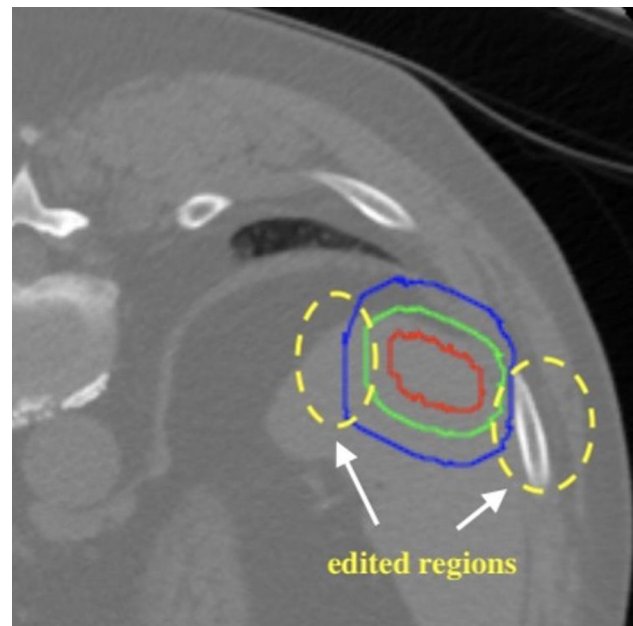


Fig. 1 gRovai displays RoIs on 2D medical image. It extracts a GTV region (red) by using NVIDIA Clara and defines automatically a CTV (green) and a PTV (blue) regions around the GTV

ranges of the CT values and the spatial gradient or curvature values are controllable in GUI sliders.

Conclusion

gRovai using AI has been developed to extract and edit RoIs and calculate DVH for Geant4-based radiotherapy simulation. NVIDIA Clara is used as an AI server to extract tumor or organ regions in a medical image dataset. The extracted region is used to estimate RoIs automatically, and then the RoIs can be edited by users. The RoIs are used to calculate a DVH with a dose distribution calculated by using Geant4 radiotherapy simulation. gRovai has also capability to visualize CT data in a 3D volume rendering with VTK. It will be available as standalone software. Therefore, it is useful and available not only for students or beginners but also for researchers using Geant4-based radiotherapy simulation.

References

- [1] Kimura A, et al., gMocren: Visualization software for Monte Carlo simulators for radiotherapy, *Journal of Advanced Simulation in Science and Engineering*, Vol.2, No.1, pp. 45–62, 2015.
- [2] NVIDIA Clara, <https://developer.nvidia.com/clara>

Segmentation of bones from head CBCT using federated learning with clinician-in-the-loop

X. Xu¹, H. H. Deng², T. Kuang², P. Yan¹, J. J. Xia²

¹Rensselaer Polytechnic Institute, Department of Biomedical Engineering, Troy, United States ²Houston Methodist Research Institute, Department of Oral and Maxillofacial Surgery, Houston, United States

Keywords Segmentation, Head CBCT, Federated learning, Clinician-in-the-loop

Purpose

Segmentation of bones from head cone-beam computed tomography (CBCT) is an essential step in craniomaxillofacial surgical planning. Currently, while deep learning-based methods have been used in clinical practice and achieved promising results, clinicians still need to manually improve the segmentation results, especially in thin bone, teeth, and other critical regions. There are two issues that hinder further advancement in this area. First, the deep learning-based methods highly rely on large-scale training datasets, which are difficult to obtain in most clinical settings where sharing data among the institutes is infeasible due to the privacy concern of patients. Second, once the training is completed, the conventional deep learning-based methods no longer receive feedback from the clinicians, whose opinion is crucial to improve the model performance to a clinically acceptable level. To address the above two issues and achieve better performance in bone segmentation, in this study, we propose a novel federated learning method equipped with a clinician-in-the-loop (CITL) training strategy for head CBCT segmentation.

Methods

The proposed method is built upon a federated learning framework [1], which consists of a server node and multiple client nodes as shown in the left of Figure 1. Each client node denotes a clinical site owning a local training dataset that cannot be shared with the server node and other client nodes. The server node and the client nodes iteratively update the deep learning segmentation model through multiple rounds of communications. At the beginning of each round, the client nodes download the global model from the server node. The client nodes then independently train their copies of the global model, namely local models, using their own set of images and annotations. After training the local models, the client nodes upload them back to the server node. At the end of each round, all the uploaded local models are aggregated to generate a new global model. The federated learning procedure continues by repeating such communication rounds in a loop manner until the global model converges. Since the local models are trained independently on different client nodes, federated learning does not require data sharing between the server and client nodes and thus can leverage the knowledge contained in the distributed clinical data without violation of the data privacy.

To allow the clinicians to provide feedback to the federated learning models, we further design a CITL training strategy, which contains two phases as shown in the right of Fig. 1. In Phase 1, we firstly train a local model on a base training set, which is manually annotated by the clinicians beforehand. In Phase 2, the trained local model is used to make predictions on the daily upcoming CBCT images. The clinicians are required to make necessary revisions and approvals on the model predicted segmentations, which are subsequently sent to the downstream tasks in craniomaxillofacial surgical planning. The feedback, i.e., the refined segmentations, made by the clinicians are then used to finetune the local model, aiming to improve the segmentation performance in the difficult regions.

Results

We conducted experiments on a clinical dataset containing 65 subjects. Each case contained a manual annotation of four facial bony structures (i.e., midface, mandible, upper and lower teeth), which were served as the ground truth for model training and evaluation. We randomly divided the dataset into training, validation, and testing sets containing 32, 12, and 21 subjects, respectively. Each subset was evenly divided into two parts to mimic two clinical sites in the FL framework. For the training set, 22 subjects were used as the base training data for Phase 1 training and the rest 10 subjects as the daily upcoming data for Phase 2 CITL finetune. We used 3D U-Net [2] as the segmentation network in our method. Dice similarity coefficient (DSC) was used as the metric to quantitatively assess model performance.

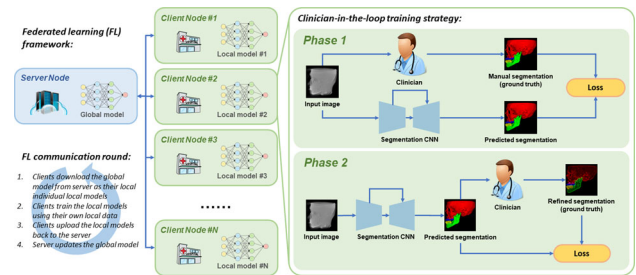


Fig. 1 Facial bones segmentation in head CBCT images using federated learning framework equipped with clinician-in-the-loop training strategy

Table 1 DSC of facial bones segmentation generated by different methods

Methods	ROIs [mean(std)]				Global
	Midface	Mandible	Upper teeth	Lower teeth	
Single-site	84.67(1.73)%	84.57(1.42)%	81.90(4.07)%	81.25(2.47)%	83.10%
Multi-site	84.43(1.77)%	84.44(1.64)%	82.79(2.59)%	81.96(1.63)%	83.40%
Single-site + CITL	84.82(1.67)%	84.71(1.32)%	82.19(4.11)%	81.62(1.75)%	83.33%
Multi-site + CITL	84.61(1.79)%	84.59(1.69)%	83.11(2.35)%	82.27(1.64)%	83.65%

The first two rows of Table 1 show the results of the model trained using single-site data via standard deep learning (“Single-site”) and multi-site data via federated learning (“Multi-site”). The model trained with the multi-site data achieved an average DSC of 83.40%, an 0.3% increase over the model trained with the single-site data (83.10%). This result demonstrated that the federated learning framework could effectively utilize additional data distributed in different clinical sites to facilitate segmentation. The last two rows of Table 1 show the results of the standard deep learning-based single-site and the federated learning-based multi-site models when they were trained with the proposed CITL strategy. Benefiting from the CITL strategy, both the single- and multi-site models achieved an overall improvement in segmentation accuracy, which promptly suggested the effectiveness of using the clinician feedback for model finetuning.

Conclusion

In this work, we propose a federated learning-based segmentation method equipped with a CITL strategy to handle the task of facial bones segmentation from head CBCT images. By using the federated learning paradigm, our method can utilize the data from multiple clinical sites to facilitate the model training without violating the data privacy, resulting in higher segmentation accuracy compared with the standard deep learning model trained on single-site data. Furthermore, our proposed CITL strategy allows the segmentation network to learn from clinician experts’ feedback, which is demonstrated effective in improving the segmentation performance.

References

- [1] McMahan H. B., Moore E., Ramage D., Hampson S., Arcas B. A. (2017) Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics, pp. 1273–1282.
- [2] Ronneberger O, Fischer P, Brox T. (2015) U-Net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention, pp 234–241.

CNN-based automated segmentation of deep vessels in non-contrast-enhanced CT images: validation on multi-institutional databases

Y. Chen¹, S. Mazen¹, K. Uemura¹, Y. Otake¹, M. Takao², S. Iwakoshi³, T. Tanaka³, N. Sugano², Y. Sato¹

¹Nara Institute of Science and Technology, Division of Information Science, Ikoma, Japan ²Osaka University, Graduate School of Medicine, Suita, Japan ³Nara Medical University, Department of Diagnostics and Interventional Radiology, Kashihara, Japan

Keywords Lower limb, vessel segmentation, convolutional neural network, non-contrast-enhanced CT

Purpose

The segmentation of deep vessels in the lower limbs, particularly the femoral artery and vein, is helpful for the planning of surgical procedures in the hip and femoral regions, such as hip joint surgeries. Several automated segmentation approaches were proposed to accelerate the segmentation time and reduce the burden. However, those approaches are mainly applied to contrast-enhanced CT images or MR images. For example, Huang et al. proposed a convolutional neural network (CNN)-based method for the segmentation of lesions in the deep vein of the lower limbs in MR images of 58 cases, in which vessel region boundaries are clearly imaged, and obtained a mean Dice coefficient (DC) of 0.74 ± 0.17 [1]. However, MR imaging requires a long scan time, and the contrast enhancement in CT depends on the adequate and uniform filling of vessels, which is often not possible due to technical issues. In addition, unlike the abdominal vessels, femoral vessels are tightly surrounded by lower limb skeletal muscles, which makes it hard to segment vessels without contrast enhancement. Even though non-contrast-enhanced CT images are widely used in hip surgeries, up to our knowledge, automated segmentation of the deep vessels from those images has not been investigated. Therefore, the purpose of this study was to investigate the performance of a CNN-based U-Net [2] for the segmentation of the deep vessels from non-contrast-enhanced CT images.

Methods

In this study, we addressed the automatic segmentation of femoral artery and vein from non-contrast-enhanced CT images using CNNs. In our experiment, a two-stage 2D U-Net was used for the vessel segmentation. In the first stage, the patient skin was segmented to isolate the body from surrounding objects. In the second step, another U-Net was used to segment the femoral artery and vein.

The performance of the models was investigated on two databases obtained from two different institutions, hereinafter termed Institution1 and Institution2. The CT from Institution1 included 25 images with 512×512 in-plane resolution and a slice thickness of 1–2 mm. The images in this database were divided into 5 and 20 images for training and testing, respectively. The second database included 17 pairs of contrast- and non-contrast-enhanced images with 512×512 in-plane resolution and a slice thickness of 1 mm, which showed good

Table 1 Accuracy of deep vessel segmentation in non-contrast-enhanced CT images from two databases

Dataset (No. of testing images)	Structure	DC (Mean \pm STD)	ASD (Mean \pm STD; mm)
Institution1 (20)	Artery	0.710 ± 0.108	1.795 ± 2.091
	Vein	0.751 ± 0.096	1.318 ± 1.311
Institution2 (17)	Artery	0.624 ± 0.090	1.893 ± 0.738

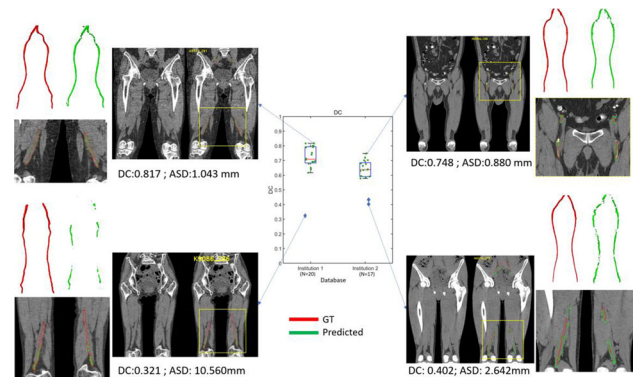


Fig. 1 Quantitative evaluation (Dice coefficient; DC) of the automated segmentation of the deep artery at the lower limbs in two different datasets. Segmentation errors were observed mainly at the intermediate part of the femur (left-down; right) and at the hip (left-up)

alignment between the contrast- and non-contrast-enhanced images. The ground-truth (GT) labels of the vessels were derived to include only the common (at the iliac region) and popliteal (at the femoral region) arteries and veins. The GT labels in Institution1 Database of the arteries and veins were annotated by a computer-science student and verified by an orthopedic surgeon. The GT labels in Institution2 Database of only the arteries were obtained by firstly segmenting the artery using a semi-automated (thresholding) approach from contrast-enhanced CT. Next, the contrast-enhanced image was registered to the non-contrast-image using a 3D image intensity-based non-rigid registration. The deformation field was used to align the segmented artery to the non-contrast-enhanced image. This database was only used for testing the model trained on Institution1 training datasets. The performance of the proposed approach was evaluated using DC and average symmetric surface distance (ASD).

Results

Table 1 summarizes the vein and artery segmentation accuracy in the two databases. The accuracy was better in Institution1 compared with Institution2 datasets. In Institution1, the DC for the vein and artery were 0.751 ± 0.099 and 0.710 ± 0.108 , and ASD were 1.318 ± 1.311 mm, 1.795 ± 2.091 mm respectively. In Institution2, the DC for the artery was 0.624 ± 0.090 and ASD was 1.893 ± 0.738 mm. Figure 1 shows representative examples for the artery segmentation in the two databases. The enlarged parts from Institution1 datasets show that the failures mainly happen at the intermediate femoral part, where the muscles are more densely aligned than those in the region around the proximal part of the femur, leading to smaller contrast with the vessels compared with other regions. The reason behind the degraded performance in Institution2 would be that the GT labels, which were obtained semi-automatically, cover only the internal part of the vessels (i.e. under-segments the vascular wall). Manual correction of those GT labels, besides the integration of the muscle information into the model training pipeline, is currently investigated for further improvements.

Conclusion

In this study, we investigated the performance of a two-stage CNN-based U-Net for the segmentation of the deep vessels from non-contrast-enhanced CT images. Compared with the results in [1] (mean DC 0.74 ± 0.17), which aimed at the segmentation of venous lesions clearly observable in MR images, our CNN shows promising performance even when applied to the whole deep vessels (veins and arteries), which are not clearly imaged in non-contrast CT. Furthermore, validated on two datasets, the experiments showed acceptable generalization capability for segmenting deep vessels.

References

- [1] Huang C, Tian J, Yuan C, Zeng P, He X, Chen H and Huang B (2019). “Fully Automated Segmentation of Lower Extremity Deep Vein Thrombosis Using Convolutional Neural Network.” *BioMed Research International*. 2019. 3401683
- [2] Hiasa Y, Otake Y, Takao M, Ogawa T, Sugano N and Sato Y (2019) “Automated Muscle Segmentation from Clinical CT using Bayesian U-Net for Personalized Musculoskeletal Modeling.” *IEEE Transactions on Medical Imaging*. 39(4): 1030–1040.

Automated segmentation of contrast-enhancing brain tumor using only post-contrast t1-weighted

N. Kashiwagi¹, C. Avare², R. Shiroishi³

¹Osaka University Graduate School of Medicine, Suita, Japan

²Avicenna.AI, Research and Development, La Ciotat, France ³Canon Medical Systems Corporation, Clinical Application Research department, Otawara, Japan

Keywords glioma, magnetic resonance imaging, convolutional neural network, segmentation

Purpose

Because of the lack of established biomarkers for primary brain tumors, magnetic resonance imaging (MRI) is the standard diagnostic tool to monitor brain tumors. In the Response Assessment in Neuro-Oncology (RANO) criteria, which is the current standard, contrast-enhancing tumor is measured by the 2D product of maximum bidimensional diameters. For a more accurate assessment, measuring the area or volume of tumor is expected, and which will be made practical by automated segmentation.

Automated segmentation of brain tumor has been explored using multi-contrast magnetic resonance imaging (MRI) [1]. However, in clinical practice, available contrasts can be restricted due to the limitation of examination time, differences of acquisition protocols, and image corruption or artifacts. Therefore, we focused on contrast-enhancing tumor, which is primary target for monitoring brain tumors.

The purpose of this study was to demonstrate automated segmentation of contrast-enhancing brain tumor with a convolutional neural network (CNN) architecture using only post-contrast t1-weighted.

Methods

This retrospective, single-center study was approved by our Institutional Review Board. Between September 2018 and August 2020, we

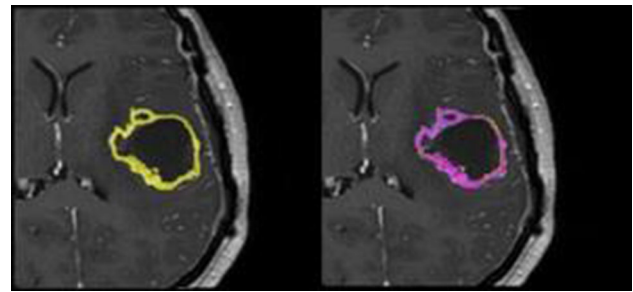


Fig. 1 Manually delineated ground truth of contrast-enhancing tumor (yellow in left image) and automatically segmented contrast-enhancing tumor region (magenta in areas overlapping with ground truth and cyan in other areas in right image)

identified 25 MR examinations from 16 patients with histological diagnoses of glioma. The patients included 4 male and 12 female with a mean age of 52.9 years (range: 17–90 years). By reviewing all the 25 post-contrast t1-weighted images, 966 slices with the presence of tumor were obtained. The images were acquired with 0.9 mm slice thickness, 240 mm × 240 mm field of view, and 512 × 512 reconstructed matrix using two 3 T and one 1.5 T clinical MR scanners. The ground truth for contrast-enhancing tumor was established by manual delineation by two radiologists using 3D Slicer.

CNN training was performed using 772 slices with U-net using focal loss function [2]. The focal loss function was introduced to address high class imbalance in segmentation task. The trained CNN was validated using the other 194 slices.

CNN performance was evaluated using Dice/F1 score (for spatial overlap) and Intraclass Correlation Coefficient (ICC) (for measurement of area) between manually and automatically segmented area. We also calculated Dice score and ICC between the manual segmentations of the two readers.

Results

Manually delineated ground truth (yellow in Fig. 1) and an automatically segmented contrast-enhancing tumor region (magenta in areas overlapping with ground truth and cyan in other areas) were obtained as illustrated figure.

Average Dice score was 0.758 (SD: 0.216) and 0.766 (0.198) for training and validation, respectively. When the tumor was confined to larger than 100 mm² in size, average Dice score was 0.820 (0.109). ICC between manually and automatically segmented area was 0.884 and 0.868 for training and validation, respectively. Regarding inter-rater difference, average Dice score and ICC was 0.832 (0.019) and 0.895.

Conclusion

CNN using only post-contrast t1-weighted can automatically segment contrast-enhancing brain tumor with almost comparable performance to manual segmentation. Although further validation and flexibility expansion in contrast image selection and segmentation target selection are needed, this study demonstrated the potential of CNN for brain tumor segmentation in limited contrast image setting.

References

- [1] Yan B, Cao M, Gong W, Wei B (2021) Multi-scale brain tumor segmentation combined with deep supervision. *Int J CARS* 2021

- [2] Lin T, Goyal P, Girshick R, He K, Dollar P (2020) Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 42, no. 02, pp. 318–327

Intestinal obstruction point finding assistance based on intestinal luminal area segmentation

H. Oda¹, Y. Hayashi², T. Kitasaka³, A. Takimoto¹, A. Hinoki¹, H. Uchida¹, K. Suzuki⁴, M. Oda⁵, K. Mori^{2,6,7}

¹Nagoya University, Graduate School of Medicine, Nagoya, Japan
²Nagoya University, Graduate School of Informatics, Nagoya, Japan
³Aichi Institute of Technology, School of Information Science, Toyota, Japan
⁴Aichi Medical University, Department of Radiology, Nagakute, Japan
⁵Nagoya University, Strategy Office, Information & Communications, Nagoya, Japan
⁶Nagoya University, Information Technology Center, Nagoya, Japan
⁷National Institute of Informatics, Research Center for Medical Bigdata, Tokyo, Japan

Keywords small bowel segmentation, intestinal contents, computer-aided detection, distance maps for segmentation

Purpose

This paper proposes an intestinal obstruction point finding assistance system from CT volumes of the ileus (non-mechanical intestinal obstruction) or the (mechanical) intestinal obstruction patients. These intestinal obstructions are diseases disrupting intestinal contents' movement. While a clinician manually traces intestinal regions to find obstruction points on non-fecal-tagged CT volumes by changing slices back and forth, this is quite a hard task for non-expert clinicians. A computer-aided detection (CADe) system is desired to be developed to assist such a burden and difficult task.

The obstruction points are usually thin or strangulated. The intestine is inflated by intestinal content retention. Based on such characteristics, the reference [1] showed a method to identify intestinal luminal regions on CT volumes to find obstruction points. This method presented areas around the endpoints of connected components as the obstruction regions. Although this method showed promising results in obstruction region identification, there was no scheme for controlling the balance of over or under presentation of obstruction point candidates.

To overcome this problem, we introduce a mechanism to select intestinal luminal connected components to be analyzed. This selection is based on the sizes of each connected component. The size-based parameter controls the balance of over or under presentation of obstruction point candidates.

In this paper, we present the method to select intestinal luminal connected components regarding their sizes. Intestinal luminal regions are segmented on distance maps inferred by a 3D fully convolutional network (FCN). The Watershed algorithm is applied to the distance map to segment intestinal luminal regions. Seed points for the Watershed algorithm are determined based on the size parameters that a user specifies. We connect fragmented connected components into a component to obtain an intestinal path. Around the areas of either end of the intestinal path are considered obstruction point candidates. Finally, we evaluate the proposed method by investigating the effect of the size parameters.

Methods

Overview

The proposed method identifies intestinal luminal regions on CT volumes to find obstruction points. Distance maps are estimated by the 3D FCN. The Watershed algorithm is applied to the distance map to segment intestinal luminal regions as *fragmented connected components*. Seed points of the Watershed algorithm are assigned based

on the size parameters. *Obstruction point candidates* are presented from the fragmented connected components.

Estimating distance maps by 3D FCN

We utilize a 3D FCN to estimate distance maps from the input CT volume, which are positive inside the intestines. The values on the distance maps represent the distance from outside the intestine, which are normalized into the range [0, 1]. Those distance maps often become wrongly positive when complexed structures appear inside the lung regions. Since wrongly segmented intestinal luminal regions in the lungs are much smaller than the surrounding air regions, we remove connected components smaller than e [%].

Acquiring fragmented connected components by Watershed algorithm

We utilize the Watershed algorithm on the distance maps to obtain *fragmented connected components* covering the intestinal luminal areas. First, we obtain connected components of the intestinal luminal areas as regions whose distance values are t ($0 < t < 1$) or higher. With the size parameters s and d ($s > 0$ and $d > 0$), we select connected components whose volume is the top s largest or not smaller than a sphere with a d mm diameter. Then, we divide each selected connected component into multiple *fragmented connected components* using the Watershed algorithm. The maximum number of local maxima on the distance maps keeping h mm between the local maxima are chosen for each selected connected component. Those local maxima are utilized as seed points of the Watershed algorithm.

Presenting obstruction point candidates from fragmented connected components

Following the previous method [1], the contiguity of fragmented connected components is represented as a graph. The longest path is regarded as an intestinal path for each connected component of the graph. Fragmented connected components at each intestinal path's endpoints are presented as *obstruction point candidates*.

Results

We utilized our proposed method on 110 CT volumes under the IRB approval of Aichi Medical University (Japan), with the patient-level four-fold cross-validation. The 3D FCN was trained using intestine labels manually traced by a trained medical student.

Performances were calculated with various parameter settings: $s \in \{5, 10, 15\}$ and $d \in \{30, 20, 10\}$ mm under $e = 60\%$, $h = 15$ mm, and $t = 0.2$. Obstruction points annotated by a surgeon were used for evaluating with metrics: 1) the minimum distance (MD) from the obstruction point and its nearest obstruction point candidate, and 2) the number of obstruction point candidates (NOPC). Those metrics are tradeoffs, and smaller is better.

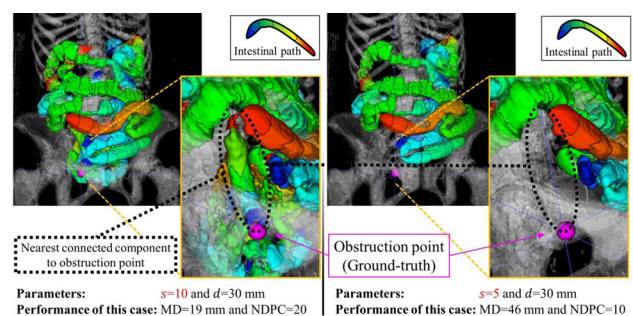


Fig. 1 Examples of different parameter settings ($s = 5$ or $s = 10$) under $d = 30$. When $s = 10$, connected component nearest to obstruction point was processed, and MD became 19 mm. When $s = 5$, connected component nearest to obstruction point was removed, and MD became 46 mm

Table 1 Performances with various parameters. Parameters for examples shown in Fig. 89 are colored in red

Parameters		Performances	
s	d [mm]	NDPC	MD [mm]
5	30	11,9 ± 2,5	22,4 ± 20,2
5	20	13,2 ± 3,5	22,3 ± 20,1
5	10	17,4 ± 6,4	21,3 ± 19,4
10	30	18,7 ± 3,5	19,6 ± 16,4
10	20	19,6 ± 4,1	19,5 ± 16,3
10	10	22,7 ± 6,7	19,1 ± 16,0
15	30	23,5 ± 5,5	18,8 ± 16,1
15	20	24,0 ± 5,9	18,8 ± 16,0
15	10	26,2 ± 7,9	18,7 ± 16,0
Previous method ¹		15,0 ± 6,2	20,1 ± 18,2

As shown in Table 1, setting s smaller strongly reduced the NOPC. In the case of Fig. 1, MD was 19 mm under $s = 10$ since a connected component of the segmentation results closest to the obstruction point was processed. In contrast, MD was 46 mm under $s = 5$ because the closest connected component was excluded. To ensure that many obstruction points are found, s should not be set too small.

Conclusion

We presented the method to select intestinal luminal connected components regarding their sizes. Experimental results showed that our proposed method correctly specified the balance of over or under presentation of obstruction point candidates.

Acknowledgments Parts of this work were supported by the Hori Sciences & Arts Foundation, JSPS KAKENHI (17H00867, 21K19898), JSPS Bilateral Joint Research Project, AMED (JP191k1010036, JP201k1010036), and JST CREST (JPMJCR20D5).

Reference

- [1] Oda H, Hayashi Y, Takayuki K, Tamada Y, Takimoto A, Hinoki A, Uchida H, Suzuki K, Itoh H, Oda M, Mori K (2021) Intestinal region reconstruction of ileus cases from 3D CT images based on graphical representation and its visualization. SPIE Medical Imaging 2021: 115971 N

Extraction of respiratory bronchioles and alveolar ducts from micro-CT volumes with distance-based tubular structure filter

T. Shiina¹, H. Oda², T. Zheng¹, S. Nakamura², Y. Hayashi¹, M. Oda^{1,3}, K. Mori^{1,4,5}

¹Nagoya University, Graduate School of Informatics, Nagoya, Japan

²Nagoya University, Graduate School of Medicine, Nagoya, Japan

³Nagoya University, Information and Communications, Nagoya, Japan

⁴Nagoya University, Information Technology Center, Nagoya, Japan

⁵National Institute of Informatics, Research Center for Medical Bigdata, Tokyo, Japan

Keywords respiratory system, acinus, graphical representation, 3D pathology

Purpose

This paper proposes an extraction method of ductal microstructures from lung micro-focus X-ray CT (μ CT) volumes. In recent years, high-resolution CT systems called μ CT have made it possible to capture the microstructures of the human lung. Three-dimensional analysis of lung microstructure using lung μ CT volumes will help the further investigation of anatomical structures of the lung.

However, observation of the *intra-acinar airways* (the respiratory bronchioles and the alveolar ducts) from the lung μ CT volumes is challenging. The lower respiratory tracts become thinner as the branch order increases, leading to difficulty understanding the distribution of the peripheral bronchioles from μ CT volumes. The alveolar ducts (airways beyond the bronchioles) are difficult to distinguish from the alveoli since the alveolar duct walls are composed of the alveolar walls. Thus, visualization of the intra-acinar airways is desirable to facilitate observation for physicians.

While there are many works to elucidate the anatomical structures of the lung, few are for the intra-acinar airways. Haefeli-Bleuer, et al. created silicone rubber casts and measured some quantitative parameters, including the lengths and the volumes of the inter-acinar airways [1]. Mori et al. proposed an anatomical labeling method of the bronchial branch by region growing and creating tree structures of the bronchus from chest CT volumes [2]. However, region growing does not work properly for extracting the intra-acinar airways. Unlike the bronchus, the walls of the intra-acinar airways are interrupted on μ CT volumes. Also, image noises of μ CT volumes are much stronger than clinical CT volumes.

This paper proposes an extraction method of the intra-acinar airways from μ CT volumes. We design a distance-based tubular structure filter to tackle the problem. We then perform the graph-based refinement to extract the complex tree structures of the intra-acinar airways.

Methods

Overview

From an input μ CT volume of a human lung specimen, we extract the skeleton of the intra-acinar airways by two steps: (1) initial extraction by distance-based tubular structure filter and (2) graph-based refinement. The output consists of the extracted intra-acinar airway segments. In our method, we apply the tubular structure enhancement filter to the distance map images generated from an input volume and do not directly apply it to an input volume. This scheme is designed to extract bronchiole regions from a noisy μ CT volume.

Preprocessing and lumen region extraction

The input μ CT volume is down-sampled by spline interpolation with factor z in each axis. We then extract the *lumen regions* as regions whose intensities are between t_1 and t_2 .

Stage 1: Initial extraction by distance-based tubular structure filter

We perform the Euclidean distance transformation on the lumen regions to obtain the distance map. The distance map represents the distances from the nearest wall of the intra-acinar airways. The values in the distance map are large around the centerlines of the intra-acinar airways, as the centers should be far from the walls. However, the distance values are also large around the centers of the other structures, such as alveoli.

We apply Sato's tubular filter with scales S [voxels] on the distance map. Since the filter is responsive to the tubular structures, it exclusively enhances the intra-acinar airways' centerlines. Thresholding for the filter responses (threshold t_3) and the thinning operation (6-neighbor connectivity) roughly generate the intra-acinar airway skeletons.

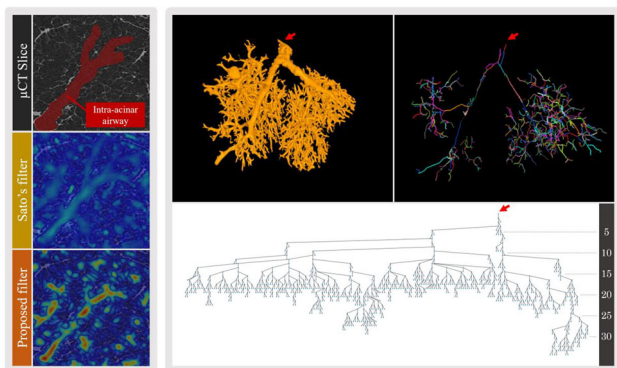


Fig. 1 (Left Half) Comparison of filter responses indicated by rainbow colors. From top to bottom: slice of Case 1, responses of Sato's tubular filter and responses of distance-based tubular structure filter. (Right Half) Extraction result of Case 1. Red arrows indicate starting point of extraction. Upper left: Initial extraction result of intra-acinar airway. Upper right: Final extraction result. Different colors represent different segments. Lower: Graphical representation. Numbers on right represent depth of segments

Stage 2: Graph-based refinement

We represent the skeletons extracted in Stage 1 as a tree. A node represents a branching point on the skeleton. An edge represents the connection of two neighboring nodes. The node nearest to the main bronchus is manually selected as the root. On the tree, the blind edges whose lengths are smaller than l [μm] are removed. Isolated nodes are also removed.

Results

We utilized two specimens of the human lungs. The specimens were scanned with a μCT scanner SKYSCAN 1272 (Bruker, Massachusetts, US) to evaluate our proposed method. The volumes had an isotropic size of $5.00 \mu\text{m}$. We manually selected one starting point of the respiratory bronchioles for each scanned volume. Thus, two intra-acinar airways (Case 1 and Case 2) were extracted. The following parameters are empirically set as $\alpha = 0.2$, $t_1 = 10,350$, $t_2 = 14,500$ and $t_3 = 0.9$. Parameters S and l are determined based on the anatomical findings^[1]: $S = \{3, 5, 7, 9\}$ and $l = 200$.

Figure 90 shows the filter responses and the extraction result of Case 1. The upper half of Fig. 90 indicates that our proposed filter enhances the centers of the intra-acinar airways compared to Sato's filter. As shown in the lower half of Fig. 90, we obtained graphical visualization of the intra-acinar airways by our proposed method, making it clear to understand the tree-like distribution

Table 1 shows the result of segment length per depth. As shown in Table 1, as the depth became larger, the number of segments increased once, reached the peak, and then turned to decrease. The number of segments increased because the intra-airway has a tree-like structure. From a certain branching point, it decreased because the maximum depth of the intra-acinar airways differs from each^[1]. Our proposed method successfully produced the results that conformed to the anatomical findings.

Table 1 Number of segments per depth (*i.e.*, branch order) in each case.

Depth range	1-3	4-6	7-9	10-12	13-15	16-18	19-21	22-24	25-27	28-30	31-33	34-36
Case 1	5	14	18	34	118	348	254	94	54	34	28	6
Case 2	7	22	88	222	98	42	2	0	0	0	0	0

Conclusion

We proposed an extraction method of the intra-acinar airways from μCT volumes. We designed a distance-based tubular structure filter and performed the graph-based refinement to extract the skeletons of the intra-acinar airways. The complex structures of the intra-acinar airways were graphically visualized. Future work includes applying the graphic data toward the accurate segmentation of the intra-acinar airway regions.

Acknowledgments Parts of this work were supported by the Hori Science & Arts Foundation, JPMJCR20D5, the JSPS Bilateral Joint Research Project, JSPS KAKENHI (17H00867, 17K20099), AMED (JP20lk1010036), and the Nagoya University CIBoG program from MEXT WISE program.

References

- [1] Haefeli-bleuer B, Weibel E R (1988) Morphometry of the Human Pulmonary Acinus. *The Anatomical Record* 220:401–414
- [2] Mori K, Hasegawa J, Suenaga Y, Toriwaki J, Anno H, Katada K (2000) Automated Anatomical Labeling of the Bronchial Branch and Its Application to the Virtual Bronchoscopy System. *IEEE Transactions on Medical Imaging* 19(2):103–114

Dataset containing RF signals of the Doppler twinkling artifact

D. Leonov^{1,2}, M. Kodenko^{1,3}, A. Nasibullina², N. Kulberg⁴
¹Moscow Radiology, Moscow, Russian Federation ²MPEI, Moscow, Russian Federation ³BMSTU, Moscow, Russian Federation ⁴Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russian Federation

Keywords Twinkling artifact, Color flow imaging, Phantoms, Raw radiofrequency signals

Purpose

In most studies devoted to the twinkling artifact, examinations were conducted using ultrasonic diagnostic devices without access to the signal-processing path. In such studies, the ultrasonic device can be considered a “black box”. The analysis based only on sonograms presented on the scanner's screen is not informative enough and creates problems with reproducibility since the signal processing algorithms in scanners produced by different manufacturers are unique. Thus, access to the raw signals is the primary condition for understanding the nature of the twinkling artifact.

To address this issue, we present an open-access dataset of RF signals obtained from the beamformer output of the preprocessing path of an ultrasonic research scanner and a tool for its viewing and analysis.

Methods

The output of the beamformer for CFI- and B-modes of the Sonomed-500 ultrasonic scanner (Spectromed, Moscow) served as a source of RF signals for the dataset. Most of the items in the dataset contain signals reflected from objects on which a twinkling artifact was observed in Doppler modes [1]. To understand the twinkling artifact's nature, we used artificial objects: rough and smooth cylinders made of low carbon, raw steel, and cylinders made of plastic (ABS), aluminum, and wood [2]. We secured the objects in fixed positions in the body of a specially designed phantom and filled the body with agar-agar, water, and ethyl alcohol (Fig. 1a–c). In humans, the twinkling artifact is primarily seen on kidney stones; however, artificial objects

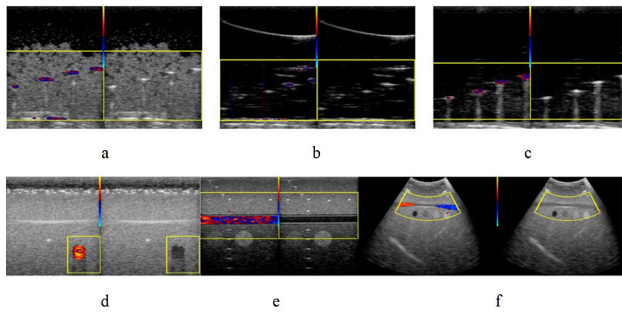


Fig. 1 Sonograms of the custom (a–c) and Gammex (d–f) phantoms: four aluminum cylinders in agar–agar (a), ethanol (b), degassed water (c), Gammex with a linear sensor at a 60° angle to the flow (d), parallel to the flow (e), with a convex sensor (f)

in the study fit better than in vitro kidney stones because they allow controlling and changing physical properties, e.g., by sanding their surfaces. A Doppler phantom Gammex 1430 LE Mini-Doppler Flow System (USA) was used to model blood vessels and obtain flowing fluid signals (Fig. 1d–f).

Results

Records of digital RF signals with twinkling artifact features were collected and placed in the public domain at https://mosmed.ai/datasets/ultrasound_doppler_twinkling_artifact. The dataset also contains signals from vessels in the Gammex phantom and reflections from tissue-imitating materials. This database will be helpful to researchers who study algorithms for B-mode and CFM signal processing. We developed an original software for working with the dataset; the source code of this program is available at <https://github.com/Center-of-Diagnostics-and-Telemedicine/TwinklingDataset-Display.git>. The code is intended to view RF signals and does not include standard CFM algorithms. It allows forming a conventional B-mode gray-scale image from the data, visualizing complex signals as graphs depending on both “fast” and “slow” time of the CFM mode, applying elements of spectral analysis. We hypothesize that this dataset could be an invaluable tool for researchers working in the field of Doppler signal processing, e.g., testing wall-filtering algorithms, training and validating artificial intelligence. It could also serve as a reference standard for the efficiency comparison of different CFM algorithms.

For more information on the dataset, its usage, the original software for working with the dataset, and in case of publishing original results obtained using the dataset, please, refer to the article [1].

Conclusion

In this work, we present a novel dataset of RF signals captured at the beamformer output of the ultrasonic medical scanner. The proposed dataset may have great practical value since it paves the way for new diagnostic tools [2], e.g., for detecting kidney stones and other objects associated with the twinkling artifact or experimenting with wall-filtering techniques.

References

- [1] Leonov DV, Reshetnikov RV, Kulberg NS, Nasibullina AA, Gromov AI. Doppler twinkling artifact observations: an open-access database of raw ultrasonic signals. *Digital Diagnostics*. 2021;2(3):261–276. <https://doi.org/10.17816/DD76511>
- [2] Leonov DV, Kulberg NS, Gromov AI, Morozov SP, Kim SY. Causes of ultrasound Doppler twinkling artifact. *Acoust Phys*. 2018;64(1):105–114. <https://doi.org/10.1134/S1063771018010128>

A study on construction of organ coverage avoidance method of acoustic shadows for automatic ultrasound diagnosis

M. Matsuyama¹, N. Koizumi¹, Y. Nishiyama¹, Y. Watanabe¹, J. Zhou¹, S. Yagasaki¹, T. Fujibayashi¹, M. Yamada¹, T. Ishikawa¹, R. Tsumura², K. Yoshinaka², N. Matsumoto³, H. Tsukihara⁴, K. Numata⁵

¹The University of Electro-Communications, Chofu, Japan ²National Institute of Advanced Industrial Science and Technology (AIST), Health and Medical Research Institute, Tsukuba, Japan ³Nihon University Hospital, Department of Medicine, Chiyoda-Ku, Japan ⁴The University of Tokyo, Department of Cardiothoracic Surgery, Bunkyo-Ku, Japan ⁵Yokohama City University Medical Center, Gastroenterological Center, Yokohama, Japan

Keywords Ultrasound image, Acoustic shadow, Robotic Ultrasound, HIFU therapy

Purpose

Ultrasound therapy has attracted attention in recent years as a non-invasive treatment method. The advantages of ultrasound treatment are that it can be performed without skin incision and that the lesion can be observed in real time. However, a drawback is that the accuracy of the examination depends on the skill of the physician. In addition, it is not always possible to visualize the entire area of interest due to acoustic shadowing caused by the reflection of sound waves from bones, calculi and gallstones [1]. Accordingly, the aim of the study was to develop a robotic system that avoids acoustic shadowing of the target organ, by extracting the organ and the region of acoustic shadows using deep learning and implementing the output information in a robotic ultrasound diagnostic system (RUDS).

Methods

The target organ was the kidney, and the source of acoustic shadow was the ribs. We propose a robot control method based on the area of overlap between the kidney and the acoustic shadow, and the positional relationship between the kidney and the center of gravity (Fig. 1). In this technique, the most recent image from RUDS is used as the input image. We then extract the kidney region and the shadow region using YOLACT [2], a model that can perform one-stage instant segmentation in real time while maintaining accuracy of detection. The kidney region is binarized first, and the coordinates of the center of gravity of the kidney are obtained from this area. In the case of overlap with acoustic shadow, the number of pixels on the left

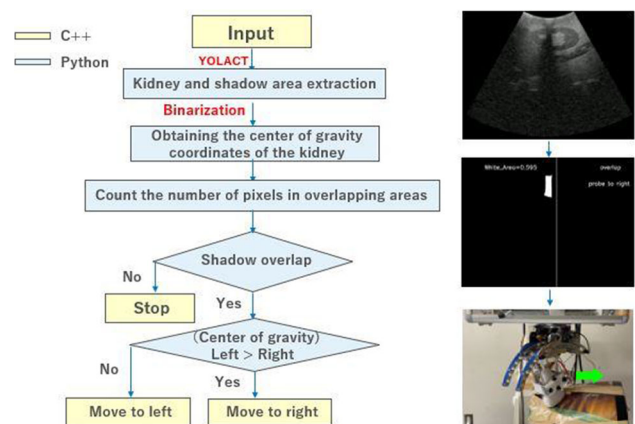


Fig. 1 Acoustic shadows avoidance control method

Table 1 Detection rate of kidney and shadows by YOLACT (mean Average Precision)

	all	50	55	60	65	70	75	80	85	90	95
Box	31.85	47.87	46.56	45.5	43.18	40.25	35.99	29.78	21.84	7.39	0.17
Mask	31.68	47.67	46.56	45.23	43.23	39.52	33.37	26.8	21.33	13.07	0.03

and right sides of the kidney is compared with respect to the center of gravity, and the RUDS probe is manipulated in the direction corresponding to the output. If there is no overlap, RUDS is stopped. This operation is repeated to avoid coverage by acoustic shadows. The relationship between the center of gravity coordinates of the kidney and the position of coverage can be used to efficiently manipulate the ultrasound probe.

To train YOLACT for kidney and shadow area segmentation, the input image size was 749×559 , the number of epochs was 800,000, and batch size was 8. The evaluation metric was mean Average Precision (mAP), and Intersection over Union (IoU) was used for detection evaluation. The data acquisition target was the experimental phantom ABDFAN, and 1280 training data and 320 test data were used. Ultrasound images that showed shadows on the left and right sides of the kidney and those showing shadows covering the kidney were used in the dataset. Acoustic shadow was created artificially. In addition, the avoidance direction of the robot motion coverage was tested in the head and tail directions relative to the phantom. The threshold of the segmentation result was set to 0.5.

Results

Table 1 lists the YOLACT detection rates for kidney and shadows. In implementing the model, it was difficult to maintain the imaging geometry of the kidney. The detection rates for the kidney and the shadow were unstable due to acoustic and non-contacted shadows other than the ribs, and were also affected by the weak real-time performance of the robot, which requires greater stability control.

Conclusion

We proposed implementation of robot manipulation using deep learning to avoid acoustic shadowing caused by ribs in ultrasound images. However, there were many false detections of ambiguous acoustic shadows, and it was necessary to deal with a variety of shapes. In the future, we would like to develop a multi-threaded program for image processing and robot manipulation to improve the real-time performance and stabilize the detection rate.

References

- [1] Dunmire B, Harper JD, Cunitz BW, Lee FC, Hsi, R, Liu Z, Bailey MR, Sorensen MD (2016) Use of the acoustic shadow width to determine kidney stone size with ultrasound. *The Journal of urology*, 195(1), 171–177.
- [2] Bolya D, Zhou C, Xiao F, Lee YJ (2019) Yolact ++: Better real-time instance segmentation. arXiv preprint arXiv:1912.06218.

Layout design of the drive mechanisms considering the delay of a surgical robot

R. Sekine¹, S. Miura¹

¹Tokyo Institute of Technology, Tokyo, Japan

Keywords Surgical robot, Robot design, Minimal invasive surgery, Master–slave

Purpose

Surgical robots are increasingly being used in the world. Robotic surgery requires less stress on the patient's body and is minimally

invasive. The da Vinci is one of the most popular surgical robots in use around the world. The da Vinci adopts a master–slave system that allows for remote operation. The master–slave system helps surgeons to operate the instruments intuitively. Many surgical robots, such as da Vinci, use a wire drive mechanism [1]. The wire drive mechanism is superior in terms of transmission efficiency and has a short time delay. Therefore, with surgical robots adopting wire drive mechanisms such as da Vinci, the surgeon is able to operate the robot in an environment with little or no time delay. However, when using wires, there are the problems of stretching and loosening. If a loose wire is used, the movement of the robot arm will be misaligned, leading to a decrease in surgical precision. In addition, since the wires tend to stretch and loosen easily, the wires need to be replaced each time, which increases the maintenance cost. Therefore, it is desirable to cut the maintenance cost of the surgical robot by reducing the number of wires used and by using other drive mechanisms. In addition to wires, the drive mechanisms used in surgical robots include flexible shafts, timing belts, and pneumatic actuators. However, they all have larger delays compared to wires. Hence, the layout design of drive mechanism is essential. When using a drive mechanism, the drive mechanism assignment affects the robot's movement. Thus, it is important to clarify where to place the drive mechanism on the robot's joint. Since industrial robots used in factories often perform predetermined and repetitive tasks, the trajectory of the robot is constant. Therefore, it is easy to determine the placement of the appropriate drive mechanism to some extent at the stage of designing the robot. On the other hand, in the case of surgical robots, unlike industrial robots, the trajectory of the robot is not constant because the operator's trajectory is not the same among trials and individuals. Therefore, surgical robots need to be designed based on human operation trajectories. The purpose of this research is to use a drive mechanism other than wires and to design a drive mechanism arrangement that satisfies the required positioning accuracy of the surgical robot.

Methods

In this research, we targeted pediatric operation, which required minimally invasive, using a surgical simulator that was able to perform tasks in a virtual space [2]. There were two manipulators in the simulator, and we controlled these manipulators to perform the task. The task was to do a needle application to the target markers in the simulator. First, a surgical simulator was used to operate the robot in an environment without the mechanical delay of the joints of the robot. A surgical robot in the virtual space was controlled using a master manipulator to perform a needle-application task that simulated suturing. At this time, the position and posture of the master manipulator were measured to obtain the operation trajectory without delay. Next, the movement trajectory of the robot was estimated by virtually considering the delay of the drive mechanism used for each joint of the surgical robot based on the acquired operation trajectory. The delay of the drive mechanism was modeled as a second-order delay system, and the Runge–Kutta method was used as the calculation process. We considered the use of four drive mechanisms: wire, flexible shaft, timing belt, and pneumatic actuator. The movement trajectory of the robot's tip was obtained for each possible combination when using each drive mechanism for each joint of the surgical robot, considering the virtual delay for each. Then, we found the deviation of the movement trajectory at the robot's tip between with and without the mechanical delay. We derived a combination of drive mechanisms to keep the required misalignment within the necessary positional accuracy. Considering the trade-off between the delay of the drive mechanism to be used and the maintenance cost, we sought a combination of drive mechanisms that satisfies both as much as possible.

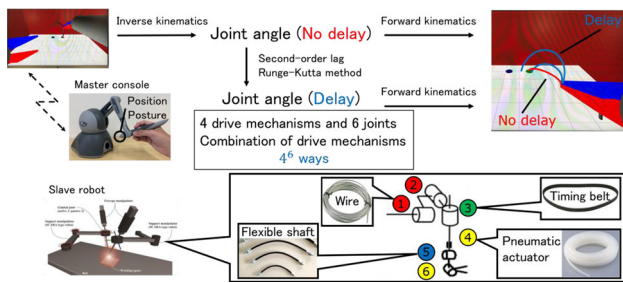


Fig. 1 Design method using surgical simulator. This figure shows the process using the simulator to determine the placement of the drive mechanism of the surgical robot

Results

As results of the derivation, four combinations of drive mechanisms were obtained. The four derived combinations were using wires and flexible shafts. There was no combination of drive mechanisms using timing belts or pneumatic actuators (Fig. 1).

Conclusion

In conclusion, we designed the arrangement of the drive mechanism of the surgical robot based on the human operation trajectory using the simulator.

References

[1] Shimachi S, Hirunyanitwatna S, Fujiwara Y, Hashimoto A, Hakozaki Y (2008) Adapter for contact force sensing of the da Vinci® robot. *Int J Med Robot Jun*;4(2):121–30
 [2] Miura S, Ohta R, Cao Y, Kawamura K, Kobayashi Y, Fujie M (2021) Using Operator Gaze Tracking to Design Wrist Mechanism for Surgical Robots. *IEEE Transactions on Human-Machine Systems* 51:376–383

Can skill-assessment suturing practice in medical school clinical practice increase the number of applicants for surgery?

S. Onishi¹, M. Murakami¹, K. Sugita¹, K. Yano¹, T. Harumatsu¹, K. Yamada¹, W. Yamada¹, M. Matsukubo¹, M. Muto¹, S. Ieiri¹
¹Kagoshima University, Department of Pediatric Surgery, Kagoshima, Japan

Keywords skill-evaluation type suturing, Shortage of Surgeons, simulator, surgical education

Purpose

Shortage of surgeon is the problem which must be resolved immediately. How to motivate medical students to be surgeon is big issue. In this study, we examined whether skill-evaluation type suturing practice can increase the number of fifth-year medical students who want to become surgeons.

Methods

Suturing practice was performed on clinical trainees who came around every two weeks, and the results were fed back on the spot. The results were fed back to the trainees on the spot. Before the

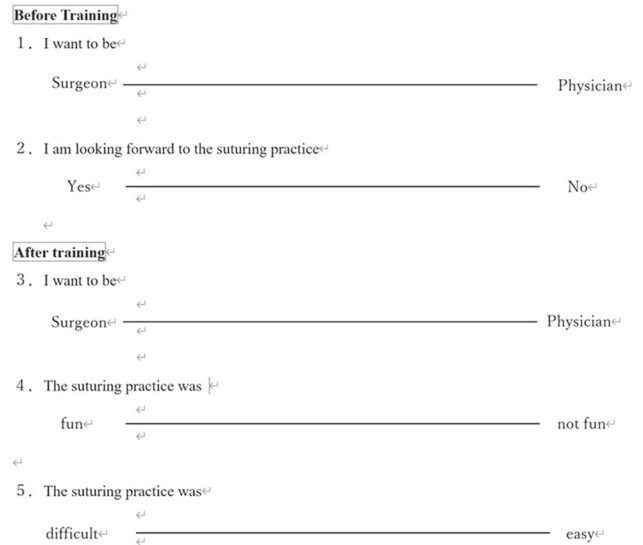


Fig. 1 The Visual Analogue Scale, with 0 being the desire for internal medicine and 100 being the desire for surgery

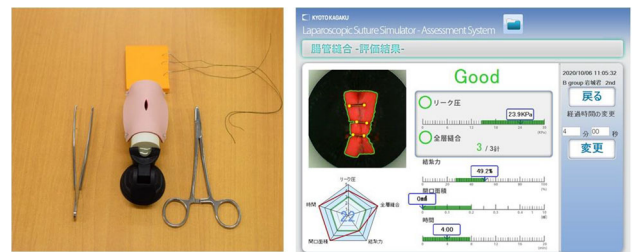


Fig. 2 A-LAP mini (KyotoKagaku), a simulator for objective evaluation of suturing skills

training, the trainees were surveyed on their level of desire for surgery using the Visual Analogue Scale, with 0 being the desire for internal medicine and 100 being the desire for surgery (Fig. 1). A-LAP mini (KyotoKagaku), a simulator for objective evaluation of suturing skills, was used (Fig. 2). The student trained suturing with needle holder and 3–0 silk suturing. After 30 min training, the students sutured 3 stiches on A-LAP mini. The countable stiches, leak area and pressure of the stiches were measured by the simulator. After the training, VAS was measured again.

Results

Fifty-five students were included in the study. The mean level of aspiration for surgery before suturing practice was 56.8 ± 27.1 . After suturing practice, the level of aspiration increased to 62.5 ± 26.3 . Of the 22 students who were “Aspirant to Physician” with a surgical aspiration level of 50 or less before the suturing practice (Table 1). In 22 students who aspirant to physician before training, 6 students increased their surgical aspiration level to 50 or more after the suturing practice. We compared these two groups, “Aspirant to Physician After training(N = 16)” and “Aspirant to Surgeon After training(N = 6)”. However, there were no significant difference between two groups.

Table 1 Surgical aspiration level of 50 or less before the suturing practice

	Aspirant-to-Physician-before-training [†] (N = 22) \bar{x}	Aspirant-to-Surgeon-before-training [†] (N = 33) \bar{x}
Training-was-fun- (VAS) \bar{x}	86,95 \bar{x}	92,55 \bar{x}
Training-was-difficult (VAS) \bar{x}	73,59 \bar{x}	70,27 \bar{x}
Total-score \bar{x}	18,32 \bar{x}	19,58 \bar{x}
Time(min) \bar{x}	3,77 \bar{x}	4,18 \bar{x}
Leak-area(mm2) \bar{x}	4 \bar{x}	4,21 \bar{x}

Conclusion

The results show that suturing practice with skill evaluation can help stimulate students' motivation to become surgeons. We could not detect the factors which motivate students to be surgeon. Further studies are needed to construct education procedure for increasing the applicant for surgeon.

Towards surgical navigation with electromagnetic tracking for robotic surgeries

L. Aguilera Saiz¹, W. ten Bolscher¹, H. Groen¹, W. Heerink¹, M. Hiep¹, T. Ruers¹

¹Netherlands Cancer Institute—Antoni van Leeuwenhoek, Surgical Oncology, Amsterdam, Netherlands

Keywords surgical navigation, electromagnetic tracking, robotic surgeries, optical tracking

Purpose

Minimally invasive procedures, like laparoscopic and robotic surgery, are progressively replacing open surgeries [1]. However, due to less overview and small field of view (FOV), the orientation and localization of both malignant tissue and healthy critical structures can be challenging. To overcome this, image-guided surgery (IGS) and surgical navigation could be used, providing the surgeon in real-time detailed 3D information of the anatomical structures. In IGS, the instruments can be tracked using different methods. Optical tracking systems (OTS) have been used in robotic surgical navigations for its high accuracy of 0.5 mm. However, the main disadvantage of these

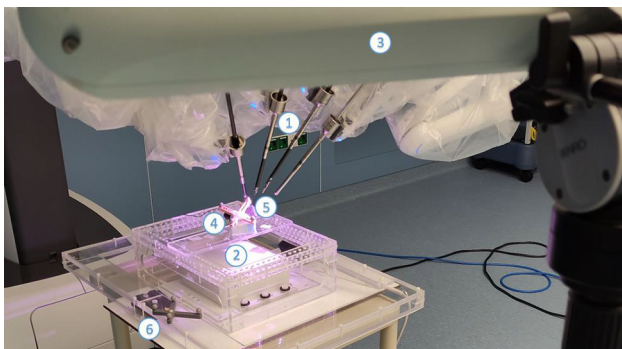


Fig. 1 Setup for accuracy measurements using da Vinci Xi surgical system (1), the Aurora NDI Planar field generator (2) and the Polaris NDI OTS (3). An electromagnetic 6DOF-sensor (4) with optical spheres (5) was moved across a grid to assess the EMT accuracy of the robot in surgical position. An optical tracker (6) attached on the frame to the field generator enables the Hand-Eye calibration

systems is the requirement of a continuous line of sight, which is not always feasible. Therefore, we propose using electromagnetic tracking (EMT) for IGS in robotic surgeries. For this purpose, we conducted a phantom study to evaluate the accuracy of different combinations of robots and EMT field generators (FG).

Methods

For this study, the da Vinci Si and Xi surgical systems (Intuitive Surgical, Sunnyvale, CA, USA) and the Aurora NDI Table Top (TT) and Planar EMT system (Northern Digital Inc., Waterloo, Canada) were used. As such, four different setup combinations were tested: Si-TT, Si-Planar, Xi-TT and Xi-Planar. For accuracy assessment, the NDI Polaris Spectra (Northern Digital Inc., Waterloo, Canada) OTS was used as reference. A 6-degrees of freedom (DOF) electromagnetic NDI reference sensor was attached to an optical NDI sensor and calibrated using the Hand-Eye method. In order to measure inside the FOV of the EMT systems, a $40 \times 32 \times 10$ cm (x, y, z) stackable box was placed in the center of the FOV (see Fig. 1). To avoid any other interference of the accuracy except for the robot, a non-ferromagnetic table was used for the setup. The robot was positioned in surgical position above the box, which replicated the location of the instruments during surgery. Five different locations (corners and center) were measured at 3 different levels, 150 consecutive measurements in each location, resulting a total of 2250 measurements within the FOV. To quantify the accuracy of the position, the Euclidean distances between the optical and the electromagnetic sensor were calculated. For the orientation, the quaternions were compared. The jitter was calculated by determining the standard deviation of the 150 measurements in each position and then calculating the root mean square.

Results

The highest position accuracy was achieved with the Si-Planar combination, with a median accuracy of 0.7 mm (IQR: 0.6–0.9) (see Table 1). Lowest position accuracy was the Xi-Planar combination, with a median accuracy error of 1.8 mm (IQR: 0.9–1.9). For the orientation, higher accuracy was obtained with the Xi-TT combination, a median accuracy error of 0.3° (IQR: 0.2–0.4); lowest was with Si-TT, median accuracy error of 1.1° (IQR: 1.0–1.2). The overall jitter error was lower than 0.05 mm and 0.1° for all combinations. Hand-Eye calibration accuracies for both TT and Planar FGs were 0.6 ± 0.1 mm.

Conclusion

Experiments for accuracy measurements of the electromagnetic field showed that the accuracy of the EMT is dependent on both the type of robot and FG. The highest position accuracy—the Si-Planar combination—is in the order of the Hand-Eye calibration. This suggests that there is not a major distortion of the robot into the electromagnetic field for the Si-Planar combination. Contrarily, lowest accuracy was achieved with the Xi-Planar combination. This indicates that the Planar FG is more sensible to the type of robot. Accuracies were around 1 mm with the TT FG for both robots, which is clinically acceptable. The Xi-TT combination has the highest orientation accuracy and the Si-TT was the least accurate. The total accuracy is determined by a number of factors, such as the influence of robotic instruments or distortion from devices in the operating room. These phantom results show that integrating EMT with da Vinci robots is feasible and accurate, the next step is the clinical evaluation to assess the value of IGS during robotic surgeries.

Table 1 Position and orientation errors from accuracy measurements

Combinati on	Euclidean distance error (mm)				Orientation error ($^\circ$)			
	Median	Q1	Q3	IQR	Median	Q1	Q3	IQR
Si-TT	0.9	0.7	1.3	0.6	1.1	1.0	1.2	0.2
Xi-TT	1.2	0.9	1.3	0.4	0.3	0.2	0.4	0.2
Si-Planar	0.7	0.6	0.9	0.3	0.3	0.2	0.5	0.3
Xi-Planar	1.8	0.9	1.9	1.0	0.4	0.3	0.5	0.2

References

- [1] Joseph JV, Arya M, Patel HR (2005) Robotic surgery: The coming of a new era in surgical innovation. *Expert Rev. Anticancer Ther.* 5(1):7–9

Intraoperative holographic image-guided lateral pelvic lymph node dissection in a transanal approach for rectal cancer

T. Yoshimoto¹, T. Tokunaga¹, Y. Saito¹, H. Kashiwara¹, K. Yoshikawa¹, T. Nakao¹, M. Nishi¹, C. Taksu¹, Y. Wada¹, S. Yamashita¹, Y. Iwakawa¹, M. Shimada¹

¹Tokushima University hospital, Digestive and Transplant Surgery, Tokushima, Japan

Keywords transanal total mesorectal exc, hologram, mixed reality, lateral pelvic lymph node diss

Purpose

We have introduced the transanal approach for lateral pelvic lymph node dissection of lower rectal cancer. In the transanal approach, it is especially important to understand the spatial anatomy, such as the localization of blood vessels and enlarged lymph nodes before surgery. We will demonstrate the surgical technique of lateral pelvic lymph node dissection using the transanal approach and report on the usefulness of simulation using virtual reality (VR) and mixed reality (MR), [1, 2].

Methods

Hologram (3D image): Polygon (stereolithography) files were created using pre-operative MDCT and MRI T2 data and exported from SYNAPSE VINCENT, then uploaded into the Holoeyes MD system (Holoeyes Inc., Tokyo, Japan). After uploading the data, the three-dimensional image was automatically converted into a case-specific hologram. The hologram was then installed into the head mount display, HoloLens (Microsoft Corporation, Redmond, WA). Preoperative simulation was performed using the HoloLens. In addition, the surgeons and assistants wore the HoloLens when they performed transanal total mesorectal excision and holography was used to confirm the dissection lines from multiple angles over the sterilized surgical field during the operation.

Surgical technique: After resection of the rectum, the lateral side of No. 283 lymph node was dissected along the levator ani muscle to identify the obturator vessels and connected to the dissecting layer on the abdominal side. The most caudal side of No. 283 lymph node was dissected to expose the coccygeal muscle. The inferior vesical vessels were identified and connected with the detached layer of the inferior bladder fascia from the abdominal cavity. Depending on the case, the proximal side of the inferior vesical vessels and obturator vessels are transperineally dissected.

Results

Case: Right-sided lymph node dissection using a transanal approach

We performed a laparoscopic low anterior resection of the rectum and right lateral lymph node dissection for Rb rectal cancer with TaTME in two teams. The main tumor was T2, however preoperative CT images showed an enlarged lymph node with a short diameter of 8 mm in the right No. 263d lymph node, and the boundary with the internal pudendal artery was unclear, suggesting the need for a combined resection. Simulation from the abdominal cavity: Regarding the blood vessels, it was confirmed that the umbilical artery and

inferior vesical artery branched from the internal iliac artery. The enlarged lymph node was located on the ventral side of the internal iliac artery, and the obturator vein and inferior vesical vein were found to branch dorsally. The proximal side of the internal pudendal artery was to be dissected from the ventral side, and the obturator vein was to be dissected from the anal side. Simulation from the transanal view: (1) Dissect the lateral side of No. 283 lymph node along the levator ani muscle, dissect the deepest No. 283 lymph nodes, and confirm the coccygeal muscle and internal pudendal artery. (2) Dissect the inferior vesical vessels and proceed to cranial side, identify the branch of the obturator vein, and dissect transanally. (3) A swollen lymph node was identified on the cranial side, and the peripheral side of the internal pudendal artery was dissected transanally.

Conclusion

Intraoperative holograms have the potential to become a new next-generation surgical support tool for use in spatial awareness and the sharing of information between surgeons.

References

- [1] Atallah S, Martin-Perez B, Larach S (2015) Image-guided real-time navigation for transanal total mesorectal excision: a pilot study. *Tech Coloproctol* 19: 679–84
- [2] Saito Y, Sugimoto M, Imura S, Morine Y, Ikemoto T, Iwahashi S, Yamada S, Shimada M (2020) Intraoperative 3D Hologram Support With Mixed Reality Techniques in Liver Surgery. *Ann Surg* 271(1): e4–e7

Metaverse and Extended reality (XR:VR/AR/MR) for endoscopic robotic tele-surgery navigation, simulation, and holographic guiding

M. Sugimoto¹

¹Teikyo University Okinaga Research Institute, Innovation Lab, Tokyo, Japan

Keywords metaverse, Extended reality, augmented reality, virtual reality

Purpose

In the medical field of coronary heart disease, non-contact, remote environments such as online medical treatment are required. Remote robotic surgery and teleconferencing are also attracting attention in the surgical field. Preoperative images such as CT/MRI, hybrid operating room and intraoperative imaging images are represented only in a flat plane when viewed on a flat monitor.

Furthermore, it is difficult to recognize the correct spatial relationship between the image and the actual surgical field and surgical instruments. As a technology to solve this problem, we developed a programmed medical device (Holoeyes MD) that utilizes Metaverse, VR virtual reality, AR augmented reality, and MR mixed reality (together, Extended reality: XR) technologies [1, 2], and commercialized it as a cloud service, Fig. 1.

Methods

After extracting the shape data of each organ from CT/MRI as polygon data, we were able to automatically generate an XR application that can present a three-dimensional organ model in real space in about five minutes.

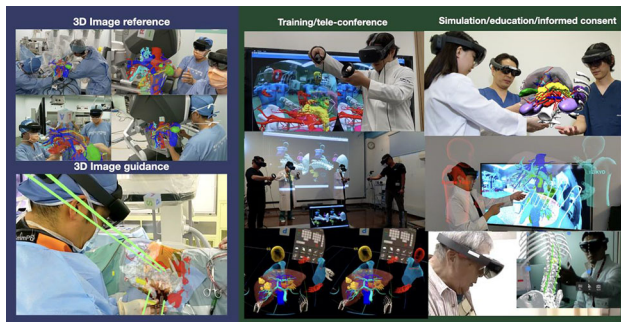


Fig. 1 Holoeyes MD

By using a head-mounted display equipped with a position sensor and transmissive wearable glasses, this application could be viewed on the sterile surgical field as if the patient were immersed in the patient's body and the organs were floating in the air. Position sensor input from infrared sensors and a 3D camera allowed gestural manipulation of the holographic polygon model display while wearing the sterile goggles.

By sharing the position of each XR goggle via 5G or WiFi, multiple people could experience the same space at the same time. In the Metaverse space, the hand and head movements of the procedure as well as the organ shape were recorded in chronological order and could be spatially traced and relived later.

The guide lines of conventional navigation systems can be superimposed on the surgical field as a hologram.

Results

This virtual shared space is called Metaverse, and is increasingly used in the surgical field as an alternative to the real world with its superior physicality, spatiality, immersion, and interactivity. This medical practice with the coexistence of presence and presence makes it easier to record and experience the skills of skilled practitioners as physical movements in the XR space, which significantly increases the efficiency of practice and contributes to the formalization of tacit medical knowledge, which is non-verbal information.

These technologies has already been used in many medical facilities for surgical planning, simulation, treatment support, medical education and training to improve techniques and avoid misidentification.

Conclusion

These are widely used as commercially available, inexpensive devices and online services in more than 200 facilities in Japan and abroad. It is already being used clinically for surgical navigation, simulation, holographic guidance, and remote online medical care in open and endoscopic surgeries, as well as remote robotic surgeries.

Our online service is inexpensive and easy to use with commercially available XR devices, suggesting that it may become a staple for surgical support and remote online medical care.

References

- [1] Sugimoto M (2021) Cloud XR (Extended Reality: Virtual Reality·Augmented Reality·Mixed Reality) and 5G networks for holographic medical image-guided surgery and telemedicine. *Multidisciplinary Computational Anatomy—Principles and Clinical Application of MCA-based Medicine*. pp 381–338,

2021. Editor: Hashizume M. Springer, Singapore. 2021. https://doi.org/10.1007/978-981-16-4325-5_52

- [2] Sugimoto M (2020) Extended Reality (XR:VR/AR/MR), 3D Printing, Holography, A.I., Radiomics, and Online VR Tele-Medicine for Precision Surgery. *Surgery and Operating Room Innovation*. pp 65–70. Editor: Takenoshita S, Yasuhara H. Springer, Singapore. 2020. https://doi.org/10.1007/978-981-15-8979-9_7

The usefulness of fusion three-dimensional computer graphics for surgical simulation of glioma

T. Koike¹, T. Kin^{1,2}, N. Shono¹, Y. Furuta¹, R. Niwa¹, K. Sato¹, T. Uchida¹, Y. Takeda¹, S. Kiyofuji¹, H. Takami¹, S. Takayanagi¹, S. Tanaka¹, N. Saito¹

¹The University of Tokyo, Department of Neurosurgery, Bunkyo-ku, Tokyo, Japan ²The University of Tokyo, Department of Medical Information Engineering, Graduate School of Medicine, Bunkyo-ku, Tokyo, Japan

Keywords Computer graphics, Deep learning, Glioma, Surgical simulation

Purpose

Glioma is often difficult to differentiate from surrounding anatomical structures in the actual surgical field because of the invasive nature of the tumor. Therefore, it is important for surgical simulation to clearly visualize and understand the spatial location of the tumor in relation to surrounding anatomical structures such as blood vessels, brain parenchyma, and white matter fibers. At our facility, we use fusion three-dimensional computer graphics (fusion 3DCG) constructed from medical image data from multiple modalities to perform surgical simulations. In this study, we report on the technical accuracy and clinical usefulness of surgical simulation using fusion 3DCG.

Methods

Fifteen cases of glioma were included. The detailed characteristics of the participants was described in Table 1. The fusion 3DCG was constructed based on computed tomography, magnetic resonance imaging, three-dimensional rotational angiography, and positron emission tomography acquired preoperatively in each case. The image processing software used was GRID (Kompath Inc., Tokyo, Japan) on a Windows personal computer. For the registration of each medical image data, we used the normalized mutual information based on our originally developed initial set value method. Some organizations used segmentation with deep learning techniques to construct their models. Tumors, arteries, and veins were extracted using multiple modalities and the multi-threshold method was used to extract microtissues. White matter fibers were modeled in 3D by diffusion tensor tractography analysis. Using the constructed fusion 3DCG, we performed virtual reality surgical simulation with virtual manipulation of craniotomy, deformation of brain parenchyma, and fence post placement. In addition, the constructed fusion 3DCG was output in Wavefront OBJ format and input into the tablet application LIVRET (Kompath Inc., Tokyo, Japan), which the surgeon can manipulate and refer to during surgery. The surgeon can then manipulate and refer to the data during the operation to verify the consistency with the fusion 3DCG intraoperative findings and the usefulness for surgical simulation.

Table 1 Characteristics of Participants

Case	Age	Sex	Diagnosis	State	Required Time (min.)
1	79	M	Glioblastoma	Primary	173
2	31	F	Anaplastic astrocytoma	Primary	154
3	41	M	Anaplastic oligodendroglioma	Primary	165
4	81	M	Glioblastoma	Primary	182
5	34	M	Anaplastic astrocytoma	Primary	179
6	46	F	Glioblastoma	Recurrence	190
7	57	F	Glioblastoma	Primary	166
8	62	M	Glioblastoma	Primary	183
9	30	F	Oligodendroglioma	Primary	118
10	32	M	Oligodendroglioma	Primary	134
11	53	F	Glioblastoma	Recurrence	193
12	32	F	Anaplastic oligodendroglioma	Primary	169
13	51	M	Oligodendroglioma	Recurrence	175
14	58	F	Glioblastoma	Primary	143
15	36	F	Anaplastic astrocytoma	Recurrence	188

Results

In all cases, it was possible to construct a fusion 3DCG using the proposed method. There were no adverse events attributable to image preparation. The surgery was performed according to the surgical plan using fusion 3DCG. The registration error between the medical image data was less than 1 mm. The proposed segmentation technique did not degrade the spatial resolution compared to 2D images, and tissues as small as 1 mm were visualized in 3D with clear boundaries. These techniques made it possible to confirm the spatial location of the anatomical structures, and it was confirmed that the spatial consistency between the fusion 3DCG and the actual surgical field was achieved. Since craniotomy and tissue deformation by virtual reality manipulation emphasizes rapidity rather than physical consistency, real-time virtual manipulation was possible and contributed to the efficiency of the surgical simulation. Using an application installed on a tablet computer, the surgeon directly manipulated these fusion 3DCGs during the surgery, allowing the surgeon to confirm the spatial relationship between the tumor and anatomical structures such as deep white matter fibers and arteries even during tumor removal

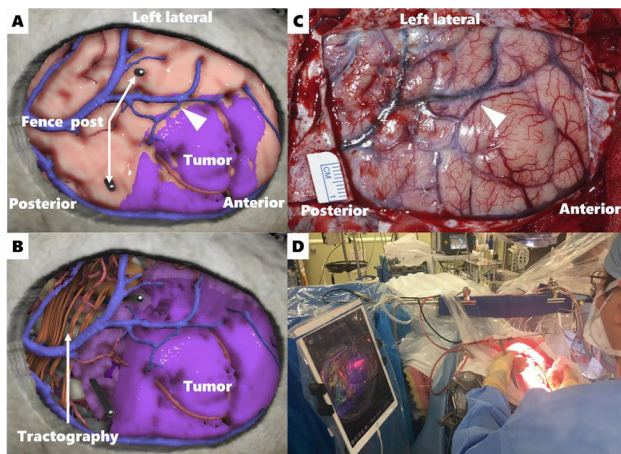


Fig. 1 The illustrative case. **A** The three-dimensional computer graphics (3DCG). The craniotomy was performed, and two fence posts were placed. **B** The 3DCG with the cerebrum in hidden state. The white matter fibers which were delineated by diffusion tensor tractography was located posterior to the tumor. **C** The actual operative field. The white arrowheads in **A** and **C** show the tiny veins on the brain surface, which were useful in determining the tumor boundary on the brain surface. **D** The 3DCG used for the preoperative study was used as the reference on the tablet using the LIVRET application

(Fig. 1). The fence post was particularly useful as a Merkmal between the tumor and the normal anatomical tissue. In summary, direct comparison of the preoperative simulation and the surgical operation contributed to the improvement of the accuracy of the surgical simulation. There were no significant complications associated with the surgery. On the other hand, the average creation time of the proposed fusion 3DCG was 167.5 min, and the issue was to shorten the required image processing time.

Conclusion

The fusion 3DCG constructed by the proposed method is a useful surgical support tool from preoperative to intraoperative stages, because it can visualize a large amount of various preoperative medical image data as a single 3DCG with high accuracy, and can perform virtual operations such as craniotomy, deformation, and device implantation.

Quantitative assessment of the forearm rotation axis

J. Oonk¹, I. Dobbe¹, G. Streekstra¹, G. Strijkers¹

¹Amsterdam UMC, Biomedical engineering & physics, Amsterdam, Netherlands

Keywords Rotation Axis, Circle fit, 4D-CT, Forearm

Purpose

The purpose of this study was to develop a method based on circle-fitting to find the axis of rotation (AoR) of the forearm. Additionally, the influence of the amount of joint positions along the range of motion (ROM) used for the circle-fit was analysed.

Methods

Both arms of 10 volunteers without a history of trauma or injury to the forearm were subjected to three CT scans: A spiral CT scan of the entire forearm for segmentation purposes, and two 4D-CT scans of the distal radio-ulnar joint (DRUJ) and proximal radio-ulnar joint (PRUJ) during forearm rotation. 3D-models of the radius and ulna were extracted from the spiral CT images by segmentation. Subsequently, the 3D models were registered to the 33 frames of each 4D-CT scan.

The coordinates of the distal radius centroid and the radial tuberosity (proximal bony landmark) in each target frame were extracted. Subsequently a circle-fit algorithm was applied to both the proximal and distal set of 3D-points. The initial step in this circle fit determined the fitting plane, which is the plane with the smallest root mean squared orthonormal distance to all 3D-points. Subsequently, the 3D-points were projected onto the fitting plane, converting the fit from 3 to 2D. Using the coordinates of the 2D-points and the equation for a circle, the centre coordinate and the radius of the circle could be calculated. Finally, the AoR was found by connecting the distal- and proximal circle fit centre coordinates by a straight line.

The methodological error of the circle fit was quantified by calculating the root mean squared distance between each 2D-point coordinate and the fitted circle. The variation in circle fit centre location when using less than 33 fit points was also evaluated. This was done to see if a scan with less than 33 target frames was also adequate to find the centre of rotation. A lower amount of target frames is desirable as it reduces the radiation dosage. Moreover, less frames makes the technique a viable option for CT-scanners without 4D-CT capabilities as the single 4D-CT scan could be replaced by a small series of static 3D-CT scans of the joint in different positions along the ROM. To evaluate the effect of the amount of target frames used in the circle-fit, a bootstrap algorithm was utilized to select 1000 random combinations of 3-, 5-, 10- and 20 distal centroid- or proximal bony landmark coordinates. A circle-fit was then applied to every combination and the resulting circle-fit centre was plotted. The mean

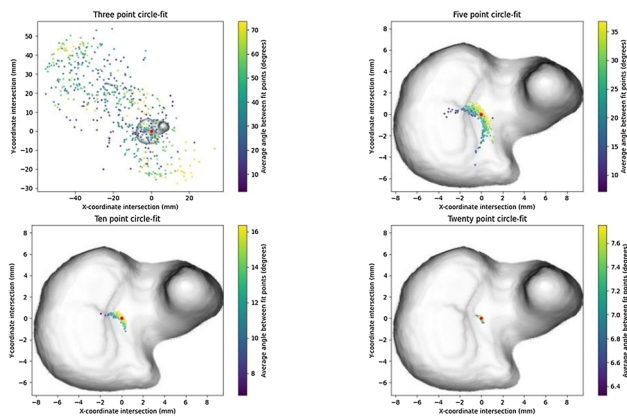


Fig. 1 The variation in the distal circle-fit centre of rotation based on the amount of points used in the circle fit. The 33-point circle fit is indicated by the red dot. The colour of the other circle-fit centre dots indicates the average angle between the fit points. A larger angle is beneficial as a bigger portion of the ROM is evaluated by the circle fit

absolute distance along the x- and y-axis between the 1000 circle fit centres and the 33-point circle fit centre was calculated to quantify the precision.

Results

The location of the AoR was similar for each case. The AoR intersected the ulnar fovea distally and the centre of the radial head proximally. The average error was 0.15 mm and 0.12 mm for the distal- and proximal circle fit respectively. The relation of the amount of points used in the circle fit and the subsequent variation of the distal- and proximal centre of rotation is shown in Fig. 1. The mean absolute distance ($[x, y]$ mm) of the 1000 circle fit centres to the 33-points fit centre were: [11.4, 10.7], [0.37, 0.41], [0.16, 0.18] and [0.06, 0.08] for the 3-, 5-, 10- and 20-points fit respectively.

Conclusion

Given the small methodological error, the circle-fit method is an accurate way to determine the AoR of the forearm. However, it should be noted that a sufficient amount of points is needed as significant variation in the centre of rotation can be seen when using a small amount of points.

Novel measurement method of residual proteins in cleanliness evaluation of reusable and reprocessed medical devices

M. Uematsu¹, Y. Miyamoto¹, H. Sakoda¹, Y. Okamoto¹, R. Nakaoka¹, Y. Haishima¹

¹National Institute of Health Sciences, Division of Medical Devices, Kawasaki-shi, Japan

Keywords cleaning evaluation, residual protein, thermal coagulation, medical devices

Purpose

The reprocessing of single-use devices (SUDs) has been widely accepted and regulated worldwide to promote the effective utilization of resources and reduce medical waste and expenses. To safely reprocess SUDs, medical device manufacturers must evaluate the cleanliness of them after washing. This evaluation is done using methods intended for reusable medical devices, since there are no specific standards for reprocessed SUDs. Accurately measuring

residual proteins is one of the major targets of cleanliness evaluation. Acceptable criteria for protein detection in reusable medical devices have been established in AAMI TIR 30, German and Japanese guidelines, and ISO 15883–5 [1].

The recovery rate of residual proteins is greatly influenced by the extraction conditions (temperature, time, and solvent) and the protein assay reagents. Although several conditions are recommended in the aforementioned standards and guidelines, there is no standardized test method. According to WHO guidelines [2], the amounts of residual proteins are evaluated after cleaning below 60 °C, because the recovery rate can decrease at higher temperatures due to thermal coagulation. In addition, high-temperature washing at 95 °C after a low-temperature wash and protein quantification is usually performed to enhance cleanliness.

In this study, we revealed the effect of thermal coagulation of proteins on their recovery rate using samples contaminated with a human blood substitute, which were heat-treated under various conditions. In addition, we developed a novel method that efficiently and reproducibly recovered proteins at greater amounts than the conventional evaluation procedures recommended by the guidelines.

Methods

Equal volumes of citrated fresh porcine blood and 0.025 mol/L CaCl₂ solution were mixed to restart coagulation. The preparation was spread thinly and widely on a stainless-steel plate and dried for 16 h at room temperature. The pseudo-blood contaminated samples were heat-treated under dry conditions using a dry oven, or under wet conditions using a programmable autoclave at temperatures ranging from 45–240 °C. Milli-Q water and 1% sodium dodecyl sulfate (SDS; pH 11.0) were used as the solvent for protein extraction. In addition, a modified sample buffer (SB) of SDS–polyacrylamide gel electrophoresis, containing 1% SDS, 10 mM Tris(2-carboxyethyl)phosphine (TCEP), and 10 mM HEPES (pH 7.0), was used as a novel alternative solvent in order to be applicable for the colorimetric assay. Extracted proteins were quantified using a BCA reagent kit (Thermo Fisher Scientific Inc., Waltham, MA). When the SB solution was used, iodoacetamide (IAM) was added to the extracts and pre-incubated at 37 °C for 20 min before the measurement. All measurements were repeated three times.

Results

The protein recovery rates from the dry-heated samples did not change significantly, regardless of the conditions used, except for heat treatment at 240 °C for 10 min. In contrast, the recovery rate by Milli-Q water from the samples treated under wet conditions $36.4 \pm 6.0\%$ (45 °C for 1 h, dissolving mode), $6.18 \pm 0.53\%$ (60 °C for 30 min, dissolving mode), and $0.422 \pm 0.055\%$ (95 °C for 10 min, dissolving mode) was significantly lower than those under dry conditions. It decreased in proportion to the heat treatment intensity. Although the reduction in the recovery rate was partially improved compared with Milli-Q water extraction, similar results were obtained by extraction using the 1% SDS (pH 11.0) solution. These results showed that under wet conditions the thermal coagulation of proteins was accelerated by water molecules mediated by heat conduction, which supports the recommendation of the WHO guidelines. As shown in Fig. 1, this phenomenon was not observed and excellent protein recovery was obtained without significant effect up to 95 °C under wet conditions. This finding originates from the reductive cleavage of the disulfide bonds of protein molecules by TCEP, which destroys the higher-order structure and promotes SDS binding. The influence of thiol groups present in reduced proteins could be completely quenched by pre-treating the samples with IAM. Standard and micro BCA reagents with excellent reproducibility and robustness are now available for protein quantification, even for samples dissolved in a solvent containing reducing agents.

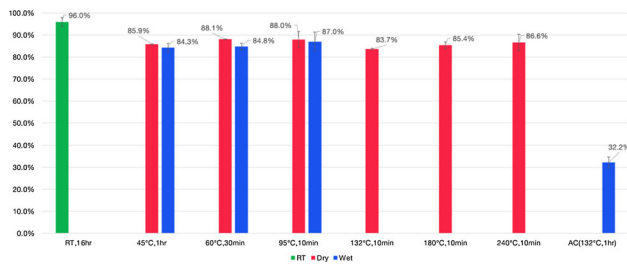


Fig. 1 Protein recovery rates from pseudo-blood contaminated samples by extraction with a sample buffer

Conclusion

In this study, we developed a novel method to recover residual proteins, including even when they were bound strongly to the surface of medical devices by thermal coagulation. It was more efficient than the conventional ones by combining extraction with a new solvent consisting of 1% SDS, 1% TCEP, and 10 mM HEPES (pH 7.0), pretreatment with IAM, and quantification with BCA reagent. In order to verify the reproducibility and robustness of this method, we plan to conduct a domestic round-robin test.

References

- [1] ISO 15883–5:2021 (2021) Washer-disinfectors—Part 5: Performance requirements and test method criteria for demonstrating cleaning efficacy.
- [2] World Health Organization and Pan American Health Organization (2016) Decontamination and reprocessing of medical devices for health-care facilities, Geneva: WHO.

Malignancy index using intraoperative flow cytometry is a valuable prognostic factor for glioblastoma treated with standard chemo-radiotherapy

T. Saito^{1,2}, Y. Muragaki², M. Nitta¹, S. Tsuzuki¹, A. Fukui¹, S. Koriyama¹, T. Kawamata¹

¹Tokyo Women's Medical University, Department of Neurosurgery, Tokyo, Japan ²Tokyo Women's Medical University, Faculty of Advanced Techno-Surgery, Tokyo, Japan

Keywords flow cytometry, glioblastoma, temozolomide, radiotherapy

Purpose

Intraoperative prediction of radiochemosensitivity is desirable for improving the clinical management of glioblastoma (GBM) patients. We have previously developed an original technique for intraoperative flow cytometry (iFC) and defined a malignancy index (MI). To determine whether MI correlates with prognosis in GBM patients who underwent the standard treatment protocol of radiotherapy and temozolomide administration, [1, 2].

Methods

This study included 102 patients with GBM that had been newly diagnosed between 2010 and 2015 who underwent our iFC analysis and received the standard treatment protocol. We evaluated MI values in each patient, then statistically analyzed the relationship between MI and prognosis using survival analyses including other clinicopathological factors (age, sex, Karnofsky performance status (KPS), extent of resection, second-line bevacizumab, O6-methylguanine-

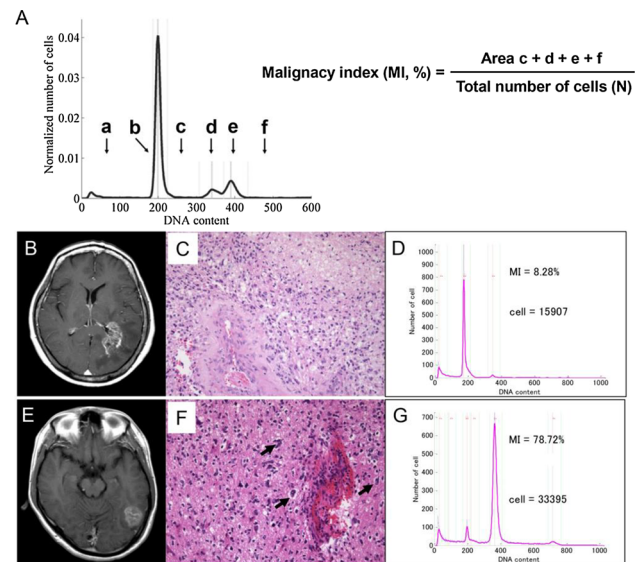


Fig. 1 A MI was determined as the ratio of the number of cells with greater-than-normal DNA content to the total number of cells (N); that is: $MI = (Area\ c + Area\ d + Area\ e + Area\ f) / N$. **B–D**, A representative case of glioblastoma (GBM) with low malignancy index (MI). **E–G** GBM case with high MI

DNA methyltransferase (MGMT) status, MIB-1 labeling index, and mutation of the isocitrate dehydrogenase 1 gene (*IDH1*).

Results

Log-rank test revealed age, KPS, extent of resection, MGMT status, *IDH1* mutation, and high MI ($\geq 26.3\%$) correlated significantly with overall survival. Multivariate analysis using Cox regression modeling identified MI as the most significant prognostic factor (hazard ratio = 2.246; 95% confidence interval = 1.347–3.800; $P = .0019$). MI showed strong correlation with *IDH1* mutation status in chi-square test ($P = .0023$). In addition, log-rank test revealed that MI affects overall survival more strongly in patients with *IDH1* wildtype than those with *IDH1* mutant.

Conclusion

MI from an iFC study can help predict the prognosis in patients with GBM who receive the standard treatment protocol. Survival may be related to sensitivity to radio-chemotherapy, Fig. 1.

References

- (1) Shioyama T, Muragaki Y, Maruyama T, Komori T, Iseki H. Intraoperative flow cytometry analysis of glioma tissue for rapid determination of tumor presence and its histopathological grade: clinical article. *Journal of neurosurgery*. Jun 2013;118(6):1232–1238.
- (2) Saito T, Hama S, Izumi H, et al. Centrosome amplification induced by survivin suppression enhances both chromosome instability and radiosensitivity in glioma cells. *British journal of cancer*. Jan 29 2008;98(2):345–355.

How diffusion tensor imaging enables direct targeting for focused ultrasound surgery

H. Hori¹, T. Hori¹, K. Abe¹, T. Taira¹

¹Moriyama Neurological Center Hospital, FUS center, Edogawa, Japan

Keywords TcMRgFUS, DTI, VIM, analysis software

Purpose

A targeting method of the Ventral Intermediate Nucleus (VIM) is used an indirect method in most facilities that have introduced the transcranial MR-guided focused ultrasound surgery (TcMRgFUS). There are various measurement methods for the indirect targeting method [1, 2], in general, the target as the VIM is located 6 mm forward from the posterior commissure (PC) and 16 mm laterally from the midline of the third ventricle. However, a sufficient therapeutic effect may not be obtained in the location of the indirect target during the TcMRgFUS procedure, due to the location of the VIM varied for each patient. In that case, the sonicate location of the focused ultrasound is slightly moved. Increasing the number of lesions is not desirable, because it may induce an unexpected adverse event due to the lesion and edema to expand. A fiber tractography analyzing the diffusion tensor imaging (DTI), which is being actively studied as a direct targeting method, expected accurately classify nuclei in the thalamus and may be a useful the targeting method in the TcMRgFUS. However, even if the DTI analysis is performed with the image processing devices (AW, ZIO, AZE, etc.) which are generally applied in most hospitals, it is undesirable usable for treatment planning due to the sufficient fibers are not visualized. An image processing software which named DSI studio is reported to useful for analysis DTI. This software is free and has been attracting attention in recent. The aim of this study is to investigate whether helpful the direct targeting method with the fiber tractography to the VIM targeting by applied the DTI analysis software called DSI studio.

Methods

Twente-five patients who underwent the TcMRgFUS at Moriyama Neurological Center Hospital was included in this study. A clinical rating scale of tremor (CRST) was applied to evaluate clinical outcome. The DTI was obtained using 1.5 T MRI system (GE healthcare, signa HDxt) and 3.0 T MRI system (Philips, Inginia Elition), and the analysis software applied DSI studio to visualizing fiber tractography. To visualizing the dentato-rubro-thalamic tract (DRTT) passing through the VIM, the region of interest (ROIs) was set at the bilateral dentate nucleus and the treated side of the red nucleus and thalamus and motor cortex respective. The relationship between the target location estimated by analysis the DSI studio and the TcMRgFUS outcomes were retrospectively investigated.

Results

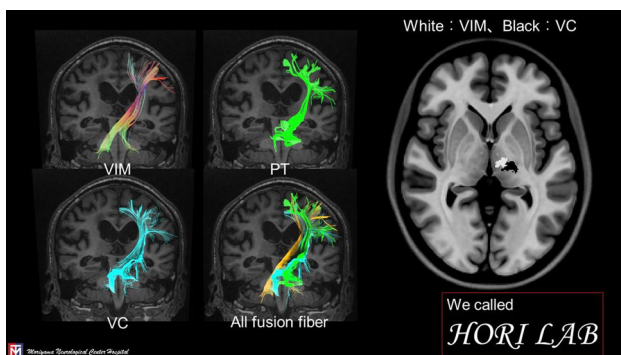


Fig. 1 The DSI studio was able to visualize for the VIM, PT and VC respectively. These fibers can be synthesized and observed. Moreover, it can be measuring the coordinates of the VIM, PT and VC by projecting these fibers onto a 2D MR image. We called HORI LAB such detailed analysis using DSI studio

The CRST score of post TcMRgFUS improved by $69.9 \pm 26.3\%$ (maximum: 100%, minimum: 22%) compared to preoperatively. The location of the VIM which is the treatment target was accurately able to visualized by using DSI studio. Moreover, DSI studio was also able to be clarify the ventral caudal nucleus (Vc) and the pyramidal tract (PT) (Fig. 1). Visualizing these nuclei location was important that avoid to adverse events. The clinical outcomes tended to be good in cases the lesions were close to the VIM. On the other hand, the adverse event was induced in case the lesion was involved the Vc or PT.

Conclusion

The DTI can be expected as a support tool for the direct targeting, by using dedicated analysis software such as DSI studio instead of a general-purpose workstation.

References

- [1] Yamada K, Kizu O, Kubota T, Ito H, Matsushima S, Oouchi H, Nishimura T (2007) The pyramidal tract has a predictable course through the centrum semiovale: a diffusion-tensor based tractography study. *J Magn Reson Imaging* 26: 519–524.
- [2] Ohye C, Higuchi Y, Shibazaki T, Hashimoto T, Koyama T, Hirai T, Matsuda S, Serizawa T, Hori T, Hayashi M, Ochiai T, Samura H, Yamashiro K (2012) Gamma knife thalamotomy for Parkinson disease and essential tremor: a prospective multicenter study. *Neurosurgery* 70: 526–536.

Removing irrelevant frames in plastic surgery videos using semi-supervised learning combined with optical flow

T. Kobayashi¹, H. Kajita², Y. Takatsume³, Y. Aoki¹

¹Keio University, Department of Electrical Engineering, Faculty of Science and Technology, Yokohama, Japan ²Keio University, Department of Plastic Surgery, Faculty of Medicine, Shinjyu-ku, Japan ³Keio University, Department of Anatomy, Faculty of Medicine, Shinjyuku-ku, Japan

Keywords Semi-supervised learning, Optical flow, Surgical videos, Plastic surgery

Purpose

It is important to improve the quality of surgery by recording and analyzing surgical procedures. However, both the quality and the quantity of the data are insufficient for the surgical procedure videos because the surgical field is hidden by hands or heads or the brightness changes due to shadows. Some research has been done to improve the quality of surgical videos, but only in a few areas such as laparoscopic surgery [1]. Therefore, a technique is needed to extract from the raw video only the frames related to the surgery for recording and analyzing the surgery video. We used first-person videos recorded with NanoCamHDI attached to the surgeon loupe and automatically classified and eliminated frames that did not show the surgical field to extract only the parts directly related to the surgery. As many types of surgeries exist for various parts of the body in plastic surgery, it is laborious to create models specific to each surgery, so a highly versatile model is necessary. We propose an accurate and versatile model that can classify various types of surgeries from a small amount of data using supervised learning, semi-supervised learning, and optical flow.

Methods

Figure 1 shows the proposed learning procedure as a process flow. We first created a dataset that included information on whether the surgical field was shown for five types of plastic surgery. We classified the unlabeled surgical videos using the model trained on these datasets. Only the results with high confidence were used as labeled images. We created a new model trained only on these labeled images. The model was used to classify unlabeled surgical videos, and the cycle of obtaining a new dataset using reliability was repeated five times to obtain the results

Datasets

Five types of surgery were filmed with a camera attached to the surgeon head: repair of congenital syndactyly, transplantation of the serratus anterior muscle to the face, facial nerve paralysis, zygomatic osteotomy, and scalp scar revision. These videos were annotated with two classes according to whether the surgical field was captured or not and cropped at 10 fps to create a dataset of 61,234 frames. Of these, 44,584 frames were related to the surgery, and 16,650 frames were unrelated. For the test data, we used a precordial keloid surgery movie, from which we used 11,755 images, 8,446 related to the surgery and 3,309 unrelated.

Reliability

The reliability of selecting the data to be trained in the next training phase from the model output was defined in two ways, as follows:

1. Classification Probability

When obtaining training data for semi-supervised learning from the supervised learning results, we defined a result with high confidence as one with a classification probability of at least 0.9 out of those classified as positive and of at least 0.9999 out of those classified as negative. When labeling iteratively in semi-supervised learning, we defined a result with high confidence as a classification probability of 0.99 or higher for both positive and negative labels.

2. Use of Optical Flow

Since the frames in which the surgical field is shown and not shown are continuous over time, we thought it would be useful to utilize time-series information, and therefore we used optical flow. We divided the video into five-second segments, and the reliability of each segment was based on whether one or more points of optical flow could be traced. Specifically, we defined a high confidence level as one in which the optical flow was obtained and classified as surgery-related and classified as irrelevant one in which the optical flow was not obtained.

Results

Table 1 shows a comparison of the results for each method. We examined three cases of semi-supervised learning: using classification probability and optical flow for reliability and not using reliability.

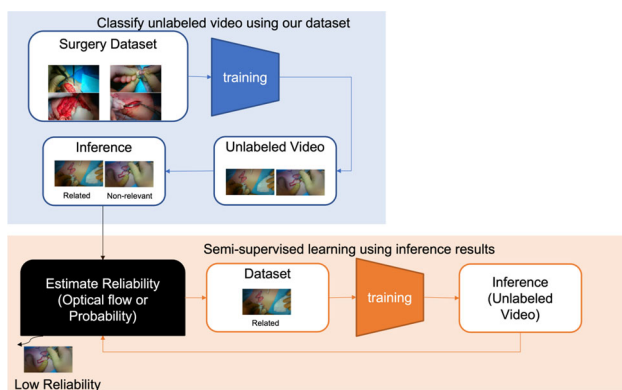


Fig. 1 Proposed process flow

Table 1 Comparison of accuracy for each method

Methods	Accuracy	Recall	Specificity
Supervised	0,818	0,752	0,992
Semi-supervised + Classification probability	0,910	0,888	0,965
Semi-supervised + Optical flow	0,976	0,974	0,981
Semi-supervised	0,820	0,756	0,980

When reliability was not used, the data labeled by the classification result of the model were directly used as the data for the next training. Recall represents the classification accuracy of surgery-related frames, while specificity represents the classification accuracy of the frames unrelated to surgery. When only the model trained by supervised learning was used, the overall accuracy was 0.818, recall was 0.752, and specificity was 0.992. As can be seen from Table 1, there was no improvement in accuracy when we simply used the best results as training data for the next model without using reliability, indicating that using results with a high confidence level plays an important role in improving accuracy.

1. When classification probability was used for reliability

The overall accuracy increased by 0.067 compared to the case in which only the supervised model was used. However, the overall accuracy after five iterations of training inference was 0.883, and there was no significant improvement in accuracy, even after iterations of training.

2. When optical flow was used for reliability

When using this model, the accuracy gradually improved as the iterations progressed. Since it is possible that the high accuracy was due to the performance of the optical flow itself, we also examined the accuracy of the optical flow. The result was 0.750. This indicates that high accuracy was achieved by combining optical flow and semi-supervised learning.

Conclusion

By using semi-supervised learning combined with optical flow, we detected frames that were irrelevant to the surgery with high accuracy. In addition, since the classification accuracy of the surgical situation is much higher than that of the model trained only with the supervised learning method, important frames that should be kept as records of the surgery are not left out. Therefore, we have proposed a model that is effective for recording and analyzing surgical videos.

References

- [1] Zohar M, Bar O, Neimark D, Hager GD, Asselmann D (2020) Accurate Detection of Out of Body Segments in Surgical Video using Semi-Supervised Learning, MIDL 923–932.

Combining, PET and DTI with intraoperative magnetic resonance imaging in the navigation are both useful for glioma surgery

A. Kuwano¹, M. Tamura², H. Kurihara¹, T. Saito¹, M. Nitta¹, T. Maruyama¹, S. Tsuzuki¹, A. Fukui¹, S. Koriyama¹, T. Kawamata¹, Y. Muragaki²

¹Tokyo Women's Medical University, Tokyo, Japan, Neurosurgery, shinjuku, Japan ²Tokyo Women's Medical University, Faculty of Advanced Techno-Surgery, Institute of Advanced Biomedical Engineering and Science, shinjuku, Japan

Keywords intraoperative MRI, DTI, PET, glioma

Purpose

In glioma surgery, the use of intraoperative magnetic resonance imaging (MRI) in the navigation system is useful for tumor removal [1]. And, we routinely use the navigation system that fuses the preoperative positron emission tomography (PET) and intraoperative MRI. Now we developed a navigation system that superimposes the fractional anisotropy (FA) color map of preoperative diffusion tensor imaging (DTI) and intraoperative magnetic resonance imaging (MRI). The current study aimed to investigate the usefulness of these systems for neurophysiological monitoring and examination under awake craniotomy during tumor removal.

Methods

In the updated navigation system using intraoperative MRI after craniotomy, the position of the surgical tool held by the surgeon can be displayed on the MRI each time. Intraoperative MRI is more accurate than preoperative MRI for surgical navigation because the brain parenchyma position shift caused by craniotomy (so called brain shift) is corrected and updated in the intraoperative MRI. Based on the actual anatomy of the surgical field and the navigation-guided MRI, the surgeon can perform tumor removal more quickly and precisely. As for PET and MRI, we evaluated the effectiveness of strategic planning for tumor removal. However, it is important to recognize that PET is a preoperative image and MRI is an intraoperative image, so there are structural and time axis discrepancies between the two. Regarding for FA of DTI and intraoperative MRI, a total of 10 glioma patients (4 patients with right-side tumors; 5 men and 5 women; average age, 34 years) were evaluated. Among them, the tumor was localized to the frontal lobe, insular cortex, and parietal lobe in 8, 1, and 1 patient, respectively. There were 3 patients who underwent surgery on general anesthesia, while 7 patients underwent awake craniotomy. The index of DTI anisotropy taken preoperatively (magnetic field: 3 T, 6 motion probing gradient directions) was analyzed as a color map (FA color map) and concurrently co-registered in the intraoperative MRI within the navigation. In addition to localization of the bipolar coagulator and the cortical stimulator for brain mapping on intraoperative MRI, the preoperative FA color map was also concurrently integrated and displayed on the navigation monitor. This functional information of white matter tracts was confirmed directly by using neurological examination and referring to the electrophysiological monitoring.

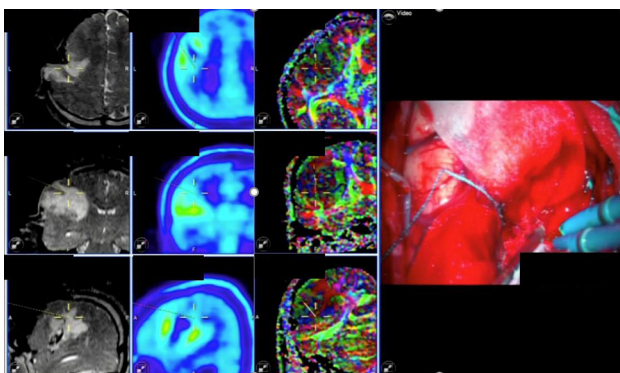


Fig. 1 Image integration of preoperative DTI-FA color map, PET and intraoperative MRI navigation system. A newly developed system in which an FA color map, as an objective measure of neural function, from preoperative DTI and PET is combined with intraoperative MRI image obtained after craniotomy, is developed

Results

Using fusion of preoperative PET and intraoperative MRI were both effective for planning strategy of tumor removal. Co-registering an intraoperative MRI and preoperative PET image was useful for planning the strategy of tumor removal, especially in high grade glioma cases. Intraoperative MRI, integrated preoperative FA color map, and microscopic surgical view were displayed on one screen in all 10 patients, and white matter tracts including the pyramidal tracts were displayed as a reference in blue, Fig. 1. Regarding motor function, motor-evoked potential was monitored as appropriate in all cases, and removal was possible while directly confirming motor symptoms under awake craniotomy. Furthermore, the white matter tracts including the superior longitudinal fasciculus were displayed in green. Importantly, it was useful not only to localize the resection site, but to identify language-related, eye movement-related, and motor tracts at the electrical stimulation site. All motor and/or language white matter tracts were identified and visualized with the co-registration and then with an acceptable post-operative neurological outcome.

Conclusion

Co-registering an intraoperative MRI and a PET image is useful for planning the strategy of high grade glioma tumor removal. And, co-registering an intraoperative MRI and a preoperative FA color map is a practical and useful method to predict the localization of critical white matter nerve functions intraoperatively in glioma surgery.

References

- [1] Y Muragaki, H Iseki, T Maruyama, T Kawamata, F Yamane, R Nakamura, O Kubo, K Takakura, T Hori (2006) Usefulness of intraoperative magnetic resonance imaging for glioma surgery. *Acta Neurochir Suppl*: 98:67–75

Utility of intraoperative flow cytometry and real-time PCR in glioma surgery

S. Koriyama^{1,2}, Y. Muragaki^{1,2}, M. Nitta^{1,2}, T. Saito^{1,2}, T. Shioyama³, A. Suzuki³, T. Komori⁴, K. Masui⁵, S. Tsuzuki^{1,2}, A. Fukui^{1,2}, A. Kuwano^{1,2}, T. Kawamata¹

¹Tokyo Women's Medical University, Department of Neurosurgery, Tokyo, Japan ²Tokyo Women's Medical University, Faculty of Advanced Techno-Surgery, Tokyo, Japan ³Nihon Kohden Corporation, Advanced Biomedical Engineering and Science, Tokyo, Japan ⁴Tokyo Metropolitan Institute of Medical Science, Department of Brain Development and Regeneration, Tokyo, Japan ⁵Tokyo Women's Medical University, Department of Pathology, Tokyo, Japan

Keywords glioma, intraoperative diagnosis, intraoperative flow cytometry, real-time PCR

Purpose

In glioma surgery, it has been reported that the higher extent of resection, the longer the prognosis, then surgical removal has been important. On the other hand, a high extent rate of resection increases the risk of functional deficit due to surrounding normal brain damage. Glioma has the nature of invasive so the tumor and the normal brain

are unclear, then the accurate and objective method of assessing the tumor boundary has been required, [1, 2].

Since the 2016 WHO classification, genetic information has become essential for the diagnosis of glioma, and there have been many reports on the correlation between genetic aberrations and prognosis. Therefore, accurate intraoperative assessment of the type and grade of glioma is a very important factor in determining the extent of resection.

Methods

We have developed an intraoperative flow cytometry system (iFC), which is currently being applied clinically as an intraoperative support device for malignant brain tumors. This iFC displays the total cell count, DNA histogram, and Malignancy Index (MI), which is the ratio of the number of cells in the “proliferative phase” (S, G2, M) to the total cell count (%) to the right of the 2C peak. Increasing MI indicates the appearance of DNA aneuploidy or an increase in the number of proliferating cells. The most important feature of the system is that it is specialized for the analysis of DNA ploidy, by thoroughly reviewing the process, analysis that used to take more than 30 min can be done in less than 10 min.

Results

Genetic mutations that correlate with diagnosis and prognosis, such as IDH mutations and 1p/19q co-deletion, can be detected in real-time-PCR: IDH mutations can be detected by HRM, while mutations in the TERT promoter region, an alternative marker for 1p/19q co-deletion, can be detected by SNP genotyping. The results are almost the same result as those obtained from postoperative pathological specimens. Along with the conventional rapid pathological diagnosis, evaluation by iFC and real-time PCR has enabled us to make an accurate diagnosis in a short time.

Conclusion

At our institute, we make surgical decisions based on such intraoperative pathological diagnosis and genetic information. In this presentation, we introduce the usefulness of iFC and real-time PCR in glioma surgery.

References

- [1] Koriyama S, Nitta M, Kobayashi T, Muragaki Y, Suzuki A, Maruyama T, Komori T, Masui K, Saito T, Yasuda T, et al. A surgical strategy for lower grade gliomas using intraoperative molecular diagnosis. *Brain Tumor Pathol.* 2018;35(3):159–167. <https://doi.org/10.1007/s10014-018-0324-1>.
- [2] Koriyama S., Nitta M., Shioyama T., Komori T., Maruyama T., Kawamata T., Muragaki Y. Intraoperative Flow Cytometry Enables the Differentiation of Primary Central Nervous System Lymphoma from Glioblastoma. *World Neurosurg.* 2018;112:e261–e268. <https://doi.org/10.1016/j.wneu.2018.01.033>.

Machine-learning classification of surgical procedures using multiple views of surgical fields

R. Fujii¹, H. Kajita², Y. Takatsume³, T. Kobayashi¹, Y. Aoki¹

¹Keio University, Department of Electrical Engineering, Faculty of Science and Technology, Yokohama, Japan ²Keio University, Department of Plastic Surgery, Faculty of Medicine, Shinjuku-ku, Japan ³Keio University, Department of Anatomy, Faculty of Medicine, Shinjuku-ku, Japan

Keywords Machine Learning, Open Surgery, Action Classification, Multi Camera View

Purpose

While the recording of surgical images is effective in sharing surgical skills, it is not widely used due to the time required, costs, and labor management. These drawbacks are especially pronounced in surgical operations. In this study, we created a video dataset by capturing a plastic surgery field using a multi-camera equipped with a shadowless lamp and then verified the effectiveness of using multi-view video for the classification task of plastic surgery [1].

Methods

Dataset

The dataset used in this study is based on 12 plastic surgery cases, which were captured using a mounted multi-camera with a shadowless lamp. Using label annotation software, each video was assigned a class label for each major procedures in plastic surgery (anesthesia, closure, design, disinfection, dressing, hemostasis, incision, and dissection). We then annotated each image with the corresponding class label of the process (Fig. 1). Because the video divided by the label has a large time difference depending on the procedure, it was additionally divided into short video clips, 10 s each. The overclasses were extracted and the underclasses excluded, and finally the video clips representing the procedures were divided into short video clips.

Network

The video clips created as described in the previous section were input to the CNN (Convolutional Neural Network) for classification. In this case, the effectiveness of using multiple-view video was confirmed by comparing selected frames having less obstruction from multiple cameras with the input from a single viewpoint.

Frame selection

The visibility score of the surgical field for each camera was evaluated and input into the CNN after a preliminary camera switch based on the score. Visibility score is derived based on the number of pixels estimated in the surgical field area in the image, and the score at time with is compared between the two cameras and reversed.

Results

The video clips used for dataset creation were stored according to the time sequence of the surgery. The division of training data, evaluation data, and test data for the accuracy evaluation was performed in the following ways:

1. Ignoring the time series and randomly dividing the data into 7:2:1.
2. Dividing the data into 7:2:1 according to the time series.
3. Divided by surgery

The results were verified as follows:

Case 1 achieved 71.8% accuracy for single-view input and 77.9% accuracy for multiple-view input, confirming the accuracy improvement achieved by using multiple-view video. Table 1 shows that the accuracy for classification is higher than the expected value (16.6%) and can be said to be sufficient for classification.

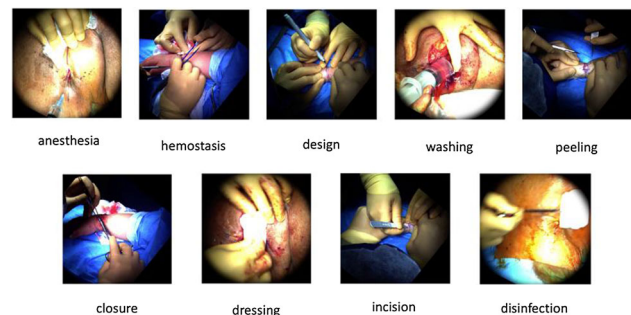


Fig. 1 Major procedures of plastic surgery

Table 1 The classification result of case 1

class	F1-score (%)		Number of videos
	single view	frame selected	
anesthesia	53,1	58,4	9
closure	73,3	85,1	39
design	82,1	84,7	37
hemostasis	56,4	60,7	15
incision	66,2	78,1	37
dissection	74,7	76,1	59
accuracy	71,8	77,9	196

In Case 2, the accuracy of multiple-view input was 57.2%, and the same improvement in accuracy was confirmed. However, the accuracy was lower compared to Case I. This is thought to be due to the domain difference caused by the change in the treatment content with the passage of time.

In Case 3, the accuracy dropped significantly to 15.3%. This is thought to be due to the imbalance of data, depending on the type of surgery, suggesting that it has not yet reached the level for practical application.

Conclusion

The purpose of this study was to confirm the effectiveness of using multiple views in a surgical process classification task. The technique was found to be effective in the actual accuracy evaluation. However, the accuracy is inadequate for practical applications. More types of surgical data may be needed for this solution.

Reference

- [1] Hachiuma R, Shimizu T, Saito H et al.: Deep Selection: A Fully Supervised Camera Selection Network for Surgery Recordings. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2020

A study on the visualization of brain tumors using a hyperspectral camera

H. Asano^{1,2}, J. Okamoto², M. Tamura², T. Maruyama³, T. Saito^{2,3}, K. Masamune², Y. Muragaki^{2,3}

¹Atom Medical Corporation, Technical Department, Sakura-ku Saitama-city, Japan ²Tokyo Women's Medical University, Faculty of Advanced Techno-Surgery, Tokyo, Japan ³Tokyo Women's Medical University, Department of Neurosurgery, Tokyo, Japan

Keywords Hyperspectral Imaging, Brain tumor, Spectral Analysis, Neurosurgery

Purpose

The surgical removal of brain tumors is a balancing process between prognosis and functional preservation. To improve prognosis, it is necessary to remove as much of the tumor as possible, but there is a risk of functional deterioration depending on the site of removal.

However, it is difficult to determine the difference in appearance between tumor and normal tissue. For this reason, various techniques are used to identify the tumor and normal tissue while performing removal. Intraoperative determination of the tumor region is based on information from intraoperative MRI and intraoperative rapid diagnosis. However, intraoperative MRI requires interruption of the surgery to take the images, and it takes time to take the images. Intraoperative rapid diagnosis is limited by the fact that only the removed points are evaluated.

To address this issue, some research has been conducted using hyperspectral cameras that can acquire spectral information as images. FABELLO et al. have performed intraoperative brain surface imaging and estimated the tumor area using clustering techniques [1]. Mori et al. proposed a method to estimate oxygen saturation from hyperspectral images [2]. These reports show the effectiveness of hyperspectral cameras, but there is still a problem that it is difficult to explain the meaning of the judgment results. In this paper, we report on the visualization of a tumor on the surface of the brain based on spectral analysis techniques.

Methods

Twelve patients who were diagnosed with malignant brain tumors and underwent surgical removal of the tumors at Tokyo Women's Medical University Hospital were included in the study. 5 patients had Glioblastoma, 1 patient had Diffuse Astrocytoma, 3 patients had Anaplastic Oligodendroglioma, 2 patients had Oligodendroglioma, 1 patient had Anaplastic Astrocytoma. This study was approved by the Ethics Committee of Tokyo Women's Medical University under approval number 130209. The research was carried out in accordance with the Declaration of Helsinki. Written informed consent was obtained from the patient. Hyperspectral imaging was performed using Eva Japan's NH-7 at a position of 40 cm from the brain surface. A Ushio halogen lamp (JDR110V75WLM/K7UV-H) with an aluminum coating on the mirror was used to project light into the imaging area. The captured data is affected by lighting, ambient light, sensitivity characteristics of the camera, and noise from the device. Therefore, the preprocessing expressed in Eq. (1) is applied to the captured data to obtain a composed spectral image.

$$CI = (RI - DR)/(WR - DR) \quad (1)$$

CI is the post-calibration image, *RI* is the original image, and *DR* is the dark current noise information of the sensor. The dark current noise was obtained by shooting with the lens covered. *WR* is the data taken of an object with known spectral characteristics. In this study, we used x-rite's color checker white balance, which is known to have almost constant spectral characteristics. For visualization by spectral analysis, two methods are used: one is based on water content, and the other is based on the ratio of oxygenated hemoglobin to reduced hemoglobin. It is known that water specifically absorbs light with a wavelength around 970 nm. When creating an RGB image for display from a hyperspectral image, light at 970 nm is assigned to the R channel, 540 nm to the G channel, and 470 nm to the B channel (Method 1). The method of visualization using the ratio of oxygenated hemoglobin to reduced hemoglobin is described next. It is known that the absorbance of oxygenated hemoglobin and reduced hemoglobin intersect at a wavelength near 800 nm. Therefore, by using Eq. (2), the ratio of oxygenated hemoglobin to reduced hemoglobin can be visualized as a color change (Method 2).

$$\begin{aligned} R &= 2 \times Y_{680\text{nm}} - Y_{880\text{nm}} \\ G &= Y_{800\text{nm}}/2 \\ B &= 3 \times Y_{880\text{nm}} - 2 \times Y_{680\text{nm}} \end{aligned} \quad (2)$$

R, G, and B represent the red, green, and blue signal intensities of the pixels in the visualized image, respectively.

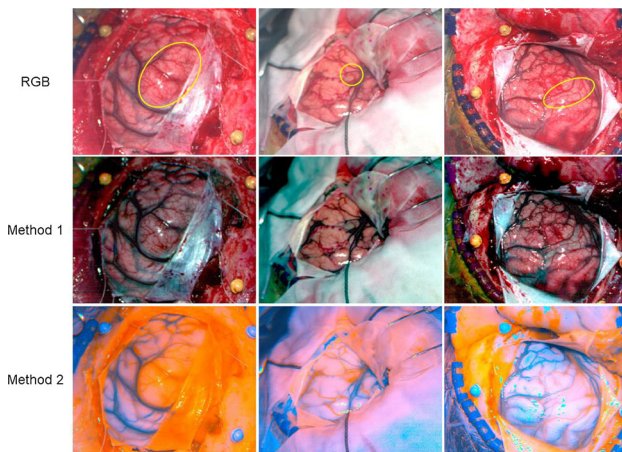


Fig. 1 Examples of imaging. The upper row shows the same visualization as in a normal RGB image, with the tumor region indicated by a yellow circle. The middle row is the result of visualization using Method 1, where tissues with high water content are shown darkly. The lower panel is the result of visualization using method 2, where the red areas have more oxygenated hemoglobin and the blue areas have more reduced hemoglobin

Results

Examples of the visualization results are shown in Fig. 1. Tumor regions were enhanced, although the results differed from each case. When the tumor tissue had more water content compared to normal tissue, it was shown to be darker than other areas by Method 1 (8 cases). In addition, when oxygenated hemoglobin was high in normal tissue and low in the tumor, method 2 caused normal tissue to be imaged in red and the tumor in blue (2 cases). On the contrary, when oxygenated hemoglobin was high specifically in the tumor, it was confirmed that normal tissue was blue and the tumor was red (7 cases). In 11 of the 12 cases captured, there was a color change specific to the tumor with either technique.

Conclusion

It was suggested that brain tumors could be visualized by using a hyperspectral camera and imaging using a spectral analysis technique. Since it takes only about 10 s to take a picture and less than 1 s to process the image, we believe that it is possible to achieve a system that can confirm brain tumors in real time during surgery. On the other hand, there are still many issues to be solved for clinical application. Since the pattern of tumor visualization is not consistent, it is necessary to verify the relationship with pathology and to clarify the mechanism. In addition, this study was limited to preoperative gray matter imaging, and verification in white matter is necessary. Furthermore, since the current number of cases is still insufficient, it is necessary to perform imaging in a larger number of cases for verification.

References

- [1] Fabelo H, Ortega S, Szolna A, Bulters D, Pineiro J, Kabwama S, J-O'Shanahan A, Bulstrode H, Bisshopp S, Kiran B, Ravi D, Lazcano R, Madronal D, Sosa C, Espino C, Marquez M, De La Luz Plaza M, Camacho R, Carrera D, Hernandez M, Callico G, Morera Molina J, Stanculescu B, Yang G, Salvador Perea R,

Juarez E, Sanz C, Sarmiento R (2019) In-Vivo Hyperspectral Human Brain Image Database for Brain Cancer Detection. IEEE Access. 7:39098–39116

- [2] Mori M, Chiba T, Nakamizo A, Kumashiro R, Murata M, Akaoshi T, Tomikawa M, Kikkawa Y, Yoshimoto K, Mizoguchi M, Sasaki T, Hashizume M (2014). Intraoperative visualization of cerebral oxygenation using hyperspectral image data: a two-dimensional mapping method. *International Journal of Computer Assisted Radiology and Surgery*, 9(6):1059–1072

Sleevemeter with contact force sensors for assisting sleeve gastrectomy

S. Cho¹, S. E. Kim¹, Y. J. Kang¹, S. Yun¹, S. H. Kim¹, M. G. Lee¹
¹Ajou Univ., Mechanical Engineering, Suwon-si, Yeongtong-gu, South Korea

Keywords Bougie, Traction, Gastrectomy, Monitoring

Purpose

The role of surgical treatment for morbid obesity is constantly increasing. As of 2018, more than 600,000 metabolic surgeries are being performed worldwide, and among them, sleeve gastrectomy has increased rapidly, making it the most common surgical procedure currently [1].

In this study, we propose a device, “sleevemeter,” that makes even beginners perform a sleeve gastrectomy easily and appropriately. The main component of the device is a bougie equipped with contact force sensors to monitor the traction force that a surgeon applies to the stomach during the operation. Using this, the remaining stomach after surgery has a thin and long shape with constant diameter. This shape allows obese patients to achieve effective weight loss. It can also prevent postoperative complications such as gastric stenosis and perforation, Fig. 1.

Methods

When performing a sleeve gastrectomy, it is important to make the remaining gastric shape thin and long with a constant diameter. To do this, after inserting a bougie into the stomach through the patient's mouth, using it as a guide, the greater curvature of stomach is excised with a stapler. At this time, if the stomach is pulled or tracted excessively and stapled along the bougie, the stomach becomes taut around the bougie. This can make the diameter of the stomach left after surgery so small that the passage can be narrowed or blocked. If the difference in diameter is large, there is a risk of perforation due to sudden pressure fluctuations as the stomach moves for digestion. Conversely, if the surgeon does not tract the stomach sufficiently and staple along the bougie, the remaining diameter of the stomach is too big. Then, the remaining volume of the stomach is large and does not lead to weight loss, which may prevent metabolic patients from improving the disease. Beginners who are afraid of complications often make these mistakes.

To solve this problem, small, thin and flexible contact force sensors are installed on the bougie surface. These bougie with the sensors, “sleevemeter,” is inserted through the patient's mouth, and the surgeon performs the sleeve gastrectomy with traction and stapling. The sensor module consists of contact sensors on the bougie, an electrical circuit that amplifies the pressure signal, and a monitor that displays the signal, warning or alarm on the screen. Pressure signals collected from the contact force sensors location are delivered to the surgeon through the monitor screen. Too high pressure from this sensor means excessive traction and will leave a narrowed gastric passage after surgery. Too small pressure means loosened traction and will keep back a large volume of stomach after surgery. This allows

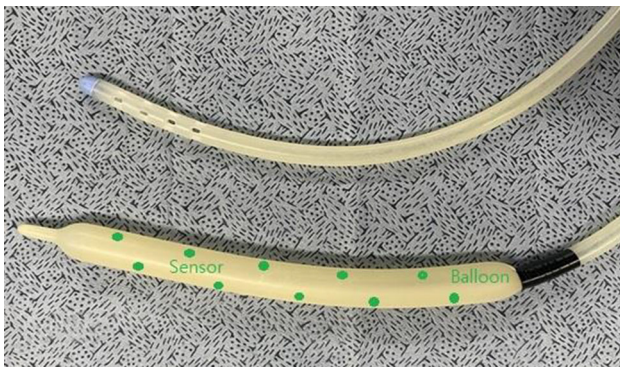


Fig. 1 Conventional Bogue (top) and proposed sleeve-meter (bottom)

the surgeon to perform the surgery in the proper traction range, which can help him maintain proper stomach shape during the surgery.

Results

Sleeve gastrectomy was performed by applying this device to the silicone model stomach and the excised swine stomach. Repeated experiments were conducted in a group consisting of beginners and experts, and the shape and volume of the stomach remaining after surgery were evaluated.

Conclusion

With using this device and method, even beginners can perform the sleeve gastrectomy more easily and effectively. Furthermore, it is also expected to prevent various complications after the surgery.

Reference

- [1] Gumbs, A.A., Gagner, M., Dakin, G. et al. Sleeve Gastrectomy for Morbid Obesity. *OBES SURG* 17, 962–969 (2007). <https://doi.org/10.1007/s11695-007-9151-x>

Development of the estimation method for organ-instruments contact during endoscopic surgery

M. Sogabe¹, R. Tamaki¹, T. Miyazaki¹, T. Kawase^{2,3}, T. Kanno⁴, K. Kawashima¹

¹The University of Tokyo, Department of Information Physics & Computing, Tokyo, Japan ²Tokyo Medical and Dental University, Institute of Biomaterials and Bioengineering, Tokyo, Japan ³Tokyo Institute of Technology, the Institute of Innovative Research, Kanagawa, Japan ⁴Riverfield Inc., Design & Development Department, Tokyo, Japan

Keywords Laparoscopic surgery, Image processing, ResNet50, Contact estimation

Purpose

The demand for minimally invasive surgical procedures using laparoscopy is increasing every year. It is widely applied to various surgical procedures, such as hepatectomy and cholecystectomy prostatectomy. Further expansion of its application is expected in the future. The number of surgeries has been increasing every year, from about 2,000 in the 1990s to more than 250,000 today. The number of patients who wish to undergo minimally invasive surgery is expected

to increase in the future, but there are still many problems to be solved, such as certification of endoscopic surgery techniques and securing of physician resources.

For ensuring human resources, a useful technology is AI-assisted. There are many challenges in laparoscopic surgery, one of which is providing a sufficient workspace and preventing damage caused by contact between instruments and organs. The problems in laparoscopic surgery are providing working space and the associated damage caused by contact between instruments and organs. In general, during laparoscopic surgery, the space surrounded by the peritoneum and organs is filled with gas, while the surrounding organs are excluded. However, compared to laparotomy, it is more difficult to grasp the contact with the organs, which may lead to excessive pressure on the organs or unnecessary contact with the organs. In addition, it is necessary to allocate human resources to hold the organ, and experience and skill are required to hold the organ.

Therefore, the development of a tool that assists the contact between the organ and the instrument will help to ensure the workspace in the operating room itself and reduce the burden on the surgeon. In response to the demand for such a tool, a sensor has been developed to detect contact with the tip of the instrument. While sensor-based contact estimation is highly accurate, it is often not resistant to high-pressure sterilization, which is essential for surgical tools. However, endoscopic image-based contact estimation does not require the preparation of new sensors and can reduce the intraoperative burden [1]. In previous studies, several methods have been devised to estimate the contact with organs and their forces from images. However, most of these methods use skin or model organs, and there is still a lack of knowledge about the actual organs to be used in laparoscopic surgery.

Methods

This report developed and demonstrated a method to determine the contact force between the surgical instruments and organs by manipulating the forceps and laparoscopic side using a pneumatically driven surgical robot developed by our group [2]. During the operation, the extracted pig organs such as livers and lungs are used to take stereo images of the contact of the surgical instruments with the organs. Specifically, a method to prevent excessive contact was realized by estimating the contact between the instrument and the organ and its intensity from the endoscope image by ResNet50 using residual blocks.

Results

Using Resnet50, three classifications were made: non-contact, weak contact, and strong contact, after which contact force was estimated and evaluated. For the classification of contact force, the state of no contact was classified as non-contact, no deformation of organs due to contact with instruments ((less than 0.5 N), and other states were classified as strong contact, and the correct answer rate was about 80%.

Conclusion

This technique is expected to lead to safer surgery by displaying the contact status in the endoscope image. It is also expected to be a stepping stone to the development of assistive robots that can autonomously hold organs based on the acquired information.

References

- [1] Jung W.J., Kwak K. S. and Lim S. C. (2021) Vision-Based Suture Tensile Force Estimation in Robotic Surgery, *Sensors*, 21(1), 110.
 [2] Miyazaki R., Kanno T. and Kawashima K. (2019) Pneumatically driven surgical instrument capable of estimating translational force and grasping force, *The International Journal of Medical Robotics and Computer Assisted Surgery*, 15(3).

Preoperative planning for cardiovascular surgery

C. Ciobirca¹, C. I. Dumitrescu², R. T. Popa¹, D. Dumitrescu¹, C. G. Tambura³, D. M. Ciobirca⁴, M. F. Tiuca⁵, S. Maces⁶, M. Moraru⁶, L. F. Barbulescu⁷, L. D. Popa⁷, S. M. Cazacu⁸

¹University of Medicine and Pharmacy of Craiova, Department of Radiology and Medical Imaging, Craiova, Romania ²University of Medicine and Pharmacy of Craiova, Department of Pharmacology, Craiova, Romania ³SPAD IMAGING INTERNATIONAL, Craiova, Romania ⁴Slatina Emergency County Hospital, ENT Department, Slatina, Romania ⁵Glan Clwyd Hospital, Vascular Department, Bodelwyddan, United Kingdom ⁶University of Medicine and Pharmacy of Craiova, Doctoral School, Craiova, Romania ⁷University of Craiova, Department of Computer Science, Electronics and Automation, Craiova, Romania ⁸University of Medicine and Pharmacy of Craiova, Department of Internal Medicine 5th year, Craiova, Romania

Keywords planning software, automatic measurements, stent simulation, cardiovascular surgery

Purpose

Various surgical minimally invasive procedures [1], or even open surgery ones [2] prove to be difficult to perform, requiring careful planning based on analysis of patient data. One of the best solutions for such planning can be virtual reality, i.e., a software 3D graphics rendering starting from a CT scan. We have developed and present further a method for virtual planning and simulation of cardiovascular surgery procedures, integrated in our own medical imaging software platform, CardioCTNav. The planning consists of 3D rendering of the area of interest, virtual angiography or coronarography, automated measurements, virtual stent simulation, and, in the future, blood flow simulation and FFR (fractional flow reserve) computation.

Methods

The preoperative planning that we present is integrated in CardioCTNav, an “in-house” custom-made cross-platform (Linux, macOS, Windows) medical imaging software application based on open-source technologies—C++ programming language, Qt, VTK, VTKDicom, CTK. It is a fast and stable, easy-to-learn and to use solution, developed by authors and offering them complete control over the implementation of the protocol described in the following.

Using CardioCTNav and starting from CT data, the surgery planning procedure consists of the following steps:

- 3D reconstruction and visualization of arteries
- Virtual angiography (or coronarography, depending on the artery with problems)—virtual 3D navigation through various arteries (e.g. coronary, mesenteric, etc.)
- Automatic measurements of the artery diameter in the investigated region
- Stent 3D visualization and simulation (diameter and length)
- Blood flow simulation and fractional flow reserve computation—under development.

For a better view of the region under investigation, the 3D reconstruction and the visualization use, besides “classical” graphics algorithms implemented in VTK library, our own custom methods and algorithms: a “free-hand” drawing mode of the transfer function, a “skin-removal” option and a variant of region-growing method for segmentation of arteries. The transfer function translates Hounsfield units from the CT data into colors and levels of opacity for the 3D rendering scene. The “free-hand” drawing mode of the transfer function shapes implemented in CardioCTNav allows a fine tuning in the selection of the tissues that are visualized. This fine tuning is necessary for virtual navigation (coronarography or angiography).

CardioCTNav software platform distinguishes between the zones inside and outside the arteries and detects the artery walls. Target points can be chosen by the user interacting with the 2D sections and they are colored accordingly to the zone where they are placed (outside or inside the artery). Also, the platform can detect an “inside” zone closest to a point that is initially placed in an “outside” zone.

Virtual angiography means that the user can navigate inside the artery using only the computer mouse device. The navigation starts from a point chosen by the user and the virtual camera is forced to remain inside the artery walls. The algorithm that implements this restriction on virtual cameras is based on collision detection and resolution directly on voxels (no segmentation is required). The path of the virtual angiography can be saved and used later, in other planning sessions.

The automatic measurements are in fact values of the artery diameter computed automatically in points either chosen by the user or along the virtual angiography path.

The stent is generated automatically along a virtual angiography path and its diameter is decided by the user.

Results

We evaluated the suggested preoperative planning procedure in a case of superior mesenteric artery aneurysm [2], and on several data sets (CT scans) that included coronary arteries, Fig. 1.

For the superior mesenteric aneurysm, the vascular surgeon had to decide between two main methods of reparation: open surgery or endovascular repair. The decision was based on a multitude of factors (e.g., the presence of mesenteric collaterals, patient frailty, etc.). Both methods required a tubular prosthetic material to be introduced inside the aneurysm that had to be specially tailored to patient physiology, and in this point our planning procedure proved to be useful.

For coronary arteries scans, we identified the obturated regions, performed virtual coronarography, automatic diameter measurements and simulated stent placement operations. In the future developments of CardioCTNav, fractional flow reserve computation and blood flow simulations will help further in the evaluation of artery’s stenosis.

Conclusion

We presented the concepts and an actual implementation in a custom software platform of a method for virtual preoperative planning of cardiovascular surgery procedures. Our method includes 3D rendering and visual “enhancement” of the region under investigation, virtual angiography / coronarography, automatic measurements, stent visualization and simulations, and, in the future developments, blood flow

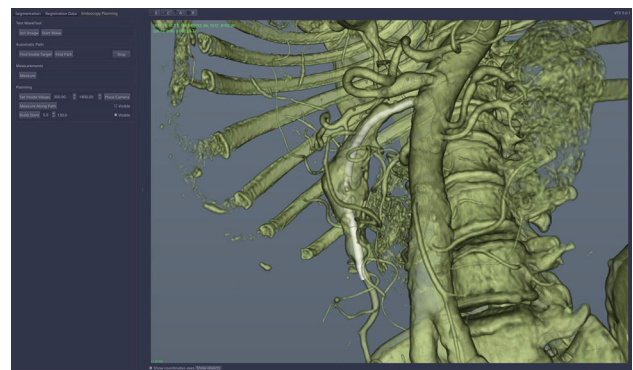


Fig. 1 CardioCTNav—stent placement simulation inside a superior mesenteric artery aneurysm

simulations and fractional flow reserve computation. The method was evaluated and validated in a special case of superior mesenteric artery aneurysm and in several CT scans with coronary arteries.

References

- [1] S. D. Preda, C. Ciobirca, G. Gruionu, A. Iacob, K. Sapolidis, L. G. Gruionu, S. Castravete, S. Patracu, V. Surlin, Preoperative Computer-Assisted Laparoscopy Planning for the Minimally Invasive Surgical Repair of Hiatal Hernia, *MDPI Diagnostics*. 10. 621. 10.3390/diagnostics10090621, August 2020
- [2] C. I. Dumitrescu, C. Ciobirca, R.T. Popa, D. Dumitrescu, C.G. Tambura, D.M. Ciobirca, R. Stavaru, M.F. Tiuca, S. Maces, L.F. Barbulescu, L.D. Popa, S.M. Cazacu, Preoperative Planning for Superior Mesenteric Artery Aneurysm, *MDPI Applied Sciences* 11, no. 21: 10311. <https://doi.org/10.3390/app112110311>, November 2021

Investigation of segmentation-based prostate volume measurement for use in PSA density calculation

K. Abe¹, H. Takeo¹, Y. Nagai², Y. Kuroki³, S. Nawano⁴

¹Kanagawa Institute of Technology, Electrical and Electronic Engineering, Atsugi, Kanagawa, Japan ²National Cancer Center Hospital East, Radiology, Kashiwa, Chiba, Japan ³Niimura Hospital, Radiology, Kagoshima, Kagoshima, Japan ⁴Shinmatsudo Central General Hospital, Radiology, Matsudo, Chiba, Japan

Keywords Prostate, Volume measurement, DeepLab, PSA density

Purpose

At present, research and development of computer aided diagnosis (CAD) is being conducted in various medical fields. Studies of the detection of organs and lesions using AI technology are also being actively pursued. [1]

Prostate cancer is a pathology with a high attack rate among men in western populations. In Japan, it is forecast to become the top-attack-rate pathology for men by 2025. While the PSA measurement test is used with high accuracy in prostate cancer detection, high PSA values are also produced in cases of benign prostatic hyperplasia and other conditions. To contend with this, PSA density (PSAD), obtained by dividing the PSA value by the volume of the prostate, is used with the objective of improving sensitivity. However, the prostate volumes used in current PSAD calculations suffer in accuracy to a degree because they rely on the ellipsoid method, which likens the prostate to an ellipse and calculates volume from the ellipse's horizontal, transverse, and vertical diameters.

In this study, we investigate prostate volume measurement using as a segmentation method a regional extraction method that employs DeepLab, and also consider a method that uses this in PSAD computation. This is expected to enable volume and PSAD calculations with higher accuracy than previous methods and enhance the accuracy of screening for prostate cancer.

Methods

(1) Preprocessing

Because the images used in this study were raw MRI images, they had contrast that disqualified them as data for research purposes. We therefore adjusted the images to obtain image uniformity through preprocessing that normalizes density and contrast.

In this study, images with the prostate manually extracted were used as label images. Initially, the authors sketched out several cases under the guidance of physicians. After this, the authors continued to

conduct extraction while referencing comments provided during the guidance phase.

As a result of removing images that do not include the prostate region, we obtained 1,680 images for use in training.

(2) Regional extraction processing using DeepLab

The DeepLab used in the present study was configured based on the DeepLab configuration of Chen et al. [2]. The development environment was Python using Anaconda. Transfer learning was carried out based on deeplabv3_resnet101, a preliminary learning model in Torchvision. Because output contained only prostate regions, we changed it to 1 channel, and because it was a resnet101 model, we set the feature vector to 2048.

(3) Prostate volume measurement and PSAD calculation

Because minute regions not connected with the prostate also occur in the DeepLab detection results, we first conducted the removal of minute objects. We then calculated volume by calculating from those images the total number of pixels and by multiplying the slice depth by the square of the pixel spacing. Next, we calculated the PSAD from this volume and the PSA value and compared the result with biopsy data.

Results

(1) Supervised DeepLab learning

Using the images described in Sect. 2 above, we conducted supervised learning of DeepLab. Learning was conducted with parameters set as follows: input images were standardized at 256'256; learning rate was set to 0.0001 (default value), batch size to 16, and the number of epochs to 25. The final accuracy during learning was 98.8% for training data and 95.2% for verification data.

(2) Volume measurement and PSAD-based cancer screening results

Table 1 shows volumes and cancer screening test results for the proposed method and for biopsy data. Figure 1 shows a graph focused on PSAD with the standard value of 0.15. The screening results for both green and purple did not change. For blue, screening was not possible with the proposed method. For red, only screening with the proposed method was successful. Table 1 shows that, compared to biopsy data, the volume measurement results of the proposed method

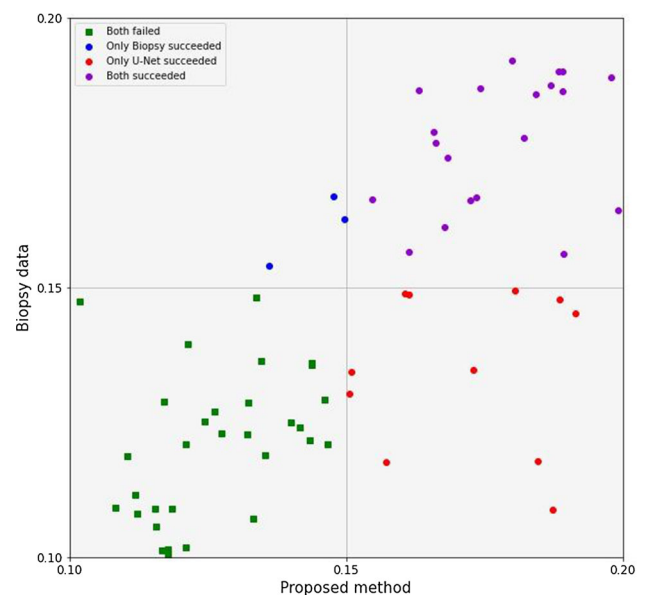


Fig. 1 Comparison of PSAD values obtained from biopsy data and from the proposed method in the vicinity of the threshold value of 0.15

Table 1 Comparison of biopsy data and the proposed method

	Biopsy data	Proposed method
Volume	14.4 - 222.4[ml]	14.2 - 160.0[ml]
Average volume	51.5[ml]	46.2[ml]
PSAD	0.05 - 2.51[ng/ml]	0.05 - 2.90[ng/ml]
Average PSAD	0.259[ng/ml]	0.288[ng/ml]
Screening accuracy	58% (109/188)	63% (119/188)

exhibited a maximum increase of 38.7% and a 79.8% decrease at minimum. Taking the average obtained a 12.9% decrease. For PSAD screening results, biopsy data produced an accuracy of 58% (109 cases out of 188 cases). In comparison, the proposed method produced an accuracy of 63% (119 cases out of 188 cases), indicating improved accuracy. Figure 1 shows improved accuracy in the vicinity of a PSAD value of 0.15. Although detection failed in 3 cases, it became possible for 13 cases, demonstrating improved accuracy.

Conclusion

In this study, we investigated prostate volume measurement using DeepLab and considered using this to improve PSAD-based cancer screening accuracy. Volume measurement using DeepLab enabled prostate shape extraction with higher precision than the ellipsoid method. This resulted in PSAD-based screening with an increase in accuracy over biopsy data of from 58 to 63%. These results suggest the usefulness of the proposed method in prostate volume measurement and in PSAD accuracy improvement.

References

- [1] Robertson S, Azizpour H, Smith K, Hartman J (2018) Digital image analysis in breast pathology-from image processing techniques to artificial intelligence. *Translational Research* Vol.194, pp.19–35
- [2] Chen L, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. arXiv: 1706.05587

Development of image processing aimed at improving benign-malignant discrimination processing of breast tumors by CAD

S. Kudo¹, K. Abe², H. Takeo², Y. Nagai³, S. Nawano⁴

¹Graduate School of Kanagawa Institute of Technology, Department of Electronics and Electrical Engineering, Atsugi City, Kanagawa Prefecture, Japan ²Kanagawa Institute of Technology, Atsugi City, Kanagawa, Japan ³National Cancer Center Hospital East, Kashiwa City, Chiba, Japan ⁴Shinmatsudo Central General Hospital, Matsudo City, Chiba, Japan

Keywords Breast tumor, CAD, Image processing, Screening

Purpose

Mammography, an essential tool in the detection of breast tumor, comprises x-ray imaging. Research by Takeshi Inuma [1], however, has shown that younger the age group, the higher the risk of harmful exposure and of reduced lifespan incurred by x-ray imaging. The ability to perform breast tumor benign-malignant discrimination at the screening mammography stage has the advantage of making it possible to eliminate bi-directional imaging, a procedure for benign-malignant discrimination conducted in the diagnostic mammography stage's second-phase detailed examination, and go immediately to biopsy.

Previously, AI-based benign-malignant discrimination developed by the authors achieved an accuracy of approximately 90% for diagnostic images (images captured during the second-phase detailed examination) and of approximately 85% for screening images (images captured during screening). In this study, the image quality of screening images is improved by the combined use of processes for improving contrast and graininess. The aim is to use such processing to bring the level of the image quality of screening images to that of diagnostic images in order to improve the performance of the benign-malignant discrimination of tumor masses at the breast tumor screening stage.

Methods

The differences between screening images and diagnostic images captured by mammography are known to arise from differences in graininess caused by variations in x-ray dosage and from variations in the amounts of radiation scattering caused by the imaging platform [2]. In this study, image processing was developed that improves the graininess and the low contrast of screening images in order to make up the difference in image quality between these two image types, thereby improving the performance of the AI-based benign-malignant discrimination processing of screening images. The procedures followed are described below.

(1) Image quality improvement through combined use of enhancement processing and morphology

After morphological processing, whose purpose is to improve graininess, enhancement processing was conducted to improve contrast. In morphological processing, an opening operation that performs erosion followed by dilation was executed over eight image density stages to reduce graininess noise.

Next, the image was enhanced through an enhancement process that subtracts the image resulting from the opening operation from the original image to obtain high frequency components, which are then added to the original image by density-dependent addition. This reduces the loss of contrast in images caused by radiation scattering.

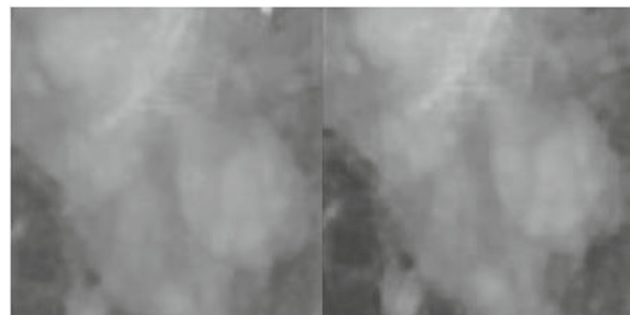


Fig. 1 Changes to breast tumor image effected by image processing (Left:Original image, Right:Processed image)

Figure 1 shows a practical application of this process to a breast tumor image.

(2) Development of AI-based benign-malignant discrimination processing

We developed AI-based benign-malignant discrimination processing using a breast tumor image database comprising benign and malignant images, 250 each, that were subjected to the proposed image processing. Training images were supplemented through data augmentation comprising 42 types of processes. The network architectures used were 4 to 6 middle layer convolutional neural networks (CNN) and VGG16, VGG19, and AlexNet transfer learning. Using the leave-one-out cross-validation (LOOCV) procedure on the test images, we calculated the average accuracy of performance tests that were performed 5 times. The AI network architecture that produced the maximum accuracy was adopted.

Results

Table 1 shows the architecture that achieved the highest accuracy among AI learning results for tumor mass benign-malignant discrimination. AI-based breast tumor benign-malignant discrimination processing that used a database containing images subjected to the proposed image processing demonstrated an improvement in accuracy of 4% over a database of unprocessed images. This performance improvement represents the achievement of high performance in terms of accuracy approaching the 90% accuracy of benign-malignant discrimination processing as applied to diagnostic images.

In addition, although AlexNet transfer learning obtained the highest accuracy for unprocessed images, for the processed images, the 6-middle-layer CNN was the network architecture that achieved the highest accuracy.

Conclusion

In this study, the authors propose and study processing for improving the image quality of breast tumor images used in the development of AI-based breast tumor benign-malignant discrimination processing. The results indicate that accuracy improvement was achieved with the AI-based breast tumor benign-malignant discrimination processing that employed the proposed image processing. Given that this performance improvement resulted in high performance in terms of approaching the accuracy of benign-malignant discrimination processing as applied to diagnostic images, the authors believe that the proposed image processing produced the image quality needed for discrimination in that it was nearly equal to that of diagnostic images. On the other hand, the fact that the AI network architecture that obtained the highest accuracy changed from transfer learning to CNN has implications to the effect that using a database comprising only mammographic images (images that, unlike ordinary photographs, are technical in nature) leads to higher accuracy. This in turn suggests the achievement of further improvement also in the reliability of the learning results. Pertaining to this point, it is necessary to pursue deeper analysis of the qualitative reliability of trained networks along with quantitative performance.

Table 1 Accuracy of AI-based benign-malignant discrimination processing

Screening images (without image processing)	Screening images (with image processing)	Diagnostic images
85%	89%	90%
AlexNet	6-middle-layer CNN	VGG16

References

- [1] Iinuma T (2004) Risk–benefit analysis of 2-year interval mammography screening, 14th Annual Meeting of the Japanese Association of Breast Tumor Screening
- [2] Soma KUDO, Kazuya ABE, Hideya TAKEO, Yuuichi NAGAI, and Shigeru NAWANO, Performance comparison and analysis of screening and diagnosis in AI-based benign-malignant discrimination processing of breast tumors, 40th JAMIT Annual Meeting (JAMIT 2021), P1–3, October 2021

Detection of micro-nodules in Pneumoconiosis in 3D CT images using 3D U-net

M. Matsuhira¹, Y. Hashimoto¹, H. Suzuki¹, Y. Kawata¹, Y. Ohtsuka², T. Kishimoto³, K. Ashizawa⁴, N. Niki⁵

¹Tokushima University, Faculty of Science and Technology, Tokushima, Japan ²Hokkaido Chuo Rosai Hospital, Hokkaido, Japan ³Okayama Rosai Hospital, Okayama, Japan ⁴Nagasaki University, Nagasaki, Japan ⁵Medical Science Institute Inc, Tokushima, Japan

Keywords CAD, 3D U-Net, Pneumoconiosis, Micro nodules

Purpose

Pneumoconiosis is an occupational respiratory disease caused by inhaling dust into the lungs, and 240,000 people in Japan undergo radiation pneumoconiosis screening annually. There is a staging system for the diagnosis of pneumoconiosis [1], and type 1/0 and above are eligible to be recognized as occupational accidents. It is important to accurately diagnose type 0/1 and 1/0. For this purpose, quantification of micro-nodules in pneumoconiosis is necessary, and detection of micro-nodules by 3D CT imaging is promising for differential diagnosis. Here, we propose a method to detect micro-nodules from 3D CT images with high accuracy using 3D U-Net [2].

Methods

This study was approved by the Institutional Review Board of Nagasaki University. 3-D CT images were acquired with a GE LightSpeed VCT. Scanning was performed at 120 kV, 167–698 mA, slice thickness 1.25 mm, 512 × 512 matrix, pixel size 0.527 mm to 0.781 mm, reconstruction interval 1.25 mm, LUNG convolution kernel. Micro nodules in 3D CT images were extracted manually. The display was set to window level 500 and window width 1500. The reading procedure was performed in the following order: right lung apex, right lung base, left lung apex, left lung base. This procedure was repeated two to three times. Based on the guidelines of the Ministry of Health, Labor and Welfare of Japan, the medical doctors diagnosed 15 stages of pneumoconiosis: type classification 0/–, 0/0, 0/1, 1/0, 1/1, 1/2, 2/1, 2/2, 2/3, 3/2, 3/3, 3/+ , 4A, 4B, and 4C. Fourteen cases with type 0/1, 1/0, 1/1, 1/2, and 2/1 were used as training data. Two cases with type 0/1 and 1/1 were used as test data. We used a computer with a GPU (Quadro GV100 (32 GB); NVIDIA Corporation), Pytorch 1.9.0 + cu111 and Monai 0.7.0 were used as libraries for machine learning.

1. Training

Lung lobe classification was performed using 3D U-Net. The training was done using 300 cases different from the above data. We targeted the micro nodules of pneumoconiosis in the right upper lobe by filling the area of the CT image except the right upper lobe field with the background value (– 2048). Pulmonary blood vessels in the right upper lobe were extracted by thresholding. The training data of the pulmonary blood vessels were added to the training data of the micro nodules. We used the classes provided by the MONAI platform. Cropping size used (24, 24, 24) and sampled to 500 for each

training image. The number of epochs is 1200. The 3D U-Net was trained using the input images and training data.

2. Test

The right upper lobe region is extracted from the test data and input to 3D U-Net. The output of 3D U-Net included labels of pulmonary blood vessels and granular shadows. The detection result was only micro nodules.

Results

The detection results of micro-nodules in right upper lobe for two cases were evaluated using the gold standard. The case of type 0/1 had TP:25, FP:4, and FN:1, the sensitivity 96.2% and the precision 86.2%. The case of type 1/1 had TP:130, FP:41, and FN:6, the sensitivity 95.6% and the precision 76.0%.

Conclusion

We developed a method for detecting micro-nodules in pneumoconiosis from 3D CT images using 3D U-Net. When we added pulmonary blood vessels information to the model, we were able to obtain results with less FP. Future work is to improve 3D U-Net to reduce FP and FN, to increase the number of cases, and to apply the method to all lung fields.

References

- [1] International Labour Organization, "Guidelines for the use of the ILO international classification of radiographs of pneumoconiosis, revised edition 2011," 2011.
- [2] Cicek O, Abdulkadir A, Lienkamp SS, et al. "3D U-Net: Learning dense volumetric segmentation from sparse annotation." Proc. Medical Image Computing and Computer-Assisted Intervention-MICCAI, pp.424–432, 2016.

Prediction scheme of the short-term urinary continence using robotic surgery images

Teramoto¹, F. Oba¹, M. Sumitomo², W. Nakamura², R. Shiroki², K. Saito¹, H. Fujita³

¹Fujita Health University, Graduate School of Health Sciences, Toyoake, Japan ²Fujita Health University, Department of Urology, School of Medicine, Toyoake, Japan ³Gifu University, Faculty of Engineering, Gifu, Germany

Keywords Robotic surgery, Deep learning, Prediction, Urinary continence

Purpose

Prostate cancer is the most prevalent cancer in males. The treatment of prostate cancer includes surgery, radiation therapy, and drug therapy, which are selected according to the patient's condition and wishes. As for surgical treatment, robotic surgery is now widely used and has many advantages, such as a short postoperative recovery period due to its minimally invasive procedure. However, urinary incontinence, which has been one of the challenges of surgical treatment, has not been completely solved. Although most robotic surgery patients recover from urinary incontinence after surgery, some patients still have poor urinary incontinence and need to change their urine pads several times a day. We have previously developed a method to predict postoperative urinary incontinence from preoperative MR images and clinical information, and obtained a prediction accuracy of 70% [1]. Here, urinary incontinence affects not only the preoperative patient condition, but also the intraoperative condition. Therefore, in this study, we developed a method for predicting urinary incontinence by analyzing intraoperative videos using deep learning and machine learning.

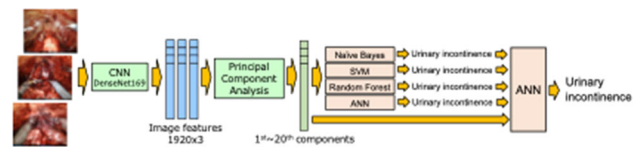


Fig. 1 An overview of the method for predicting urinary incontinence using robotic surgery videos

Methods

For this study, we collected videos during robotic surgery and postoperative urinary incontinence information from 40 prostate cancer patients who underwent robotic surgery at Fujita Health University Hospital. Robotic surgery was performed by nine surgeons using da Vinci Si or Xi (Intuitive Surgical, Sunnyvale, CA, USA). The number of times the urine pad was changed per day was used as a criterion for good or poor urinary incontinence. Patients with 0 or 1 change per day were considered to have good urinary incontinence, and those with 2 or more changes per day were considered to have poor urinary incontinence. From the intraoperative video, three still images were extracted including before and after prostatectomy. An overview of the method for predicting urinary incontinence using robotic surgery videos is shown in Fig. 1. First, three color images were fed to DenseNet169, which was pre-trained by ImageNet database, and 5760 features were extracted by global average pooling in the last convolutional layer. And then, to compress the features, 20 principal components per case were obtained by principal component analysis. They were subjected to machine learning (naive Bayes (NB), support vector machine (SVM), random forest (RF), artificial neural network (ANN)) to calculate the probability of bad urinary incontinence. To further improve the prediction accuracy, these four probabilities and 20 principal components were fed into another ANN as a cascade prediction method to calculate the probability of poor urinary incontinence.

Results

We evaluated the prediction accuracy of urinary incontinence by leave-one-out cross validation from 40 patients undergoing robotic surgery. The balanced accuracies of prediction by four machine learning methods were 51.6% (NB), 71.9% (SVM), 57.8% (RF), and 56.3% (ANN). When those results were given to the ANN to make cascade predictions, the balanced accuracy improved to 81.3%.

Conclusion

In this study, we have developed a method for predicting urinary incontinence using intraoperative videos using deep learning and machine learning. Experimental results indicate that proposed method will be useful to predict urinary incontinence in prostate cancer patients undergoing robotic surgery.

References

- [1] M. Sumitomo, A. Teramoto, R. Toda, N. Fukami, K. Fukaya, K. Zennami, M. Ichino, K. Takahara, M. Kusaka, R. Shiroki, "Deep learning using preoperative MRI information to predict early recovery of urinary continence after robot-assisted radical prostatectomy," International Journal of Urology, Vol.27, pp.922–928, 2020.

Unsupervised deep feature generation from 9ch 2.5-dimensional images for detecting cerebral aneurysm on MR angiography image

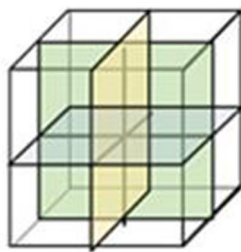
K. Hosoda¹, M. Nemoto², H. Yamaguchi¹, Y. Kimura², N. Hayashi³
¹Kindai University, Graduate School of Biology-Oriented Science and Technology, Kinokawa-shi, Wakayama, Japan ²Kindai University, Biology-Oriented Science and Technology, Kinokawa-

shi, Wakayama, Japan ³University of Tokyo Hospital, Department of Computational Diagnostic Radiology and Preventive Medicine, Bunkyo-ku, Tokyo, Japan

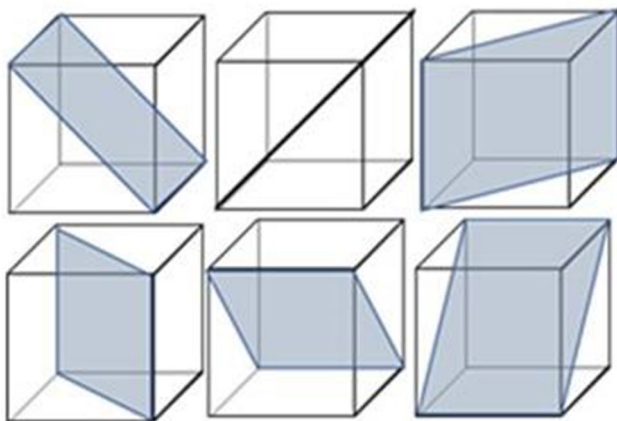
Keywords Deep Feature Generation, 9ch 2.5-Dimensional Images, Convolutional autoencoder, MR angiography

Purpose

This study proposes an unsupervised deep feature generation method by nine channels 2.5-dimensional (9ch 2.5D) image analysis with a deep convolutional autoencoder (CAE). The 3ch 2.5D image analysis with deep convolutional neural networks has promise for various applications in computer-aided detection of 3D medical images [1]. The 3ch 2.5D image is an efficient compression of a 3D image and includes axial, coronal, and sagittal slices of the 3D image. The smaller data length of the 2.5D image than that of the 3D image often enhances the convergence of deep learnings with a small training dataset. Our previous research [2] also used the 3ch 2.5D image analysis based on the unsupervised deep learning. However, the 3ch 2.5D image analysis cannot recognize objects that are not shown in the three cross-sectional images. The proposed method uses partial 9ch 2.5D image patches, including the axial, coronal, sagittal, and six oblique slices of the 3D image patches. The addition of oblique slices



(a) Conventional three slices of 2.5D image [1]



(b) Newly introduced six oblique slices

Fig. 1 Slices for 9ch-2.5D image

makes possible the analysis of objects that do not exist in the 3ch image patches. We apply the proposed method to cerebral aneurysm detection on MR angiography (MRA) images to evaluate the performance improvement by increasing the number of slices included in the 2.5D image patches.

Methods

An input MRA image is scaled to a 0.469 mm isotropic voxel as a preprocess. The voxel intensity is normalized based on the intensity of sagittal slices near the center of the head. The cerebral arterial vascular region is extracted from the preprocessed image by thresholding and morphological processes. The partial vascular 3D image patches sized $32 \times 32 \times 32$ voxels are extracted from the vascular region with 16 voxel intervals. These patches are as aneurysm candidates. Each 3D vascular image patch is converted into a 9ch 2.5D image patch consisting of three typical slices (axial, coronal, sagittal) and six slices in the diagonal direction, shown in Fig. 1. Those slices pass through the center of gravity of the 3D image patch. Multiple image features for the aneurysm candidates are extracted from the results of a deep CAE analysis for the 9ch 2.5D image patches. The deep CAE including three convolution layers, two max-pooling layers, and a full connection layer has been trained on 331 cases of aneurysm-free MRA image dataset and outputs latent variables. The feature set includes (1) the latent variables by the deep CAE, (2) the Mahalanobis distance from the normal vascular class data in the latent variable space, and the voxel value statistics on a difference image between the original 2.5D image and a reproduced image from the latent variables. The image reproduction is by a decoder with the deep CAE symmetric structure. All the aneurysm candidate patches are classified by an Ada-boosted classifier ensemble with the image features. The proposed method is evaluated by three-fold cross-validation with 450 cases of the brain MRA images with at least one aneurysm and are taken at the University of Tokyo Hospital. In the evaluation, the performance of aneurysm candidate classification by the proposed method is also compared using the features extracted from 3ch 2.5D patches.

Results

In the aneurysm candidate patch extraction, 511.14 patches were extracted from a MRA image on average. The candidate classification result by the proposed method compared to the classification by the features from 3ch 2.5D image are shown in Table 1. When the number of false positives (FPs) per case was three, the detection sensitivity of cerebral aneurysm increased from 76.7 to 88.0%.

Conclusion

We proposed the CAE-based deep feature generation by 9ch 2.5D data image analysis. We applied the proposed method to cerebral aneurysm detection. The evaluation results showed the usefulness of the proposed method with over 10% sensitivity improvement. We plan to introduce 2D projection techniques such as the maximum intensity projection and the minimum intensity projection for 9ch 2.5D image conversion. We also plan to apply the proposed method to other lesion detection tasks (Fig. 1).

Table 1 Aneurysm detection performances by three-fold cross-validation with 450 MRA images (mean \pm s.d.)

	Sensitivity at 3 FPs/case	AUC
9ch (proposed)	0.880 \pm 0.010	0.993 \pm 0.001
3ch	0.767 \pm 0.020	0.981 \pm 0.001

References

- [1] H.R. Roth, L. Lu, A. Seff, K.M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, R.M. Summers (2014) A New 2.5D Representation for Lymph Node Detection using Random Sets of Deep Convolutional Neural Network Observations. MICCAI 2014: Medical Image Computing and Computer-aided Intervention, LNCS 8673, 520–527.
- [2] K. Ushifusa, M. Nemoto, Y. Kimura, T. Nagaoka, T. Yamada, N. Hayashi (2021) Detection of cerebral aneurysms on MR angiography using generated features by unsupervised deep learning for multiple 2.5-dimensional images. IFMIA 2021: Int Forum on Medical Imaging in Asia.

Automated analysis of carotid plaque using ultrasound video images: fully automated recognition of plaque regions using U-Net

G. Hirano¹, H. Takai², Y. Higashi¹, A. Teramoto¹, S. Watanabe³, K. Sugimoto¹, S. Matsumoto³, K. Saito¹, H. Fujita⁴

¹Fujita Health University, Graduate School of Health Sciences, Toyoake, Japan ²Fujita Health University Hospital, Department of Radiology, Toyoake, Japan ³Fujita Health University, Department of Comprehensive Strokeology, School of Medicine, Toyoake, Japan ⁴Gifu University, Faculty of Engineering, Gifu, Germany

Keywords ultrasound, carotid plaque, segmentation, deep learning

Purpose

The leading cause of death worldwide is ischemic heart disease, and the second leading cause is stroke. The risk factor for these diseases is the formation of plaque due to the progression of atherosclerosis. Ultrasonography, MRI, and angiography are used to diagnose the presence and differential diagnosis of plaque. Because of the physical and mental burden on the patient of performing these multiple detailed examinations, it is ideal to examine plaque using only ultrasonography, which is used early in the diagnosis of plaque. Plaque can be classified into three major categories: low echoic plaque, isoechoic plaque, and calcified plaque. In particular, low echoic plaque with fat and hemorrhage is called unstable plaque, which can cause heart disease and stroke. Judgment of the type of plaque influences the decision of the treatment plan. However, it is currently subjective to the doctor. If plaque can be analyzed from ultrasound images, high-risk cases can be detected quickly and treated promptly, leading to improved treatment accuracies. In order to analyze plaque, it is important to accurately identify the plaque region. In this study, we developed automated extraction of carotid plaque as a preliminary study to analyze plaque using ultrasound video images.

Methods

For this study, we collected 20 cases of video images in which the presence of plaque was confirmed by carotid ultrasonography. All cases were taken in B-mode with long-axis cross-sections where plaque was largely visible. The B-mode video image was converted to 15 images per second in the video image into still images, and then they were resized to 436×398 pixels. To extract carotid plaque from ultrasound images, we introduced U-Net [1], which is one of the FCN (Fully Convolutional Networks) and widely used in the field of medical image processing. The structure of U-Net is shown in the upper part of Fig. 1. The U-Net consisted of five layers of encoders and decoders. An ultrasound image was given to the input layer of U-Net, and the plaque region was extracted (labeled) from the output.

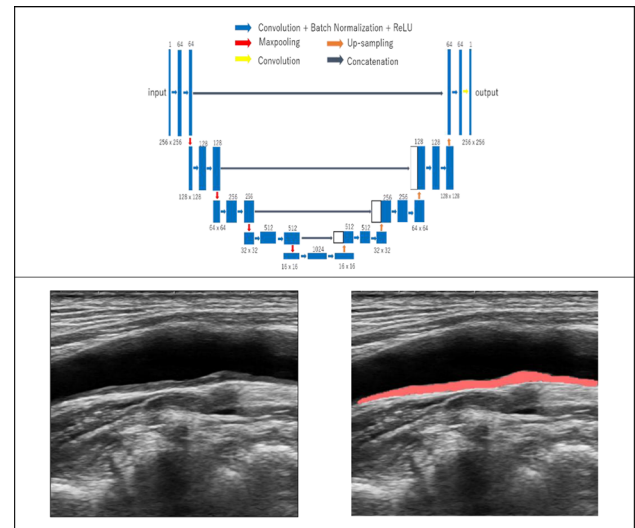


Fig. 1 The structure of U-Net(upper) and the example of an input image(bottom left) and the output image added to the input image(bottom right)

The correct label images used for training of U-Net were created using in-house software and checked by radiological technologist. The effectiveness of this method was evaluated by the dice similarity coefficient using the output image of U-Net and the correct label image. In addition, we confirmed the recognition of the shape of the plaque area. In the experiments, Intel Corei5-8500 was used for the CPU and used NVIDIA GeForce RTX 2070 SUPER for the GPU. TensorFlow and Keras were used for the deep learning frameworks. The number of epochs for training of U-Net was 100, and a batch size was set to 4.

Results

The five-fold cross-validation method was used to verify the extraction performance. The mean dice similarity coefficient of plaque area for 1945 images of 20 cases was 0.596. Examples of an input image (left) and the output image added to the input image (right) are shown in the bottom of Fig. 1. Cases in which the boundary between the plaque and the vessel wall was clear, or cases in which the plaque was spread thinly and extensively against the vessel wall, showed good extraction accuracy. Although some extraction defects occurred in cases where the plaque was similar to the vessel wall or in plaques with low intensity, it was possible to identify the shape of the plaque.

Conclusion

In this study, we developed a method to extract carotid plaque from ultrasound images using U-Net. As a result of the experiment, it was possible to extract the plaque region and identify its shape, confirming the effectiveness of this method.

References

- [1] Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015 2015, pp.234–241.

Automatic CT segmentation and classification of pancreatic cancer including uncertainty maps

M. Riera i Marín^{1,2}, J. García¹, J. Rodríguez-Comas³, M. A. González Ballester^{2,4}

¹Sycal Technologies SL, Software department, Barcelona, Spain

²Universitat Pompeu Fabra (UPF), Barcelona, Spain ³Sycal Technologies SL, Scientific department, Barcelona, Spain ⁴ICREA, Barcelona, Spain

Keywords Deep Learning, segmentation, uncertainty, pancreatic cancer

Purpose

Pancreatic cancer is one of the most lethal cancers, with less than five months of life expectancy after its diagnosis. Pancreatic Cystic Lesions (PCL) are frequent and difficult to detect on Computational Tomography (CT) scans and they can evolve to pancreatic cancer. In this work we present a method to detect and classify PCLs between potentially malignant and non potentially malignant.

We propose a Deep Learning (DL) architecture in order to segment both the pancreas and the PCLs, followed by a post processing pipeline in order to improve the segmentation results. Crucially, we also implemented a method to compute uncertainty segmentation maps, thus enhancing the interpretability and robustness of our method. Finally, a classification procedure is developed in order to classify the PCLs between potentially malignant and non potentially malignant.

Methods

Our detection method consists of an image segmentation architecture with additive attention gates [1] to segment the target pancreatic cystic lesions on the input image test. Soft-Tissue normalization is used in the interest of highlighting the organs (i.e., the pancreas). A central crop of the abdomen is applied to restrict the area of segmentation of the pancreatic organ.

For inference, Test Time Augmentation (TTA) is used in order to have multiple predictions and average them to obtain a final one. Afterwards, the predicted label is passed through a post processing pipeline, which has two main parts: (1) Preserving the biggest lobe segmented as pancreas, and (2) Eliminating any segmented cyst that is in the edges of the abdomen or that is not in contact with the pancreas.

The uncertainty maps are probabilistic maps that give an idea of how sure the pixel-wise predictions are. In order to compute these uncertainty maps the method presented in [2] is used. It consists of generating multiple predictions for the same image and superimposing them, computing the 20th and 80th percentile of all those multiple predictions.

For the cyst classification three steps take place:

- Metadata; the age and sex of the patient.
- Hounsfield Units (HU) to determine the solidity of the cyst and its possible malignancy.
- Morphological features, which help characterize the cyst, such as the position of the cyst with respect to the pancreas, the size of the detected lesion, etc.

Each of the three steps give a partial classification which in the end are merged all together to obtain a final classification. The metadata step uses information found on the literature, showing which cysts are more probable depending on the age and sex of the patient. The second step, which uses information obtained from the image through image processing, consists of a Machine Learning (ML) method called Gradient Boosting Random Forest. The third step uses multiple image processing techniques such as equalization, denoising

Table 1 Performance metrics for the cyst segmentation. These metrics are obtained per patient

segmentation	Sensitivity	Specificity
with TTA	0,74	0,77
processing on	0,88	0,84

and thresholds in order to find characteristics that give information about the cyst.

Results

Our training dataset consists of 616 CT images, belonging to 71 patients with confirmed PCLs and 30 control patients with non-diagnosed lesions in the pancreas. The inference dataset consists of 150 CT images.

Our segmentation method performs with a sensitivity of 0.74 and a specificity of 0.77 before applying the post processing, which can be observed in Table 1. Once our post processing pipeline is applied, the metrics increase to 0.88 and 0.84 of sensitivity and specificity respectively, which clearly shows the efficiency of the post processing method.

In Fig. 1, an example of the segmentation and uncertainty maps (for both the pancreas and the cyst) is shown. It can be easily observed that the highest uncertainty appears on the edges of the segmentations and that, in the pancreatic uncertainty map, there is no uncertainty at all within the cyst area.

The accuracy obtained for the classification algorithm is 0.70. Further work will go in the direction of enlarging the dataset, making it more balanced and heterogeneous with the different types of cyst contemplated in this work, leading to improved classification accuracy.

Conclusion

We developed a DL segmentation pipeline for the pancreas and the cystic lesions, including the computation of the respective uncertainty maps. Thanks to these uncertainty maps it can be observed that most of the ambiguity in the predictions appears on the edges of the cyst and the pancreas.

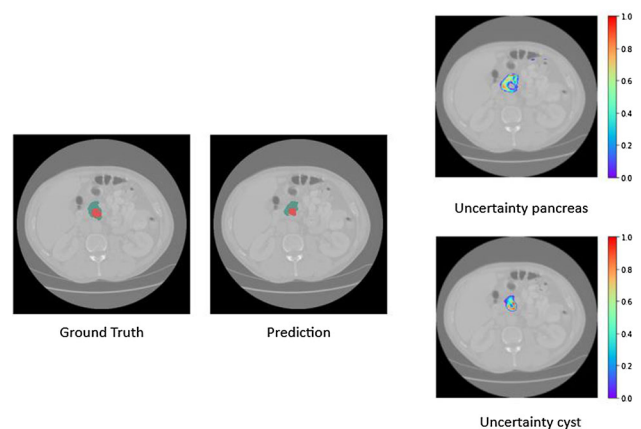


Fig. 1 Example segmentation results, including uncertainty maps. Ground truth (left figure), prediction (middle figure): pancreas (green) and pancreatic cystic lesion (red) are depicted. Uncertainty maps (right figure): uncertainty pancreas (upper figure), uncertainty cyst (lower figure)

The usage of uncertainty maps is three fold; on one hand, they can be used to obtain a qualitative evaluation of the segmentations; on the other hand, it can be used to spot areas where the ground truth might not be accurate enough leading to a possible correction; finally, these uncertainty maps can be used to improve the performance of the training while it is going on, which is one of the future and final goals of this work.

Preliminary results show that the classification tends to over classify the potentially malignant cysts and under classify the benign cysts. This is a good start considering that in this case it is better to have a FP than a FN. However, future work will focus on improved classification schemes as well as improving the dataset.

This work was funded by the Industrial Doctorates program (Generalitat de Catalunya).

References

- [1] Oktay O., Schlemper J., Le Folgoc L., Lee M., Heinrich M., Misawa K., Mori K., McDonagh S., Y Hammerla N., Kainz B., Glocker B., Rueckert D., (2018, April): Attention U-Net: Learning Where to Look for the Pancreas, arXiv: 1804.03999 [cs.CV]
- [2] LaBonte T., Martinez C, Roberts S. A., (2020, April): We know where we don't know: 3D Bayesian CNNs for credible geometric uncertainty, arXiv: 1910.10793 [eess.IV]

Prediction of extracorporeal shock wave lithotripsy outcome by combined analysis of CT image textures and patient factors

Y. Nakamae¹, M. Nemoto¹, Y. Kimura¹, S. Yamashita¹, R. Deguchi¹, Y. Kohjimoto¹, I. Hara¹

¹Kindai University, Graduate School of Biology-Oriented Science and Technology, Wakayama Kinokawa City, Japan

Keywords Urolithiasis, ESWL, Image textures, Patient factors

Purpose

Urolithiasis is a disease in which stones form in the urinary tract and is caused by lifestyle factors such as diet and obesity. Extracorporeal shock wave lithotripsy (ESWL) is a treatment to crush calculi by shock waves from the body surface. The ESWL is a first-line treatment choice for upper urinary tract calculi. It has advantages of safety advantage and non-invasiveness over transurethral lithotripsy and percutaneous lithotripsy [1]. However, ESWL has some problems: the lower treatment success rate than other treatments and the requirement of several weeks to several months of follow-up. The failure of ESWL with extended follow-up delays the application of other treatments. It increases the risk of urinary tract infection and renal function disease. Advance prediction of ESWL outcomes is clinically meaningful. We propose a method to estimate the outcome of ESWL treatment for ureteral stones by combined analysis of CT image textures and patient factors. The proposed method is experimentally evaluated using clinical data.

Methods

Figure 1 shows the flowchart of the proposed method. The input data consists of a preoperative CT image, target stone position, and patient factors: patient age, skin-to-stone distance, and ureteral wall thickness. An experimental urologist has measured these patient factors. At first, the 3D stone region on the CT image is extracted based on the

stone position. Next, 11 CT texture features are measured on the stone region. The texture features include the statistics of the CT value in the stone region and the evaluation values of the CT value gradient concentration in the direction of the stone center. The gradient concentration features are from a histogram of the deviation angle between the direction of the stone center and the CT value gradient vector with a class width of 45 degrees. The relative frequencies of each four classes are used as the features. Finally, the ESWL outcome of the target stone is predicted by a support vector machine (SVM) with the 11 CT texture features, stone volume, and three patient factors. A radial basis function kernel is applied to the SVM in this study.

An evaluation experiment uses data from 171 patients with a single ureteral stone. These patients were taken the preoperative CT image and underwent ESWL treatment at Wakayama Medical University Hospital between January and November 2009. The CT images have various pixel spacing from 0.47×0.47 mm to 0.98×0.98 mm and various slice thicknesses from 1.25 mm to 10 mm. We define a stone-free (SF) case, which is a successful stone removal case, as no residual stone larger than 4 mm was found on CT images taken within three months after ESWL. The prediction performance is evaluated by five-fold cross-validation.

Results

The evaluation experiment showed that the mean area under the ROC curve was 0.742 with 0.038 standard deviations. When the prediction specificity for stone removal failure cases was 0.733, the mean sensitivity for SF cases was 0.692 with 0.056 standard deviations.

Conclusion

We proposed the method to predict the outcome of ESWL by combined analysis of CT image textures and patient factors. The evaluation results suggest the usefulness of the proposed method. The prediction accuracies of the proposed method in this experiment were based on CT images of various resolutions. The results were comparable to the prediction accuracy of a previous study that used only high-resolution CT images [2]. We plan to validate the proposed method by higher resolution CT images. We also plan to apply deep learning with a large-scale CT dataset.

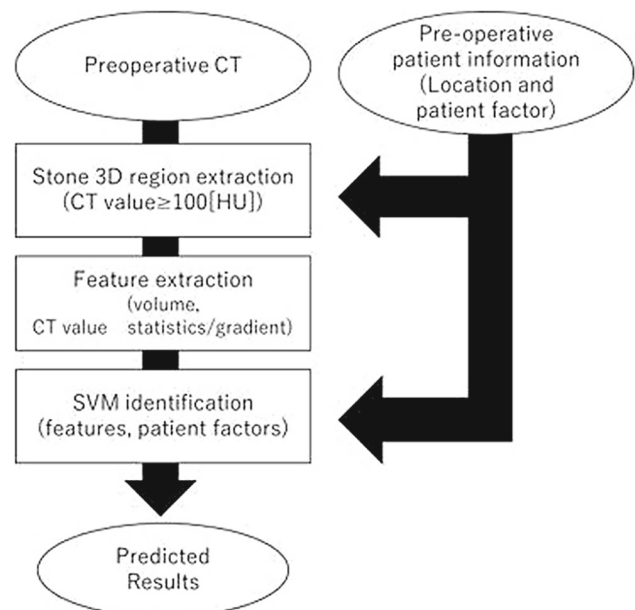


Fig. 1 The flowchart of the proposed method

References

- [1] S Yamashita, Y Kohjimoto, Y Iwahashi, T Iguchi, A Iba, S Nishizawa and I Hara (2018) Three-dimensional mean atone density measurement is superior for predicting extracorporeal shock wave lithotripsy success“, International Journal of Urology 26(2), 185–191.
- [2] M. Mannil, J. von Spiczak, T. Hermanns, C. Poyet, H. Alkadhi, C.D. Fankhauser (2018) Three-Dimensional Texture Analysis with Machine Learning Provides Incremental Predictive Information for Successful Shock Wave Lithotripsy in Patients with Kidney Stones. The Journal of urology, 200(4), 829–836.

Development of Cerebral Endovascular Treatment Remote Control System in Smart Cyber Operating Theater

T. Ishikawa¹, K. Yamaguchi¹, T. Funatsu¹, Y. Moteki¹, Y. Muragaki¹, T. Kawamata¹
¹Tokyo Women’s Medical University, Neurosurgery, Tokyo, Japan

Keywords remote treatment, neurosurgery, endovascular, operating room

Purpose

At our hospital, we have started the operation of ”Smart Cyber Operating Theater (SCOT)“, which has a platform that can collect treatment / surgical equipment and information obtained from them in real time. In SCOT, various information can be aggregated and displayed by a system (OPeLiNK system) that fuses image data and actual treatment data. By sharing such information with remote areas, we aim to ”remote treatment by supervisors“ and ”create educational content.“

Methods

Using the OPeLiNK system at SCOT, a display that aggregates perspective images, surgical field images, patient biometric information, preoperative simulation information, etc. is placed in the operating room to ”visualize“ the progress of surgery. In addition, multiple images are fused and displayed in a time-integrated state during surgery. In addition, comment registration and event setting will be done by voice during the operation. Use these to look back at post-operative conferences and create educational content. Furthermore, by constructing a similar system (strategy desk) in a remote location, the surgical status can be confirmed in a fusion display in a remote location, see Fig. 1, and an interactive function and writing on the screen can be added, [1].

Results

Currently, the OPeLiNK system at SCOT is in place, and it has become possible to display multiple images in a time-integrated state during surgery. In addition, we were able to register comments by voice and set up events, and were able to create educational content. In addition, the network between the operating room and the faculty office has been completed, and a system has been set up aiming for full-scale operation of telemedicine.

Conclusion

In SCOT, various information can be aggregated and displayed and useful.

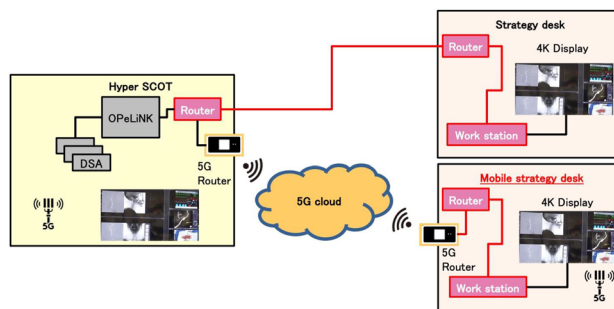


Fig. 1 Near future of mobile strategy desk

References

- [1] Orru E, Marosfoi M, Patel N V, Coon AL, Wald C, Repucci N, Nicholson P, Pereira VM, Wakhloo AK. International teleproctoring in neurointerventional surgery and its potential impact on clinical trials in the era of COVID-19: legal and technical considerations. J. Neurointerv. Surg. 2021;13: 1022–1026.

Data augmentation using CycleGAN for an artificial intelligence system to support surgeon decision-making

Misako Kimura¹, Yusuke Matsunobu², Ken’ichi Shinozuka², Toshiya Kamiyama⁴, Kohei Ebe⁵, Atsuro Fujinaga³, Hiroaki Nakanuma³, Masahiro Kawamura³, Yuichi Endo³, Tsuyoshi Etoh³, Masafumi Inomata³, Tatsushi Tokuyasu²

¹Department of Information and Systems Engineering, Graduate School of Engineering, Fukuoka, Japan ²Institute of Technology, Fukuoka, Japan ³Faculty of Information Engineering, Department of Information and Systems Engineering, Fukuoka Institute of Technology, Fukuoka, Japan ⁴Faculty of Medicine, Department of Gastroenterological and Pediatric Surgery, Oita University, Oita, Japan ⁵AI Tech Research, ATR Advanced, Olympus Corporation, Tokyo, Japan ⁶Information Aided Medical Solutions Development, Application Software Engineering, Olympus Medical Systems Corporation, Tokyo, Japan

Keywords Keywords: Artificial intelligence, bile duct injury, CycleGAN, data augmentation

Table 1 Average DICE coefficients for each landmark

Evaluation Dataset	Landmark	Original model	New model	p-value
Normal LC	CBD	0.607	0.631	0.894
	CD	0.304	0.345	0.246
	S4	0.526	0.520	0.657
	RS	0.434	0.423	0.827
Abnormal LC	CBD	0.445	0.543	0.420
	CD	0.270	0.393	0.013*
	S4	0.426	0.509	0.150
	RS	0.288	0.309	0.622

* p < 0.05

Purpose

The cognitive vulnerability of humans is thought to be a main cause of bile duct injury incidents during laparoscopic cholecystectomy (LC). This study developed an artificial intelligence system to intraoperatively identify four anatomical structures as landmarks to prevent bile duct injury using videos of LC with few abnormal features [1]. The constructed YOLOv3 model was able to show four landmarks for cases with few abnormal findings. However, we confirmed that the YOLOv3 model could not sufficiently identify the landmarks of LC cases with acute cholecystitis in a clinical performance test [2]. To apply an artificial intelligence model in severe LC cases, it is necessary to additionally prepare the annotated data for such severe LC cases; however, it can be assumed that the inter-annotator differences will be high in the annotation data because of low anatomical visibility. Because over 90% of LCs are performed on benign disease, it may not be easy to collect enough LC videos of severe cases. To address these issues, this study proposes an effective way to augment the annotated data.

Methods

In this study, we used the CycleGAN deep-learning model to generate the abnormal features of acute cholecystitis in the endoscopic camera images of normal LC cases. We collected 256 surgical videos of LC performed at Oita University Hospital from 2011 to 2021, and selected the 26 LC videos with abnormal features from the 256 videos. Ten CycleGAN models were constructed using nine LC cases with acute cholecystitis and one case with gangrenous cholecystitis out of 26 videos. To examine the effectiveness of our proposal, we constructed two YOLOv3 models, one using our annotated dataset and one using the new annotated dataset augmented by the CycleGAN models. To evaluate the performances of the YOLOv3 models, we performed an objective evaluation by Dice coefficients and a subjective evaluation by three expert surgeons on a 5-point scale.

Results

The CycleGAN models successfully changed the appearance of the anatomical structures to that of the abnormal features of acute cholecystitis. The YOLOv3 model trained on the annotation data augmented by the proposed method maintained its performance on the normal LC cases. Moreover, it performed better on the abnormal LC cases than the original model. Table 1 Optical flow algorithms which are included in our evaluation, and mean value of interpolation error (IE) over 100 test sequences. Algorithm Training data / parameter sets included in our study. Mean IE. Farneback Default parameters as used in OpenCV implementation. 32.02. Dual TV L1 Default parameters as used in OpenCV implementation. 30.02. SPyNet (Sintel) trained on FlyingChairs, fine-tuned on MPI-Sintel. 28.41. SPyNet (Base) trained on FlyingChairs. 28.00. FlowNet 2 (Sintel) trained on FlyingChairs/FlyingThings3D, fine-tuned on MPI-Sintel. 27.82. LiteFlowNet (Sintel) trained on FlyingChairs/FlyingThings3D, fine-tuned on MPI-Sintel. 27.52. PWC-Net (Sintel) trained on FlyingChairs/FlyingThings3D, fine-tuned on MPI-Sintel. 27.51. PWC-Net (Base) trained on FlyingChairs/FlyingThings3D. 27.43. FlowNet 2 (Base) trained on FlyingChairs/FlyingThings3D. 26.97. LiteFlowNet (Base) trained on FlyingChairs/FlyingThings3D. 26.97. Deepflow (Middlebury) DeepMatching, Middlebury parameters. 26.93. Deepflow (Sintel) DeepMatching, MPI-Sintel parameters. 26.71. Table 1 shows the Dice coefficients obtained by applying both models to the evaluation data. For the normal LC images, the average Dice coefficients for the CBD and CD were improved; however, no statistically significant difference was confirmed. For the abnormal LC images, although there is a difference in the change in the average Dice coefficient depending on the landmark, no statistically significant difference was confirmed except for the results for the CD landmark.

Conclusion

We used CycleGAN to augment an annotated dataset consisting of LC cases with few abnormal features to train a YOLOv3 model for

landmark identification in LC cases with acute cholecystitis. In both the objective and subjective evaluations, the effectiveness of the CycleGAN-based data augmentation with respect to the accuracy of landmark identification and for preventing BDI were confirmed. However, sufficient data to verify the effectiveness of the proposed method has not been prepared, and we plan to increase the evaluation images to verify our proposal in the future.

References

- [1] Tokuyasu T, Iwashita Y, Matsunobu Y, Kamiyama T, Ishikake M, Sakaguchi S, Ebe K, Tada K, Endo Y, Etoh T, Nakashima M, Inomata M (2020) Development of an artificial intelligence system using deep learning to indicate anatomical landmarks during laparoscopic cholecystectomy. *Surgical Endoscopy*. <https://doi.org/10.1007/s00464-020-07548-x>
- [2] Inomata M, Fujinaga A, Nakanuma H, Endo Y, Etoh T, Tokuyasu T, Kitano S (2021) Development of endoscopic surgery navigated by artificial intelligence. *Surgical Technology International*. <https://doi.org/10.52198/21.STI.39.CR1432>

Evaluation of Optical Flow Methods on Neurosurgical Microscope Video Data

M. Philipp^{1,2}, M. Gutt-Will³, S. Saur¹, A. Mathis³, A. Raabe³, F. Mathis-Ullrich²

¹Department of Information and Systems Engineering, Graduate School of Engineering, Fukuoka Institute of Technology, Fukuoka, Japan ²Faculty of Information Engineering, Department of Information and Systems Engineering, Fukuoka Institute of Technology, Fukuoka, Japan ³Faculty of Medicine, Department of Gastroenterological and Pediatric Surgery, Oita University, Oita, Japan ⁴AI Tech Research, ATR Advanced, Olympus Corporation, Tokyo, Japan ⁵Information Aided Medical Solutions Development, Application Software Engineering, Olympus Medical Systems Corporation, Tokyo, Japan ¹Carl Zeiss Meditec AG, Oberkochen, Germany ²Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany ³Department of Neurosurgery, University Hospital Bern, Bern, Switzerland

Keywords Optical flow, Neurosurgery, Evaluation, Accuracy, Video data

Purpose

During neurosurgical interventions, the treatment of fine nerve structures in the brain or spine requires a surgical microscope. In the microscope's field of view, neurosurgical treatment steps are focusing mostly on a small region of the image. Significant contextual information regarding the treatment step can be extracted from this image region. The information is represented by spatio-temporal variations in the microscope video.

In computer vision, spatio-temporal variations are often captured by optical flow. Optical flow represents the apparent pixel-wise motion field between consecutive video frames. In practice, optical flow estimation on neurosurgical video data is impeded by characteristic visual effects such as strong blur and poor textures (see Fig. 61). However, to obtain meaningful spatio-temporal features from neurosurgical videos, accurate optical flow calculation is crucial. Currently, there are no accurate optical flow methods that are designed specifically for neurosurgical domain. Thus, in this work we



Fig. 1 Test sequence in our evaluation data with corresponding plots for two calculated optical flow, LiteFlowNet and Farneback, and interpolation error (IE). Optical flow is converted to polar space for visualization using HSV space, whereas hue encodes pixel-wise vector direction and saturation is used to show pixel-wise vector magnitude

Table 1 Optical flow algorithms which are included in our evaluation, and mean value of interpolation error (IE) over 100 test sequences

Algorithm	Training data / parameter sets included in our study	Mean IE
Farneback	Default parameters as used in OpenCV implementation	32.02
Dual TV L1	Default parameters as used in OpenCV implementation	30.02
SPyNet (Sintel)	trained on FlyingChairs, fine-tuned on MPI-Sintel	28.41
SPyNet (Base)	trained on FlyingChairs	28.00
FlowNet 2 (Sintel)	trained on FlyingChairs/FlyingThings3D, fine-tuned on MPI-Sintel	27.82
LiteFlowNet (Sintel)	trained on FlyingChairs/FlyingThings3D, fine-tuned on MPI-Sintel	27.52
PWC-Net (Sintel)	trained on FlyingChairs/FlyingThings3D, fine-tuned on MPI-Sintel	27.51
PWC-Net (Base)	trained on FlyingChairs/FlyingThings3D	27.43
FlowNet 2 (Base)	trained on FlyingChairs/FlyingThings3D	26.97
LiteFlowNet (Base)	trained on FlyingChairs/FlyingThings3D	26.97
Deepflow (Middlebury)	DeepMatching, Middlebury parameters	26.93
Deepflow (Sintel)	DeepMatching, MPI-Sintel parameters	26.71

evaluate how state-of-the-art optical flow algorithms perform on neurosurgical video data w.r.t. accuracy.

Methods

In the optical flow research community, accuracy is usually evaluated by calculating the L2 distance between estimated and ground truth optical flow. However, this evaluation approach is applicable only to classical optical flow datasets for which ground truth exists, such as *FlyingChairs/FlyingThings3D* or *MPI-Sintel*. For neurosurgical video data, no (public) ground truth are available. Generating a neurosurgical dataset with corresponding ground truth through phantom studies or rendering engines (Blender) suffers from limited transferability to reality. For solid conclusions in the neurosurgical domain, optical flow accuracy needs to be evaluated directly and (ideally retrospectively) on *real-world clinical data*. The frame interpolation method by Baker et al. [1] fulfils this requirement, and we use it for evaluation of optical flow algorithms.

The method compares the accuracy of different optical flow algorithms based on their frame interpolation error. For an image test sequence $\{I_1, I_2, I_3\}$, an interpolated image I_t is calculated at the timestep between I_1 and I_3 using an estimated optical flow. In case of ideal interpolation, I_t is equivalent to I_2 . The interpolation error of I_t compared to ground truth I_2 serves as basis for comparing different optical flow algorithms. We re-implemented the frame interpolation algorithm from [1] and, as proposed there, chose mean-squared error for color images as interpolation error metric.

We collected surgical microscope video data from five cranial tumor surgeries and randomly select 100 test sequences, each consisting of images I_1, I_2, I_3 . In our evaluation we compare the following

state-of-the-art optical flow algorithms: Farneback, Dual TV L1, Deepflow, and methods relying on convolutional neural networks (CNNs), namely SPyNet, FlowNet 2, PWC-Net, LiteFlowNet (see Table 1 for details). Each CNN-based method is included in two variants, differing w.r.t. training datasets. One variant (Base) is trained on FlyingChairs and/or FlyingThings3D, the other (Sintel) is additionally fine-tuned on MPI-Sintel. Mean interpolation error (IE) is presented in Table 1. For statistical analysis of the interpolation error over the 100 test sequences, we conduct a one-way repeated measures ANOVA and Fisher post-hoc test ($\alpha=0.05$).

Results

The one-way repeated measures ANOVA shows that the choice of the optical flow algorithm significantly influences the interpolation error, with $F(11, 1089)=49.47, p<0.001$. The Fisher post-hoc test reveals that Farneback, followed by Dual TV L1 exhibit statistically significant higher interpolation error values compared to all other algorithms. We find a top group consisting of PWC-Net, LiteFlowNet and FlowNet 2 (all trained on FlyingChairs/FlyingThings3D) and Deepflow. The Fisher post-hoc test does not provide further sub-grouping within the top group.

For the CNN-based methods, the post-hoc test indicates that the interpolation error depends on a choice of the training dataset. However, fine tuning on the MPI-Sintel dataset does not bring any benefit. Instead, the accuracy drops for most of the networks. Thus, we conclude that training on MPI-Sintel data is not useful when networks are applied in neurosurgical domain.

In addition to the quantitative analysis, we conduct a qualitative assessment to gain more trust in the interpolation error for the domain

of neurosurgical video data. Figure 1 shows a situation with a suction (left instrument) and scissors (right instrument), whereas the suction features larger motion. The optical flow estimated by Farneback and LiteFlowNet (Base) are compared. Farneback exhibits highest interpolation error on this sequence, while LiteFlowNet (Base) as member of the top group, shows the lowest interpolation error. In the qualitative comparison of the optical flow images, LiteFlowNet seems to calculate the suction's motion more physically plausible than Farneback. Farneback does not capture the instrument motion at the entrance in the image plane (lower left corner) and displays more splatter around the instrument's shape. Qualitative evaluation of these two algorithms supports the outcome from the interpolation method that we used for numerical evaluations. Farneback has an error of 32.83, while LiteFlowNet has around 31% less, see Fig. 1.

Conclusion

In our work, we compare various state-of-the-art optical flow algorithms with respect to their accuracy on neurosurgical microscope video data. We use the frame interpolation method for quantitative evaluations. Based on the interpolation error, statistical analysis

identifies a top group among the tested algorithms, containing FlowNet 2, PWC-Net and LiteFlowNet and DeepFlow. For the CNN-based algorithms, we observed a tendency that fine-tuning on the dataset MPI-Sintel deteriorates the interpolation error compared to training on FlyingChairs/FlyingThings3D. Our qualitative comparisons agree with the results of the statistical analysis. As next step, we would like to evaluate the algorithms from the top group in a concrete medical application for computer-assisted neurosurgery.

Reference

- [1] Baker S, Scharstein D, Lewis JP, Roth S, Black MJ, Szeliski R (2011) A Database and Evaluation Methodology for Optical Flow. *Int J Comput Vis* 92: 1–31. <https://doi.org/10.1007/s11263-010-0390-2>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.