Int J CARS (2023) 18:S1–S123
https://doi.org/10.1007/s11548-023-02878-2

ABSTRACTS

CARS 2023—Computer Assisted Radiology and Surgery Proceedings of the 37th International Congress and Exhibition Munich, Germany, June 20–23, 2023

Preface

In addition to the traditional CARS topics and presentations, this year's CARS Congress is revisiting and widening its scope towards both, the Digital Operating Room of the Future (DORF) and Model Guided Medicine (MGM).

Based on the results of a think tank on these themes with members of the CARS Congress Organising Committee in October 2022 and follow up discussions with members of the CARS community, presentations on the DOR have been assigned to two sessions in the traditional DOR workshop at CARS organised in cooperation with ISCAS and SPIE. This is complemented by a special workshop on the OR 2040, with selected topics relating to a previous workshop on OR 2020, organised by Kevin Cleary and Song K Mun in Washington, USA [1].

Specifically, the goal of this newly established workshop for CARS 2023 "OR 2040" is to identify, not only the clinical and technical requirements for the next generation operating room, but also how it can be embedded into the hospital of the future. The horizon of 2040 is selected as a target timeframe, but with a possible glance also at what might be expected beyond this period. Below are some of the themes the OR 2040 initiative focus on in the corresponding two workshop sessions:

Session 1: Surgical Robotics and Telesurgery.
Session 2: Surgical Informatics and Interventional Suites.

Both workshops, the DOR and the OR 2040 are complemented with panels to discuss not only how intelligent systems can improve patient care in the OR but also how they will impact the work profile of healthcare workers. Synergistic interaction and collaboration between humans and intelligent machines are expected to be the long term focus of both workshops in the CARS Congress.

Model Guided Medicine is being covered in the Opening Session of CARS 2023 with brief presentations and position statements from selected panellists and a follow up discussion including the audience. Questions which surface from the theme of MGM relate to why, how, where and when MGM methods and tools will impact an increasingly AI based (biased!) decision making process in health care.

For example, how can MGM become an enabler for moving from a data driven machine learning/AI to a model driven machine learning/AI in Medicine? In particular, how can certain AI concepts such as transparency, predictability, cause-effect reasoning, cooperativeness, agent and safety driven, data and model interoperability be promoted with MGM? Should model driven machine learning be the basis for a transparent machine intelligence and replace a rather black box based artificial intelligence? Finally, what role will a model based domain evidence play when it comes to verifying, validating and to evaluating AI algorithms?

In the model based approach, the focus is on selecting and combining all significant knowledge notations from a given domain of discourse into one or more structured models. A return to the knowledge- or information-centric phases or even to the source data may be necessary in order to enhance the fidelity of any of these models [2]. Model driven AI aims at achieving a balanced decision making process by selecting one, several or none of the available models in the domain of discourse.

CARS 2023 and beyond are planned to promote and enable Q&A discussions on MGM generally and model driven AI specifically. In this process it can draw on modelling expertise in the CARS community which traditionally included many presentations with focus on models coming from application areas such as:

Computational models of anatomy, physiology and pathology, e.g. FEM, CFD,
Clinical decision making models, e.g. Bayesian, TNM,
Machine learning models, e.g. CNN, fuzzy sets,
Physical phenomena models, e.g. electromagnetic, implants,
Software engineering models, e.g. CASE, MDA, QFD,etc.

To integrate this rich pool of expertise in CARS on modelling into intelligent clinical decision making is a challenge requiring mind sets which have been cultivated at CARS for the last 40 years and are now in need to be given attention to, more than ever before. This applies in particular to addressing questions and problems relating to

complexity and transparency (or lack thereof) of IT systems and their application in health care systems such as for the OR or AI assisted clinical decision making.

Assuming that many engineers and scientists consider certain state of the art AI algorithms and systems to be incomprehensibly complex, how can we expect patients, physicians and health care providers be well advised to actually using them [3]? Some critical questions need to be addressed such as:

1. What basic value system (e.g. of patients, physicians/medical staff, health care providers, researchers, public solidarity base, profit or control maximization), if any, should be reflected in AI based IT systems that are designed to assist in clinical decision making, specifically in the domain of CARS?
2. Why do we need to re-examine communication behavior of humans with intelligent and networked machines?
3. How should IT systems be designed that record and (transparently) display a reproducible path on clinical decision making?
4. How can possible negative side effects in the use of AI based IT systems be minimized?
5. Who assumes responsibility for damages incurred through the use of AI systems in health care, specifically in the domain of CARS?
6. Where and when can different concepts and models relating to AI based IT systems be realized in a controlled (certified?) and verifiable manner?

From the CARS 2023 Opening Session to the Closing Session and beyond, we want to make a start into this challenging adventure of integrating AI related systems and Model Guided Medicine into CARS, with a distinguished list of individuals in the panels and subsequent think tanks, who have participated in shaping CARS what it is - unique.

Heinz U. Lemke
Munich, June 2023.

References

[1] K. Cleary, A. Kinsella, S.K. Mun (2005), OR 2020 workshop report: Operating room of the future, Article in International Congress Series, May 2005, Volume 1281, May 2005, Pages 832–838, Elsevier

[2] H.U. Lemke (2022) Moving from data, information, knowledge and models to wisdom-based decision making in the domain of Computer Assisted Radiology and Surgery (CARS), International Journal of Computer Assisted Radiology and Surgery, Vol.17, pages 1513–1517, Sept. 2022. Int J CARS 17, 1513–1517 (2022). https://doi.org/10.1007/s11548-022-02731-y

[3] J. Weizenbaum (2003) …der Mensch, nicht die Maschine ist das Maß, In: Niederlag W, Lemke HU, Bondolfi A, Rienhoff O, Ethik und Informationstechnik am Beispiel der Telemedizin, Health Academy, 02/2003, Dresden, 2003, 58–63.

CARS 2023

Computer Assisted Radiology and Surgery

Proceedings of the 37th International Congress and Exhibition Munich, Germany, June 20–23, 2023

CARS 2023 Honorary President

Leonard Berliner, MD
Staten Island University Hospital, New York, NY (US)

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)
Randy E. Ellis, PhD
Queen's University, Kingston, ON (CA)
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)
Kensaku Mori, PhD
Nagoya University (JP)
Yoshihiro Muragaki, MD, PhD
CAMED, Kobe (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)
CAR '91	**Auguste Wackenheim, MD** Strasbourg (FR)
CAR '93	**J. Oscar M.C. Craig, MD** London (GB)
CAR '95	**Alexander R. Margulis, MD** San Francisco (US)
CAR '97	**Takahiro Kozuka, MD** Osaka (JP)
CAR '98	**Herbert Kaufmann, MD** Berlin (DE)
CARS '99	**Raffaella de Dominicis, MD** Florence (IT)
CARS 2000	**Pierre Rabischong, MD, PhD** Montpellier (FR)
CARS 2001	**Toyomi Fujino, MD, PhD** Tokyo (JP)
CARS 2002	**Kiyonari Inamura, PhD** Hyogo (JP)
CARS 2003	**H.K. Huang, DSc, FRCR (Hon.)** Los Angeles, CA (US)
CARS 2004	**Karl-Heinz Höhne, PhD** Hamburg (DE)
CARS 2005	**Martin Lipton, MD, FACR, FAHA** Los Angeles, CA (US)
CARS 2006	**Hans G. Ringertz, MD, PhD** Stockholm (SE)
CARS 2007	**Madjid Samii, MD** Hannover (DE)
CARS 2008	**Javier Herrero Jover, MD, PhD** Barcelona (ES)
CARS 2009	**Bernard L. Crowe, BA, MPH** Canberra, NSW (AU)
CARS 2010	**Osman M. Ratib, MD, PhD** Geneva (CH)
CARS 2011	**Michael M. Vannier, MD** Chicago, IL (US)
CARS 2012	**Davide Caramella, MD** Pisa (IT)
CARS 2013	**Nassir Navab, PhD** Munich (DE)
CARS 2014	**Makoto Hashizume, MD, PhD, FACS** Fukuoka (JP)
CARS 2015	**Javier Herrero Jover, MD, PhD** Barcelona (ES)
CARS 2016	**Davide Caramella, MD** Pisa (IT)
CARS 2017	**Javier Herrero Jover, MD, PhD** Barcelona (ES)
CARS 2018	**Volkmar Falk, MD, PhD** Berlin (DE)
CARS 2019	**Pierre Jannin, PhD** Rennes (FR)
CARS 2020	**Hubertus Feussner, MD** Munich (DE)
CARS 2021	**Dirk Wilhelm, MD** Munich (DE)
CARS 2022	**Yoshihiro Muragaki, MD, PhD** Kobe (JP)

Program Committee

Purang Abolmaesumi, PhD
University of British Columbia, Vancouver, BC (CA)
Asma'a Al-Ekrish, MD
King Saud University, Riyadh (SA)
Paul R. Algra, MD, PhD
Medical Center Alkmaar (NL)
Russell J. Andrews, MD
NASA Ames Research Center, Los Gatos, CA (US)
Christos Angelopoulos, DDS
Columbia University, New York, NY (US)
Yoshinori Arai, PhD, DDS
Nihon University, Atsugi (JP)
Jumpei Arata, PhD
Kyushu University, Fukuoka (JP)
Eiichiro Ariji, DDS, PhD
Aichi-Gakuin University School of Dentistry, Nagoya (JP)
Paulo M. Azevedo-Marques, PhD
University of Sao Paulo, Ribeirao Preto (BR)
Frits H. Barneveld Binkhuysen, MD, PhD
Meander Medical Center, Soest (NL)
John S.H. Baxter, PhD
University of Rennes (FR)
Erwin Bellon, PhD
University Hospitals Leuven (BE)
Elizabeth Beckmann, BSc
Lanmark, Beaconsfield (GB)
Maximilian Berlet, MD
Klinikum rechts der Isar der TUM, Munich (DE)
Dirk Beyersdorff, MD
University Hospital Hamburg-Eppendorf (DE)
Silvio Diego Bianchi, MD
University of Torino (IT)
Ulrich Bick, MD
Charité—University Medicine Berlin (DE)
Uldis Bite, MD
Mayo Clinic, Rochester, MN (US)
Davide Caramella, MD
University of Pisa (IT)
Mario Ceresa, PhD
Pompeu Fabra University, Barcelona (ES)
Ruey-Feng Chang, PhD
National Taiwan University, Taipei (RC)
Kiyoyuki Chinzei, PhD
AIST Tsukuba (JP)
Kevin Cleary, PhD
Children's National Medical Center, Washington, DC (US)
D. Louis Collins, PhD
McGill University, Montreal, QC (CA)
Mario A. Cypko, PhD
Free University Berlin (DE)

Computer Assisted Radiology and Surgery 37th International Congress and Exhibition

Chair: Ulrich Bick, MD (DE)

Automatic segmentation of the hypothalamus on high-resolution T1-weighted MR images

I. Vernikouskaya[1], H. P. Müller[2], J. Kassubek[2], V. Rasche[1]

[1]Ulm University Medical Center, Department of Internal Medicine II, Ulm, Germany [2]University of Ulm, Department of Neurology, Ulm, Germany

Keywords hypothalamus, segmentation, U-Net, T1-weighted MRI

Purpose

The hypothalamus is a small structure of the brain with a crucial role in the regulation of the human body. Some neurodegenerative diseases, such as dementia syndromes and amyotrophic lateral sclerosis (ALS) may be related to hypothalamic volume variation [1]. Because of the small size of the hypothalamus and lack of image contrast in its vicinity, hypothalamic morphological landmarks are not always clear on magnetic resonance imaging (MRI). This leads to the scarcity of neuroimaging studies of this structure due to work-intensive operator-dependent manual delineations and lack of the automated segmentation tools.

In this work, we propose a fully automatic approach relying on convolutional neural network (CNN) to segment the hypothalamus on high-resolution MRI scans. We compare the segmentation performance of 2D-CNNs of U-Net architecture with four state-of-the-art encoders.

Methods

108 high-resolution 2D T1-weighted MRI datasets from 66 patients with sporadic ALS and 42 healthy controls were acquired at an 1.5 T MRI scanner (Symphony, Siemens Medical, Erlangen, Germany). Morphological data were obtained using a MPRAGE sequence (144 sagittal slices, no gap, $1.0 \times 1.2 \times 1.0$ mm^3 voxels, $256 \times 192 \times 256$ matrix, TE = 4.2 ms, TR = 1600 ms). The images were pre-processed (rigid brain normalization to the anterior commissure—posterior commissure line) and stored as a series of 50 slices per subject at a resolution of $0.0625 \times 0.0625 \times 0.5$ mm^3 with matrix size of 1024×1024 pixels. Manual delineation of the hypothalamus was performed by two experienced operators using the TIFT software [1] to generate a ground truth data.

All available data from both groups were split into training (47 ALS, 24 controls), validation (4 ALS, 3 controls), and test (15 ALS, 15 controls) data. ALS and controls in the test group were gender and age matched (p-value(gender) = 0.481, p-value(age) = 0.921).

Four types of 2D-CNNs of U-Net architecture (VGG16, ResNet50, Inceptionv3, and EfficeintNetB0), pre-trained by ImageNet database, were trained on downscaled image pairs of 512×512-pixel size. All models were trained on a GeForce GTX 1060 6 GB GPU for 25 epochs with early stopping with a batch size of 4 samples per pass. The loss function was the sum of the categorical Cross Entropy and Jaccard loss and Adaptive Moment Estimation (Adam) was used as the optimizer. The intersection over union (IoU) metric was optimized during training. Training was stopped when the cross-validation loss was observed to have ceased improving for 10 consecutive epochs and the model with the lowest validation loss was chosen for prediction.

The performance of each CNN model was evaluated in terms of IoU. Further, true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) detections were calculated for each volume to estimate average Precision (positive predictive value), Recall (sensitivity of binary classification), and F1-score (harmonic mean of both metrics) of each model. Average prediction time per image was assessed. Comparison between ground truth hypothalamus volume and hypothalamus volume segmented with various encoders was done using paired t-test or Wilcoxon signed-rank test as appropriate depending on the results of Shapiro–Wilk test for normality. Value of $p < 0.05$ indicated significance. Average and standard deviation, as well as Pearson correlation coefficients (r) are reported.

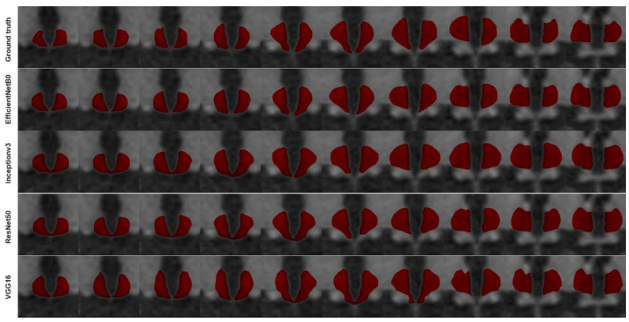

Fig. 1 Comparison of predictions by four different U-Net architectures with the ground truth in ten consecutive test image slices

Results

Figure 1 shows the comparison of segmentations predicted by four different U-Nets overlaid on the MR images in ten consecutive slices of a single patient test dataset. Disagreements between the ground truth and predicted segmentations were observed at the edge slices (anterior, posterior), where false pixels were predicted to some extent by all networks.

Table 1 summarizes the obtained results. All investigated network architectures achieved similar performance in terms of IoU with the highest value for EfficientNetB0 (0.88). High Recall values for all models indicate that high fraction of pixels that should be predicted as hypothalamus was also predicted as hypothalamus. The highest Precision, i.e. most of the pixels predicted as hypothalamus were true predictions, was achieved with EfficientNetB0 model (0.87). F1-score was also highest for EfficientNetB0 (0.87). The fastest prediction per image was achieved with EfficientNetB0, permitting segmentation of the whole hypothalamus in 1.43 s on a GPU.

No significant difference between the ground truth volume and the volume segmented by EfficientNetB0 was observed, whether all other networks significantly overestimated the segmented volume (Fig. 2). Significant linear correlation with the ground truth was obtained for all investigated models, with the lower r-value for VGG16.

Conclusion

In this study, we propose a fully automatic hypothalamus segmentation approach based on the use of a convolutional network and applied it to ALS patients' group and controls. Automatic approach permitted extremely fast structure segmentation as compared to even semi-automated approaches requiring 20–40 min processing time per hypothalamus [2]. We demonstrate that our method outperforms the human inter-rater reliability scores (ICC of 0.78 and 0.82) and approaches intra-rater levels (ICC of 0.82–0.97) reported earlier for 3 T MRI [2]. Although our best performing model still misses few hypothalamus pixels, the average model IoU (0.88) exceeded the dice coefficient of 0.77 (obtained with a best performed consensus model built from three model outputs trained on three different views of the brain (sagittal, coronal, and axial) as previously proposed by Rodrigues et al. [3]).

Table 1 Comparison of prediction results by four different U-Net architectures

	Ground truth	EfficientNetB0	Inceptionv3	ResNet50	VGG16
IoU	–	0.88 ± 0.03	0.87 ± 0.02	0.87 ± 0.02	0.85 ± 0.04
Precision	–	0.87 ± 0.05	0.79 ± 0.05	0.82 ± 0.05	0.82 ± 0.05
Recall	–	0.86 ± 0.05	0.93 ± 0.04	0.87 ± 0.04	0.85 ± 0.11
F1-score	–	0.87 ± 0.03	0.85 ± 0.03	0.85 ± 0.03	0.83 ± 0.07
Prediction time per image [ms]	–	29	31	35	42
Hypothalamus volume cm^3	0.82 ± 0.10	0.80 ± 0.10	0.96 ± 0.12	0.87 ± 0.11	0.85 ± 0.16
Pearson correlation	–	0.799	0.799	0.806	0.706

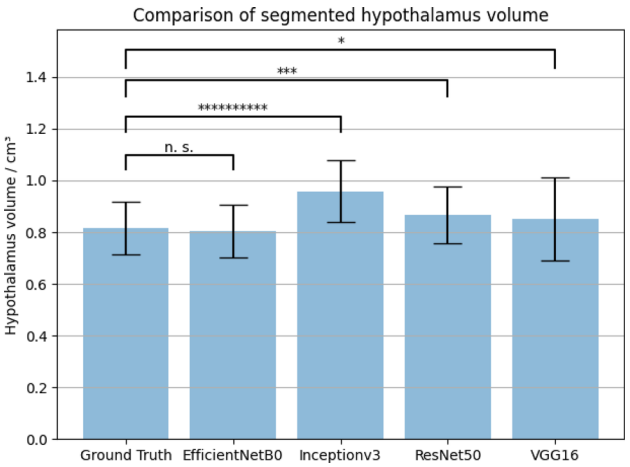

Fig. 2 Comparison of hypothalamus volume segmented with different U-Nets versus ground truth

References

[1] Gorges M, Vercruysse P, Müller HP, Huppertz HJ, Rosenbohm A, Nagel G, Weydt P, Petersen A, Ludolph AC, Kassubek J, Dupuis L (2017) Hypothalamic atrophy is related to body mass index and age at onset in amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 88(12):1033–1041.

[2] Wolff J, Schindler S, Lucas C, Binninger AS, Weinrich L, Schreiber J, Hegerl U, Möller HE, Leitzke M, Geyer S (2018) A semi-automated algorithm for hypothalamus volumetry in 3 Tesla magnetic resonance images. Psychiatry Res Neuroimaging 277:45–51.

[3] Rodrigues L, Rezende T, Zanesco A, Hernández A, Franca Jr M, Rittner L. (2020) Hypothalamus fully automatic segmentation from MR images using a U-Net based architecture. In: Proceedings of the 15th International Symposium on Medical Information Processing and Analysis. International Society for Optics and Photonics, 113300 J.

Multi-network approach for image segmentation in non-contrast enhanced cardiac 3D MRI of arrhythmic patients

I. Vernikouskaya[1], D. Bertsche[1], P. Metze[1], L. M. Schneider[1], V. Rasche[1]

[1]Ulm University Medical Center, Department of Internal Medicine II, Ulm, Germany

Keywords left atrial appendage, arrhytmia, segmentation, two-stage pipeline.

Purpose

Image fusion techniques integrating 3D anatomical models obtained from pre-procedural imaging into the live XR fluoroscopy can be applied to guide left atrial appendage (LAA) closure. Cardiac MRI (CMR) gains in importance for non-invasive evaluation of LAA morphology as alternative to gold-standard transoesophageal echocardiography (TOE) or computed tomography angiography (CTA). Segmenting the left atrium including LAA from non-contrast enhanced CMR for multi-modal image fusion is, however, challenging due to the low contrast between the atrial tissue, blood pool, and surrounding anatomical structures including other cardiac chambers and aorta, as well as large anatomical variations of LAA. Additionally, arrhythmia often impairs the image quality in ECG synchronized acquisitions. Therefore, there is a strong need for an advanced image segmentation method to be applied to CMR for image fusion in patients with atrial fibrillation.

Deep learning methods have the potential to provide faster and more accurate segmentation compared to conventional approaches [1]. Fully-convolutional neural network (CNN), especially U-Net, has become the method of choice for cardiac segmentation. Recently, there is also a growing interest in applying neural networks in a multi-stage pipeline which breaks the segmentation problem into subtasks [2].

Here we propose a CMR segmentation method based on CNN implemented in a two-stage pipeline. Specifically, we use state-of-the-art object detection network YOLOv4 for a region-of-interest (ROI) localization of the relevant cardiac structures followed by the 2D U-Net for subsequent voxel classification within the identified ROI.

Methods

In the first stage of our proposed two-stage approach, we used YOLOv4 object detection algorithm to detect bounding boxes of the area including left atrium with LAA (LA), right atrium with vena cava (RA), and aortic arch (AO) on each 2D image slice. Subsequently, the image slices and corresponding masks were automatically cropped based on the detected regions and padded with black pixels to ensure an equal image size of 160 × 160 pixels for the second-stage CNN. Finally, the second-stage CNN was trained on created cropped image samples with ResNet50 and VGG16 backbones to do the segmentation of LA, RA, and AO classes. The predicted masks were then transformed in target region of the original size volume using bounding box labels for 3D visualization.

The two-stage approach (YOLOv4 + U-Net) with VGG16 or ResNet50 encoder was compared with the single-stage approach trained on the image samples of original size (U-Net) in 9 patient's test datasets (1065 images).

Table 1 Comparison between ResNet50, VGG16, YOLOv4 + ResNet50, and YOLOv4 + VGG16 in terms of IoU, dice index, and Hausdorff distance with paired Student's t-test

	IoU			DICE			Hausdorff (mm)		
	RA	AO	LA	RA	AO	LA	RA	AO	LA
ResNet50	0.65 (0.12)	0.76 (0.07)	0.77 (0.06)	0.78 (0.09)	0.86 (0.05)	0.87 (0.04)	18.85 (9.97)	10.76 (4.17)	25.51 (30.53)
VGG16	0.67 (0.08)	0.74 (0.09)	0.76 (0.08)	0.80 (0.06)	0.85 (0.06)	0.86 (0.06)	22.92 (17.28)	26.18 (38.31)	17.02 (17.17)
p-value	0.26	0.14	0.66	0.22	0.16	0.59	0.45	0.26	0.15
YOLOv4 + ResNet50	**0.71 (0.07)**	**0.77 (0.05)**	**0.79 (0.06)**	**0.83 (0.05)**	**0.87 (0.04)**	**0.88 (0.04)**	**18.75 (12.49)**	10.78 (8.28)	**7.99 (2.87)**
YOLOv4 + VGG16	0.66 (0.12)	**0.77 (0.06)**	0.78 (0.06)	0.79 (0.09)	**0.87 (0.04)**	0.87 (0.04)	20.88 (19.36)	**10.21 (6.47)**	9.15 (4.11)
p-value	0.11	0.68	0.14	0.09	0.70	0.13	0.50	0.89	0.25

Non-contrast enhanced CMR image data were acquired with a navigator-gated 3D Dixon sequence on a 3 T MRI scanner. The data were reconstructed at a resolution of 1.33 mm^3 with a 2D matrix size of 384 × 384 pixels. In total, 29 patient datasets with 115 slices per patient on average were used for training and test.

Volumetric segmentation models of RA, LA, and AO were generated from 3D CMR volumes using 3DSlicer software. Each single structure was converted to a series of axial DICOM binary label maps, which were merged into a single label map.

To create the training datasets for YOLOv4 algorithm, we used MIT licensed open-source image annotation tool LabelImg (https://github.com/tzutalin/labelImg) to generate a series of MR images and corresponding text files containing the coordinates and sizes of built bounding boxes around the defined cardiac structures.

The segmentation accuracy achieved with the different network architectures and approaches was evaluated using three widely used measures: the average IoU metric, the volumetric dice coefficient, and the Hausdorff distance. Using these measures, the difference between each class of predicted masks for all 2D slices of a single patient and corresponding ground truth masks was calculated. Also processing time and memory performance of each method were estimated.

Results

The bounding boxes containing the area of interest could be predicted by YOLOv4 algorithm correctly in 2569 out of 2643 (or 97%) image slices.

Training of single-stage U-Net and second-stage U-Net with either encoder required approximately 3 min and 1 min per epoch or 90 ms and 30 ms per image respectively. ResNet50 and VGG16 consumed 4 GB and 4.4 GB GPU memory respectively, YOLOv4 + ResNet50 and YOLOv4 + VGG16 − 3.9 GB and 2.6 GB respectively. At test time, the two-stage approach could generate the entire segmentation output within 1 s on average (or 9 ms per slice) using about 4 GB GPU memory with ResNet50 encoder and within 1.9 s on average (16 ms per slice) using 3.8 GB GPU memory with VGG16. Prediction of the whole 3D volume with single-stage ResNet50 took 3.5 s (30 ms per slice) using 4.3 GB RAM, whereas prediction with VGG16 took 8 s (68 ms per slice) using 4.4 GB RAM.

ResNet50 and VGG16 backbones performed similarly in terms of IoU, dice index, and Hausdorff distance for U-Net and YOLOv4 + U-Net approaches (Table 1). The highest performance was achieved for the combination of YOLOv4 algorithm with ResNet50 for all three classes.

Figure 1 demonstrates segmentation outputs for a single MR slice with four investigated approaches.

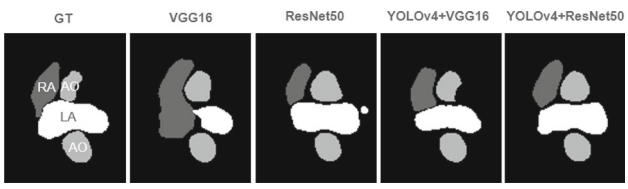

Fig. 1 Comparison of segmentation outputs for a single MR slice for VGG16, ResNet50, YOLOv4 + VGG16, YOLOv4 + ResNet50 vs. ground truth (GT)

Conclusion

This study demonstrates the potential of CNN in automatic 3D segmentation of left atrium, right atrium, and aorta from non-contrast enhanced axial CMR images in arrhythmic patients in general and the application of a two-stage approach for improving the segmentation in particular. The obtained results indicate that defining an ROI with the first-stage CNN lead to improvement of segmentation, save memory, and increase the speed of automatic segmentation.

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] Zhuang X, Li L, Payer C, Štern D, Urschler M, Heinrich MP, et al., G. Yang (2019) Evaluation of algorithms for Multi-Modality Whole Heart Segmentation: An open-access grand challenge. Med Image Anal 58:101,537.

[2] Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, Rueckert D (2020) Deep Learning for Cardiac Image Segmentation: A Review. Front Cardiovasc Med https://doi.org/10.3389/fcvm.2020.00025.

Deep Learning Reconstruction vs. Hybrid-Type Iterative Reconstruction: Capabilities for Radiation Dose Reduction on High-Definition CT

Y. Ohno[1,2], N. Hamabuchi[1], D. Takenaka[1,3], H. Kimata[4], Y. Ito[4], K. Fujii[4], N. Akino[4], Y. Oshima[1], H. Nagata[2], T. Ueda[1], Y. Ozawa[1], T. Yoshikawa[1,3], H. Toyama[1]

[1]Fujita Health University School of Medicine, Radiology, Toyoake, Japan [2]Fujita Health University School of Medicine, Joint Research Laboratory of Advanced Medical Imaging, Toyoake, Japan [3]Hyogo Cancer Center, Diagnostic Radiology, Akashi, Japan [4]Canon Medical Systems Corporation, Otawara, Japan

Keywords CT, Lung, Deep learning reconstruction, Iterative reconstruction.

Purpose

Since multidetector-row CT (MDCT) became widely clinically applied in the mid-2000s, academic and social interests in radiation dose reduction for MDCT examinations without any decline in diagnostic capability have been continuing and led to recommendations for its use to all imaging vendors, although its use in routine clinical practice of dose reduction for various MDCT techniques has varied among institutions and scanners. Moreover, hybrid-type and model-based iterative reconstruction (IR) methods have been introduced and used for CT examinations and continuously improved since the 2010s [1]. In addition, deep learning reconstruction (DLR) has been introduced by almost all major imaging vendors, and tested for CT examinations of a variety of organs. During the same periods, the number of detector rows for MDCT has been increased for wider coverage within one rotation as well as reduction of detector collimations [2]. High definition CTs (HDCTs) with and without a photon counting system have been clinically applied and tested since 2015. One of these HDCTs, an ultra-high-resolution or super-high-resolution CT (UHR-CT or SHR-CT) produced by Canon Medical Systems is widely available in routine clinical practice. This CT system has three different scan modes: normal resolution (NR: 0.5 mm × 80 rows/896 channels), high-resolution (HR: 0.5 mm × 80 rows/1792 channels) and super-high-resolution (SHR: 0.25 mm × 160 rows/1792 channels), and improvements in spatial resolutions for UHR-CT have been reported by several investigators. Moreover, UHR-CT makes it possible to use larger matrix sizes such as 1024 or 2048 for certain CT examinations and it has been suggested this may be useful for some clinical purposes. However, one of the drawbacks of UHR-CT might be the relative reduction in signal-to-noise ratio (SNR) and contrast-noise ratio (CNR) due to a decrease in the detector collimation size, even when the same radiation dose protocol with standard reconstruction algorithms is used. To the best of our knowledge, however, no one has tried reducing the radiation dose for HDCT or evaluated the capability of DLR for image quality improvement on reduced- and ultra-low-dose chest CTs in comparison with clinically applicable IRs for patients with a variety of chest diseases. We hypothesized that, in comparison with hybrid-type IR, the DLR algorithm makes it possible to improve image noise and reduce radiation dose while maintaining the capability for evaluation of radiological findings for chest reduced- and ultra-low-dose CT examinations of a variety of chest disease patients, so that DLR can be put to clinical use with the appropriate reconstruction time and few plastic artifacts as compared with model-based IR as it is used at many institutions. The purpose of this study was to compare the capabilities of DLR for image quality improvement and lung texture evaluation with those of hybrid-type iterative reconstruction (IR) for standard-, reduced- and ultra-low dose CTs (SDCT, RDCT and ULDCT) obtained with high-definition CT (HDCT) and reconstructed at 0.25 mm, 0.5 mm and 1 mm section thicknesses with 512´512 or 1024´1024 matrixes for patients with various pulmonary diseases.

Methods

Forty patients with various pulmonary diseases underwent SDCT (CTDIvol: mean ± standard deviation, 9.0 ± 1.8 mGy), RDCT (CTDIvol: 1.7 ± 0.2 mGy) and ULDCT (CTDIvol: 0.8 ± 0.1 mGy) at a HDCT. All CT data set were then reconstructed with 512 × 512 or 1024 × 1024 matrixes by means of hybrid-type IR and DLR. Sigal-to-noise ratio (SNR) of lung parenchyma, overall image quality and probabilities of all lung textures were assessed for each CT data set. SNR, overall image quality and detection performance of each lung texture reconstructed with DLR and hybrid-type IR were then compared by means of paired t-tests, Wilcoxon's signed rank tests and ROC analyses for all CT data at each section thickness.

Results

Data for each radiation dose showed DLR attained significantly higher SNR than hybrid-type IR for each of the CT data ($p < 0.0001$).

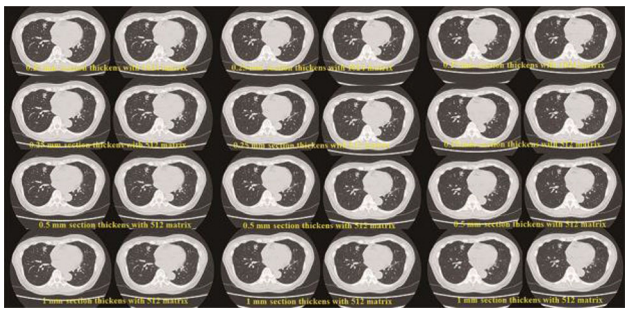

Fig. 1 70-year-old female with invasive adenocarcinoma in the right middle lobe

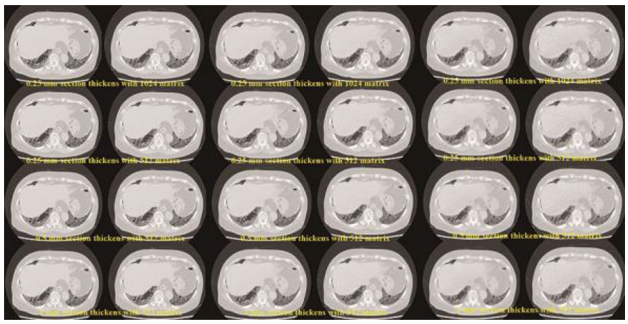

Fig. 2 65-year-old female with interstitial lung disease due to mixed connective tissue

Data for CTs with a 512 × 512 matrix and 1 mm, 0.5 mm and 0.25 mm section thicknesses demonstrated overall image quality of DLR was significantly better than that of hybrid-type IR for each radiation dose ($p < 0.0001$). Assessments of all findings except consolidation and nodules or masses, areas under the curve (AUCs) for ULDCT with hybrid-type IR for each section thickness ($0.91 \leq AUC \leq 0.97$) were significantly smaller than those with DLR ($0.97 \leq AUC \leq 1$, $p < 0.05$) and the standard protocol ($0.98 \leq AUC \leq 1$, $p < 0.05$).

Conclusion

DLR is potentially more effective for image quality improvement and lung texture evaluation than hybrid-type IR on all radiation dose CTs obtained at HDCT and reconstructed at 0.25 mm, 0.5 mm and 1 mm section thicknesses with 512 × 512 or 1024 × 1024 matrixes for patients with a variety of pulmonary diseases, Figs. 1 and 2.

References

[1] Ohno Y, Takenaka D, Kanda T, Yoshikawa T, Matsumoto S, Sugihara N, Sugimura K (2012) Adaptive iterative dose reduction using 3D processing for reduced- and low-dose pulmonary CT: comparison with standard-dose CT for image noise reduction and radiological findings. AJR Am J Roentgenol. 2012;199(4):W477–85.

[2] Akagi M, Nakamura Y, Higaki T, Narita K, Honda Y, Zhou J, Yu Z, Akino N, Awai K (2019) Deep learning reconstruction improves image quality of abdominal ultra-high-resolution CT. Eur Radiol. 2019;29(11):6163–6171.

Deep Learning Reconstruction for MR Imaging: Comparison of Capability for T-factor Evaluation with Thin-Section CT in NSCLC Patients

Y. Ohno[1,2], D. Takenaka[1,3], Y. Ozawa[1], K. Yamamoto[4], M. Shinohara[4], M. Ikedo[4], M. Yui[4], Y. Oshima[1], N. Hamabuchi[1], H. Nagata[2], T. Ueda[1], H. Ikeda[1], T. Yoshikawa[1,3], H. Toyama[1]

[1]Fujita Health University School of Medicine, Radiology, Toyoake, Japan [2]Fujita Health University School of Medicine, Joint Research Laboratory of Advanced Medical Imaging, Toyoake, Japan [3]Hyogo Cancer Center, Diagnostic Radiology, Akashi, Japan [4]Canon Medical Systems Corporation, Otawara, Japan

Keywords MRI, Lung, Compressed sensing, Deep learning reconstruction.

Purpose

Lung cancer affects an estimated 2 million new patients each year and is associated with 1.76 million deaths per year making it the leading cause of cancer-related death in the world. Some investigators have suggested that multimodality treatment, including preoperative systemic therapy for preoperatively diagnosed locally advanced non-small cell lung cancer (NSCLC), may be useful because of the unfavorable prognosis for surgery alone. This means that preoperative assessment of locoregional staging has become very important for management of NSCLC. Although it had suggested since 1991 that unenhanced or contrast-enhanced (CE-) magnetic resonance (MR) imaging and angiography could be useful for mediastinal, chest wall or pulmonary vasculature invasions, unenhanced or CE—thin-section multidetector-row CT (MDCT) with multiplanar reconstruction (MPR) images have been used since 2005 for differentiating T3 or T4 tumors from T1 or T2 tumors in routine clinical practice because its potential which is superior to that of routine MDCT. In view of these findings, there has been a continual need during the last decade for spatial, temporal or contrast resolutions of MR imaging, while 3D fast spoiled gradient echo (GRE) sequences with or without fat suppression techniques have been suggested by all MR vendors as useful for various clinical purposes. Moreover, compressed sensing (CS) or deep learning reconstruction (DLR) have been proposed as useful for improving temporal resolution and image quality on MR sequences in different body fields [1,2]. However, there have been no reports regarding the utility of DLR for image quality and T-factor assessment improvements on T2-weighted imaging (T2WI), short inversion time (TI) inversion recovery (STIR) imaging and unenhanced- and CE-3D fast spoiled GRE imaging with and without CS in comparison with thin-section MDCT for NSCLC patients. Moreover, there have been no reports either addressing question of the appropriate sequence for T factor evaluation of NSCLC patients using currently applied MR sequences. The purpose of this study was thus to determine the utility of DLR compared with that of thin-section MDCT for improving image quality on MR imaging and the appropriate sequence for T-factor assessment of NSCLC patients.

Methods

As subjects for this study 213 pathologically diagnosed NSCLC patients who underwent thin-section MDCT and MR imaging as well as T-factor diagnosis were retrospectively enrolled. For quantitative image quality assessment, regions of interests (ROIs) with the same diameter were placed over each tumor, the intercostal or trapezius muscles on both sides of the same slice plane and the trachea, and signal-to-noise ratios (SNRs) of tumor and muscle and contrast-to-noise ratio (CNR) between the tumor and muscle on each MR protocol were calculated. For qualitative assessment of image quality, board-certified chest radiologists independently and visually evaluated overall image quality, artifact incidence and diagnostic confidence level of each MR protocol by using a 5-point scoring system. All final visual scores for every patient were determined by consensus of the two readers. The same board-certified chest radiologists who evaluated overall image quality also evaluated the probability of the occurrence of T3 or T4 cases by using the following 5-point visual scoring system. To determine the capability of DLR for each sequence, SNR and CNR obtained with each sequence with and without DLR were compared by using the paired t-test. For a comparison of qualitative image quality, interobserver agreements for overall image quality, artifacts and diagnostic confidence level for each method were assessed by means of weighted kappa statistics. Then, Wilcoxon's signed rank test was then used for a comparison of overall image quality and artifact level attained by each sequence with and without DLR. To compare qualitative diagnostic performance among all sequences with and without DLR, ROC analyses were performed to compare the diagnostic capability of all methods for differentiation of T3 or T4 from T1 or T2. Next, sensitivity, specificity and accuracy for mediastinal or chest wall invasion attained by all methods were compared by means of McNemar's test. In addition, kappa statistics were used to determine interobserver agreement for T-factor evaluation for each method. Diagnostic accuracy for T-factor was then compared among all sequences and thin-section CT with the aid of McNemar's test.

Results

SNRs and CNRs of T2WI, STIR imaging, unenhanced thin-section Quick 3D imaging and CE-thin-section Quick 3D imaging with DLR were significantly higher than those of sequences without DLR ($p < 0.05$). All interobserver agreements for overall image quality

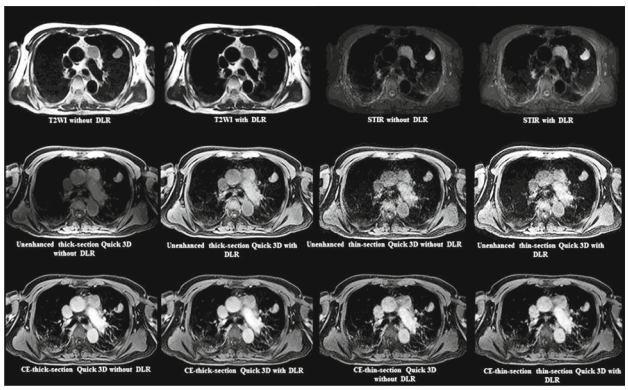

Fig. 1 71-year-old male patient with adenocarcinoma in the left upper lobe and mediastinal lymph node metastasis (N2 disease)

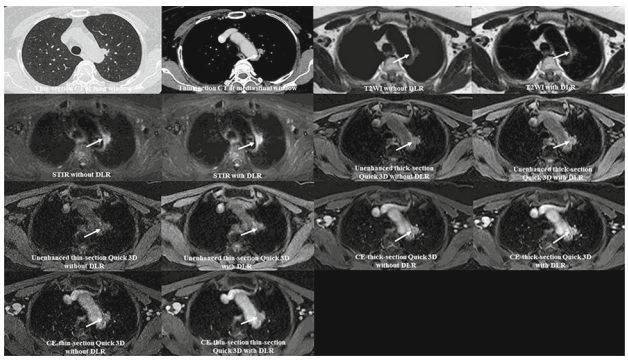

Fig. 2 75-year-old female patient with adenocarcinoma in the left upper lobe featuring aortic invasion and classified as a T4 case

($0.64 \leq \kappa \leq 0.82$, $p < 0.0001$) and artifacts ($0.62 \leq \kappa \leq 0.72$, $p < 0.0001$) were determined as significantly substantial or almost perfect. The application of DLR median resulted in significantly improved overall image quality and artifacts of T2WI, STIR imaging, unenhanced thin-section Quick 3D imaging, CE-thick-section Quick 3D imaging and CE-thin-section Quick 3D imaging ($p < 0.05$). Interobserver agreement for the probability of occurrence of T3 or T4 cases determined with each method were rated substantial or almost perfect ($0.72 \leq \kappa \leq 0.86$, $p < 0.0001$). Areas under the curves (AUCs) were significantly larger, while sensitivities and accuracies of STIR imaging with and without DLR and CE-thin-section Quick 3D imaging with and without DLR were significantly higher than those of thin-section CT, T2WI with and without DLR, unenhanced thick-section and thin-section Quick 3D imaging with and without DLR ($p < 0.05$). Interobserver agreements for and diagnostic accuracy of T-factor evaluation for all methods were almost perfect ($0.81 \leq \kappa \leq 0.92$). Diagnostic accuracy of STIR imaging with and without DLR, CE-thick-section Quick 3D imaging with and without DLR and CE-thin-section Quick 3D imaging with and without DLR was significantly higher than that of thin-section CT, T2WI with and without DLR, unenhanced thick-section Quick 3D imaging with and without DLR and unenhanced thin-section Quick 3D imaging with and without DLR ($p < 0.05$). In addition, CE-thin-section Quick 3D imaging with DLR was significantly more accurate than STIR imaging with and without DLR ($p < 0.05$), see also Figs. 1 and 2.

Conclusion

DLR is thus considered useful for image quality improvement on MR imaging. STIR imaging and CE-Quick 3D imaging with or without CS were validated as appropriate MR sequences for T-factor evaluation in NSCLC patients.

References

[1] Ueda T, Ohno Y, Yamamoto K, Iwase A, Fukuba T, et al. (2021) Compressed sensing and deep learning reconstruction for women's pelvic MRI denoising: Utility for improving image quality and examination time in routine clinical practice. Eur J Radiol. 2021;134:109,430.

[2] Ikeda H, Ohno Y, Murayama K, Yamamoto K, Iwase A, Fukuba T, Toyama H (2021) Compressed sensing and parallel imaging accelerated T2 FSE sequence for head and neck MR imaging: Comparison of its utility in routine clinical practice. Eur J Radiol. 2021;135:109,501.

Language Image Assisted labeling of video data from medical intervention rooms using Grounded Language-Image Pre-training (GLIP)

P. Schülein[1], A. Hadzic[1], M. Vetter[1]

[1]University of Applied Sciences Mannheim, ESM-Institute, Mannheim, Germany

Keywords intervention room dataset, few shot learning, assisted dataset labeling, Vision Language Models.

Purpose

Manual labeling of datasets is time-consuming and scales with the set of objects and the image data. Especially for datasets from domains where no publicly available large datasets exist, like the medical intervention room domain, networks for target applications must either be trained using synthetic data or datasets must be laboriously labeled by hand. While in other areas the data can be outsourced through Amazon Mechanical Turk or other labeling providers, this is not possible due to privacy requirements in the medical space.

With the emergence of new deep learning architectures such as CLIP [1] or GLIP [2], which have a strong zero- and few-shot capability, the manual labeling of datasets from the medical domain can potentially be reduced.

Together with the Department of Radiology and Nuclear Medicine at the University Hospital Mannheim, a clinical study is conducted to explore possible assistance systems during TACE interventions. RGB cameras are used to record these interventions and in addition to identifying possible assistance systems, also serve as a dataset for researching them.

The amount of data that is recorded thereby is large. Labeling of the entire image data by hand is unrealistic. Therefore, this work investigates whether the zero- and few-shot capability of the vision language network GLIP is sufficient for the labeling of a medical intervention dataset or whether a sufficiently good labeling is possible by fine-tuning the network with study data. The overall goal is to create a dataset from the medical intervention space that includes not only people and objects but also the activities of the health professionals.

Methods

To investigate the accuracy of GLIP for a dataset from the intervention space, images from 8 interventions of the study are used as the training/validation data and images from 3 other interventions are used as test data. Here, the distribution of the split is 192/21/50 (train/val/test) images. In total an amount of 24 different objects present during angiographic intervention are labeled. Among others this includes persons, masks, gloves, syringes, monitors, sterile blankets, x-ray protection vests or radiation shielding glass. The object instances labeled by hand are 9643 in this case.

As described in the GLIP paper, the zero-shot accuracy, then the 1/3/5/10 shot and finally the full shot accuracy is investigated by using the 192 hand labeled train images. The GLIP-T(4) model with the pre-trained weights provided by the authors is used as baseline.

In further experiments the full shot finetune serves as a baseline for the next experiments. With the trained weights 2000 additional training images are labeled by the network itself without verification by hand. Here the total amount of object instances could be increased to 68,894. In further experiments this new trainset is used to train the few/full shots again using the baseline GLIP-T(4) model weights. It is investigated whether a consistent accuracy is achieved on the test dataset compared to the hand labeled images.

The last experiment combines the hand labeled and self-labeled trainsets to investigate whether the accuracy can be increased further, while starting again from the provided baseline weights.

Results

The results of the experiments performed are shown in Table 1. The results of 0/1/3/5/10/full—shot trainings at different intersection over unions (IoU) for different training datasets are visualized per row.

All experiments were started with the same initial weights and trained to convergence as described in [2]. In all cases the full shot trainings yielded the best accuracy.

Conclusion

The accuracy of the results on N = 192 training data is insufficient in zero- and few-shot cases. This is certainly due to the underrepresentation of medical image data as well as medical text in the original training of the network.

Although the accuracy could not be sufficiently improved by few-shot training, the full-shot experiment yields promising results especially at IoU = 0.5. Likewise, the accuracy of the network could even be slightly improved by self-labeling without correction using the trained network with N = 192. By combining the image data labeled

Table 1 Results of zero/few shot and full shot training on the test dataset based on different training data

Number of training data	IoU	Zero-shot	Fine-tuning				
			1	3	5	10	All
N = 192	IoU = 0.50:0.95	14.6	38.3	41.0	42.6	43.6	51.8
	IoU = 0.75	14.7	40.7	42.6	45.4	47.0	56.1
	IoU = 0.50	23.5	59.3	62.5	63.4	64.2	74.4
N = 2000	IoU = 0.50:0.95	–	37.8	39.4	39.9	40.8	52.7
	IoU = 0.75	–	40.3	42.7	43.0	43.7	57.6
	IoU = 0.50	–	58.0	60.5	59.9	60.6	76.6
N = 2192	IoU = 0.50:0.95	–					52.8
	IoU = 0.75	–					56.3
	IoU = 0.50	–					76.1

by hand and the image data labeled by the network, the accuracy could be improved as well. In summary, such a self-labeling training loop can further improve the accuracy of object recognition in medical intervention rooms compared to the training data labeled by hand while increasing the used training dataset.

In future work, the GLIP network will be adapted to the medical domain to enable further research in assistance systems in the intervention space.

Acknowledgement
This research project is part of the Research Campus M2OLIE and funded by the German Federal Ministry of Education and Research (BMBF) within the Framework "Forschungscampus – public-private partnership for Innovations" under the funding code 13GW0389C.

References

[1] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G. (2021) Learning transferable visual models from natural language supervision. InInternational Conference on Machine Learning, Jul 1 (pp. 8748–8763). PMLR.

[2] Li LH, Zhang P, Zhang H, Yang J, Li C, Zhong Y, Wang L, Yuan L, Zhang L, Hwang JN, Chang KW. (2022) Grounded language-image pre-training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10,965–10,975).

Rib born detection of scan path planning for fully-automated ultrasound robotic system

K. Okuzaki[1], N. Koizumi[1], K. Yoshinaka[2], J. Zhou[1], T. Fujibayashi[1], R. Tsumura[2]

[1]The University of Electro-Communications, Department of Mechanical and Intelligent Systems Engineering, Chofu, Japan
[2]National Institute of Advanced Industrial and Science, Health and Medical Research Institute, Tsukuba, Japan

Keywords robotic ultrasound, scan path planning, rib born detection, fully automated.

Purpose

Ultrasound (US) diagnosis is widely used because it is safe, real-time, non-exposure, non-invasive, and inexpensive compared to CT and MRI. However, scanning with the US probe is dependent on the physician. Because of the high dependency on surgeon-physicians, sufficient ultrasound examinations cannot be provided in areas where there is a shortage of surgeon-physicians due to the declining birthrate and aging population, or in remote areas or developing countries where surgeon-physicians are not available. In addition, due to the spread of COVID-19, contact infection between surgeon-physician and patient in US diagnosis has become a problem. For these reasons, there is a demand for robotic development of US robots systems that assist the surgeon-physician or do not require the surgeon-physician's intervention is underway.

Scanning path planning is an essential elemental technology for fully automated US robotics. In order to obtain cross-sectional image of the target organ, this it is necessary for the robot to recognize where on the body surface to contact place the US probe. In order to place the US probe on appropriate position and find the target image, the robot needs to recognize the human body by detecting landmarks on the body surface. Prior studies required surgeon-physician intervention at this stage or failed to adequately account for individual body size differences.[1][2] In this study, we focus on the abdominal region such as liver US imaging, which is one of the most important areas in US diagnosis. During the abdominal US scanning. The ribs are critical the most important body surface landmarks in abdominal ultrasound screening. Most abdominal cross section images are standardized to be scanned with reference to the ribs. In other words, rib position detection is crucial for scanning path planning. Then, this study aims to develop an estimation method of ribs area for the scanning path planning with robotic US systems.

Methods

We propose a method for detecting ribs based on RGB-D images and respiratory variation. It is difficult to estimate the location of ribs on RGB images because the ribs are not visible depending on individual body differences. Therefore, we hypothesized that it would be possible to detect rib regions based on changes in body surface position due to breathing. It is anatomically known that deep breathing lifts the chest against the abdomen, bordering the ribs. Then, we considered that the ribs region could be detected by comparing the height of the chest during respiratory cycle. We generate a depth difference image by taking the difference between the depth image taken at the resting inspiratory position and the depth image taken at the maximum inspiratory position, which clearly shows the rib position. That can be realized without the intervention of a surgeon-physician by applying the instruction to the patient to "take a big breath," which is often used in X-ray examinations in current clinical practice. In addition, by

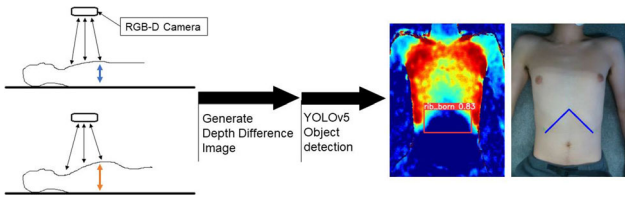

Fig. 1 Rib born detection method and result image

applying training using the YOLOv5 object detection model to this depth difference image, the rib position can be detected (Fig. 1).

Results

In an experiment conducted with nine subjects, the proposed method of rib detection on depth difference images marked an IoU of 0.951 and an average confidence of 0.77. The average error between the ground-truth and predicted positions was 15 pixel $\simeq$ 13 mm. The results were superior to the rib detection from the RGB image, which was performed as a comparison.

Conclusion

The proposed Depth Difference Image method, which measures respiratory variation, was able to accurately estimate the ribs, an important landmark for abdominal ultrasound US screening, without contact and examiner physicians' intervention. The coordinates of the ribs can be easily obtained by using an object detection model, which is suitable as input for the initial positioning of the ultrasound robotic US system.

References

[1] Q. Huang, J. Lan and X. Li, Robotic Arm Based Automatic Ultrasound Scanning for Three-Dimensional Imaging, in IEEE Transactions on Industrial Informatics, vol. 15, no. 2, pp. 1173–1182, Feb. 2019, https://doi.org/10.1109/TII.2018.2871864.

[2] J. Tan, B. Li, Y. Leng, Y. Li, J. Peng, J. Wu, B. Luo, X. Chen, Y. Rong, C. Fu "Fully Automatic Dual-probe Lung Ultrasound Scanning Robot for Screening Triage," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2022, https://doi.org/10.1109/TUFFC.2022.3211532.

Improvement of a skeleton segmentation model of bone scintigrams with a transformer and component tree loss function

Q. A. Nguyen[1], J. Cousty[2], Y. Kenmochi[2], S. Higashiyama[3], J. Kawabe[3], A. Shimizu[1]

[1]Tokyo University of Agriculture and Technology, Institute of Engineering, Koganei, Tokyo, Japan [2]GREYC, CNRS, Université de Caen Normandie, ENSICAEN, France [3]Osaka Metropolitan University, Department of Nuclear Medicine, Sumiyoshi, Osaka, Japan

Keywords Skeleton segmentation, Deep learning, Bone scintigram, Component tree loss.

Purpose

Bone scintigraphy is widely used in clinical radiological examinations and is effective in diagnosing bone metastases from neoplastic diseases. Bone scintigrams of the human body are two-dimensional images obtained from the bone scintigraphy process from the anterior and posterior sides of the patient. In bone scintigraphy diagnostic imaging, the Bone Scan Index (BSI) is useful for quantitatively evaluating the spread of bone metastases. As the calculation of BSI requires the bone metastasis area and the bone area on whole-body anterior–posterior images, an accurate skeleton segmentation of the whole-body anteroposterior images is essential. However, there are cases where the recognition fails owing to high concentrations of positive accumulation and the irregular shape of organs when using the conventional method. In this study, we introduce a novel deep learning-based model with a transformer that combines the component tree loss function for whole-body bone scintigram skeleton segmentation. The proposed method aims to improve the dice similarity coefficient (DSC) of bones for both anterior and posterior images. Additionally, the component tree loss function is expected to reduce the difference in the number of connected components and holes between the predicted results and ground truths.

Methods

TransBtrflyNet network. The TransBtrflyNet model is a combination of BtrflyNet and TransUNet [1]. The skeleton segmentation input is a pair of anterior and posterior images. The network consists of an encoder, concatenation part, and a decoder. Following the TransUNet framework, a CNN-Transformer hybrid model was used for the encoder. The size of the input image was 576 $\times$ 256 pixels. First, the anterior and posterior images were fed simultaneously into the CNN blocks to generate feature maps. The transformer was applied to 1 $\times$ 1 patches extracted from the CNN feature maps of anterior and posterior images. The outputs from the transformer block were then reshaped for the next steps. The concatenation part has multiple convolutions, max-pooling layers, deconvolutions, and a skip connection, to enhance features. By concatenating the outputs from the transformer blocks of anterior and posterior features, the model can learn the characteristics of both the anterior and posterior images. The output from the concatenation part is deconvoluted into anterior and posterior features for the decoder process. Skip connections were used to concatenate the features from the encoder and decoder. Finally, the segmentation heads output the segmentation labels for anterior and posterior images.

Loss function. The loss function consists of cross-entropy loss (L_{CE}), deep supervision (DSV) loss (L_{DSV}) which is calculated based on the dice loss, and component tree (CTr) loss (L_{CTr}) which is designed to reconnect discontinuous components by selecting or discarding image maxima [2]. The loss function is expressed as follows:

$$L_{CTr} = \sum_{i=1}^{i \le l} \max(m - \mathbf{sm}_{r\_i}, 0)^p + \sum_{i=l+1}^{i \le k} \mathbf{sm}^p_{r\_i} \quad \text{with} \quad \mathbf{r} = \text{argsort}(\mathbf{im}), \tag{1}$$

where the constant margin m $\subseteq$ R is set to 1, which is the maximum value of the softmax function. The symbol $l \subseteq N^+$ is the target number of maxima. We selected the value of l based on the number of connected components of the ground truth. The values of l differed for each bone. Here, **sm** $\subseteq R^+$ and **im** $\subseteq R^+$ represent a saliency and an important measure of the maxima respectively. We used a combination of dynamic and volume for (**sm**, **im**).

$$L_{TransBtrflyNet} = (0.5L_{CE\_A} + 0.5L_{DSV\_A} + 0.0001L_{CTr\_A} + 0.5L_{CE\_P} + 0.5L_{DSV\_P} + 0.0001L_{CTr\_P})/2, \tag{2}$$

where A and P are abbreviations of anterior and posterior.

Results

The segmentation targets included 12 anterior bones (skull, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, sacrum, pelvis, ribs,

scapula, humerus, femur, sternum, and clavicle) and 10 posterior bones (skull, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, sacrum, pelvis, ribs, scapula, humerus, and femur). Experiments were conducted on 2470 bone scintigrams obtained from 1235 cases, and the respective results were compared and discussed. The threefold cross-validation method was used for performance evaluation.

We defined the difference between the predictions and ground truths for the number of connected components (delta-CCs) and holes (delta-holes). We compared the results between the TransBtrflyNet and TransBtrflyNet + CTr models in terms of DSC, delta-CCs, and delta-holes. Wilcoxon signed-rank test was used for statistical analysis.

Dice similarity coefficient. We successfully improved the DSC of three anterior bones: skull ($p < 0.01$), pelvis ($p < 0.01$), and sternum ($p < 0.05$). However, the DSC of lumbar ($p < 0.01$), and clavicle ($p < 0.01$) of TransBtrflyNet + CTr was inferior to that of TransBtrflyNet.

For posterior results, the DSC of three bones of TransBtrflyNet + CTr: sacrum ($p < 0.05$), whole scapula ($p < 0.05$), and scapula overlapping ribs ($p < 0.01$), significantly outperformed that of TransBtrflyNet. Skull ($p < 0.01$), cervical ($p < 0.01$), and humerus ($p < 0.01$) showed statistically lower results. Besides the DSC, we evaluated the performance of the two models based on the delta-CCs and delta-holes.

Delta-CCs and delta-holes. For the results of TransBtrflyNet + CTr, the delta-CCs were reduced in three anterior bones: skull ($p < 0.05$), humerus ($p < 0.05$), and clavicle ($p < 0.05$); the delta-holes were reduced in one anterior bone, lumbar ($p < 0.05$), and one posterior bone, scapula ($p < 0.01$). The TransBtrflyNet + CTr model did not produce any inferior results for both delta-CCs and delta-holes.

Figure 1 and Table 1 show an example in which the delta-holes of lumbar of TransBtrflyNet + CTr statistically exceeded that of TransBtrflyNet. The number of connected components of femur also decreased. The average DSC of the model with the CTr loss was better than that of the model without the CTr loss.

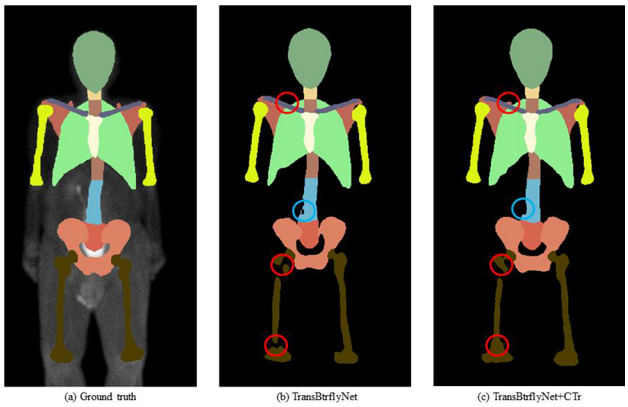

Fig. 1 Example of skeleton segmentation by TransBtrflyNet and TransBtrflyNet + CTr. **a** Ground truth, **b** TransBtrflyNet, and **c** TransBtrflyNet + CTr. Red circle: delta-CCs, blue circle: delta holes

Table 1 DSC, delta-CCs, and delta-holes results of Fig. 1

DSC/delta-CCs/delta-holes	TransBtrflyNet	TransBtrflyNet + CTr
All bones	0.8977/0.3333/0.0833	*0.9089/0.0833/0*
Lumbar	*0.9165*/0/1	0.8916/0/*0*
Femur	0.7867/3/0	*0.8125/1*/0
Scapula	0.8956/1/0	*0.8994*/0/0

Conclusion

We present a deep learning-based method that integrates a transformer for the skeleton segmentation of bone scintigrams. We also introduced the CTr loss function to reduce the number of connected components and holes. The TransBtrflyNet + CTr model showed promising results compared to the TransBtrflyNet model.

In the future, we will try various combinations of (**sm**, **im**) of CTr loss corresponding to the characteristics of each bone to anatomically improve the segmentation results.

References

[1] Chen, Lu, Yu, Luo, Adeli, Wang, Lu, Yuille, Zhou (2021) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation.

[2] Perret, Cousty (2022) Component Tree Loss Function: Definition and Optimization. Discrete Geometry and Mathematical Morphology. DGMM 2022. Lecture Notes in Computer Science 13,493:248–260.

Four-dimensional ultrasound reconstruction of carotid artery based on vascular pulsation force and image segmentation

C. Cui[1], Y. Lin[1], Z. Li[2]

[1]Shanghai Jiao Tong University, School of Mechanical Engineering, Shanghai, China [2]Shanghai General Hospital, Department of Ultrasound, Shanghai, China

Keywords carotid artery ultrasound, Image phase estimation, 3D ultrasound segmentation, 4D ultrasound reconstruction.

Purpose

Carotid atherosclerosis is one of the main causes of cardiovascular diseases. Ultrasound images of carotid artery morphology are the main reference for doctors when evaluating the degree of carotid atherosclerosis. Compared with two-dimensional (2D) and three-dimensional (3D) ultrasound, four-dimensional (4D) ultrasound can help doctors identify the structural space of the carotid artery more intuitively. Considering the periodicity of the carotid artery pulsation, 4D ultrasound can be acquired by reconstructing 2D ultrasound images with a certain pose and phase. In this study, a novel highly efficient and accurate 4D carotid ultrasound imaging method combining a vascular pulsation force and image segmentation is proposed.

Methods

Considering that 4D ultrasound reconstruction requires pose and phase information of ultrasound images, a portable and compact mechanical device was designed to collect ultrasound images in parallel, with the phase information estimated according to the vascular pulsation force and image information. A 3D ultrasound segmentation method was developed to segment the vessel lumen and wall such that the carotid artery could be presented in a dynamic

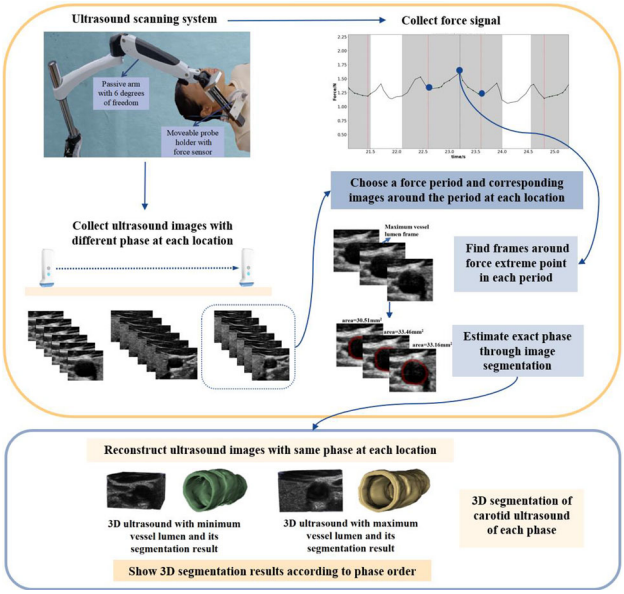

Fig. 1 Workflow of the proposed 4D ultrasound reconstruction method

manner according to the phase order. The entire reconstruction process is illustrated in Fig. 1.

Ultrasound image acquisition strategy: To collect ultrasound images with sufficient information, a 2D ultrasound probe with a force sensor was installed on a linear sliding table. The table was secured to a six-degree-of-freedom passive arm that could adjust the initial position and contact force between the probe and neck. The initial position enabled a section of the carotid artery to be completely scanned along a fixed direction; the initial force was approximately 1 N. The target carotid artery region was divided into multiple locations at an interval of 1 mm; then the probe scanned for 2 s at a speed of 10 frames per second at each location, and moved to the next location for 0.5 s. Force sensor and ultrasound data were simultaneously collected.

Ultrasound image phase estimation: Several methods have been used to obtain phase information, such as using an extra electrocardiogram [1] or comparing the vessel lumen area through image segmentation [2]. However, these methods require either additional equipment or the processing of all 2D images, resulting in high computational costs. Because the contact force between the ultrasound probe and neck can reflect the changes in the vessel lumen, the adjacent time section of the lowest and highest points in the force waveform should include ultrasound images with the minimum and maximum vessel lumens. Three images around the extreme points of the force waveform were selected for further segmentation to obtain the exact phase. One thousand and eight images with labels were trained using nnu-Net to segment the selected pictures such that the minimum and maximum vessel lumens could be found at each location. The intermediate phase of images at each position was obtained by regularly choosing images between those with minimum and maximum vessel lumens.

Four-dimensional reconstruction and segmentation: Images with the same phase were stitched into a 3D model according to their location. An effective U-shaped cross-shaped window (CSWin) Transformer (U-CSWT) network was used to segment the vessel lumen and wall at each phase. The U-CSWT is a symmetric network that consists of an encoder, a decoder and a skip connection. The basic unit of a U-CSWT is the CSWin transformer block. Consisting of a multihead self-attention module, the block has a strong capability of long-distance context modeling. The encoder branch gradually shrinks the image while extracting high dimensional feature information and the decoder branch fuses low-level and high-level features and restores the image to the original size. Datasets containing 212 3D ultrasound images were trained using the network. The 3D reconstructed images of all phases were segmented and saved according to their phases such that the doctor could observe the carotid artery in a dynamic manner.

Results

The proposed ultrasound phase acquisition method was verified through comparison with the doctor's assessment from their manual segmentation result. As shown in Fig. 1, the proposed method can recognize the maximum vessel lumen ultrasound image accurately. The dice similarity coefficients between the predicted results and reference masks were 94.4 $\pm$ 3.0% for the media-adventitia boundary (MAB) and 90.8 $\pm$ 5.1% for the lumen-intima boundary (LIB), and the volumetric error of the vessel wall volume (VWV) was approximately 0.81%. Figure 1 shows the 3D results segmented by the U-CSWT. Compared with other segmentation methods, U-CSWT produces more accurate segmentation results.

Conclusion

A novel 4D ultrasound image reconstruction method based on an effective combination of force signals and image information was proposed. Compared with spatiotemporal reconstruction from freehand sonography, the proposed system can perform both 3D and 4D ultrasound reconstruction using a compact and portable device without additional optical navigation equipment, which is cost effective and convenient. Through the image segmentation of the ultrasound frames around the extreme point of the force waveform, the phase of the ultrasound images can be accurately estimated. Images with the same phase were reconstructed according to their location, and the 3D vessel lumen and wall were segmented using U-CSWT. Finally, a 4D dynamic ultrasound of the carotid artery was presented, providing more intuitive and multidimensional information for doctors to determine the state of the carotid artery.

References

[1] Thurgood H, Witte R, Laksari K. (2021). 4D Reconstruction and Identification of Carotid Artery Stenosis Utilizing a Novel Pulsatile Ultrasound Phantom[J]. Current Protocols, 1(10): e264.

[2] Liang H, Ning G, Dai S, Ma L, Luo J, Zhang X, Liao H. (2022). Spatiotemporal reconstruction method of carotid artery ultrasound from freehand sonography[J]. International Journal of Computer Assisted Radiology and Surgery, 17(9): 1731–1743.

Registration of digital subtraction angiography images using non-linear displacement vector fields with Swin-Unet-transformer

B. Kraus[1], L. Huischen[1], M. Vetter[1]

[1]Mannheim University of Applied Sciences, Institut für eingebettete Systeme, Regelungs- und Medizintechnik (ESM), Mannheim, Germany

Keywords transformer, displacement vector field, digital subtraction angiograph, deep learning.

Purpose

Digital subtraction angiography (DSA) is an interventional radiological method for visualizing blood vessels which uses a pre-contrast image subtracted from subsequent images including contrast medium inflow. Thus, all structures except vessels are masked. The quality of the resulting vessel map is heavily dependent on the quality of registration between the mask and contrast flow images. Patient movement, breathing and hydrostatic effects of blood and contrast flow are known to introduce artefacts in subtraction. Aiming for compensation of these deforming artefacts we propose a deep learning network based on a Swin-Unet [1] transformer to calculate the non-linear displacement vector fields between image pairs. While being robust against the change of image characteristics between picture with and without vessels, the proposed method is fast to compute compared to classical methods.

Methods

For the task of calculating the 2D displacement vector field between mask and subsequent images a deep learning model can benefit from global and local semantic feature learning. The proposed method consequently uses an adapted Swin-Unet, a network originally developed for medical image segmentation. This network utilizes hierarchical transformer blocks with shifting windows to compose an encoder/bottleneck/decoder structure with skip connections [1]. The Model is adjusted to be able to perform pixelwise regression of a 2D displacement vector field resulting in an output of shape image height * image width * 2.

The Unet-like structure of the network allows for spatial information to be distributed via skip connections. The attention mechanism of the Swin-transformer blocks harnesses the contextual information in and between each image patch.

A broad spectrum of train data is required to generalize the wealth of deformations and DSA datasets are not publicly available in abundance. Therefore, a semi-synthetic dataset with high variance is created. 3968 frontal thorax CT scans from the BIMCV COVID-19 + dataset [2] are chosen and randomized B-spline transformations are applied, resulting in a displacement vector field with a mean vector length of 14 mm and a standard deviation of 8 mm. In the following step a snippet taken of segmented retina blood vessels from the IOSTAR Dataset is randomly cropped, scaled, and adjusted in its pixel intensity. Then, it is added to the augmented thorax image. A separate set of vessel maps and 905 thorax scans is used exclusively for validation purposes.

This dataset generation prioritizes producing high variance within the domain of the training data rather than accurate and realistic simulation and is therefore similar to the structured domain randomization approach for training object detection networks.

Consequently, a pair of training input is comprised of an image with a vessel mimicking contrast flow and a mask version of the image without the vessel. Since the displacement vector field needed to achieve perfect mask registration is known from the dataset generation step, it can be used as target for supervised training.

In training the deep learning model is presented with the image pair and tasked to calculate the corresponding displacement vector field despite information in one image being partly occluded by the inserted vessel. MSE-loss between target and calculated displacement vector field is minimized using a stochastic gradient descent optimizer with a cosine learning rate schedule. The calculated displacement vector field can be applied to the mask image to acquire a vessel map by performing the subtraction of the images in logarithmic intensity space (Fig. 1).

Additionally, a similar training without the addition of vessels is performed to observe the influence the occlusions have on the quality of the calculated displacement vector fields.

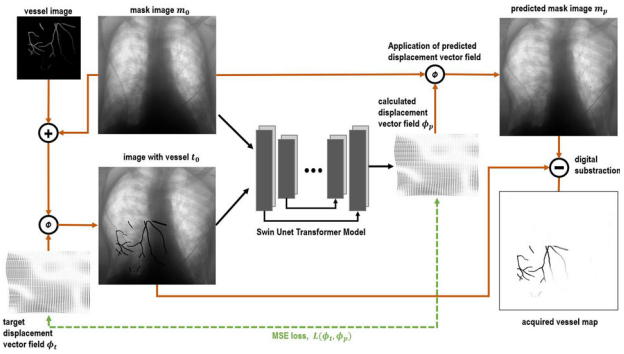

Fig. 1 Learning strategy for a Swin Unet Transformer Model using semi-synthetic data with example results

Results

After training for 640 epochs with a batch size of 6, the validations MSE between calculated and target displacement vector fields converges at 1.67 [pixel-size]^2 for the training with vessels and 1.64 [pixel-size]^2 for training without vessels on 384 by 384 images. This results in a mean absolute pixel intensity error of $\sim$ 0.01 when performing the subtraction of two registered images with total intensities scaled from 0 to 1 for comparability (omitting the occlusions by vessels). At the method's core, the Swin-Unet transformer model achieves a pure inference throughput of 16 fps on a NVIDIA GeForce RTX^{TM} 3090 with a batch size of one on those image dimensions.

Minor artefacts in square patterns can sometimes be observed during optical examination of images to which the calculated displacement field vector was applied to. Vessel maps resulting from this subtraction process show a lower degree of DSA typical artefacts. However, the edges of the vessels tend to have a small degree of grainy noise. This is most likely due to the square image patch partition intrinsic to the Swin-Transformer approach and additional training diminishes this effect.

Conclusion

This work offers promising results towards a fast and accurate deep-learning-based method for calculating non-linear 2D displacement vector fields for a registration setup trained only on a semi-synthetic dataset. However, further steps for improving image quality and reducing artefacts are needed. In future work, the proposed method needs to be evaluated against real data from angiographic interventions to prove the validity of the synthetic dataset generation approach. To further improve its results, additional training or fine-tuning of the proposed machine learning model with real data can be employed. The proposed method can also be adapted to 3D registration problems.

References

[1] Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M (2021). Swin-unet: Unet-like pure transformer for medical image segmentation. arXiv preprint arXiv:2105.05537

[2] Vayá MD, Saborit JM, Montell JA, Pertusa A, Bustos A, Cazorla M, Galant J, Barber X, Orozco-Beltrán D, García-García F, Caparrós M (2020). BIMCV COVID-19 + : a large annotated dataset of RX and CT images from COVID-19 patients. arXiv preprint arXiv:2006.01174

Predicting structural brain trajectories with discrete optimal transport normalizing flows

M. Masias[1], M. Á. González Ballester[1], G. Piella[1]

[1]University Pompeu Fabra, Information and communication technologies department, Barcelona, Spain

Keywords Optimal transport, Discrete normalizing flows, Sliced-Wasserstein distance, Brain aging.

Purpose

Extensive clinical research has been conducted to characterize brain changes through lifetime. These studies are based on the recruitment and analysis of brain images at different time points. The number of subjects undergoing all the acquisitions is usually deficient, requiring a great deal of effort and a considerable amount of time to obtain complete enough databases. Consequently, most of these studies have opted to infer population-level trends using regression models or atlases that best fit cross-sectional data at different points. Thanks to them we know how the average brain will progress over the years. Yet, changes might be heterogeneous in the population, both spatially and temporally, traits that the previous methods might not be able to capture. There is therefore a need to find more precise predictive methods, with a range of clinical implications, from the imputation of longitudinal data in incomplete datasets to the detection of small variations w.r.t. the expected trajectories.

The use of generative models may be suitable for the described problem, as they are able to generate high-quality synthetic data. Although providing very accurate results, the commonly proposed approaches do not offer explicit control over the trajectory of latent variables. The manifold on which predicted system dynamics are embedded remains unknown and conditioning on additional clinical information, by design, can only be introduced as a constant factor. Recently, [1] proposed to consider the challenge of estimating these paths in the absence of longitudinal data as one of dynamic optimal transport (OT), taking advantage of Normalizing Flows (NF) and its generalization to approximate any pair of unknown distributions through the minimization of sliced-Wasserstein distance. In this work, we exploit this idea to explicitly model individual brain trajectories, using cross-sectional data in a reduced space, while allowing us to get back individual predicted synthetic images at different ages.

Methods

T1-weighted MR images from the IXI-Database were used. The dataset contains healthy volunteers in an age range between 20 and 70 years old, recruited at two different scanners: 322 volunteers at a 1.5 T Philips scanner, and 185 volunteers at a 3 T Philips system. All images were normalized, bias-field corrected, and transformed to the MNI space using rigid and affine registration. Forty consecutive central axial slices were selected per patient, and they were divided into different decades according to the patient's age.

The dimensionality of input images was first reduced using principal component analysis (PCA). The first 50 components were used, keeping 70% of the total variance. The discrete OT-NF was implemented in such a learned low-dimensional space. For the NF we implemented a RealNVP architecture, composed of 5 coupling layers (CL), each of them encoding for a specific decade of transformation. Thus, each of these CL constituted not only the transformation core of the network but also the estimator for each of the age ranges. Each CL was composed of 3 alternating checkerboard masking scheme transforms.

To train the discrete OT-NF, here we adopted the strategy proposed by [2]. Monge's formulation of OT states that given a pair of probability measures μ and ν, from X and Y metric spaces, there exists a diffeomorphism T: $X \rightarrow Y$ such that it allows reaching an infimum for a given cost function c:

$$\inf_T \left\{ \int c(x, T(x)) | T^*(\mu) = \nu \right\}$$

Being $T^*(\cdot)$ the forward operator $T(\cdot)$. The latter diffeomorphism is called the optimal transportation map. However, the existence and uniqueness of such a function cannot be guaranteed. Therefore, the authors propose a relaxation of Monge's OT problem by replacing the equality $T^*(\mu) = \nu$ with the minimization of the distance between $T^*(\mu)$ and ν. In this manner, given a cost function c, now the OT problem in the discrete form can be relaxed to the following training loss:

$$\min_T \left\{ d(T^*(\mu), \nu) + \lambda \sum (x, T(x)) \right\}$$

In this case, for $d(\cdot,\cdot)$ we used sliced-Wasserstein distance (SWD) between the output of the last CL and the target distribution. The Euclidean distance between adjacent transformed points was chosen as the regularizer c, and λ was set to 1. Since the shortest Euclidean distance between two points is a straight line, in a first attempt to preserve the manifold of the transformation, we added an extra loss consisting of the SWD between the output of some of the CL and the corresponding experimental age distribution.

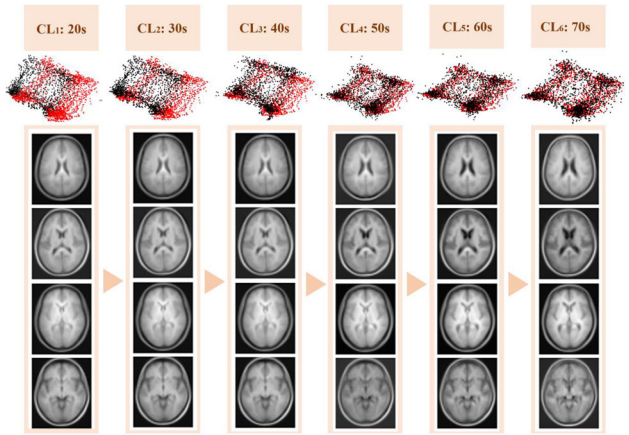

Fig. 1 Detail on the discrete OT-NF architecture, with changing distribution, represented using first two PCs; and the corresponding predicted brain images per decade/CL

Results

Once trained, the discrete OT-NF was used to transform the input distribution, in black (i.e. brain images at age twenty); to the target distribution, in red (i.e. brain images at age seventy), as shown in Fig. 1.

Conclusion

In this work, we propose an approach for structural brain image prediction with explicit control of sample trajectories in the absence of longitudinal data for training, taking advantage of discrete OT-NF. We present here some results, in which OT paths were estimated in a low-dimensional space obtained by PCA. The quality of the retrieved predicted images was not yet comparable to those proposed in the current literature, and only general aging trends could be discerned in the resulting images, the most notable of which were the decrease in the brain tissue with the associated increase in the volume of the ventricle. Future work includes the use of more powerful dimensionality reduction techniques, more efficient Wasserstein distance variants, and/or the definition of physiologically meaningful trajectories. Validation of results with real longitudinal data is also

required. Despite all the aforementioned, the proposed framework allows synthesizing brain images while simultaneously monitoring brain progression, opening the possibility of conditioning by insults/stimuli affecting only specific time points.

References

[1] Tong A, Huang J, Wolf G, van Dijk D, Krishnaswamy S (2020) Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics. Proc Mach Learn Res, 2020 Jul;119:9526–9536

[2] Florentin Coeurdoux, Nicolas Dobigeon, and Pierre Chainais.(2022). Learning optimal transport between two empirical distributions with normalizing flows. arXiv:2207.01246

Effect of patient-specific scapula and humerus morphology on shoulder biomechanics: A comparison of rotator cuff tear and osteoarthritis patients

A. Oswald[1], J. Menze[1], M. Jacxsens[2], T. Rojas[3], M. Zumstein[4,5], K. Gerber[1]

[1]University of Bern, School of Biomedical and Precision Engineering, Bern, Switzerland [2]Kantonsspital St Gallen, Clinic for Orthopedic Surgery and Traumatology of the Musculoskeletal System, St Gallen, Switzerland [3]Clinica Santa Maria, Clinic of Traumatology and Orthopedics, Providencia, Chile [4]Inselspital, University Clinic for Orthopedic Surgery and Tramatology, Bern, Switzerland [5]Orthopaedics Sonnenhof, Shoulder, Elbow and Sports Orthopaedics, Bern, Switzerland

Keywords Patient specific biomechanics, Scapular morphology, Glenohumeral joint forces, Musculoskeletal modelling.

Purpose

Clinical metrics pertaining to scapular morphology have been observed to differ between osteoarthritis (OA) and rotator cuff tear (RCT) patient groups. Such metrics include the critical shoulder angle (CSA), defined as the angle between a line connecting the superior and inferior margins of the glenoid, and a line connecting the inferior glenoid margin and the most lateral boarder of the acromion on a true anterior–posterior radiograph. It has been shown that RCT patients are associated with a CSA $> 35°$, compared to OA patients who are more likely to have a CSA $< 30°$. Using computational shoulder models, it has been shown that higher CSAs induce higher shear and smaller compression forces in the glenohumeral joint during active abduction. In comparison, reduced CSAs increase the compressive forces acting on the glenohumeral joint in abduction and potentially are a factor contributing to mechanical overload. It is thus hypothesized that the ratio of shear to compressive force (instability ratio) is increased in RCT patients compared to OA patients in the glenohumeral joint, however patient-specific analysis has never been performed. The purpose of this study is to develop an automated pipeline for creating a patient-specific computational biomechanical model of the shoulder, and to use these models to compare the glenohumeral joint reaction forces between RCT and OA patients.

Methods

In this study, 10 patients with a RCT (mean CSA of 39.6°), and 10 patients with OA (mean CSA of 17.3°) were included. Analysis was performed on CT images (in plane resolution of 0.555 mm in the sagittal and coronal planes, and a slice thickness of 0.312 mm) with approval of the local ethical commission (no. 2016–01,858). From each CT, the scapula and proximal humerus were manually segmented by a clinical expert. The CSA was calculated using a true anterior–posterior projected plane from the CT scans. The computational modelling was done using the AnyBody Modeling System (ver 7.3.4, AnyBody Technology A/S, Aalborg, Denmark) [1]. This simulation tool uses inverse dynamics analysis to estimate muscle and joint forces based on a prescribed movement. The modelling parameters for inertia, geometry and muscle contraction in the original model are based on data from the Delft shoulder group. For personalization of the AnyBody model, the patient scapulae were registered to the original AnyBody scapula (OABS) using the deterministic atlas algorithm from Deformetrica [2]. In this software, large deformation diffeometric mappings (LDDMM) are used to morph the OABS to each patient scapula. Only smooth and invertible deformations (diffeomorphisms) are permitted, so that there are no overlaps or inversions within the original geometry. For increased morphing accuracy at the patient-specific glenoid, points picked by a clinical expert, on the inferior-, superior-, anterior- and posterior-most aspects of the glenoid rim were included as fixed points in the optimization. The Hausdorff distance with 26′000 points per scapula was used to measure the global accuracy of the morphing.

The patient specific humerus radius was calculated using a simple sphere fit to the medial aspect of the segmented humeral head. This distance was used to define the glenohumeral joint center for a patient-adapted joint configuration in the musculoskeletal model.

The glenohumeral compressive (medial–lateral), vertical shear (inferior-superior), and horizontal shear (anterior–posterior) joint forces were calculated for each patient over a 0°-90° abduction in the scapular plane. The forces were then normalized to % body weight (%BW) for each patient.

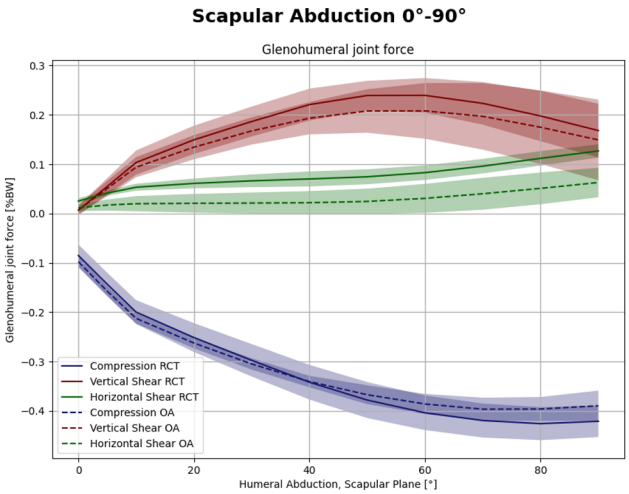

Fig. 1 Comparison of glenohumeral joint forces between OA and RCT patient groups in a 0°–90°-bodyweight abduction motion in the scapular plane. The forces are given in %BW

Results

The mean Hausdorff distance for the registration of all ten OA patients to the AnyBody scapula was 2.37 mm, with a median over all the points of 0.24 mm. The mean Hausdorff distance for the ten RCT patients was 2.40 mm, with a median over all points of 0.26 mm.

The joint reaction forces showed an increase in vertical and horizontal shear forces in the RCT patient group compared with the OA patients. The horizontal shear forces showed a difference of 6.5%BW, with a maximal force in the RCT group of 12.7%BW and a maximal force in the OA group of 6.2%BW. In the vertical plane, the difference was 3.8%BW with a maximal force of 23.9%BW in the RCT

group, and a 20.1%BW maximal force in the OA group. The compressive forces were similar between the groups, with a maximal compressive force of 42.5%BW in the RCT group, and a maximum force of 39.8%BW in the OA group, giving a 2.7%BW difference, mostly at higher humerus angles (Fig. 1).

This is consistent with previous research showing an increase in shear forces for patients with higher CSAs. These results also corroborate well with previous research showing an increasing instability ratio for patients with increasing CSAs.

Conclusion

This work presents an automated pipeline to import patient-specific scapula and glenohumeral morphology extracted from CT imaging into the AnyBody modelling system. The framework is robust to large morphological variation in the patient scapulae and requires little manual input to process large amounts of data. Preliminary results show good conformance in glenohumeral joint forces compared to previous studies.

References

[1] Damsgaard, M., Rasmussen, J., Christensen, S. T., Surma, E., & De Zee, M. (2006). Analysis of musculoskeletal systems in the AnyBody Modeling System. *Simulation Modelling Practice and Theory*, *14*(8), 1100–1111.

[2] Durrleman, S., Prastawa, M., Charon, N., Korenberg, J. R., Joshi, S., Gerig, G., & Trouvé, A. (2014). Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage*, *101*, 35–49.

Modeling of the major temporal arcade in retinal fundus images using genetic programming

I. Cruz-Aceves[1], C. Chalopin[1], M. A. Hernandez-Gonzalez[1], L. M. Lopez-Montero[1]

[1]Centro de Investigación en Matemáticas, Ciencias de la Computación, Guanajuato, Mexico

Keywords Genetic programming, Machine learning, Major temporal arcade, Retinal fundus images.

Purpose

The monitoring of the morphology of the major temporal arcade (MTA) could facilitate the diagnosis of different types of pathology, such as diabetes and hypertension. Consequently, the quantitative analysis of the MTA which is composed of the superior and inferior temporal arcades plays an important role for systems that perform computer-aided diagnosis in Ophthalmology. In this paper, a novel method based on genetic programming to parameterize the architecture of the MTA in retinal fundus images is proposed.

Methods

In literature, the problem of parameterization of the major temporal arcade has been mainly addressed in two different steps: vessel segmentation and numerical modeling. Oloumi et al. [1] introduced the first approach for solving the present problem, where Gabor filters were employed for vessel segmentation, and the Generalized Hough transform was used for the MTA parabolic parameterization. The main drawback using the Hough transform is the fact that the major temporal arcade is not strictly symmetric, which is critical to obtain suitable results. To improve the performance of the state-of-the-art methods based on parabolic modeling, in this paper a novel method based on multiscale Gaussian matched filters (MGMF) for vessel segmentation and genetic programming for the parameterization of the MTA is proposed. The MGMF is a specialized method for automatic vessel segmentation consisting of a multilayer perceptron and a directional Gaussian filter bank with a multiscale width. The multiscale width parameter is used to detect the MTA since it is the thickest vascular structure in the retina. On the other hand, genetic programming (GP) is an evolutionary strategy used for searching a solution in the mathematical expression space [2]. In the second step of the proposed method, GP is used for determining the mathematical expression that best fits the MTA. In this step, the end points of the segmented MTA, and the location of the optic disk are used to guide the GP search. A training set of retinal fundus images is used for tuning the MGFM and the GP parameters in order to be directly applied on the test set of images. It is important to point out that the proposed method is the first approach using genetic programming to parameterize the MTA in retinal fundus images.

Results

The computational experiments were performed in a computer with AMD Ryzen 5 5500U (Radeon Graphics) 2.10 GHz processor and 8 GB of RAM, using Python 3.11.1 with the package gplearn. The DRIVE database (https://drive.grand-challenge.org/) of 40 retinal fundus images was used to perform the numerical modeling of the MTA with 20 images for training and the remaining 20 images for testing. In general, DRIVE database is used for vessel segmentation, in the present work, the delineation of the MTA was performed by a specialist in Ophthalmology of the High Specialty Medical Unit (UMAE-T1, IMSS, Leon, Gto., México). For the experiments, the parameter values for genetic programming were defined using the training set of images as population size of 1000, number of genera tions of 30, crossover rate of 0.7, and mutation rate of 0.05. In order to compare the proposed method with the state-of-the-art using the manual delineation of the MTA, in Table 1, the numerical results are presented in terms of Hausdorff distance and execution time. The proposed method presents superior performance with respect to the state-of-the-art parabolic methods in terms of distance (in pixels) and it also presents competitive performance in terms of execution time. The comparative methods are from the state-of-the-art, where the hybrid method is based on the Univariate Marginal Distribution

Table 1 Performance comparison of MTA detection (average of 30 runs) in terms of the Hausdorff distance and execution time using the test set of images

Method	Hausdorff Distance	Execution time (secs)
Hybrid (UMDA + SA)	105.8 ± 27.54	1.68
Generalized Hough	64.49 ± 0.00	3.92 (per pixel)
MIPAV Software	59.91 ± 0.00	212
Gabor + Hough [1]	34.90 ± 16.60	200
Proposed method	32.46 ± 13.21	34.78

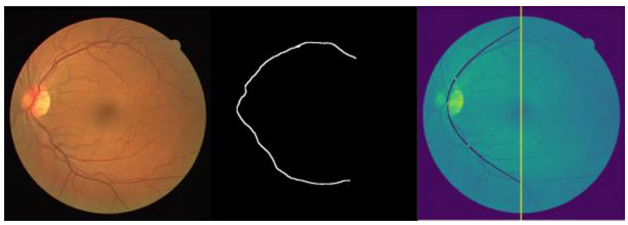

Fig. 1 From left to right, test retinal fundus image, manual delineation of the MTA performed by the specialist, and result of the MTA parameterization using the proposed method

Algorithm with a search guided with the metaheuristic of Simulated Annealing. The Generalized Hough is a matlab script freely available for detecting parabolas in Matlab central, and the MIPAV software (Medical Image Processing, Analysis, and Visualization) is an appli cation for quantitative analysis containing a script for parabola detection.

In Fig. 1, the parameterization of the MTA obtained with the proposed method is presented. The model is based on the location of three points (superior, middle, and inferior) along the MTA, where the vertical yellow line represents the search limit of the points based on the location of the fovea.

Conclusion

In this paper, a new method for the numerical modeling of the MTA using genetic programming has been proposed. The method consists of the steps of automatic MTA segmentation and vessel modeling. In the segmentation step, the multiscale Gaussian matched filter has been used to enhance and segment the MTA since the vessel width is used as a discriminant feature. In the step of numerical modeling, the evolutionary algorithm of genetic programming has been used to parameterize the MTA, and it was compared with four state-of-the-art specialized methods based on parabolic modeling obtaining superior performance in terms of the Hausdorff distance (32.46 + − 13.21 pixels). Moreover, in terms of computational time, the proposed method is competitive with 34.78 s per image.

References

[1] Oloumi F, Rangayyan RM, Ells AL (2012) Parabolic Modeling of the Major Temporal Arcade in Retinal Fundus Images, IEEE Transactions on Instrumentation and Measurement, vol. 61, no. 7, pp. 1825–1838,https://doi.org/10.1109/TIM.2012.2192339.

[2] Koza JR (1994) Genetic programming as a means for programming computers by natural selection, Stat Comput 4, 87–112 (1994). https://doi.org/10.1007/BF00175355.

The CathPilot: Evaluating the Performance, Safety, and Feasibility for Peripheral Vascular Interventions

Y. Alawneh[1], J. Zhou[1], A. Sewani[1], M. Keshavarz[1], M. Tahmasebi[1], T. Roy[2], A. Kayssi[3], A. Dueck [3], G. Wright[4], A. Tavallaei[1]

[1]Toronto Metropolitan University, Department of Biomedical Engineering, Toronto, Canada [2]Houston Methodist Hospital, Department of Vascular Surgery, Houston, United States [3]Sunnybrook Health Sciences, Department of Vascular Surgery, Toronto, Canada [4]Sunnybrook Health Sciences, Department of Physical Sciences, Toronto, Canada

Keywords Peripheral interventions, Catheter manipulation, Endovascular, Angioplasty.

Purpose

Conventional peripheral endovascular revascularization procedures suffer from high technical failure rates (15–20%) and complication rates. This is primarily attributed to challenges in the support, steering, and visualization of current guidewires leading to the inability to cross the arterial occlusions [1]. In an effort to address these challenges, a novel steerable catheter called the CathPilot has been developed [2], Fig. 1. This study aims to evaluate the safety and feasibility of the CathPilot and compare its performance to that of conventional catheters in peripheral vascular interventions.

Methods

The CathPilot performance was validated ex vivo with chronic total occlusion samples (CTO), and its safety and feasibility were demonstrated in vivo in porcine models. The CTO samples were harvested from the popliteal and tibial arteries of two patients with chronic limb ischemia (CLI) or late-stage PAD who had failed revascularization attempts. We experimented with both fresh and fixed samples. For the fixed tissue crossing experiments, the CTOs were stored in a 10% buffered formalin and cut into segments 21.9 ± 7.0 mm in length (6 segments in total with an average ID of 3.1 ± 0.8 mm). The fresh CTOs were stored in 0.9% saline at 4 °C and cut into 47.6 ± 3.6 mm segments (11 segments in total with an average ID of 3.3 ± 1.1 mm). These experiments were conducted within 48 h of amputation. For both experiments, four users attempted crossing the occlusion with a non-steerable catheter and the other with the CathPilot. Each user had 10 min to achieve the task. Success rates, procedure times, and radiation doses were recorded. The CathPilot was tested in vivo on a male swine (56 kg) to assess safety and feasibility. Heparin anticoagulant (100 IU/Kg direct injection and 2000 IU in IV) was administered. Procedures were performed by an experienced vascular interventionalist with more than 15 years of experience. A balloon catheter (7 mm OD, Abbott Armada 35 PTA) was advanced from the carotid artery and deployed midway through the aorta to create blood flow dynamics similar to arterial occlusion. The CathPilot was advanced, from femoral access, through a 12 Fr delivery sheath (Cook Medical G56279) and deployed at the target site inferior to the renal arteries. With the frame of the CathPilot anchored onto the anatomy, it was steered and moved for 15 min, with its motion confirmed via fluoroscopy. It was flushed with saline (∼ 10 ml each time) every 5 min. After the procedure, the pig was euthanized. The artery was excised and examined for any signs of

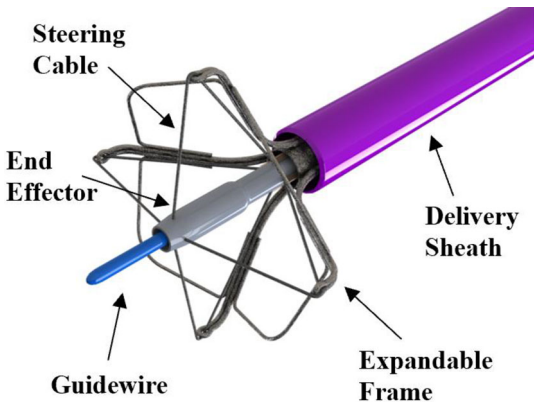

Fig. 1 The CathPilot steering system labelled

damage or trauma to the endothelial layer at the site of CathPilot deployment.

Results

The results of the ex vivo experiments showed that the CathPilot significantly outperformed conventional non-steerable catheters in crossing occlusions for both fixed and fresh tissue samples. For the fresh and fixed lesions, the CathPilot had 100% and 83% success rates, respectively, while the non-steerable catheter only had a success rate of 0% and 9.5%, respectively. The CathPilot also demonstrated significantly shorter procedure times and lower radiation doses in the ex vivo experiments. In the in vivo study, the CathPilot was fully functional with no signs of coagulation on the device, and there was no visible damage or trauma to the endothelial layer at the deployment site.

Conclusion

These findings demonstrate the potential value of the CathPilot technology in improving the success rate of crossing peripheral arterial occlusions. The study also shows that the CathPilot concept is safe and feasible. The CathPilot promises to significantly reduce failure and complication rates of peripheral revascularization procedures and the technical challenges and complications associated with conventional catheters.

References

[1] Farber A, Menard MT, Conte MS, Kaufman JA, Powell RJ, Choudry NK, et al. (2022) Surgery or Endovascular Therapy for Chronic Limb-Threatening Ischemia, New England Journal of Medicine, Dec. 2022, 387(25):2305–2316.

[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. IEE/ASME Trans Mechatronics.

3D cerebral fluidography for assessing cerebrospinal fluid dynamics using diffusion-weighted imaging at human 7 Tesla MRI

S. Fujiwara[1], K. Ogasawara[1], K. Chida[1], Y. Ogasawara[1], J. I. Nomura[1], S. Oshida[1], K. Fujimoto[1], S. Tsutsui[1], K. Setta[1], Y. Yoshioka[2]

[1]Iwate Medical University, Department of Neurosurgery, Yahaba, Japan [2]National Institute of Information and Communications Technology (NICT) and Osaka University, Center for Information and Neural Networks (CiNet), Suita, Japan

Keywords MRI, diffusion weighted imaging, CSF, neurofluid.

Purpose

Diffusion-weighted imaging (DWI) was well-known as a tool for simultaneously estimating the microcirculation of blood in the capillary network and molecular water diffusion in brain tissues. The signal intensity seen on DWI relating to the cerebrospinal fluid (CSF) was also mentioned in the early days, [1], however, little was discussed about the accuracies in qualification and quantification of CSF dynamics in those days because it was before the concept the "glymphatic pathway, [2] The present study aimed to investigate whether 3D cerebral fluidography based on diffusion-weighted imaging (DWI) can visualize cerebrospinal fluid (CSF) dynamics in the brains of healthy subjects using a 7 Tesla magnetic resonance imaging (7 T-MRI) scanner.

Methods

This prospective, observational study was conducted in accordance with the Declaration of Helsinki. Thirteen healthy subjects participated in this study from April 2015 to July 2017. The inclusion criteria for this study were as follows: provision of written informed consent to participate from the participant; age < 50 years and $>$ 20 years; no neurological deficits; no abnormal lesions on anatomical MRI; no history of brain disorders, including those following surgical operation or irradiation, stroke, infection, remarkable atrophy, or demyelinating disease.

All MRI procedures were performed using a 7 T-MRI scanner (Discovery MR950; GE Healthcare, Milwaukee, WI, USA) with a 32-channel head coil. For each subject, a single-shot spin-echo echo-planar imaging (SE-EPI) sequence with 2-mm isotropic voxels was performed as DWI in the axial direction from the section showing the fastigium of the fourth ventricle to that showing both the primary motor area and the central sulcus.

DWI with multiple ten b values (0–2000s/mm^2) was performed for ten subjects using a 7 T-MRI scanner. 3D cerebral fluidography based on the DWI signal variations in different three-orthogonal motion probing gradient directions was developed for visualizing the movement of CSF voxel-by-voxel. Then, apparent diffusion coefficient (ADC) was estimated using the functional fitting procedure for quantifying the CSF movement at the following twelve anatomical space: the subarachnoid space at a section showing the centrum semiovale over the roof of the lateral ventricle (SAS); anterior horns and trigones of the lateral ventricle (aLV and tLV, respectively) on the left and right sides; left and right foramina of Monro (FM); cistern of the velum interpositum (CVI); left and right Sylvian cisterns (SC); the quadrigeminal cistern (QC); and the fourth ventricle at the section showing the interpeduncular sulcus (FV).

Results

Among the 13 healthy subjects who underwent imaging using 7 T-MRI scans, 3 subjects were excluded because of severe image deterioration due to susceptibility artifacts at the frontal lobe region near the frontal sinus and temporal base near the pyramidal bone (n = 2)

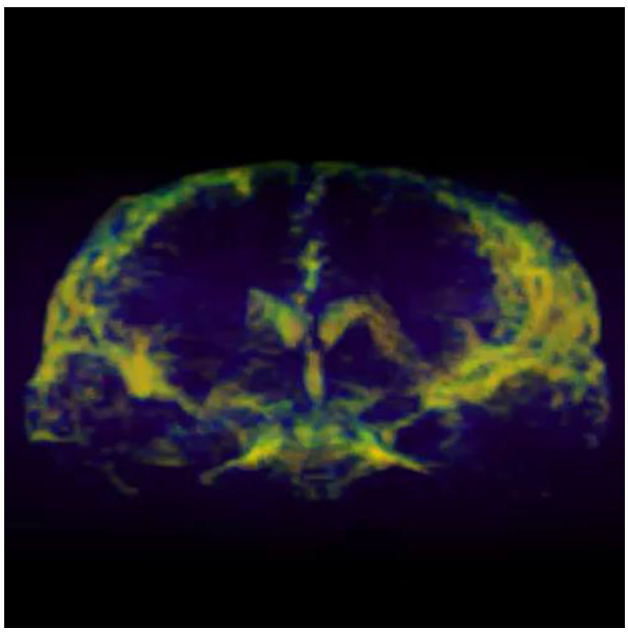

Fig. 1 3D cerebral fluidography (front view)

and signal loss on some images caused by the strong magnetic field inhomogeneity (n = 1). Finally, a dataset of 10 subjects (10 men; mean age, 31.8 ± 3.2 years; range, 28–39 years) was analyzed.

ADC was significantly lower at SAS and aLV than that at all anatomical spaces. On the other hand, ADC at FM, SC or FV was significantly higher than that at the other spaces, which was identical to the findings of 3D cerebral fluidography (Fig. 1).

Conclusion

3D cerebral fluidography could show differences of CSF dynamics among anatomical spaces, and ADC have the potential to quantitatively assess the differences of CSF dynamics. DWI at 7 T-MRI may play an important role to clear a mechanism of the glymphatic system.

References

[1] Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988; 168:497–505.

[2] Iliff JJ, Nedergaard M. Is there a cerebral lymphatic system? Stroke; a journal of cerebral circulation. 2013; 44:S93–95.

Semantic segmentation dataset authoring with simpler labels

L. Uramoto[1], Y. Hayashi[1], M. Oda[1,2], T. Kitasaka[3], K. Mori[1,4,5]

[1]Nagoya University, Graduate School of Informatics, Nagoya, Japan [2]Nagoya University, Information and Communications, Nagoya, Japan [3]Aichi Institute of Technology, Toyota, Japan [4]Nagoya University, Information Technology Center, Nagoya, Japan [5]National Institute of Informatics, Research Center of Medical Big Data, Tokyo, Japan

Keywords Semantic segmentation, Datasets, Weak labels, Laparoscopy.

Purpose

Semantic segmentation, the task of labeling each pixel in an image with its correct class, is a key problem in scene understanding. In order to create more reliable surgical robots, it is essential to develop computer vision models understanding the semantics of surgical images. Deep learning based semantic segmentation for laparoscopic images has been studied. While effective, it requires large amounts of annotated data. Creating ground truth labels is time-consuming; images consist of multiple classes with complex shapes. In laparoscopic datasets this is further complicated by a need for expert annotators identifying organs, tissue types, and tools correctly.

Solutions to the dataset authoring problem have been researched. Simple labels, like lines [1], have been proposed. While addressing the workload problem, such approaches do not address the domain knowledge issue, requiring an expert annotator for every frame. We address the domain knowledge issue by presenting a novel approach to dataset creation, along with a training scheme for arbitrary segmentation models. In the proposed approach, only a part of the dataset is labeled with strong labels, with the rest being labeled with simpler labels.

Methods

We propose a dataset authoring scheme, where parts of the data are labeled with strong labels, and parts with proposed weak labels. A non-expert, unable to create strong labels, could still create simplified, weak labels. Training on the weakly labeled data would be done by transforming the segmentation output to a simplified form to match the weak label.

Our proposed weak labels are created by merging similar classes, simplifying the labeling process. Consider a segmentation problem with n classes. The ground truth labels are vectors **y** of length n. Choosing $m \leq n$ groups and assigning each class into exactly one group, there exists matrix **C** transforming **y** from classes to groups. The matrix **C** is defined elementwise, with the element i, j being 1 if class j is a member of group i, and 0 otherwise. When each class belongs to exactly one class, if **y** is a vector with non-negative elements summing to 1, then so is **Cy**. Thus, given a segmentation output **z** for a pixel and its ground truth **y**, loss functions can be used on the transformed pair (**Cz, Cy**).

In addition to annotating a necessary portion of data with strong labels, human annotators can use weak labels on other parts of the data. A comparison of the strong and the proposed weak labels is shown in Fig. 1. Denote the annotated labels by **y'**, and ground truth by **y**. Instead of labeling **y**, an annotator can label parts of the data using weak labels **y' = Cy**. This would mean, that an annotator could label all tools as "tool" instead of identifying the tool types, simplifying the process. A training sample now consists of a triplet (**x,y',C**), where **x** is an input, **y'** a label, and **C** a transformation matrix. Given a loss function L, training on a dataset of the proposed form is done with L(**Cz, y'**).

We propose a 3-stage training scheme for datasets of this form. First, the model is trained on the strongly labeled data. Then, the learning rate is lowered, and the model trained on the weakly labeled data, including the strongly labeled data transformed to weakly labeled data. Finally, the learning rate is lowered again, and the model trained once more on only the strongly labeled data.

Results

To test the feasibility of the proposed framework, we performed experiments on the CholecSeg8k [2] dataset, using the DeepLabV3 model. CholeSeg8k consists of clips of 16 cholecystectomy videos, and their segmentation ground truths, covering 8080 frames. The dataset contains 13 classes, many present in only a few videos. Images containing these rare classes were removed, and the remaining 6209 images consisting only of the main 8 classes used. The videos were sampled to 5 folds, guaranteeing the presence of each class in all folds. The classes were simplified to 4 groups based on perceived similarities: "tissue" (abdominal wall, fat, and GI-tract), "organ" (liver and gallbladder), "tool" (grasper and L-hook), and "background" (background). While closer to organs, GI-tract was assigned

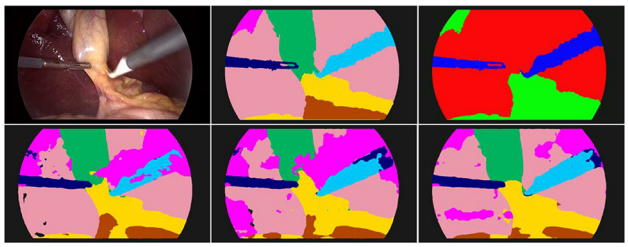

Fig. 1 Top: A laparoscopic image (left), with its strong (center), and proposed weak (right) labels. A grasper (dark blue) and an L-hook (light blue) are operating on a gallbladder (green). The liver (pink), abdominal tissue (magenta), fat (yellow), and GI-tract (brown) can be seen in the background. The weak labels reduce these to "organ" (red), "tool" (blue), and "tissue" (green). Discriminating between fat and the GI-tract is difficult, but in our framework the task is simplified. Bottom: Segmentation results using 50% of the strongly labeled data and 0% (left), 25%(center), and 50% (right) weakly labeled data

Table 1 Quantitative results

Weak	Strong			
	25%	50%	75%	100%
Traditional (0%)	0.622	0.748	0.789	0.828
0%	0.629	0.735	0.803	0.83
25%	0.655	0.751	0.809	
50%	0.707	0.805		
75%	0.727			

The amount of strongly and weakly labeled data used in training are indicated by columns and rows respectively. Impossible combinations are left empty

to the "tissue" group, due to being difficult to distinguish from fat tissue.

Weak labels were created by transforming the ground truth **y' = Cy**, with **C** defined by our groups. Performance was measured with mean Dice score over a fivefold cross-validation run. One fold was used for testing, and four for training. Parts of the training data were simplified to weak labels. The amounts of strongly and weakly labeled training data used for training were varied.

We chose Combo Loss, combining cross entropy and Dice, as the loss function, and Adam as the optimizer. The proposed training scheme, with an initial learning rate of 0.0005, was used. The learning rate was reduced by 25% and 50% in the second and third phases respectively, and models trained for 10 epochs in the first and the second phase, and 5 epochs in the final phase.

Table 1 shows quantitative results of our experiments. Having access to new, weakly labeled data improved performance. At 50% strongly and 50% weakly labeled data, the performance starts coming close to that of a full strongly labeled dataset. We conclude that the presented approach is feasible and can be used to simplify authoring semantic segmentation datasets.

Conclusion

In this paper, we presented a novel approach to semantic segmentation dataset creation. Our approach addresses the domain knowledge problem, allowing for less trained annotators to participate in the creation of large datasets. We showed that with our simplified labels, authoring a large laparoscopic dataset is possible without the need for an expert annotator for each frame.

References

[1] Fuentes-Hurtado F, Kadkhodamohammadi A, Flouty E, Barbarisi S, Luengo I, Stoyanov D (2019) Easylabels: weak labels for scene segmentation in laparoscopic videos. International Journal of Computer Assisted Radiology and Surgery 14(7):1247–1257.

[2] Hong W, Kao C, Kuo Y, Wang J, Chang W, Shih C (2020) CholecSeg8k: A Semantic Segmentation Dataset for Laparoscopic Cholecystectomy Based on Cholec80. arXiv preprint arXiv:2012.12453

Discriminant Analysis of Parkinson-diseases-related patterns in the substantia nigra by nonnegative feature expression of neuromelanin images

H. Itoh[1], M. Oda[1,2], S. Saiki[3], K. Kamagata[3], N. Hattori[3], S. Aoki[3], K. Mori[1,2,4,5]

[1]Nagoya University, Graduate School of Informatics, Nagoya, Japan [2]Nagoya University, Information and Communications, Nagoya, Japan [3]Juntendo University, School of Medicine, Tokyo, Japan [4]Nagoya University, Information Technology Center, Nagoya, Japan [5]National Institute of Informatics, Research Center for Medical Bigdata, Tokyo, Japan

Keywords Parkinson disease, Substantia nigra, Neuromelanin image, Tensor decomposition.

Purpose

This work proposes a novel method to analyse the three-dimensional pattern in the substantia nigra towards automated prediction of Parkinson disease (PD). Parkinson disease is the second most common progressive neurodegenerative disorder. The characteristic of Parkinson disease is a progressive loss of dopaminergic neurons in the substanita nigra pars compacta. Several advanced imaging markers have emerged as tools for visualising neuro-anatomic and functional processes in PD. Neuromelanin-sensitive MRI provides a neuromelanin image (NMI) with neuromelanin-sensitive contrast, and the T1 high-signal-intensity area in the midbrain represents neuromelanin-rich parts. Since neuromelanin exists only in dopaminergic neurons of the substantia nigra pars compacta in the midbrain, we can capture early PD-related changes in the substantia nigra such that left–right imbalance from neuromelanin images. However, even for a neurologist, the classification between PD and non-PD substantia nigra from just reading NMI is difficult [1]. Therefore, we tackle the statistical analysis of NMIs between PD and non-PD to extract discriminant patterns in the substantia nigra.

Methods

Figure 1 shows our pipeline for feature extraction, classification and analysis of patterns in the substantia nigra for classifying PD and non-PD cases. In this work, a volume of interest (VOI) is a region of the substantia nigra in a NMI. Expressing collected VOIs as a fourth-order tensor, we compute its nonnegative tensor decomposition [2]. This decomposes a VOI into a linear combination of third-order rank-1 tensors, i.e., of three-dimensional patterns. Therefore, a set of coefficients of rank-1 tensors expresses a pattern of VOI as a feature vector. While previous work uses CANDECOMP/PARAFAC decomposition [1], negative values in both coefficients and rank-1 tensors hinder the interpretability of its decomposition results. By

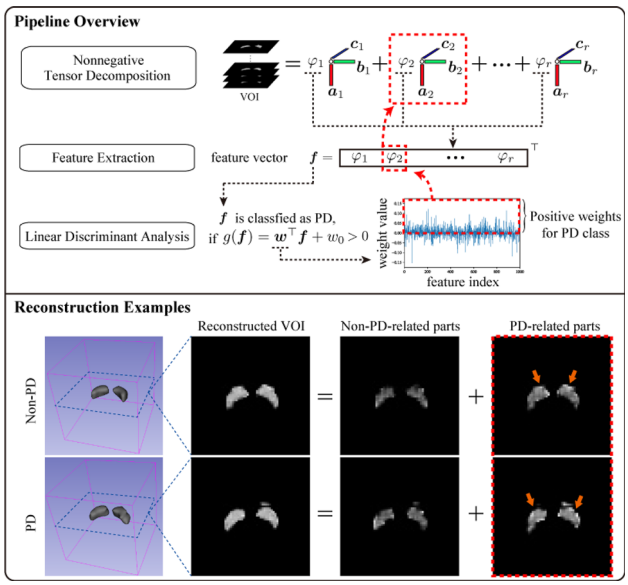

Fig. 1 Overview of the proposed pipeline and example of analytical results. We reconstructed PD-related pattern with rank-1 tensors of the nonnegative tensor decomposition by finding PD-related features from the weights of the decision function in the LDA

introducing nonnegativity to a decomposition, we can increase the decomposition's interpretability. A nonnegative value expresses the existence of a pattern in the decomposition.

Using extracted feature vectors, we compute a weight vector and thresholding value of a decision function in Fisher's linear discriminant analysis (LDA). For nonnegative feature values, a positive-value weight in the decision function indicates a PD-related feature. Therefore, we can select PD-related features and reconstruct PD-related parts by using only the selected features for visual interpretation in our analysis.

Results

For the analysis, we collected 155 NMIs of 73 non-PD and 82 PD patients, where a board-certified radiologist with ten years of experience specialising in Neuroradiology annotated regions of substantia nigra. Based on these voxel-wise annotations, we extracted the VOIs of the substantia nigra from 155 NMIs by setting the centre of a VOI to be the centre of gravity in a region of the substantia nigra. Each VOI is expressed as a volumetric image of $64 \times 64 \times 64$ voxels. After we normalised the value range of VOI to [0, 1], we computed the fourth-order nonnegative tensor decomposition of 1000 rank-1 tensors for the set of 155 VOIs and then obtained 155 1000-dimensional feature vectors. For these feature vectors, we searched a weight vector in the LDA. For the set of feature vectors, the LDA achieved the correct classification of 80% and 78% for PD and non-PD samples, respectively.

We searched PD-related and non-PD-related features from the LDA's weight vector. Using these two sets of features, we reconstructed VOIs and compared them with reconstructed VOIs without feature selection. Figure 1 shows an example of the comparison. In Fig. 1, the marked parts suggest the reconstructed ventrolateral tier, where the severe loss of neurons is happening in Parkinson disease. The LDS further weights the importance of patterns in these regions by its weights for classification. From these results, we concluded that the proposed method finds neuropathologically correct and discriminative patterns in the substantia nigra.

Conclusion

This work proposed the method to analyse the three-dimensional pattern of the substantia nigra in the PD and explored its validity towards automated prediction of the PD. The experimental results showed that our proposed method achieved interpretable feature extraction and decision function via the feature-selected reconstruction of VOIs. Even though this work still needs the evaluation of classification accuracy for an external testing dataset, we will tackle this point in our future work. Parts of this research were supported by the Japan Agency for Medical Research and Development (AMED, No.22dm0307101h0004) and the MEXT/JSPS KAKENHI (No. 21K19898).

References

[1] Itoh H, Tao H, Oda M, Saiki S, Kamagata K, Hattori N, Aoki S, Mori K (2022) Pattern Analysis of Substantia Nigra in Parkinson Disease by Fifth-Order Tensor Decomposition and Multi-sequence MRI, Proc. Multiscale Multimodal Mdedical Imaging 2022, LNCS 13,594: 63–75

[2] Cichocki A, Zdunek R, Phan AH, Amari S (2009) Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. Wiley, Hoboken

Engineers in medicine part 1: A pilot comparative profile of engineering students vs non-engineering students in admission to medical school

N. Groth[1], C. Y. Lin[1], E. vanSonnenberg[1], G. Fogerty[1], J. Winkleman[1]

[1]University of Arizona College of Medicine—PHoenix, Medicine, Phoenix, United States

Keywords Engineer, Medical School, Admissions, Profile.

Purpose

With a growing concordance between medicine and engineering, physicians with backgrounds in engineering, artificial intelligence, and informatics likely will have an increasing impact in the advancement of medicine. We have implemented several projects to understand the career paths and interests of engineers as they transition to medicine. The purpose of this study is to gain insight into the profiles of engineers who have matriculated into our medical school, and to compare their profiles to non-engineering major students.

Methods

This is an anonymous retrospective study that utilized data provided by a single medical school admissions department from 2013 to 2021. There was a total of 49,454 applicants spanning the 9 years of data. 1711 applicants met the study requirements to be classified as engineers. Applicants were separated, based on undergraduate degree, and any student with an engineering degree was classified as an engineer. Those students with more than one degree must have at least one pertaining to engineering to be included in the engineering cohort; all others were classified as non-engineers as a separate cohort. Acceptances were offered to 1434 total applicants, 79 of whom had engineering degrees. The applicants with engineering majors were compared to non-engineering students, assessing several factors including: acceptance rate, undergraduate GPA, science GPA, MCAT scores, MCAT Chemistry/Physics section score, combined clinical

hours, and shadowing hours. The data analysis was done with Rstudio under the "Spotted Wakerobin" release, with tidyverse packages, and the graphs were created with a ggplot2 package.

Results

From 2013 to 2021, our school saw a trend of an overall increase in applications per year across both engineering and non-engineering students. With the exception of 2016, which showed an acceptance rate of 2.5% for engineer applicants and 3% for non-engineering applicants to medical school, an overall nonsignificant higher acceptance rate of 5% for engineers occurred, compared to the 3% for non-engineering applicants (X2, $p = 0.218$). The engineering majors with the most offers of admission were biomedical and biosystems engineers 53/79 (67%), followed by chemical engineers 10/79 (13%), and mechanical engineers 5/79 (6%). The results of the t-test showed that there was no significant difference between the two cohorts in undergraduate GPA, science GPA, MCAT scores, MCAT chemistry/physics scores, combined clinical hours, and shadowing hours, which all yielded p-values > 0.05 per comparative metric.

Conclusion

In this pilot study, we hypothesized and expected that there would be a significant difference in acceptance rates, undergraduate GPA, science GPA, and MCAT scores between applicants to medical school with an engineering background vs those without an engineering background. Although engineers did show a generally higher acceptance rate, our results did not support our hypotheses, as we found no statistically significant differences between the two cohorts in any of the comparison metrics. It is possible that the small sample size of the study may have limited the power and accuracy of our results. To further investigate this issue, a multi-center study would be in order, and we are presently in the planning stages of that study. This larger study will likely provide more robust and reliable results, and should help to elucidate the potential impact of an engineering background on successful admission to medical school.

Engineers in medicine part 3: what are their engineering majors and what specialties do they pursue?

M. Rosztoczy[1], E. vanSonnenberg[1], D. Crawford[1,2], P. Kang[1], P. Yue-Cheng[1,3], M. Mamlouk[1,4,5], G. Fogerty[6], J. Winkleman[1], N. Groth[1]

[1]University of Arizona College of Medicine—Phoenix, Medicine, Phoenix, United States [2]Washington University School of Medicine in St. Louis, Department of Radiology, St. Louis, United States [3]Stanford University School of Medicine, Department of Radiology, Stanford, United States [4]The Permanente Medical Group, Kaiser Permanente Medical Center, Department of Radiology, Santa Clara, United States [5]University of California, San Francisco, (5) Department of Radiology and Biomedical Imaging, San Francisco, United States [6]University of Arizona College of Medicine—Phoenix, Medicine, Phoenix, United States

Keywords Engineer, Residency, Medical School, Graduate.

Purpose

The intersection of engineering and medicine is a dynamic and rapidly emerging concordance that involves the translation of innovative ideas from engineers to the implementation of clinical solutions by physicians. As the role of engineering in medicine continues to evolve and expand, it is important to understand the career paths and areas of focus for engineers who have made the transition into the medical field. We are currently engaged in several projects analyzing the interconnections and potential for engineers to pursue a career in medicine. This project focuses on the subcategories of engineering degrees the medical students majored in, and medical specialties those with undergraduate degrees in engineering pursued.

Methods

29 medical students who graduated from our medical school in the last eight years with undergraduate degrees in engineering were investigated. Their specific engineering major was noted, and their residencies were used to determine what field of medicine they pursued. The most popular fields were determined, along with the divide between primary care, surgical specialties, and the remainder of the specialties: Psychiatry, PM&R, Pathology, Emergency Medicine, Diagnostic Radiology, Radiation Oncology, Anesthesiology, Neurology, Dermatology (P3ERRAND). A Fisher's Exact Test was used for statistical analysis to determine if a specific type of engineering was related to specialization in one of the three subcategories.

Results

Overall, engineering majors represented 2–11% of each class, totaling 29/539 (5%) of graduates in the period of 2014–2021. The most common engineering major was biomedical engineering 18/29 (62%), followed by electrical engineering and chemical engineering 3/29 (10%) each. Bioengineering was the third most common 2/29 (7%), and two students had unspecified engineering majors 2/29 (7%). There was one mechanical engineering major 1/29 (3%). All 29 graduates successfully matched into their chosen residency programs through the NRMP.

The most frequently entered specialty was Surgery 7/29 (24%), followed by Radiology 6/29 (21%), and then Emergency Medicine 5/29 (17%). Divided into subcategories, 10/29 (34%) entered a surgical field, 7/29 (24%) pursued primary care specialties, and 12/29 (41%) were classified as (P3ERRAND).

A Fisher's Exact Test showed statistically significant results for the biomedical engineering majors; these graduates were less likely to specialize in either primary care or Surgery ($p < 0.001$). The correlation between biomedical engineering and P3ERRAND was nonsignificant, yielding a p-value of 0.36. The results of the other specific engineering majors were not statistically significant, all yielding p-values > 0.01.

Conclusion

These preliminary data suggest that the most frequent undergraduate degree in engineering of students who attend our medical school is biomedical engineering. The most common subcategory of residency pursued was Surgery. However, biomedical engineering majors who made up the majority of engineers graduating from our medical school were less likely to pursue primary care and Surgery with statistical significance. The results of this pilot study could be further validated with an increased sample size, although each graduating class had an engineering student present. The generalizability and the accuracy of these results can be furthered with implementation of a multi-center study. We are currently in the planning stages of that study.

Engineers in medicine part 2: What is medical school like for engineering majors?

J. Winkelman[1], E. vanSonnenberg[1], G. Fogerty[1], J. Foote[1], N. Groth[1]

[1]University of Arizona College of Medicine Phoenix, Medicine, Phoenix, United States

Keywords Engineers, Medical School, Medical Education, Engineers in medicine.

Purpose

Scant research has focused on the effects of various college majors and their influence on the success and challenges of medical school. This study seeks to gain insights on those students who majored in engineering, their perspectives, and how they perform, compared to their non-engineering classmates in medical school.

Methods

17 students in our current medical school who majored in engineering during college were surveyed. The survey was created on Qualtrics, an online survey software platform. The survey consisted of 21 multiple choice questions. Most questions were Likert scale questions, asking participants to indicate the degree to which they agree or disagree with a statement, ranging from strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree.

Questions were largely divided into two subsets. The first subset asked participants to compare their medical school academic performance and stress levels to their peers in medical school. The second subset of questions asked participants to contrast their medical school achievements and stress levels to their experiences when studying engineering during their undergraduate studies. Questions also were organized to analyze if respondents in different years of medical school had differences in responses to one-another. Chi square test was used to determine if there were differences in the categorical distribution, and whether the distributions differed from what would be expected to happen by chance.

Results

The engineering students compared to their non-engineering medical school classmates believe, with statistical significance, that they are less stressed (11/17 (64.7%), P = 0.0005), and are academically advantaged (14/17 (82.4%), P = 0.0001). The engineering students, when asked to compare their experience in medical versus engineering schools, state with statistical significance that they spend more time on their medical studies than on their college engineering studies (12/17 (70.6%), P = 0.0001). When comparing differences by year in medical school, second year students stated, with statistical significance, that medical school is harder than engineering, compared to first and fourth year students (second years: 5/6 (83.4%), all other years: 0/11 (0%), P = 0.03). Currently, not statistically significant, but trending, is that the engineering students as a whole (not divided by year) say the first two years of medical school were easier compared to engineering (P = 0.16). When students were asked to compare medical school to their engineering studies, they claim there is no difference in their academic performances (P = 0.71) or stress levels (P = 0.41, not statistically significant).

Conclusion

Our survey results indicate that students in medical school who majored in engineering view themselves differently, and in some respects, more favorably, compared to their non-engineering classmates. This suggests there may be significance to students with engineering backgrounds vis-à-vis how these students learn and react to the stresses of medical school. They believe they are both academically advantaged and less stressed compared to their classmates, although they say medical school is more time consuming compared to their engineering studies. There is no statistically significant agreement among students surveyed on if medical school is academically easier or less stressful compared to their engineering studies. A limitation of the study is that it includes only one institution as a pilot. We are planning a multi-institutional study as an outgrowth of this pilot study to provide further power and validity.

Endoscope manipulator with gimbals rotation and linear rail mechanism for transcanal endoscopic ear surgery

M. Izawa[1], K. Yokoyama[1], T. Kawai[1], T. Fujita[2], N. Uehara[2], T. Yamashita[2], A. Nishikawa[3], H. Suzuki[4]

[1]Osaka institute of technology, Graduate School of Robotics and Design, Osaka, Japan [2]Kobe University Graduate School of Medicine, Otolaryngology, Kobe, Japan [3]Osaka University, Graduate School of Engineering Science, Toyonaka, Japan [4]Chuo University, Faculty of Science and Engineering, Bunkyo-ku, Japan

Keywords Surgical robot, Endoscope manipulator, Third arm, Transcanal endoscopic ear surg.

Purpose

A locally operated surgical support robot holding an endoscope is introduced in a sterile area for otologic surgery, and a surgeon can perform robot-assisted transcanal endoscopic ear surgery (TEES) with low physical load and view stabilization. Although serial linkages manipulator RobOtol with a mechanical pivot point standing on the floor provides two hands surgery solely [1], the cantilevered mechanism overhangs on the body because of the mechanical structure. Therefore, we have developed small locally operated detachable end-effector manipulator as the endoscope robot connected to a bed rail (E-LODEM) for TEES leading to precise and safe motion in the ear canal [2]. A new small endoscope manipulator which has motor driven three degree-of-freedoms was developed that can act as a third arm for the surgeon. The mechanism of the manipulator is with gimbals rotational linkages for pivoting motion of the endoscope, and with a linear guide rail for its inserting and removal motion.

The present study describes the newly proposed two types of manipulators in detail. In addition, we report the mechanical performance of proposed manipulators evaluated while performing in simulated TEES.

Methods

The first developed manipulator consists of bevel geared direct current (DC) servo motors for the yaw and pitch axes and a belt pulley mechanism with a DC servo motor for the insertion axis attached to an endoscope (Manipulator 1). The distance between the pivot of the mechanism and the ear canal entrance is 200 mm. The operation range is $\pm$ 1° adjusted by mechanical pins for the rotational axes, and 0 to 100 mm for the linear axis. The positional accuracy is 0.1 mm, 0.2 mm, and 0.6 mm respectively. The dimensions are 200 mm $\times$ 120 mm $\times$ 200 mm. The mass is 1.5 kg. The manipulator is leader–follower controlled by the assistant. Two leader devices embedded encoders are a joystick controlled for the yaw and pitch axes, and a rack-and-pinion geared linear slider controlled for the insertion axis. A motion scale of the leader–follower control was set at 2:1 in the simulated surgery. The leader–follower control is performed when a footswitch is pressed.

The second developed manipulator consists of geared DC servo motors for the yaw and pitch axes and a rack-and-pinion mechanism with an ultrasonic motor for the insertion axis attached to an endoscope (Manipulator 2). The operation range is $\pm$ 1° adjusted by ratchet gears for the rotational axes, and 0 to 30 mm driven by the ultrasonic motor in the inside of ear and 30 to 100 mm driven by manually in the outside of ear for the linear axis. The endoscope can be manually removed from the external auditory canal in the full range of linear motion releasing clam pin. The positional accuracy is 0.1 mm, 0.2 mm, and 0.1 mm respectively. The dimensions are 274 mm $\times$ 195 mm $\times$ 201 mm. The mass is 2.0 kg. The manipulator is on–off controlled by the foot of the surgeon. The foot-operated

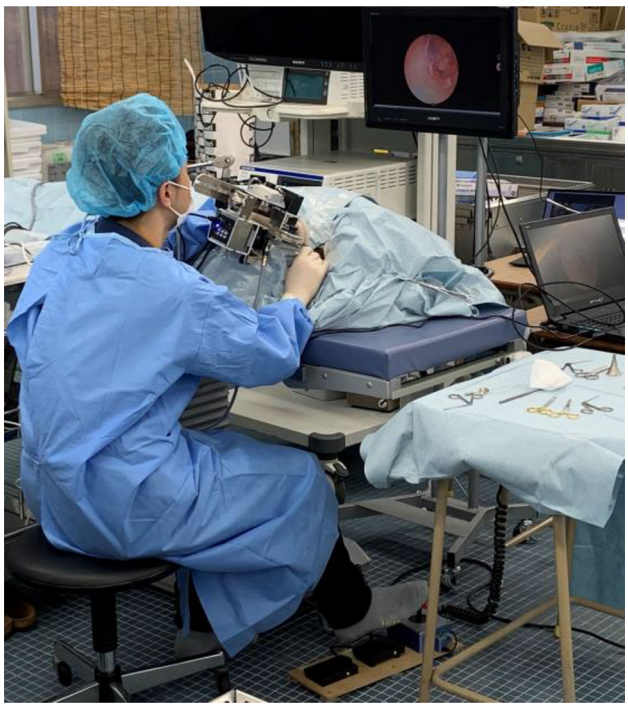

Fig. 1 Simulated surgery performed by an otologic surgeon using an endoscope manipulator for TEES with gimbals rotation and linear rail mechanism driven by DC servo and ultrasonic motors

device constructed of a joystick switch in four directions for the yaw and pitch axes, and two switches in the forward and backward for the insertion axis.

We performed a simulated approach in a trans-auricular endoscopic otoscopic surgery on a cadaver using two types of manipulators. The experimental situation is shown in Fig. 1. The manipulator with the endoscope is placed in front of the left ear of the cadaver using a fixed arm connected to a bed rail. The surgeon operated with a scalpel in his right hand and a suction tube in his left hand. This simulated surgery was performed at the Clinical Anatomy Training Center, Kobe University School of Medicine, based on the "Guidelines for Cadaver Dissection in Education and Research of Clinical Medicine" published by the Japan Surgical Association and the Japanese Association of Anatomists.

Results

The endoscope held by Manipulator 1 and the surgical instruments held by the surgeon in both hands were inserted simultaneously into the ear canal. The assistant was also able to operate the manipulator intuitively by leader–follower control. However, Manipulator 1 was difficult to position due to bevel gear backlashes and mechanical deflection of the fixed arm.

The simulated surgery using the Manipulator 2 was able to open the tympanic membrane and observe the middle ear. The surgeon could perform solo surgery by operating the manipulator with the foot switch device. However, due to the small gap between the shoulder of the cadaver and the manipulator, it was difficult for the surgeon to insert the instrument in his left hand from the desired position.

Conclusion

We performed a simulated surgery on a cadaver using two types of small endoscope manipulators with gimbals rotational linkages and a linear motion rail developed for robot assisted TEES performed by a surgeon in a sterile environment. The designed operating range and the positional accuracy of these prototypes were satisfied with observing the middle ear using the endoscope and two surgical tools. The results of this study show that the proposed manipulator can be used for such applications. Future studies will construct downsized manipulator and stable fixed arm for initial positioning. This study was partially supported by Kakenhi (JP22H00589).

References

[1] Vittoria S, Lahlou G, Torres R, Daoudi H, Mosnier I, Mazalaigue S, Ferrary E, Nguyen Y, Sterkers O (2021) Robot-based assistance in middle ear surgery and cochlear implantation: first clinical report. Eur Arch Otorhinolaryngol, 278(1):77–85.

[2] Fujita T, Nishikawa A, Kawai T, Suzuki H, Uehara N, Yamashita T, Kakigi A, Nibu K (2022) Research on robots for TEES and their control. The 4th World Congress on Endoscopic Ear Surgery, Panel Discussion/Round Table 33.

Postoperative CT scan comparison after corrective Osteotomies in Wrist and Forearm Surgery using a laser osteotome—a cadaver study

M. Hofer[1], E. Coppo[2,3], M. Morawska[4], M. Müller-Gerbl[5], P. Honigmann[1,2]

[1]Universität Basel, Medizinische Fakultät, Medical Additive Manufacturing, Department of Biomedical Engineering, Basel, Switzerland [2]Kantonsspital Baselland, Hand and peripheral Nerve Surgery, Liestal, Switzerland [3]Universität Basel, Medical Additive Manufacturing,Department of Biomedical Engineering, Basel, Switzerland [4]Advanced Osteotomy Tools AG, Basel, Switzerland [5]Universität Basel, Medizinische Fakultät, Basel, Switzerland

Keywords robotic assisted surgery, laser ablation, osteotomy forearm, hand surgery.

Purpose

Conventional bone surgery is associated with direct contact of instruments with the bone causing friction, heat, pressure and hence, damaging the bone and surrounding soft tissues. We introduced the technology of navigated cold ablation robot-guided laser osteotomy in a previous publication showing no carbonization, evaluating new feasible cutting patterns, while achieving sufficient primary stability and compression with lag screws only [1]. The purpose of this ongoing study is to standardize the workflows and to show accuracy in pre-clinical cadaver tests performing one and two plane osteotomies in the field of wrist- and forearm surgery comparing pre- and postoperative CT scans.

Methods

Technique

Osteotomies were performed with CARLO® which is a miniaturized ablation laser with an optical system controlled by a navigation system. The system uses Yttrium Aluminum Garnet doped with Erbium (Er:YAG), with a wavelength of 2943 nm, that corresponds to peak absorption coefficient for water and hydroxyapatite. The energy of a laser pulse hitting the bone tissue heats up the water content of the bone and vaporizes it. The increase in local pressure causes "micro-explosions", breaking up the bone structure. The debris is being expelled immediately and at high velocity, providing a clean-cut line with preservation of the bone microstructure [2]. CARLO® also has an Optical Coherence Tomography (OCT)-based depth control system to visualize the current cutting level and avoid soft-tissue

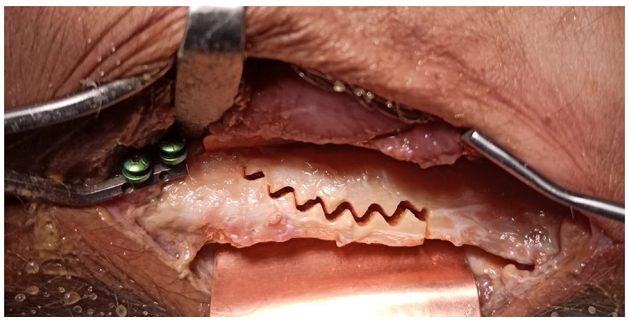

Fig. 1 Corrective osteotomy of the distal radius lengthening it by 4 mm and correcting radial inclination by 10°

damage. CARLO® enables new cutting patterns that are impossible to achieve with conventional methods.

Cadaver tests

A total of 20 corrective laser-osteotomies were performed on the distal metaphyseal radius and ulna. The Osteotomies were planned patient specific and ad hoc on a 3D CT-model using CARLO®'s planning software. On the radius ten two plane adjustments were conducted lengthening the radius by 4 mm and correcting palmar angulation, respectively radial inclination by 10° using a curved sine cut (Fig. 1). On the ulna one plane corrective osteotomies were performed by shortening the ulna by 2.5 mm using a sawtooth or sine cut. Standard surgical approaches were carried out to expose the bones. The navigation device was mounted at the proximal end of the surgical field using two self-drilling screws. The osteotomies were stabilized using lag screws only. Pre- and postoperative CT-scans were taken to compare the virtual surgical planning with the postoperative results and assess reliability.

Results

Efficient workflows have been established thanks to the modification of the planning software making patient specific preoperative planning of the osteotomy more user friendly. Definite accuracy measurements concerning the one and two plane adjustments have not been finally evaluated yet, since the trials are still ongoing. Preliminary trials however showed minor deviations to presurgical planning, achieving similar results as today's usual methods using 3D printed patient specific cutting guides. Additionally, the osteotomy sites in dia- and metaphyseal regions did not show any carbonisation and the lag screws provided good compression and stability, potentially making a plate for fixation obsolete.

Conclusion

First cadaveric results show promising results concerning the accuracy of one and two plane osteotomies in wrist and forearm surgery using a Cold Ablation Robot-Guided Laser Osteotome. The fast and patient specific preoperative planning could make the expensive and time-consuming process of designing patient specific guides redundant. Future steps are stability testing of the osteosynthesis followed by certification and first use in patients.

References

[1] Honigmann P, Hofer M, Hirsch S, Morawska M, Müller-Gerbl M, Thieringer FM, Coppo E. Cold ablation robot-guided laser osteotomy in hand, wrist and forearm surgery-A feasibility study. Int J Med Robot. 2022 Oct;18(5):e2438. https://doi.org/10.1002/rcs.2438. Epub 2022 Jul 8. PMID: 35,770,622; PMCID: PMC9541476.

[2] Baek, K.; Deibel, W.; Marinov, D.; Griessen, M.; Dard, M.; Bruno, A.; Zeilhofer, H.; Cattin, P.; Juergens, P. A Comparative Investigation of Bone Surface after Cutting with Mechanical Tools and Er:YAG Laser. *Laser Surg Med* 2015, *47*, 426–432, https://doi.org/10.1002/lsm.22352.

Automated Surgical Procedure Capture and Analysis System for Surgical Performance Evaluation

K. Yoshimitsu[1], T. Ishikawa[1], Y. Muragaki[1], K. Masamune[1]

[1]Tokyo Women's Medical University, Institute of Advanced Biomedical Engineering & Science, Tokyo, Japan

Keywords Surgical Procedure, 3D LiDAR, Automated system, neurosurgical endovascular.

Purpose

The missions imposed on the surgical operating team each have an "ideal surgical procedure" to a greater or lesser degree. Various factors that create a gap with the ideal form, and in many cases, the accumulation of those gaps. In other words, it leads to a decrease in surgical quality, efficiency, and productivity as a demerit method. Conventionally, many surgical support systems, both hardware and software, have been developed with the aim of reducing the burden on surgical staff and realizing efficient surgical support. However, efforts to improve the quality and performance of surgery by unraveling the gap between the ideal and the ideal state throughout surgery have not been realized.

The purpose of this research is to establish a method to immediately score the performance of the clinical team after the operation and to show that the technique and work efficiency are improved.

The final goal of this research is to present a performance score at the end of surgery. This system is like the automatic scoring function of karaoke, "Today's case is 96/100 points!". Thus, we aim to improve the performance of the clinical team as a whole.

Methods

In order to achieve the above-mentioned purpose, we decided to build a system in two stages that is able to sense automatically and identify the procedure and task content of each medical staff during a case. The first step is to build a system to identify "who, what, and when". The second step is to construct an evaluation function that scores the clinical team's performance based on the measurement data. In this paper, the first stage was carried out. We targeted neurosurgical endovascular treatment cases, a large number of cases of surgery to be measured, and the procedure is standardized, and the difference in technique and team performance could be evaluated remarkably when comparing skilled and inexperienced surgeons. We investigated two products of infrared depth sensors for staff movement line tracking sensors. 3D LiDAR sensor was chosen that has a rich track record of synchronizing with other modalities and system cooperation, and installed 20 such sensors on the ceiling surface in the corridor in hallway, the command room, and the operating room to eliminate blind spots. This system enables to know where the person is on the operating room floor. Since the analysis system can define areas, it is possible to analyze what is being done by linking areas and tasks. By linking the face recognition system to individual identification, the system could collect data on when and who is doing what by integrating the above.

Results

Figure 1 shows the flow line of surgeons and nurses working in the operating room.

The figure shows the change in the coordinates of the location in the operating room over time with a flow line, and to express it with a

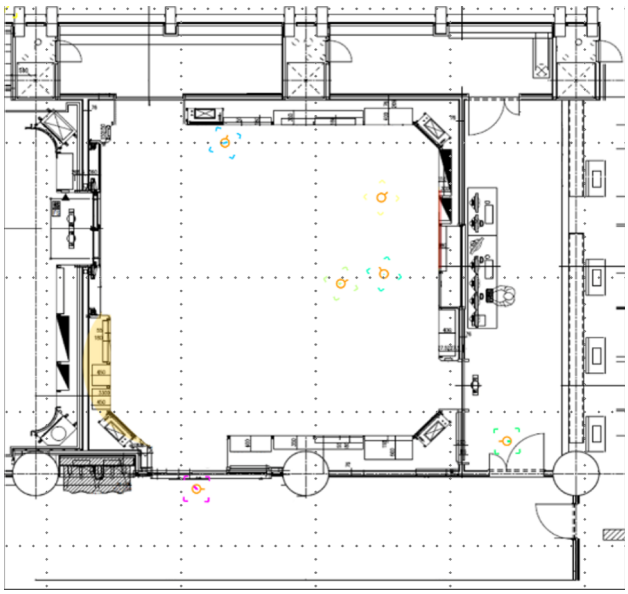

Fig. 1 Real-time capture screen of the flow of surgeons and circulating nurses in the surgical theater

heat map according to the length of stay. In addition, the data could be obtained to determine how much time was spent on which task from the correlation between location and task.

The constructed system was able to measure when and who was doing what. On the other hand, in order to grasp more detailed and real-time surgical conditions and task details, small tasks and highly accurate analysis can be realized by utilizing equipment status and advanced sensing devices. Highly accurate analysis of surgical conditions could be used to automate various devices operating in the operating room using the analysis results themselves as input, which is considered to be of great help in further improving surgical efficiency.

Conclusion

The performance of the clinical team is immediately scored on the spot after the surgery, and the operating room is equipped with 20 infrared depth sensors installed on the ceiling for the purpose of establishing a method to show improvements in technology and work efficiency. We built a system that can collect data on when, who is doing what on the floor. In the future, we would carefully consider the analysis method and proceed with the development of applications that lead to the improvement of surgical efficiency.

CARS in 2023: Redefining Radiology and Surgery

R. Andrews[1]

[1]NASA Ames Research Center, Nanotechnology & Smart Systems, Los Gatos, United States

Keywords artificial intelligence, personalized healthcare, smart devices, wearables.

Purpose

Radiology traditionally has been defined as "the science dealing with x-rays or nuclear radiation, especially for medical uses". Surgery has been defined as "the art, practice, or work of treating disease, injuries, or deformities by manual or operative procedures". [Dictionary.com] In 2023, radiology is more appropriately defined as "imaging"; surgery might be more appropriately defined as "interventional diagnosis or treatment" (as opposed to pharmacological diagnosis or treatment).

We argue that the traditional boundaries defining "radiology" and "surgery" no longer are appropriate in 2023—in part thanks to computer-based advances over the past several decades, e.g. deep learning (DL) and artificial intelligence (AI)—and that even "imaging" and "intervention" suggest boundaries that are being dismantled as diagnosis and treatment techniques advance and integrate.

Methods

Medical imaging has progressed from radiological films ("x-rays") to non-radiological imaging (e.g. ultrasound) to "unseen by human eyes" digital computerized image analysis such as DL/AI interpretation of chest CTs and mammograms. No longer is an image created; rather, the data collected (e.g. pixels) are directly analyzed to provide information ("normal" vs "abnormal", "normal" vs "fibrocystic disease" vs "carcinoma"). Thanks to DI/AI—where thousands (potentially millions) of patients" imaging data are accumulated and assessed (greatly exceeding the "data bank" an individual radiologist can accumulate in a lifetime)—accuracy can be markedly enhanced.

Surgery now includes techniques quite remote from traditional manipulation, excision, or implantation. Examples include not only radiosurgery and interventional "radiology", but more unusual techniques such as transient blood–brain barrier opening with MRI-guided focused ultrasound.

Where is "Computer-Assisted Radiology and Surgery" headed in 2023 and beyond?

We argue that CARS, cutting-edge technology for diagnosis and treatment, will bypass the traditional healthcare institutions (office-, clinic-, or hospital-based imaging and interventions) in favor of gathering data directly from the patient for diagnosis and using similar patient-based techniques for treatment.

Results

Wearable devices ("wearables") meet the criteria for truly patient-based, "point-of-care" (POC), diagnosis and treatment. Continuous individual data, pooled from thousands of patients, are used in digital clinical trials—yielding (thanks to DL/AI computer analysis) precise, personalized diagnoses and treatments. Wearable devices, those in close contact/proximity to the individual, include smart rings/wristbands/wristwatches, skin patches, smart clothing/glasses/masks, ingestibles, and implantables.

A graphene patch for bioimpedance monitoring can penetrate tissue 20 mm or more; such patches have been shown to be an accurate blood pressure monitor. Microneedle patches—some biodegradable—can deliver drugs, cells, and subcellular moieties over hours to days to weeks. Liquid metal (flexible) fibers have been embroidered onto fabric for "smart garments".

An obvious application is diabetes: glucose can be monitored continuously by a skin patch, with feedback-guided delivery of insulin and/or glucagon by an implantable device. Multi-sensor rings and bracelets are employed for personalized protocols addressing conditions from weight management/obesity to sleep disorders. Optimizing sleep with such wearables to enhance athletic performance has been pursued by professional sports organizations like the National Basketball Association in the USA.

Additional applications of wearables include monitoring infants for developmental motor milestones, monitoring the response to medications of patients with Parkinson's disease or epilepsy, and monitoring/treating gastrointestinal tract disorders with smart ingestibles.

Smartphones are key to many wearable applications. They can be combined with various biosensing wearables for analysis of bodily fluids, e.g. mouthguard (saliva), contact lens (tears), skin patch (sweat). The technology is advancing from monitoring a single analyte (e.g. pH) to multiplex monitoring. Data from individual patients

can be transmitted by smartphone (in some applications, smartphones equipped with optical or electrochemical sensors) to a central data repository for DL/AI assessment. Thus both diagnosis and treatment can be continuously upgraded thanks to the combination of individual and "crowd-sourced" data.

Wearables offer continuous personalized diagnosis and treatment that is not constrained by traditional institution- and physician-based "intermediate" institutions. The efficiency of wearable "closed-loop" diagnosis and treatment at the level of the individual patient will—thanks to DL/AI—transform traditional Radiology and Surgery in ways we cannot even imagine at present.

Conclusion

As more conditions previously considered for traditional surgery become treatable with pharmacologic and cell-based technologies, "smart" wearables communicating amongst themselves (at the individual level) and with central data repositories (via smartphone) promise to be the next generation of CARS.

Multipurpose Prediction Model for Type-A Patients in Heart Surgery using Bayesian Networks

M. Cypko[1], A. Abd el Al[1], M. Montagner[1], V. Falk[1], A. Meyer[1]

[1]Deutsches Herzzentrum der Charité—Berlin, Department of Cardiothoracic and Vascular Surgery | CVK, Berlin, Germany

Keywords clinical decision support, bayesian network, expert system, heart surgery.

Purpose

In cardiac surgery, there is a need for decision support at various stages of the clinical patient journey in acute Type-A patients. Some prognostic approaches exist with promising results but use intransparent multiple linear regression models. Especially in the context of critical treatment decision making, decision recommendations should be transparent. Knowledge-based machine intelligence approaches, particularly Bayesian networks, are recognized for transparent and reproducible decision support [1]. Knowledge-based machine intelligence approaches, in particular Bayesian networks, are recognized for transparent and reproducible decision support. Therefore, we performed a proof-of-concept here from a methodological and application-oriented perspective.

Methods

First, we use Bayesian networks to develop a clinical risk model for Type-A patients in heart surgery focusing primary on five sets of clinical outcome variables: *In-house mortality*, *30-day mortality*, *One-year mortality*, *Surgery*, and *MACCE*. Second, we highlight another strength of the flexibility of Bayesian networks by applying our Type-A model in a clinical workflow with four phases: *Aortic Emergency Telephone*, *Hospitalization*, *Directly After Surgery* and *Post-Surgical Care*, see Table 1. Each clinical phase is distinguished by a different set of available patient data and a different clinical interest in the model predictions. To keep the balance between modeling effort and model quality, we developed a semi-automatic modeling and validation pipeline, see Fig. 1.

Table 1 Overview of four defined phases, each with different available datasets and interest in decision support

Phase (P) of Prediction in Patient Journey	P1: Aortic Emergency Telephone	P2: Hospitalization	P3: Directly after Surgery	P4: Post-Surgical Care
Available Patient Examination Data	Checklist Questionary	P1 data + Imaging, Echo and Clinical Examination	P2 data +Intra-OP	P3 data + Post-OP
DHZB's Interest of Prediction	30-Day Mortality + 1-Year Mortality	Surgical Procedure + In-Hospital Death	MACCE + 30-days Mortality + 1-Year Mortality	+ 30-days Mortality + 1-Year Mortality

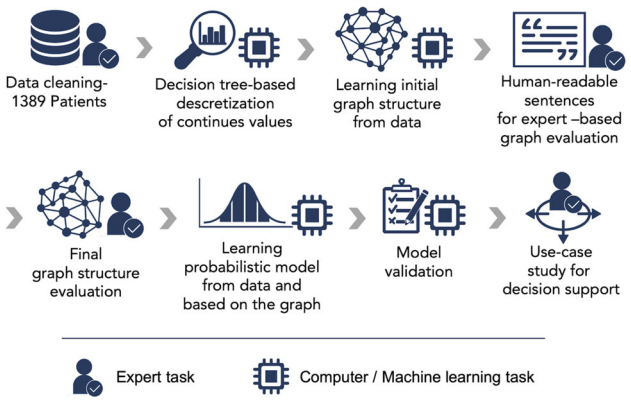

Fig. 1 Methods and their sequence in the semi-automatic development and validation of a knowledge-based Bayesian network decision model

Results

The final decision model consists of 64 decision-relevant variables including basic patient characteristics, symptoms, examination results, clinical interventions, and clinical outcome. The model development required over 30 h of expert graph modeling and included a dataset of 1389 patients for learning the probabilistic model (splitted into 70% training data and 30% validation data). The general model validation (representing the last clinical phase) achieved a prediction of *30-days mortality* and *One-year mortality* with an AUC of 0.92 and 0.87 respectively. Qualitative validation with a clinician indicates the value of model predictions in different clinical phases, as well as demonstrates problems and uncertainties in using the decision model.

Conclusion

While current state-of-the-art predictor approaches are based on multivariate logistic regression models with AUCs below 0.80, our model significantly exceeds those results (cf. [2]). However, a detailed examination of our model shows that some large subnetworks have equal probability distributions (i.e., parts remained unlearned). This indicated too little diversity in the data, which could be also confirmed by low precision in our practical tests of decision support.

Our network provides an important basis for the development and discussion of clinical guidelines (e.g. applicability of emergency calls) and (in-)transparent decision support. In the next steps, missing probability distributions could be completed by experts and later replaced by new data. Further, a model refinement and validation with more experts is planned to investigate the general applicability of the model.

References

[1] Cypko MA, Stoehr M, Kozniewski M, Druzdzel MJ, Dietz A, Berliner L, Lemke HU (2017) Validation workflow for a clinical bayesian network model in multidisciplinary decision making in head and neck oncology treatment. International Journal of Computer Assisted Radiology and Surgery 12(11), 1959–1970

[2] Ghoreishi M, Wise ES, Croal-Abrahams L, Tran D, Pasrija C, Drucker CB, Griffith BP, Gammie JS, Crawford RS, Taylor BS

(2018) A novel risk score predicts operative mortality after acute type a aortic dissection repair. The Annals of Thoracic Surgery 106(6), 1759–1766

Digital IR Suite of the Future—Stage 1—design and requirements using quality function deployment (QFD)

L. Berliner[1], V. Gallo[1], J. Scheiner[1], S. Hanna[1], C. Amato[2], B. DeForest[1], P. Ciaravino[1], M. Raden[1]

[1]Staten Island University Hospital—Northwell Health, Radiology, Staten Island, United States [2]Cannon Design, Architecture, Los Angeles, United States

Keywords Interventional Radiology Suite, Quality Function Deployment, Digital Operating Room, Project Management.

Purpose

There have been tremendous advancements in the field of Interventional Radiology (IR) in the last 40–50 years. A project is underway to design a comprehensive and technologically advanced Digital IR suite to meet the needs of IR in the mid-twenty-first century.

This report, regarding the first stage of this project (Stage 1) has been undertaken to determine the requirements, priorities, and innovative design considerations. This will be followed by a second report regarding Stage 2—ensuring project success in the implementation phase. Ultimately, the construction of the IR facility (Stage 3) will be presented.

This report is largely concerned with accepted engineering and project management practices that will be utilized to specify the needs and resources required for this project. This includes Quality Function Deployment (QFD) and Success Resource Deployment (SRD) (variations of which have been incorporated into a variety of project management systems, including Six Sigma quality programs). QFD is a structured and quantitative method to identify customer and/or stakeholder needs accurately and fairly. Desired features proposed by the QFD process are analyzed in terms of importance and feasibility, so that the priorities and resources of each stakeholder is considered and agreed upon in the overall process. All team members present their needs on equal footing in a structure called a House of Quality (HOQ), or QFD Matrix (Fig. 1) which avoids "ad hoc" decision making. (SRD will be presented in the Stage 2 report.)

A comprehensive HOQ, and supporting narrative, summarizing the team's conclusions with respect to requirements and priorities of the IR suite will be presented.

Methods

QFD provides methods for the development of a "wish list" of required and desired features, and a ranking of their importance, as compiled by all stakeholders (interventionalists, nurses, architects, engineers, and administrators) involved in the project.

In Stage 1, a Primary Team creates a list of features of the IR suite that will be important, if not critical, including advances in imaging, navigation, treatment, and communications technology. Architectural design considerations in the early planning stage, by an architect familiar with hospital design and construction, ensures that new concepts in room design will be incorporated and that technical feasibility will be properly analyzed. This also ensures that technical and regulatory requirements are fulfilled.

Areas under investigation by team members include:

1. What are the requirements of IR now and in the foreseeable future?

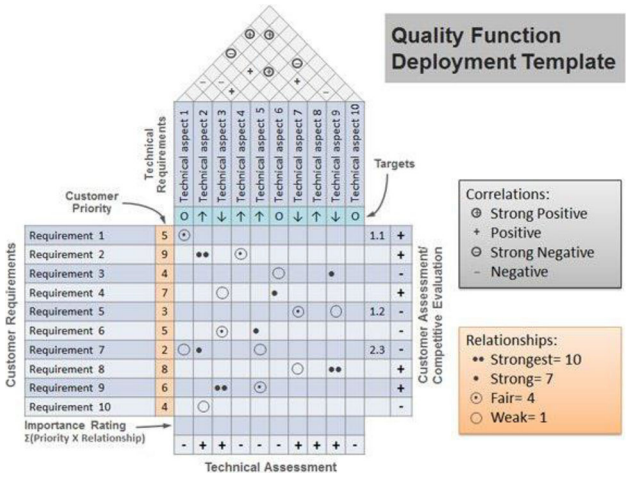

Fig. 1 QFD "House of Quality" (HOQ) that is used for the planning of a project. The HOQ template displays how project requirements and priorities are presented graphically. *Source*: https://cdn.free-power-point-templates.com/articles/wp-content/uploads/2013/05/quality-function-deployment.jpg

2. Which procedures and technologies are currently available or on the horizon for IR?
3. Which computer applications and Information Technology (IT) will assist in pre-procedure planning, patient assessment and patient modeling, intra-operative navigation, and patient monitoring?
4. What kind of advanced and upgradeable IT infrastructure will be required?
5. How will machine intelligence or Artificial Intelligence (AI) be used in IR?
6. What architectural innovations could facilitate optimal implementation?
7. What are the patient experience, care, safety, monitoring, and throughput issues that need to be considered?

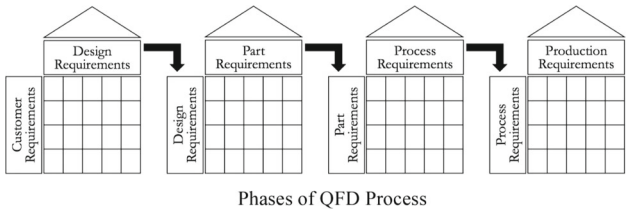

Fig. 2 A "waterfall" of QFD Houses of Quality allows a large multi-dimensional project to be broken down into more manageable components. Presented here is an example of a manufacturing process. An additional preliminary "Visioning" Phase, before "Design Requirements" would be beneficial. The outcome of the visioning sessions would define the ideal prototype (or project DNA) that defines the Basis of Design Requirements. Source: https://core.ac.uk/download/pdf/214330943.pdf

The submitted lists of requirements are then divided into related categories and each category is then consolidated into a list of related topics. Team members will then rate each item in terms of its importance. This information will be entered into a series (or "waterfall") of graphic spread sheet-type structures ("House of Quality" or QFD Matrix). The use of QFD waterfall (Fig. 2) allows a large multi-dimensional project to be broken down into more manageable components.

Results

The resulting waterfall of QFD Matrices will be presented and will include: current and potential future IR procedures; the imaging and therapeutic equipment that will be necessary; IT communications and image distribution requirements; a centralized computer lab for procedural planning including image and patient-model guidance or navigation; connections to the electronic health record; connections to other hospital departments, such as pathology; patient considerations such as waiting time for a procedure and pre- and post-procedure care; and, administrative concerns such as funding, staffing, and regulatory issues.

Some areas of interest include robotics, navigational systems including angiographic navigational systems, interventional oncology, advanced IT infrastructure, new visualization systems, machine intelligence in the OR (i.e., "intelligent" surgical instruments and devices), and the development of model-guided medicine and surgery (beyond image-guided medicine and surgery). Another innovation is to incorporate breast interventions into the more comprehensive realm of IR to allow optimum use of resources in Women's Health. Another consideration is whether the IR suite should be OR capable, as IR procedures are becoming more surgical in nature.

Some preliminary architectural considerations for the Digital IR Suite include (1) flexibility of the procedure room walls and structures to allow adaptable configuration of the work-space; (2) optimal physical arrangement of CT, angiographic, ultrasound, MR, ablation, and other equipment; (3) optimal design of patient prep and recovery areas; (4) innovations in design to allow efficient and maximized use of the facilities for In-patient and Out-patient procedures; (5) the impact of architectural design on IR performance and the patient and staff experience; and (6) additional requirements for a free-standing IR facility.

Conclusion

Methodology, related to Stage 1 of project development, which is based on sound engineering and business principles will be presented: (1) how to go about designing a Digital IR Suite of the future; (2) what a future IR suite should look and feel like; (3) which technical components will be required; (4) which procedures could be done now and in the future; (5) how to future proof the new design to enable continuous technology and operational changes. A summary of the Team's conclusions regarding this wide range of technical and architectural requirements and priorities of the IR suite of the future will be presented.

Stage 2, the early implementation phase, which is designed to plan for a successful project will be presented separately.

Stage 3, concerning the actual construction and utilization of the facility, will be the subject of future reports.

Digital IR Suite of the Future – Stage 2—ensuring success during implementation using success resource deployment (SRD)

L. Berliner[1], V. Gallo[1], J. Scheiner[1], S. Hanna[1], C. Amato[2], B. DeForest[1], P. Ciaravino[1], M. Raden[1]

[1]Staten Island University Hospital—Northwell Health, Radiology, Staten Island, United States [2]Cannon Design, Architecture, Los Angeles, United States

Keywords Interventional Radiology Suite, Success Resource Deployment, Digital Operating Room, Project Management.

Purpose

Tremendous advances have been made in the field of Interventional Radiology (IR) in the last half century. A project is underway to design a comprehensive and technologically advanced Digital IR suite incorporating these advances, to meet the needs of IR in the mid-twenty-first century.

The first stage in this project (Stage 1) concerning the design, requirements, and priorities for a Digital IR Suite is addressed in a separate report. This report presents methodology for the strategic guidance of Stage 2 of the project—the implementation phase. Techniques, based on sound manufacturing and engineering principles, need to be applied to ensure project success during this phase of the project.

One methodology that works well with the output of Quality Function Deployment (QFD), as described in a separate report on Stage 1, is Success Resource Deployment (SRD). SRD was designed to help develop a strategy, with associated tasks and assignments, to organize and monitor resources thereby ensuring success of a project. In addition to traditional SRD methods,

Hoshin Kanri (HK) or Strategy Deployment (as translated from the original Japanese), provides a set of project management tools that may be utilized as an SRD tool to facilitate efficient, effective, and successful project management.

This report will present the use of SRD in ensuring successful project management in designing and preparing for construction of a Digital IR suite. Some of the project features that will be implemented will be enumerated.

Methods

In Stage 1, a cohesive Primary Team was organized. As it becomes established that the project will be feasible and will be entering Stage 2, a broader and more compressive, system-wide Secondary Team is organized from a wider range of departments throughout the institution.

The Primary and Secondary Teams will need to become familiar with the working strategy and the necessary engineering tools of project development, including QFD and SRD. QFD provides a system in which the key stakeholders can specify the requirements that they desire and/or require and apply a rating to each requirement in terms of its importance. The resulting "House of Quality", or waterfall of grouped Houses of Quality, allows all stakeholders to have documented input into the project and to express their priorities. An SRD methodology, such as Hoshin Kanri (SD), by utilizing the findings of QFD, provides the organizational tools to ensure success of the project.

The primary structure of Hoshin Kanri is the X-Matrix or Strategic Deployment Matrix (Fig. 1), which is populated by information and priorities established by the QFD House of Quality in Stage 1. This X-Matrix is used to organize long-term and short-term goals (Steps 1 and 2); establish the resources and tasks that will be required (Step 3); provide a timeline and means of measuring progress and success (Step 4) and assign the individuals responsible for specific tasks (Step 5).

Following completion of the X-Matrix, an Action Plan (Fig. 2), based on the entries in the X-Matrix, is generated. The Action Plan allows accurate monitoring of the project and allows assessment of where additional resources may be needed.

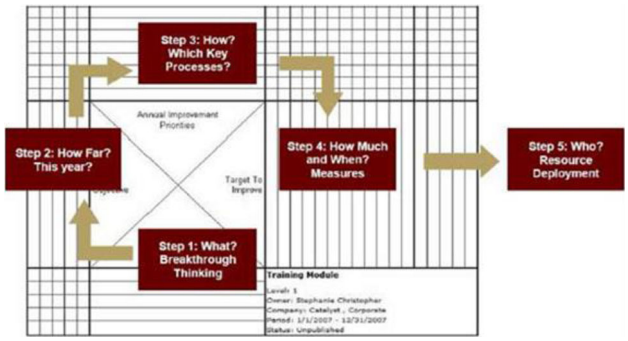

Fig. 1 The Hoshin Kanri X-Matrix establishes long term goals (Step 1), short term goals (Step 2), the tasks required to meet the goals (Step 3), when the goal is to be achieved and how it will be evaluated (Step 4), and who will be responsible for each goal (Step 5). Source: https://www.mcts.com/hoshin-kanri.htm

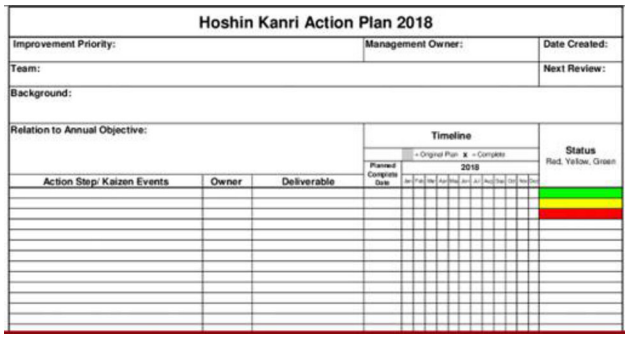

Fig. 2 The Hoshin Kanri Action Plan is based on the output of the X-Matrix. Source: https://flevy.com/browse/strategy-marketing-and-sales/templates-for-hoshin-kanri-strategy-deployment-1227&language=cn&language=en&language=cn&language=sp&language=cn

Results

Stage 2 will be undertaken after the completion of Stage 1, and results will be presented upon completion:

1. Review of the QFD House of Quality waterfall and having all stakeholders agree on the desired long-term and short-term goals (Steps 1 and 2).
2. Present the tasks and resources that will be required for the success of the project. These will include Clinical, Administrative, and Board of Trustees approval of equipment, design, and construction; a comprehensive Business Plan including profit/loss analysis and financial planning; Certificate of Need preparation and application (Step 3)
3. Presentation of a realistic timeline and means of assessing project progress (Step 4).
4. Presenting who will be responsible for each component of the project (Step 5).
5. Presentation of the completed Action Plan.

The tasks enumerated in the Action Plan will cover a vast variety of subjects: final selection of current and future IR procedures to be performed; configuring a centralized computer lab for procedural planning and navigation; selection of the specific imaging and therapeutic equipment that will be necessary; finalizing optimal architectural design; configuring the IT communications and image distribution requirements; establishing connections to the electronic health record, as well as other hospital departments, such as pathology; ensuring that pre- and post-procedure patient care issues are fulfilled; and, addressing administrative concerns such as funding, staffing, and regulatory issues.

Some areas of interest include robotics, navigational systems including angiographic navigational systems, breast interventions, interventional oncology, advanced IT infrastructure, new visualization systems, machine intelligence in the OR (i.e., "intelligent" surgical instruments and devices), and the development of model-guided medicine and surgery.

Conclusion

We have presented a methodology for the development of a Digital IR Suite of the future. Stage 1 (concerning the establishment of project vision, requirements, and priorities) is presented separately.

In this report, we describe the necessary engineering and business management tools that will help ensure the success of Stage 2, or the early implementation phase. The foreseeable requirements, in terms of procedures and technological support requirements have been indicated, and specific devices and equipment will be selected. The methodology for generating an Action Plan to ensure success has been presented. The results will be presented separately when completed.

Stage 3 (concerning the actual construction and utilization of the facility) will be the subject of future reports.

Multi-task mission planning for autonomous service robots within the OR wing

L. Bernhard[1], A. Knoll[2], D. Wilhelm[3]

[1]Klinikum rechts der Isar, Research Group MITI, München, Germany [2]Technical University of Munich, Chair of Robotics, Artificial Intelligence and Real-Time Systems, München, Germany [3]Klinikum rechts der Isar, Department of Surgery, München, Germany

Keywords mobile robotics, service robotics, OR wing workflow, OR wing logistics.

Purpose

Mobile service robots have emerged as a promising means for combatting the severe nursing shortage that paralyzes today's healthcare systems. While various use cases for other parts of the hospital have been studied in related work, the operating room wing (OR wing) is highly promising for mobile robotic support, also due to being particularly affected by the severe lack of personnel. In our contribution to CARS 2022, we have presented methods and studies regarding the

dimensioning and composition of mobile robotic fleets within the OR wing [1]. Now, we focus on the problem of dynamically assigning tasks to robots, which is a common problem in mobile robotic fleet management. The goal is to process incoming task requests from the users and integrate them into multi-task mission plans of the currently available members of the robotic fleet, while optimizing for driving durations and adhering to special constraints introduced by the OR wing use case.

Methods

The described problem is related to the well-known Vehicle Routing Problem (VRP), which is a combinatorial optimization problem. However, for the OR wing use case, several additions must be made to meet the requirements. On the one hand, several known variants of the VRP, such as the Capacitated VRP, the Multi-Depot VRP, the Open VRP and the VRP with Pickups and Deliveries must be combined. Secondly, further constraints must be added to address task prioritizations and alternative Pickup locations. The overall goal of the problem is to allocate incoming tasks to the mission plans of multiple service robots, while optimizing the robot driving durations and adhering to above constraints. In the following, the described problem is referred to as the Vehicle Routing Problem for the OR wing (VRP-OR). To the best knowledge of the author, the VRP-OR has not yet been described and studied in scientific literature, including other domains.

To solve the VRP-OR, a greedy algorithm is proposed. Greedy algorithms are based on the idea of repeatedly making the locally optimal decision, which leads to a locally optimal solution that may or may not coincide with the global optimum. In general, greedy algorithms are associated with a low computational complexity, which is well-suited for addressing the challenging time requirements of fleet management within the OR wing.

A simulation environment, which was described in previous work [1], was used to evaluate the developed mission planning algorithm. A fleet of 6 all-rounder robots moving at 1.2 m/s and serving 6 operating rooms was simulated for the purposes of this study. The intervention plan of each operating room was compiled from three consecutive surgical interventions. 8 different simulation scenarios were defined by varying the mission size, i.e. the number of task requests constituting a mission, from 1 to 8. Therein, a mission size of 1 corresponds to a trivial single-task scheduling approach with instantaneous assignment, which was included as a baseline reference for comparison. The following types of tasks were considered: transportation of sterilely packaged materials, transportation of heavy loads (e.g. containers, devices), intra- or postoperative disposal of waste, delivery of lab samples to tube mail stations, adjustment of medical devices, assistance during the sterile clothing procedure.

Results

The obtained task driving durations for the studied mission lengths are shown in Fig. 1. For the reference scenario with mission length 1, a mean of 29.29 s and a median of 25.82 s were obtained. From there, values decrease in an inverse exponential-like fashion. For a mission length of 4, a mean task driving time of 15.02 s and a median of 8.08 s were achieved. This corresponds to a decrease by 48.7% and 68.71%, respectively, relative to the reference scenario. Accordingly, for a mission length of 8, a mean value of 11.41 s and a median value of 3.48 s were achieved, which corresponds to a decrease by 61.04% and 86.54%, respectively. A maximum computation time of only 38.9 ms was observed for performing a full scheduling cycle.

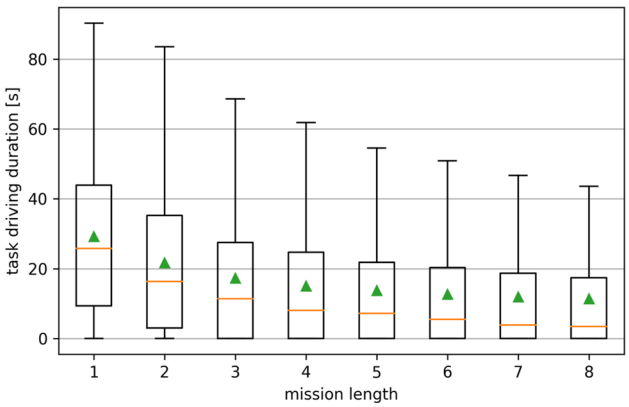

Fig. 1 The obtained robot driving durations for the simulated scenarios with different mission lengths are shown

Conclusion

A greedy-based solution algorithm for the hitherto unstudied VRP-OR was proposed. Compared to a simple single-task approach, task driving durations can be considerably shortened by means of multi-task mission planning. This is due to the elimination of redundant robot journeys and the merging of task requests involving locations situated in close proximity to each other. This effectively reduces the waiting times of the users (i.e., members of the surgical teams), improves the total duration of surgical interventions and benefits the overall acceptance of mobile service robotics for this application scenario.

At the same time, very short computation times are achieved, that are not possible with optimal solution strategies, due to the NP-hardness of the underlying optimization problem. This allows for very short reaction times and fast rescheduling, which is required to deal with the highly dynamic workflows associated with OR wings.

References

[1] Bernhard L, Amalanesan AF, Baumann O, Rothmeyer F, Hafner Y, Berlet M, Wilhelm D, Knoll A (2022) Mobile service robots for the operating room wing: balancing cost and performance by optimizing robotic fleet size and composition. Int J CARS. https://doi.org/10.1007/s11548-022-02735-8

The surgical report of the future—development of a crucial document

M. Berlet[1,2], T. Vogel[2], J. Fuchtmann[2], L. Wagner[2], L. Bernhard[2], M. Kähler[2], H. Feussner[1,2], M. Kranzfelder[2], H. Friess[1], D. Wilhelm[1,2]

[1]Klinikum rechts der Isar, Technical University of Munich, Department of Surgery, München, Germany [2]Klinikum rechts der

Isar, Technical University of Munich, Research Group Minimally-Invasive Therapy and Intervention (MITI), München, Germany

Keywords surgical reporting, natural language processing, cholecystectomy, automated documentation.

Purpose

Thorough and comprehensive documentation is essential for any surgical procedure. Narrative full-text description of the chronological sequence of surgical steps, difficulties encountered, and adverse events is the current gold standard. Even though this form of documentation has proven its worth, there are numerous shortcomings. From a surgeon's perspective, report writing is time-consuming and must be done immediately after the procedure in order to remember all details, which can interfere with the surgical schedule of a given facility. On the reader's part, the entire text must be worked through to obtain information of particular interest, which in turn can be tedious. Likewise, it is difficult to extract structured information such as data for reimbursement, quality management, and science from the full-text. Therefore, a lot of additional work such as assigning diagnosis and procedure codes must be performed by the surgeon himself or by documentation assistants, requiring additional effort. Thus, most current projects dealing with "intelligent surgical assistance" aim to sample structured data from sensors, video recordings, and vital parameters in combination with machine learning to ultimately use phase and event recognition for controlling and training assistive systems. [1] Derived from resulting information, automated surgical reports could be generated as well. [2] Since structured data is not sufficiently human-readable in most cases, viable output is crucial to fulfill the primary function of surgical reporting, which is, after all, to transmit information between medical professionals.

In the following, we propose a strategy to create individual templates that could be modified by assistant documentation programs depending on the individual course of a surgery in the future, using Natural Language Processing (NLP). - To improve the time aspect, report templates in the form of text modules are in daily clinical use for the most common surgeries since reporting is conducted digitally. Advantageously, the surgeon only has to change parts of these text snippets if the respective course deviates from the standard procedure. In addition, proven templates can be shared with other surgeons in the same department, respecting local and individual techniques and pReferences. Text templates from various medical institutions and different surgeons reporting the same type of procedure may differ significantly in the sequence of steps, specific technique, or equipment used. These inter-individual and inter-facility differences are major challenges in developing applications for assistive and automated surgical documentation. Our method aims to create basic templates, which however are highly individualized, and can be automatically adapted via future OR workflow analysis systems based on structured sensor, video, and audio information recorded during surgery.

Methods

To investigate the current situation of surgical reporting, we chose laparoscopic cholecystectomy (LCHE) as a model surgery and analyzed reports written on this procedure using natural language processing and calculation of Levenshtein distances using the *Natural Language Tool Kit (NTLK)* for the programming language "Python" and the *Stringdist* package for the statistical language "R". A total of 70 full-text reports were evaluated in this way. Similarities and differences among surgeons in terms of report writing were analyzed by automated cluster analysis. Using NLP, a frequency ranking of word stems was created and divided into the categories "actions", "anatomical structures", "instruments" and "description/properties" (Fig. 1a). Based on the 60 most frequent words, we manually created a customizable basic template (c), which is currently being tested for acceptance in an online study for the surgeons in our department at Klinikum rechts der Isar (**d**).

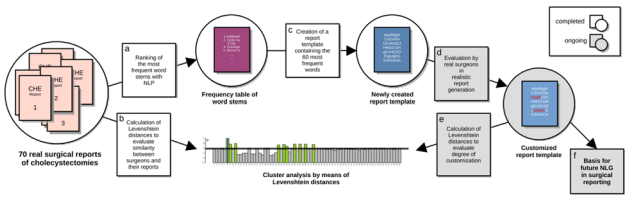

Fig. 1 Proposed workflow to create a basic report template for future automated surgical reporting; each step could easily be integrated in daily clinical routine. Gray-colored elements of the study are still ongoing. *NLG* Natural Language Generation

Results

Analysis of Levenshtein distances between the surgical reports combined with cluster analysis revealed a degree of individuality regarding wording, style and level of detail that allowed an individual author to be identified quite reliably, indicating a strong influence of personalized templates to be in use already (**b**). In addition, the amount of structured and repetitive text varied substantially between the various authors. Manual creation of a basic template aggregating all 70 reports and containing the 60 most frequent words (frequency 42.86%-100% of all reports) was feasible by this method. It is noteworthy that already the 8 most frequently used word stems were sufficient to identify the type of underlying surgery. ("Gallbladder": 100.00%, "Cystic duct": 98.57%, "Clip": 98.57%, "Drainage": 97.14%, "Wound dressing": 94.29%, "Cystic artery": 92.86%, "Irrigation": 92.86%, "Trocar": 92.86%) Local specialties were satisfactorily represented by the word ranking list, so that a template could be created that corresponds to the standard procedures and equipment of our surgical department.

Conclusion

The proposed strategy of creating customizable templates for future automated surgical documentation, considers local specialties satisfactorily. The standard template created is currently being tested for its acceptance by surgeons in our department in an online study, where those find a medical case for which they are required to write a report based on the proposed text module. The resulting reports will reveal where a future application would need to insert or delete text to properly customize the template. This study is still a proof of concept. However, other surgeries, even less standardized ones, could be involved the same way and text templates could be proposed to surgeons in clinical practice to create input and targets for future machine learning applications. Thereby, the proposed study design allows data collection during daily clinical routine. Surgeons simply

need to use the proposed templates and can modify them as required.The next step is then to use structured data derived from sensors in the operation room to generate input for the surgical report and to automatically customize the text module in this sense. Instead of extracting structured data from written text, as we do today, conversely in the future, structured data will be used to generate surgical documentation, which in turn must be readable by humans. This reversal of information flow solves most of the problems of current documentation by combining comprehensive surgical reporting with the provision of structured data.

References

[1] Kranzfelder M, Schneider A, Fiolka A, Koller S, Reiser S, Vogel T, Wilhelm D, Feussner H. Reliability of sensor-based real-time workflow recognition in laparoscopic cholecystectomy. Int J Comput Assist Radiol Surg. 2014 Nov;9(6):941–8. https://doi.org/10.1007/s11548-014-0986-z

[2] Berlet M, Vogel T, Ostler D, Czempiel T, Kähler M, Brunner S, Feussner H, Wilhelm D, Kranzfelder M. Surgical reporting for laparoscopic cholecystectomy based on phase annotation by a convolutional neural network (CNN) and the phenomenon of phase flickering: a proof of concept. Int J Comput Assist Radiol Surg. 2022 Nov;17(11):1991–1999. https://doi.org/10.1007/s11548-022-02680-6

Beyond telerobotics: promising applications of autonomous robots within the OR wing

L. Bernhard[1], C. Yang[2], O. Ratib[3], D. Wilhelm[4], C. Amato[2]

[1]Klinikum rechts der Isar, Research Group MITI, München, Germany [2]CannonDesign, Los Angeles, United States [3]University Hospital of Geneva, Department of Radiology and Medical Informatics, Geneva, Switzerland [4]Klinikum rechts der Isar, Department of Surgery, München, Germany

Keywords Autonomous Robotics, Surgical Robotics, OR of the Future, Hospital Design.

Purpose

We assume that the acceptance and use of robotic systems, AI and automation in healthcare will depend on their performance, but also on the availability of trained personnel. While the quality of mechatronic systems is continuously improving and approaches for autonomous concepts are being developed and presented, the availability and training of skilled personnel is at a critical juncture considering today's significant nursing shortage. According to the World Health Organization, the world will be short of 12.9 million healthcare workers by 2035. Today, there is a shortage of 7.2 million physicians, nurses, and allied health workers worldwide. Due to the above we must therefore assume that robotic systems will have to replace missing personnel in the near future, in order to sustain care providing and ensure treatment quality. The use of (partly) autonomous robots in healthcare environments potentially enables staff to do more work, easier, faster and more accurately than before within similar sized workspaces. In the context of remote-controlled telerobotics, studies have already shown the benefits of this technology [1]. Thus, similar benefits are also expected for (semi-)autonomous assistive robots in a broader context.

However, it is still unclear, where and how autonomous robotic systems should reasonably be used, and when human competence should be preferred. With this extended abstract, we present a new concept for rating the potential of robotic concepts for the operating room (OR) wing. Based on that, the most relevant and promising concepts are identified and the spatial/procedural requirements for integrating such systems into the OR wing environment are shortly discussed.

Methods

The robotic concepts presented and rated in the following result from internal, multidisciplinary discussions and experiences from our own projects on robotic systems and on concepts for the healthcare of tomorrow (digitalization, hospital of the future (CARS 2021), operating theatre of the future (CARS 2022)). Furthermore, a literature research related to the research question was carried out.

Our study explicitly focuses on (partially) autonomous assistance robots, as only these are expected to achieve a significant personnel relief, which we assume to add the most significant value. Beside this, however, we will consider other criteria, as elaborated in the following. As they have already become a standard component of today's operating theatres, teleoperated systems are explicitly not the main focus of our work.

Results

One might suggest that due to the increasing shortage of healthcare workers autonomous robotic systems will have a high potential anyway. From a scientific perspective this might indeed be true, as the race has been started for the development of human-equivalent systems. In the healthcare environment, however, we have to consider different aspects, from which we assume the following criteria to become the most critical concerning how and where to involve robotic systems in the future:

Staff Demand: We derive the need for a robotic solution from the availability of trained personnel, but also from how we expect the healthcare system to evolve over the next decade. Due to the ability of performing manual tasks, robots will preferentially be used in areas where these represent a core competence, however, only with a moderate level of complexity and risk.

Aptitude and Qualification: To apply robots in areas where humans show a clear superiority to what tasks demand, might not be reasonable. We rate human beings in general as having a high ability for improvisation and learning, good communications skills and high flexibility. On the other hand humans often tend to be fatigued by repetitive tasks and to show a low resilience to stress.

Ethical Framework: The healthcare system might be seen as one of the most critical fields of our society regarding ethical aspects. We believe personal care delivery and welfare to remain a core aspect of healthcare, especially in the context of surgeries and interventional therapies, whereas other fields, such as training support and transport might be less critical. We are committed to keeping healthcare running so that in areas with staff shortages or where personnel costs become unacceptable, robots could become a mandatory tool [2].

Spatial Considerations: Most of robotic systems differ significantly from humans, be it for their weight, for their dimensions or for requiring a safety environment. Accommodating robotic systems in healthcare environments require comprehensive planning to ensure its various applications (surgical robots, care robots, exoskeletons, AGV's, etc.) are integrated in an efficient, non-obstructive and coordinated way. This includes maintenance (charging stations, repair, etc.) a central operations/intelligence center, and other considerations. While some devices will be easy to implement, others require various degrees of architectural, infrastructure and IT considerations. For example, an Intuitive DaVinci surgical robot can be rolled into most operating rooms and will only require power and data to function. On the other hand, an automatic guided vehicle system (AGV) will require dedicated vertical elevators, dedicated collision-free corridors, strengthening structural floor slabs to carry the supply system heavy-loads, floor tracks and so on.

Table 1 Ranking of robotic concepts for the OR wing

	Degree of autonomy	Degree of risk	Staff demand	Aptitude and qualification	Ethical framework	Spatial considerations	Sum of ranks
Mobile systems							
Robotic circulating nurse	A4	R1	7	6	8	6	27
Self-driving patient bed	A4	R2	6	8	6	7	27
Off-stage AGV	A4–A5	R1	4	7	10	9	30
Cleaning robot	A5	R1	5	8	10	10	33
Care robot	A3	R2	6	2	2	3	13
Static systems							
Robotic scrub nurse	A4	R1–R2	6	7	8	6	27
Robotic surgeon	A3	R3	1	2	2	6	11
Camera guidance robot	A4	R3	3	4	6	8	21
Robotic assistance	A3–A4	R1–R2	5	6	7	5	23

Fig. 1 Future OR design envisioned by the authors

Taking into account the criteria just mentioned, the potential of selected robotic use cases can be ranked according to Table 1. For each concept, the above criteria are rated from 1 to 10. The sum of these ratings yields a final score, as a measure for the relative potential. Additionally, ratings regarding autonomy (A0-A5, with A0 being no automation and A5 being full automation) and risk (R1-R3, with R1 indicating low risk and R3 indicating high risk) are given. The autonomy rating was based on similar classifications in the context of self-driving cars. The risk rating was inspired by the risk classes defined by the MDR.

Conclusion

Based on the ranking presented in the previous, it can be concluded that many robotic concepts for the OR wing show great potential. This is especially true for mobile robots that execute logistical and preparation tasks, such as robotic circulating nurses, self-driving patient beds, off-stage AGVs and cleaning robots, which are also mostly associated with a low risk rating. However, the integration of these systems will certainly require modifications of the target environment, since the mobile robots must be able to execute their tasks safely, robustly and efficiently. At the same time, it must be ensured that the environment is still suitable for human workers: Even if some visions propagate robotic systems as a substitute for all human activity, we see the collaboration between humans and robots as more likely for healthcare. Figure 1 shows a first conceptual visualization of a future OR as a workspace for human–robot collaboration. In our presentation, we will elaborate on such spatial considerations and discuss the potential of different robotic concepts of the OR wing in more detail.

References

[1] Green CA, Lin JA, Huang E, O'Sullivan P, Higgins RM (2022) Enhancing robotic efficiency through the eyes of robotic surgeons: sub-analysis of the expertise in perception during robotic surgery (ExPeRtS) study. Surgical endoscopy. https://doi.org/10.1007/s00464-022-09315-6

[2] Floridi L, Cowls J (2022) A Unified Framework of Five Principles for AI in Society. In: Carta S (ed) Machine Learning and the City, vol 27. Wiley, pp 535–545

27th Annual Conference of the International Society for Computer Aided Surgery

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

Towards Immersive Collaboration of Remote Surgical Training

Z. Shao[1], Y. Tong[2], W. Si[1]

[1]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Institute of Advanced Integration Technology, Shenzhen, China [2]The Chinese University of Hong Kong, Department of Computer Science and Engineering, Hong Kong, China

Keywords Surgical training, Virtual reality, Remote interaction, Immersive collaboration.

Purpose

Virtual reality based surgical training has been widely used to improve novice surgeons' skill due to its low-cost and reproducible features, but social distancing restrictions due to COVID-19 challenged our ability to educate incoming surgery interns who depend on early simulation training for basic skill acquisition [1]. The appearance of remote training can well tackle this issue, but it still suffers from lack of the fidelity of human–computer interface and interactive framework for remote collaboration. Therefore, we propose a collaborative framework for remote surgical training and its tailored high-fidelity interface, which supports virtual surgical scenes construction and real-time remote collaboration for immersive demonstration and collaborative training.

Methods

This paper presents a collaborative framework with high-fidelity interface for remote surgical training, Fig. 1. Compared with the telementoring framework in [2], we consider about the trade-off between the real-time performance and the fidelity of human–computer interface and take the following approaches as improvement: (1) To solve the problem of relatively low fidelity in [2], we use real surgical instruments with high degree of freedom of movement, and then perform real-time motion capture through a series of highly sensitive sensors to achieve unrestricted and accurate surgical operation. (2) To avoid problems may be caused by redundant computational cost of server, we propose a framework of distributed interaction, display devices and remote simulation servers to distribute the operations to each client. (3) To further enhance immersion in remote interaction, we display our scenes through virtual reality with full information of each participant, enabling immersive demonstration and collaborative training.

We perform real-time motion capture through a series of sensors to obtain the kinematic data of the surgical instruments in each scene and then mapped the surgical instrument motion information to the virtual space. To reduce the computational burden on the server, we apply the simulation on local computer. Afterwards, we propose a new synchronization method and co-localization method to achieve multi-person scene synchronization for surgical training. To keep all participants' scenes synchronized, real-time operation and location information are shared through remote data transfer. We use Transmission Control Protocol (TCP) to share the coordinates and motion information of objects in all virtual scenes to solve the synchronization problem of remote interaction. We then maintain real-time data transmission during the training to synchronize the movement of surgical instruments in all scenes, ensuring the consistency among scenes for all clients. Besides, we used immersive head-mounted display 'HTC-Vive' for virtual reality display. We set up a virtual-real alignment step in advance and change the camera position in the virtual scene changes as the trainee moves the device during training to accurately overlay the virtual scene and information into the real scene. Moreover, the surgical instruments in the virtual space can be manipulated and interacted with virtual objects, allowing participants to receive the accurate visual feedback while operating. Instructors can consequently show their operations in the trainees' scenes, enabling trainees to learn the procedures and details of the surgery in short time through immersive demonstration.

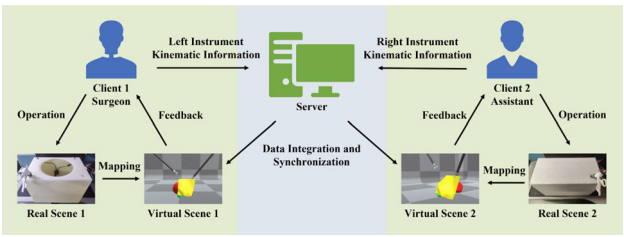

Fig. 1 The framework of our system. The kinematic information in each scene is mapped to the virtual one and uploaded to server. Server receives the data from all clients, integrates data and then sends it to all clients synchronously. Afterwards, the simulation of the whole scene is done locally and displayed in the virtual reality

Results

To verify the effectiveness and applicability of our system, we carried out the following experiments: (1) We measured the measurement accuracy of the surgical instruments, reaching an average error of surgical instruments for both hands of 1.99 mm. (2) We tested the real-time synchronization of scenes under remote connection to ensure real-time performance when remotely connected with the average delay of 86 ms. (3) We conducted remote collaboration tests of peg-transfer, a basic laparoscopic task, to evaluate the practical applicability of the system. The results show that our framework has good performance of accuracy and data transmission. Moreover, the remote collaboration tests prove that our system is practical in general situations.

Conclusion

Our system has been proved its advantages of low cost, no social distancing restrictions, and acceptable real-time performance, meanwhile it can provide the high-level features, remote demonstration as well as collaborative training for multi-person surgical scene, compared with traditional virtual surgical training systems. It is fully functional at its base and has good scalability to integrate more features in the future.

Acknowledgement

This work was supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: T45-401/22-N), in part by Shenzhen Fundmental Research Program under Grant (JCYJ20200109110208764, JCYJ20200109110420626), in part by National Natural Science Foundation of China (U22A2034) and in part by Guangdong Basic and Applied Basic Research. Foundation (2023A1515030268, 2021A1515012604).

References

[1] Nagaraj, M. B., AbdelFattah, K. R., Scott, D. J., Farr, D. E. (2022). Creating a Proficiency-based Remote Laparoscopic Skills Curriculum for the COVID-19 Era. Journal of Surgical Education, 79(1), 229–236.

[2] Shabir D, Abdurahiman N, Padhan J, Trinh M, Balakrishnan S, Kurer M, Ali O, Al-Ansari A, Yaacoub E, Deng Z, Erbad A,Mohammed A, Navkar NV (2021) Towards development of a tele-mentoring framework for minimally invasive surgeries. The International Journal of Medical Robotics and Computer Assisted Surgery, 17(5), e2305.

3D Holographic Volume Rendering with Mid-Air Gestural Interaction for Intuitive Diagnosis

Y. Qiu[1], W. Si[1]

[1]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Institute of Advanced Integration Technology, Shenzhen, China

Keywords intuitive and interactive mid, gesture interaction, free-hand visualization, medical data for diagnosis.

Purpose

With the great popularity of XR-based medical scenarios, it becomes more and more important for surgeons to interact naturally and intuitively with patients' diagnostic data. The rendering quality of 3D virtual models and the smoothness of human–computer interaction have become particularly important. However, existing methods require manual manipulation to align surgical instruments and holograms, which is cumbersome and increases system latency. Therefore, we propose a 3D holographic volume rendering system along with the head monitors (i.e., HoloLens2) that assists intuitive diagnosis with gesture-based interaction, enabling the surgeons to intuitively operate the 3D volume data in the mid-air. Users are able to use gestural interaction to operate virtual surgical instruments, observe patients' lesions from any perspective and any cross-section perform simulated surgical operations naturally in real time.

Methods

In this work, we visualize the patients' 3D raw medical data (such as CT & Ultrasound) with Microsoft HoloLens2, while we provide the surgeons the mid-air interface for free-hand gestural interaction [1,2]. In more detail, the density of each voxel was obtained using the ray matching. For each point in the data, we divide the ray from the vertex position into steps in the direction toward the eyes. The density of the point is obtained from the texture coordinates of each step to determine the color of the 3D voxel. Once the shader has the density of the voxel, it uses the density to calculate the RGBA value in the transfer function. By adjusting the value of the transfer function, the material and transparency of each tissue are realistically restored. The user is able to visualize the internal structure of the target region, enabling an intuitive multi-sensory interactive surgical scenario. Microsoft Hololens2 acts as a camera to capture real-world pose changes. Holographic viewing of 3D medical volume data can be positioned and tracked precisely, thus enabling interaction between the entities and the 3D medical volume data. They are scalable and manually rotatable in the XR environment, using the feasibility of augmented reality to make gestural interaction reliable and flexible.

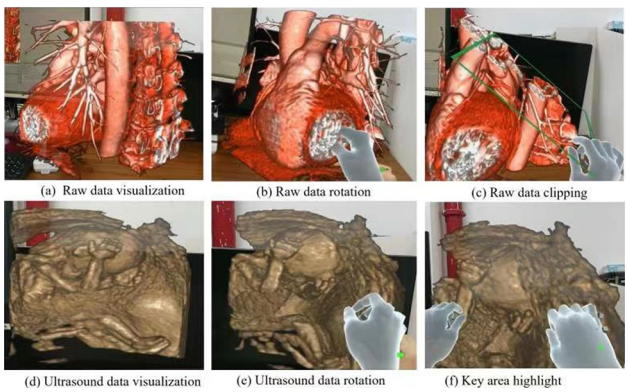

Fig. 1 Gesture interaction with 3D volume data

Results

In the system we have implemented, mid-air 3D volume data can be displayed holographically in real space, unlike screens. the surgeon can view the model from any angle and without the constraints of environment or geographic location. Gesture interaction allows the user to pick up and rotate the 3D model at will, zooming in on key areas so that they can be displayed more clearly. Clipping enables the surgeon to view various cross-sections of the target region and analyze the patients' lesion without visual dead space (see Fig. 1). The system allows the transmission and manipulation of multiple medical data, such as ultrasound and CT raw data, to meet the needs of different clinical diagnoses. Surgeons can use bare hands to interact with 3D medical data in real space.

Conclusion

In conclusion, we present the XR-based medical interaction system with a head-mounted display. The experiments demonstrate its potential, the high-quality 3D patients' data can be presented realistically and frees the surgeon's hands to manipulate the 3D objects in the XR environment. Although the 3D volume data is clear enough to show the details of the target region, the large patients' data really spares a lot of computing resources, which is limited for Microsoft HoloLens2. For future work, we will explore further algorithms and hardware to make 3D volume data with varying degrees of clarity for medical scenarios with different needs, with the aim of reducing the computational complexity load and improving the smoothness of gestural interaction in HoloLens2 to bring a more natural user experience for doctors and help them with more intuitive diagnosis.

Acknowledgement

This work was supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: T45-401/22-N), in part by Shenzhen Fundmental Research Program under Grant (JCYJ20200109110208764, JCYJ20200109110420626), in part by National Natural Science Foundation of China (U22A2034) and in part by Guangdong Basic and Applied Basic Research. Foundation (2023A1515030268, 2021A1515012604).

References

[1] Rezende, L. S., Sá, P. H., Macedo, M. C., Apolinário, A. L., Winkler, I., & SG, M. A. M. (2020, November). Volume Rendering: An Analysis based on the HoloLens2 Augmented Reality Device. In *2020 22nd Symposium on Virtual and Augmented Reality (SVR)* (pp. 35–38). IEEE.

[2] Jang, J., Frier, W., & Park, J. (2022, October). Multimodal Volume Data Exploration through Mid-Air Haptics. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 243–251). IEEE.

Robust and Efficient AR-based Navigation for Laparoscopic Liver Surgery

Y. Li[1], L. Yu[2], W. Si[3], P. A. Heng[1]

[1]The Chinese University of Hong Kong, Department of Computer Science and Engineering, Hong Kong, China [2]The University of Hong Kong, Department of Statistics & Actuarial Science, Hong Kong, China [3]Shenzhen Institute of Advanced Technology, Chinese

Academy of Sciences, Institute of Advanced Integration Technology, Shenzhen, China

Keywords Liver Registration, Surgical Navigation, Laparoscopic surgery, CNN.

Purpose

Augmented reality (AR) based surgical navigation has gained increasing attention in recent years due to their potential to reduce risks and improve outcomes. However, the current method of manual alignment of pre-operative models with intra-operative video in laparoscopic liver surgery is not robust, as it heavily relies on the operator's experience. The complexity of intra-operative dynamics also greatly influences navigation accuracy and can lead to errors in fast-paced and high-stakes surgical situations. Besides, current AR-based navigation framework for laparoscopic liver surgery often suffers from lack of efficiency due to their reliance on slow and large networks for tasks such as video segmentation and surface reconstruction, which can compromise the effectiveness of the surgery. In order to overcome these limitations and improve the robustness and efficiency of AR-based navigation for laparoscopic liver surgery, we introduce robust and lightweight deep learning models in our AR-based navigation approach, which can replace manual alignment and improve computational speed.

Methods

In our work, we use lightweight and advanced networks to improve the robustness and real-time performance of our AR-based navigation for laparoscopic liver surgery. The overall AR process consists of three steps: video segmentation, surface reconstruction, and registration. First, the laparoscopic video is segmented to separate the liver from other structures in the field of view. Next, the surface of the liver is reconstructed using the segmented video and the disparity estimation model. In these steps, we incorporate a lightweight semantic segmentation model, DeepLabV3 with MobileNetv2, to improve the efficiency of video segmentation and a lightweight disparity estimation model, STereo TRansformer (STTR)[1], to improve the quality of surface reconstruction. These models are chosen for their high robustness and computational efficiency, as well as their outstanding performance in other surgical vision tasks. Finally, the pre-operative model is registered with the reconstructed surface using the deep learning model. For the registration, we propose a novel registration network, Shuffle-Transformer-Net (STNet), to improve robustness and efficiency in registering the pre-operative model with the intra-operative liver surface. STNet takes the intra-operative surface and the pre-operative model as inputs, analyze the features of these inputs, and outputs a displacement field that registers the pre-operative model to the intra-operative surface. The STNet has an encoder-decoder structure, with the ShuffleNetv2 as the encoder and a Vision Transformer (ViT) in the bottleneck layer to extract both local and global features. The use of the ViT makes the STNet more robust and able to handle the complexity of surgical scenes, leading to less manual intervention and better automatic AR-baesd navigation.

Results

To evaluate the performance of our proposed STNet, we conduct experiments using a dataset comprising a simulated pre-operative liver model and 135 deformed partial intra-operative liver surfaces. Here 70% of the data is used for training, 20% of the data was used for validation, and 10% of the data was used for testing. The STNet achieves an average registration error of 13 mm the experiments. In comparison, the state-of-the-art method for registration in laparoscopic liver resection, V2S-Net[1], has an average error of 16 mm. These results demonstrate the improved accuracy of our proposed STNet for registering the pre-operative model with the intra-operative liver surface. Additionally, we find that the use of a lightweight semantic segmentation model, DeepLabV3 with MobileNetv2, and a lightweight disparity estimation model, STTR, improves the overall performance of the AR-based navigation. In terms of computational efficiency, the parameters of the whole AR-based navigation using the V2S-Net method [2] are 35.2 M, while our proposed method using STNet has a much smaller parameter size of 6.5 M. The frame rate of the V2S-Net method is 10 Hz, while our method achieves a faster frame rate of 13 Hz.

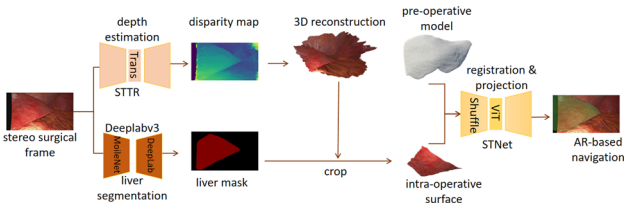

Fig. 1 Framework of our AR-based navigation. The stereo surgical video undergoes video segmentation to obtain liver mask and depth estimation to obtain a disparity map. The 3D reconstruction is obtained from the disparity map. Using the liver mask, we can crop the liver intra-operative surface from the 3D reconstruction. Finally, the intra-operative surface is registered with the pre-operative model and the model is fused into the image

Conclusion

In conclusion, we present an AR-based navigation for laparoscopic liver resection that uses a novel registration network, STNet, which achieves high robustness and efficiency. Our experiments show the proposed method can well integrate pre-operative information in intra-operative scene, which can provide surgeon with comprehensive and intuitive guidance. In the future, we plan to address the issue of tool occlusion in order to improve the quality of surface reconstruction in our AR-based navigation. We also aim to develop a tool position monitoring method that provides alerts when the instrument is close to a critical structure in order to prevent accidental injury to the patient. Additionally, we will explore the use of deep learning methods to address the issue of initial model and intra-operative reconstructed surface alignment in order to further improve the accuracy of our navigation (Fig. 1).

Acknowledgement

This work was supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: T45-401/22-N), in part byShenzhen Fundmental Research Program under Grant (JCYJ20200109110208764, JCYJ20200109110420626), in part by National Natural Science Foundation of China (U22A2034) and in part by Guangdong Basic and Applied Basic Research. Foundation (2023A1515030268, 2021A1515012604).

References

[1] Li Z, Liu X, Drenkow N, et al. Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 6197–6206.

[2] Pfeiffer M, Riediger C, Leger S, et al. Non-rigid volume to surface registration using a data-driven biomechanical model[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020: 724–734.

The influence of surgical navigation on the complexity of open liver surgery

K. Olthof[1], J. Smit[1], M. Fusaglia[1], N. Kok[1], K. Kuhlmann[1], T. Ruers[1,2]

[1]The Netherlands Cancer Institute—Antoni van Leeuwenhoek, Surgical oncology, Amsterdam, Netherlands [2]University of Twente, Nanobiophysics Group, Enschede, Netherlands

Keywords Surgery, Navigation, Liver, Vanishing lesions.

Purpose

Locating small and vanishing lesions during surgery using ultrasound (US) remains a challenge in liver surgery, particularly in patients with widespread bilobar metastases. When these lesions are smaller than the resolution of intraoperative ultrasound or appear isoechoic, the surgeon estimates their location based on diagnostic imaging and experience. The use of three-dimensional (3D) models segmented from preoperative imaging could provide a more intuitive understanding of the patient anatomy for the surgical team. In addition, surgical navigation—a technique that gives the surgeon a virtual representation of their instruments with relation to the patient-specific liver model, may help to guide the surgeon towards small targets more easily. The aim of this prospective study is to determine whether integration of anatomical models and surgical navigation to the clinical workflow decreases the complexity of the surgical procedure.

Methods

This study included candidates 18 years or older who were scheduled for open surgical treatment of colorectal liver metastases. A preoperative contrast-enhanced CT or MR scan was used to create a 3D model of the liver, which included the parenchyma, hepatic vasculature, biliary tree, lesions, and cysts. Surgical procedures were conducted according to standard protocol. Initially, surgeons were not provided with the 3D model unless requested. Similarly, surgical navigation using an in-house developed system was only initiated upon request from the surgeon. When navigation was requested, a landmark-based registration using an electromagnetically-tracked US transducer was performed as described in [1]. Landmarks were identified as vessel bifurcations or other clear anatomical structures (e.g., cysts or other tumors) in the US image. Navigation information was visualized as a virtual scene depicting the liver model and tracked surgical instruments, as well as the US video stream augmented with a projection of the registered 3D model. Postoperatively, the projection error was calculated for each navigated procedure as the Euclidean distance between clearly identifiable landmarks on US and the augmented delineations of the registered 3D model. The influence of the 3D model and navigation on the complexity of the procedure was evaluated using a questionnaire. Surgeons were asked to grade the complexity of the procedure (on a 10-point scale) with and without the aid of 3D models and with or without surgical navigation, and to specify how these techniques influenced the complexity of the surgery. Additionally, local recurrence rates were reported at a follow-up of 3 months.

Results

Fifteen patients were included in this study. Patient characteristics, information on the surgical procedure and 3-month follow-up are provided in Table 1. The surgeons requested the aid of the liver model

Table 1 Patient characteristics, information on the surgical procedure and 3-month follow-up

Patient	Age	Type of surgery	N treated lesions	Mean tumor size (mm)	Model requested	Navigation requested	TRE (mm)	3 m follow-up
1	61	Chip + burn*	4	9.5 ± 3.9	Yes	Yes	9.2	No recurrence
2	76	Chip + burn	4	12.5 ± 3.7	Yes	Yes	6.3	No recurrence
3	74	Hemi-hepatectomy	8	8.6 ± 5.1	Yes	No	NA	NA
4	62	First stage chip + burn**	7	7.9 ± 3.7	Yes	Yes	6.8	No recurrence liver, positive lymph node
5	54	Chip + burn	6	12.7 ± 4.1	Yes	Yes	7.2	5 new liver lesions
6	66	Chip + burn	9	10.1 ± 6.6	Yes	No	NA	No recurrence
7	64	Chip + burn	6	11.7 ± 4.2	Yes	No	NA	No recurrence
8	65	Resection	3	9.7 ± 2.1	Yes	No	NA	No recurrence
9	55	Chip + burn	3	8.0 ± 1.0	Yes	Yes	3.2	No recurrence
10	61	First stage chip + burn	5	8.6 ± 1.7	Yes	Yes	3.3	No recurrence
11	36	Chip + burn	12	9.2 ± 5.6	Yes	No	NA	No recurrence
12	55	First stage chip + burn	17	7.4 ± 3.3	Yes	No	NA	No recurrence
13	65	Chip + burn	18	7.3 ± 3.3	Yes	Yes	9.3	No recurrence
14	55	Chip + burn	15	5.4 ± 1.6	Yes	Yes	7.8	5 new liver lesions
15	53	Chip + burn	14	9.9 ± 11.0	Yes	Yes	1.9	No recurrence
Mean	60.3 ± 9.7		9.1 ± 5.3	9.2 ± 2.0			6.1 ± 2.7	

*Local resection of larger and superficial lesions combined with ablation of smaller deeper lesions

**Clearance of one hemiliver in a first non-curative intervention, where the remaining tumors are resected by a hemihepatectomy after a period of liver regeneration

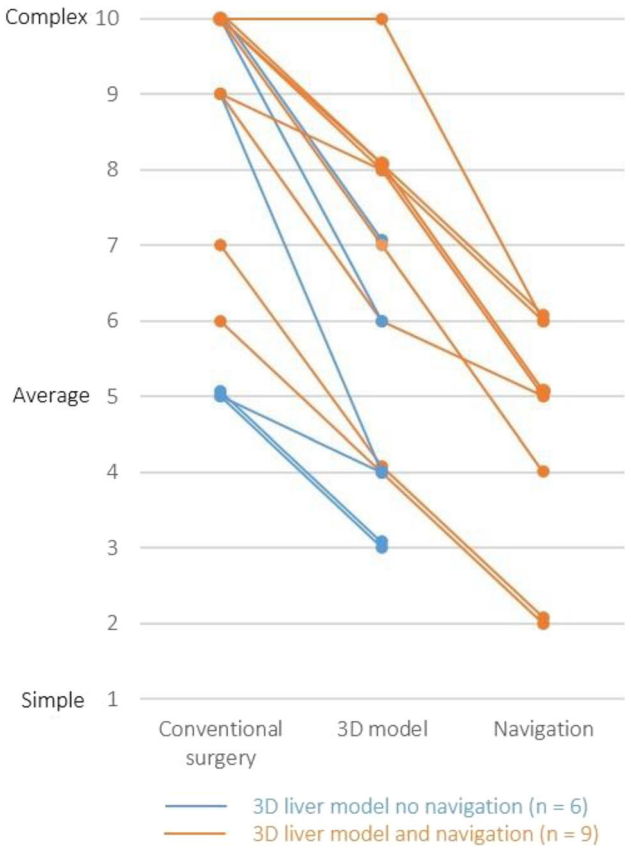

Fig. 1 Influence of the use 3D model and surgical navigation on the complexity of the procedure

in all procedures. Navigation was used in nine of these cases to locate lesions that were not found using US alone. The mean projection error for navigated lesions was 6.1 ± 2.7 mm. One patient died due to a complication related to insufficient liver remnant. No local recurrences were found during postoperative follow-up, but two patients developed new metastases in other liver segments. The results of the survey indicated that solely using 3D models reduced the complexity, with an average decrease of 2.3 (n = 15) on a 10-point scale compared to conventional surgery. Additionally, surgical navigation was found to be particularly useful in localizing small lesions, further simplifying the procedure with an average decrease in complexity of 2.4 (n = 9) compared to surgery with solely a 3D model (Fig. 1).

Conclusion

This study shows that the integration of patient-specific 3D models and surgical navigation decreases complexity of the surgical procedure in patients with bilobar spread liver metastases. The use of 3D models provides surgeons with a better understanding of patient anatomy by clarifying the spatial relationship between critical structures and small lesions. In addition, keeping track of treated lesions provides an overview of the procedure in patients with a high number of lesions. Surgical navigation aids localizing of difficult-to-detect lesions and confirming their locations. Surgeons reported that decisiveness had improved and that some lesions could not be located without navigation. Initial results are promising, a longer follow-up and larger patient group are required to endorse these results.

References

[1] Smit J.N., Kuhlmann K.F.D, Ivashchenko O.V., Thomson B.R., Langø T., Kok N.F.M., Fusaglia M. & Ruers T.J.M. (2022) Ultrasound-based navigation for open liver surgery using active liver tracking. International Journal of Computer Assisted Radiology and Surgery 17:1765–1773.

Reducing False Positives in Automatic Hepatic Tumor Segmentation: a Deep Learning Approach

T. Natali[1], A. Zhylka[1], K. Olthof[1], J. Smit[1], T. Baetens[2], T. Ruers[1,3], M. Fusaglia[1]

[1]The Netherlands Cancer Institute—Antoni van Leeuwenhoek, Surgical Oncology, Amsterdam, Netherlands [2]The Netherlands Cancer Institute—Antoni van Leeuwenhoek, Radiology, Amsterdam, Netherlands [3]University of Twente, Nanobiophysics, Enschede, Netherlands

Keywords Segmentation, Navigation, Deep-Learning, Surgery.

Purpose

Surgical resection is the treatment option with longest-term prognosis when it comes to treating liver cancer (either in the form of primary liver cancer or colorectal metastases). During surgery, ultrasound is the standard modality for identifying and characterizing these lesions. Although ultrasound provides a real-time high quality image of the scanned area, its assessment is challenging and highly dependent on the operator. Accurate segmentation of hepatic tumors during liver surgery can improve the safety and outcomes of the procedure, by reducing the number of positive resection margins and undetected lesions. Deep-learning based automatic segmentation methods can generate accurate labels in real-time, but are affected by a high false positives rate. In this study we propose a two networks approach that can generate accurate intraoperative tumor borders with low false positive rate, which in combination with surgical navigation will ease the detection and tumor margin delineation procedure during liver surgery.

Methods

We propose an approach which concatenates two deep learning networks: a binary classifier and a segmentation network since performing classification prior to segmentation was shown to improve precision [1]. Thus, binary tumor classification is performed on single 2D ultrasound images, followed by segmentation by the second network of the positively classified ones. ResNET-50 was trained for classification task, while for the second step a U-Net was trained with the nnU-Net framework [2], to accurately identify the boundaries of the hepatic tumors. The segmentation network inputs a single hepatic ultrasound image and outputs a binary map indicating the tumor location, Fig. 1.

Two datasets were used to train and test the networks. For the binary classifier, 1600 ultrasound images from 45 patients and 88 sweeps were acquired and classified. Of these, 50% of the images depicted tumors. Images were classified by a medical technician with more than 5 years of experience in the assessment of intraoperative hepatic ultrasound images. Tumors of varying echogenicity (excluding isoechoic lesions) and sizes (ranging from 5 to 58 mm) were included. The segmentation network was trained and tested on 1000 ultrasound images all containing tumors. Tumor regions were manually labelled and used as ground truth. Labels were delineated by a radiologist with more than 20 years of experience. The performances

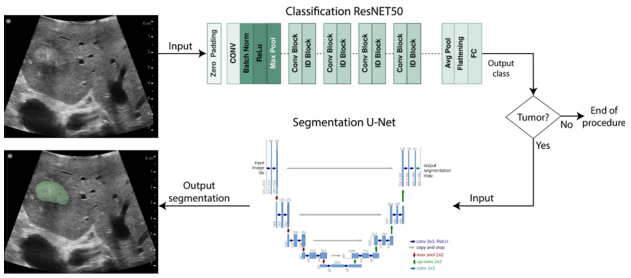

Fig. 1 Workflow schematic. The proposed method uses two CNNs. First, images are classified with a ResNET50 network to identify those containing tumors. Afterwards, the images labelled as tumors are input to the segmentation U-Net network which will output a binary mask with the tumor delineation

of the networks were computed on a test set consisting of 1148 images from 5 US sweeps from 5 patients.

For the classifier network, the dataset was divided into training and validation sets, with the proportion of 90%, 10%, respectively. The classifier was trained starting from a pre-trained ImageNET. The training was then finalized on our training set using a cross-entropy loss function and Adam optimization, with a learning rate of 0.001 and a batch size of 32. The model was trained for 100 epochs. For the segmentation network 950 images for training and 50 for validation were used. As loss function in the training, a combination of cross-entropy and Dice loss was used with a learning rate of 0.0001 and a batch size of 32. The model was trained for 1000 epochs. Best models were determined based on validation performances.

Evaluation of both networks was performed on the test dataset. Precision, recall, and F1 score computed per image and per acquisition were used to evaluate the classifier. In order to make the performance of the segmentation network comparable to that of the classifier, any non-empty segmentation in an US image was considered as positive classification result (e.g., if at least a pixel is predicted as tumor than the image is classified as "tumor"), Table 1. Based on that, precision of the tumor indication was compared with and without using a classifier.

Results

The results showed that the addition of the binary classifier was able to reduce the number of false indications of tumor presence compared to the segmentation network alone. The average precision on the test set from of the classification network resulted to be 0.85 which is a great improvement from the 0.21 achieved with the simplified output of the segmentation network. These results suggests that the predictions of the classifier are more robust to false positives, and that this method can be effectively used to improve automatic segmentation quality during intraoperative hepatic US sweeps. On the other hand, the classification network also introduced false negatives that were not present when using only the segmentation network: from an average recall of 0.90 it dropped to 0.71. The F1 score of the combined networks is 0.67 while when using only the segmentation network, the F1 score is 0.32.

Table 1 Result of segmentation and classification

	Precision	Recall	F1
Simplified segmentation	0.21	0.9	0.67
Classification	0.85	0.71	0.32

Conclusion

The combination of the segmentation and classification network increased in precision and F1 scores of the presented automatic hepatic tumor segmentation method. Validation of the method is required on a larger test set, including up to 50 US acquisitions. Its potential clinical application will be assessed in the surgical procedure along with the generalization to other US transducers and manufacturers. Ultimately we aim to provide the surgeons with an automated tool that can be used real-time on any ultrasound transducer for the automatic detection and segmentation of hepatic tumors during navigated surgeries.

References

[1] Zhang S, Liao M, Wang J, Zhu Y, Zhang Y, Zhang J, Zheng R, Lv L, Zhu D, Chen H, Wang W, Fully automatic tumor segmentation of breast ultrasound images with deep learning. Journal of Applied Clinical Medical Physics. n/a(n/a): p. e13863.

[2] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods, 2021. 18(2): p. 203–211

Comparison of Robotic and Open Pancreaticoduodenectomy with Portal-Superior Mesenteric Vein Resection for Advanced Peri-ampullary Tumors

Y. J. Chen[1], H. C. Lin[1]

[1]Taichung Veterans General hospital, Surgery, Taichung, Taiwan

Keywords pancreaticoduodenectomy, Robotic, Open, Vein resection.

Purpose

The aim of this study was to compare the surgical, oncological, and survival outcomes of pancreaticoduodenectomy (PD) with the portal vein (PV)/superior mesenteric vein (SMV) resection using either robotic or open surgery [1].

Methods

Data on patients with periampullary lesions who underwent PD was collected and analyzed. The surgical risks, oncological outcomes, and survival outcomes of patients who underwent PV/SMV resection were compared to those of patients who did not undergo PV/SMV resection.

Results

A total of 301 patients were enrolled in the study from July 2016 to September 2022. Out of these, 35 (12.0%) underwent vein resection and the remaining 266 (88.0%) did not have vein resection. After a vein was resected, it was either repaired directly or repaired using a GORE-TEX® vascular graft as a replacement. Of the patients in the group that underwent vein resection, 8 (22.8%) had the procedure done using a robotic approach while 27 (77.2%) had it done using the open method. The operation time was significantly longer in the robotic group compared to the open group, with a median of 760 min

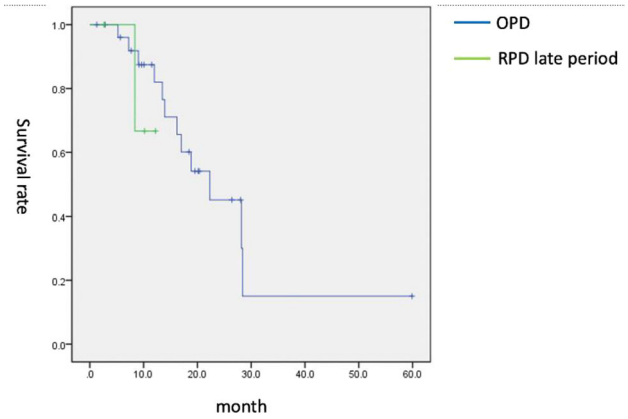

Fig. 1 There was no statistically significant difference in survival rate between the open group and the late robotic group (p = 0.446)

versus 444 min ($p < 0.001$). The median blood loss was higher in the robotic group (1600 c.c. vs. 1150 c.c.), but this difference was not statistically significant (p = 0.708). The study found that there were no significant differences in postoperative complications, such as pancreatic fistula (POPF), post pancreatectomy hemorrhage (PPH), delayed gastric emptying (DGE), chyle leakage, wound infection, re-operation, and length of stay (LOS), between the two groups in terms of surgical outcomes. Additionally, there was no significant difference in oncologic outcome with regard to tumor size, the rate of curative resection (R0), or the number of lymph nodes that were harvested. However, the operative mortality was higher in the robotic group, with 2 out of 8 patients, which led to a worse survival outcome in that group. The study found that when the robotic group was separated into two subgroups, the early and late periods, the operation time and blood loss were lower in the late period, although there was no statistically significant difference. Furthermore, there was a better survival outcome in the late period of the robotic group. There was no significant difference in survival between the open group and the late robotics Group (p = 0.446), Fig. 1.

Conclusion

Pancreaticoduodenectomy with vein resection is technically feasible using both open and robotic surgery in selected patients. However, additional PV/SMV resection increases the surgical risks of the procedure in the early period of robotic surgery. The study suggests that it may be possible to achieve better survival outcomes after the learning curve is passed.

References

[1] Fancellu A, Petrucciani N, Porcu A, Deiana G, Sanna V, Ninniri C, Perra T, Celorai V, Nigri G (2020) The impact on survival and morbidity of portal-mesenteric resection during pancreaticoduodenectomy for pancreatic head adenocarcinoma: A systemic review and meta-analysis of comparative studies. Cancer12(7): 1976

Instrument segmentation with TernausNet during temporal bone surgery

C. Yuan[1], D. Shin[2], T. N. Le[2], V. Y. Lin[2], J. M. Chen[2], L. A. Kahrs[3], J. T. Lui[4]

[1]University of Toronto, Engineering, Toronto, Canada [2]Sunnybrook Health Sciences Centre, Department of Otolaryngology–Head & Neck Surgery, Toronto, Canada [3]University of Toronto Mississauga, Department of Mathematical & Computational Sciences, Mississauga, Canada [4]Cumming School of Medicine, University of Calgary, Section of Otolaryngology–Head & Neck Surgery, Department of Surgery, Calgary, Canada

Keywords machine learning, neural network, cochlear implantation, surgical tool.

Purpose

Surgical procedures like mastoidectomy and cochlear implantation are on the verge of higher-level assistive technology in the operating room. Like bone removal during hip or knee arthroplasty, robot technology exists that performs the mastoidectomy in an image-guide manner. Previously, instrument tracking and control has been done by methods like electromagnetic or optical tracking as well as robot encoders. The main purpose of this contribution is to provide a machine learning based segmentations of instruments for future vision-based tracking during temporal bone surgery. Previous research used convolutional neural networks (CNNs) on microscopic images during mastoidectomies. Tool detection was performed using the You Only Look Once version 4 (YOLOv4) and You Only Look at CoefficienTs (YOLACT) CNNs [1]. Nevertheless, no instrument segmentation results were provided. Our research question is concerned with the finding of suitable methods and evaluation of their accuracy.

Methods

Video footage was captured during 15 cochlear implantations with a resolution of 1920 × 1088 pixels. On 744 images 4645 labels of seven classes of surgical instruments (burr drill, suction, Rosen needle, forceps, retractor, irrigation, and snap) were created by two annotators. An exemplary image with ground truth labels of three classes is shown in Fig. 1 (top). For binary segmentation, all tool masks were concatenated into one singular foreground and the resulting background. Another 114 images lacked surgical instruments and received no labels. Empty masks were generated for the images lacking labels to utilize negative examples for model training. Data was split into 80/10/10 for training, validation and testing sets. Augmentations (random square-shaped cropping, random rotation, horizontal flip, vertical flip, transpose, grid distortion, motion blur) were performed on the training set and results in 4298 samples. Binary and multiclass segmentation methods were implemented as TernausNet with VGG16 encoder [2]. The cost functions utilized for binary and multiclass segmentation were binary cross-entropy and categorical cross-entropy, respectively. The input size of the images was 512 × 512 pixels. To optimize the training procedure, the Adam optimizer was utilized, which minimizes the cost function by using squared gradients to scale the learning rate while updating the parameters based on the moving average of the gradient rather than just the gradient itself. The learning rate was initially set as $1e-4$ but was dynamically lowered using a learning rate scheduler with a patience of 5 epochs, a reduction by a factor of 10, and a minimum learning rate of $1e-5$. Training was limited to a total of 100 epochs, with only the best epoch being saved to the final model based on the validation loss. A batch size of 16 was utilized, with a total of 74 iterations per epoch, calculated by dividing the length of the original

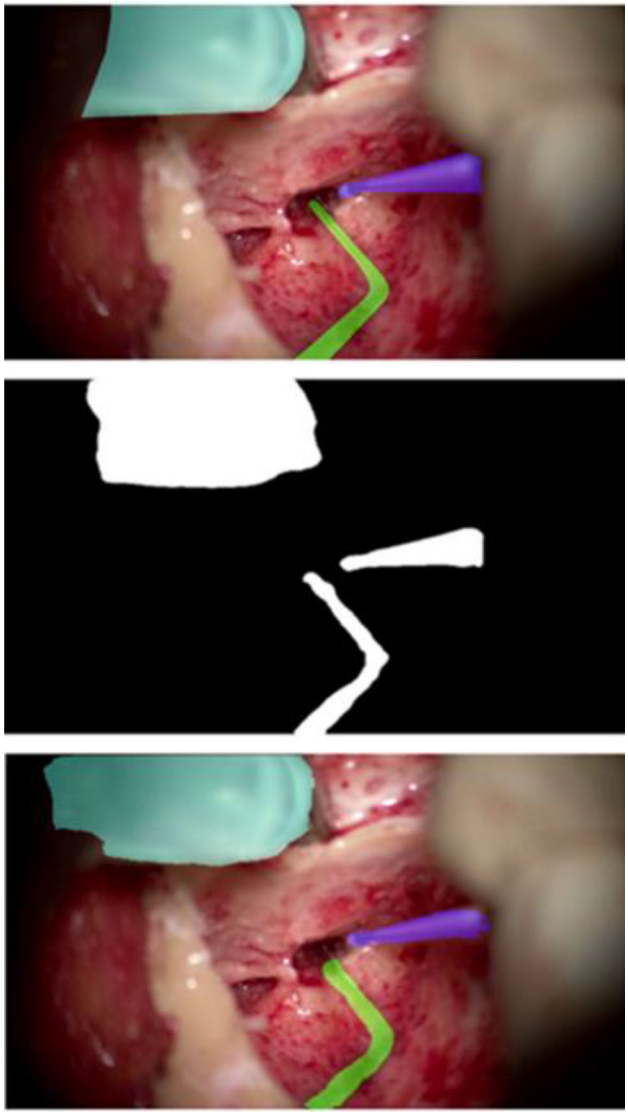

Fig. 1 Exemplary ground truth (top), binary segmentation (middle) and multiclass segmentation results (bottom) of the otologic burr (purple), irrigation (light blue), and suction (green)

training set by the batch size then multiplying by two. Total training time was approximately 3 h per model utilizing TensorFlow as a deep learning framework. All networks were trained and tested using a setup comprised of an AMD Ryzen 9 5900X CPU and a NVIDIA GeForce RTX 3090 GPU.

Results

During training the F1 (Dice) Score of TernausNet got 95% for binary and 62% for multiclass segmentation. Analysis of validation F1 Scores result in 87% for the binary and 61% for multiclass segmentation. Similar segmentation results are achieved for the Intersection over Union (IoU, Jaccard Index) metric during training (91% binary, 56% multiclass) and validation (77% binary, 55% multiclass). An example of the binary and the multiclass segmentation is shown in Fig. 1 (middle and bottom).

Conclusion

Machine learning based segmentation of surgical instruments were demonstrated in microscope footage of cochlear implantation surgery. Previously, only instrument detection (YOLOv5) with bounding boxes and not pixel-level segmentation was published [1]. Our work allows the extension of conventional tracking sensors with a robust, vision-based instrument tracking for surgical robots and other microscopic guidance systems. Availability of large open-access datasets will be crucial for future implementations.

References

[1] Choi J, Cho S, Chung JW, Kim N (2021) Video recognition of simple mastoidectomy using convolutional neural networks: Detection and segmentation of surgical tools and anatomical regions. Computer Methods and Programs in Biomedicine 208:106,251 (1–10).

[2] Iglovikov VI, Shvets AA (2021) TernausNet. Computer-Aided Analysis of Gastrointestinal Videos (Editors Bernal J, Histace A), Chapter 15:127–132.

Development of a total hip arthoplasty support system that includes dummy and virtual lower-limb models

A. Hanafusa[1], H. Kato[1], S. Hara[1], A. Kato[1], S. Mohamaddan[1], M. Takagi[1], T. Baba[2], Y. Homma[2], Y. Oishi[3]

[1]Shibaura Institute of Technology, Bio-science and Engineering, Saitama, Japan [2]Juntendo University School of Medicine, Department of Orthopedic Surgery, Tokyo, Japan [3]Surgical Alliance Inc., Tokyo, Japan

Keywords Total hip arthroplasty, Surgery support system, Dummy lower-limb model, Virtual lower-limb model.

Purpose

Total hip arthroplasty (THA) is a surgical procedure used to replace a hip joint damaged by symptoms such as osteoarthritis, osteonecrosis of the femoral head, rheumatoid arthritis, with an artificial hip joint. Several methods of THA are available, including the direct anterior approach (DAA), which prevents cutting of muscles or ligaments as much as possible. Although this method is minimally invasive[1], with quick postoperative recovery and low patient burden, the procedure is complicated and has not gained wide usage. A lower-limb traction device was developed to pull and maintain the limb position and install implants accurately[2]. However, the current device does not have a traction limit and entails a risk of complications due to excessive traction. Therefore, the purpose of the study is to develop a system to detect excessive traction and improve the safety of surgery by measuring the load applied to the patient's body. To determine the limitation of load, dummy and virtual lower-limb models that can estimate forces around the hip muscles were also developed.

Methods

Figure 1 shows the configuration of the system. The traction device developed by Surgical Alliance Inc. has four degrees of freedom: traction-relaxation, flexion–extension, abduction–adduction, and rotation of the lower-limb. A six-axis force sensor made by Leptrino Inc. was mounted between the traction plate and the boot, where the patient's foot was inserted to measure the force and moment applied to the boot. Traction devices with a force sensor have already been approved as medical devices and can be used in actual surgery. A dummy lower-limb model was developed to test the traction and force sensors and to estimate the muscle force around the hip. Clay was added around the bone model, and the length and weight of the model were 0.84[m] and 8.05[kg] which are similar to those of adult males.

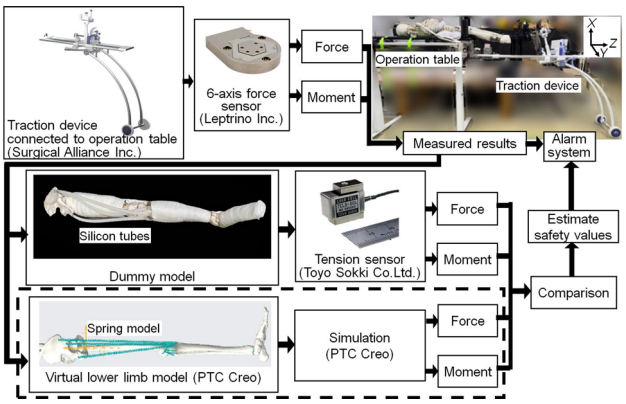

Fig. 1 Configuration of the total hip arthroplasty support system that includes dummy and virtual lower-limb models

In addition, three silicone tubes with outer and inner diameters of 15.0 and 6.0 [mm] were attached between the pelvis bone and the model that imitated the rectus femoris, sartorius, and semitendinosus muscles. The positions of the muscles were determined based on their anatomical origin and insertion. In addition, a small tension sensor made by Toyo Sokki Co. Ltd. was inserted into the middle of the tube to measure the tension force. The virtual lower-limb model was designed using the PTC Creo CAD system. The bone shape was 3D scanned from a human skeletal model, and the muscles and ligaments were modeled using a spring model. Knee joint and ankle joint were modeled by revolute and sphere joints respectively. Three muscles around the hip were defined as dummy models. The spring coefficient was defined using the Young's modulus of the silicon tube acquired by the tensile test. In addition, the operation table and traction device were modeled to move the lower-limb model. As the Creo CAD system includes the function of inverse dynamic analysis, the forces of the springs can be calculated during the movement of the foot connected to the traction table.

Traction experiments were performed using a traction device with a force sensor and lower-limb dummy model. The boot was pulled 50 [mm] in Z direction, as shown in Fig. 1, and the traction force was measured. In addition, the tension force of the silicon tube, which imitates the rectus femoris muscle, was measured. The experiment was repeated ten times. Using the virtual lower-limb CAD model, the foot part was also pulled 50 [mm] and the necessary force to pull and spring forces of muscles were calculated.

Results

The measured average maximum force in the Z direction was 42.7 [N] and the tension force of the silicon tube was 4.04 [N]. Conversely traction force calculated by the virtual model was 197 [N] and the spring forces for the rectus femoris, sartorius, and semitendinosus muscles were 48.2, 52.6 and 62.4 [N] respectively.

The pelvis was not tightly fixed and slipped during the experiment, but was completely fixed by the CAD model. Moreover, looseness was observed when the silicon rubber tube was connected, and the foot was inserted into the boot. This may cause a relatively small measured force compared with the calculated force. Additionally, the spring coefficients should be reviewed. However, during surgery, the traction force is approximately 200 [N], which is larger than that of the dummy lower-limb model.

Conclusion

A traction device with a force sensor approved as a medical device was developed. A lower dummy model that includes silicon rubber tubes as muscles around the hip joint was developed. In addition, a virtual lower-limb model that included a spring model as a muscle model, an operation table, and a traction device was modelled using the CAD system. These models were developed to estimate the force applied to the lower-limb tissues, particularly the muscles around the hip joint.

From the results of the experiment, the traction and tension forces of the muscle can be measured and calculated. However, because there were differences between the measured and calculated forces, both models require improvements. In addition, other movements of the traction table, such as rotation, extension, and adduction used during the surgery should be evaluated.

For future studies, by referring to the measured force and moments during surgery, the limit values of movement should be determined. An alarm and safety system should be built into the support system for safe surgery.

References

[1] Bergin PF, Doppelt JD, Kephart CJ, Benke MT, Graeter JH, Holmes AS, Haleem-Smith H, Tuan RS, Unger AS (2011) Comparison of minimally invasive direct anterior versus posterior total hip arthroplasty based on inflammation and muscle damage markers. J Bone Joint Surg Am 93 (15):1392–1398.

[2] Bannno S, Baba T, Tanabe H, Homma Y, Ochi H, Watari T, Kobayashi H, Kaneko K (2020) Use of Traction table did not increase complications in total hip arthroplasty through direct anterior approach performed by novice surgeon. Journal of Orthopaedic Surgery, Volume 28, Issue 2.

Fibular osteotomies with an electromagnetic navigated surgical cutting guide: a proof of concept study

A. F. de Geer[1,2], L. M. N. Aukema[1], M. J. A. van Alphen[1], W. H. Schreuder[1], R. L. P. van Veen[1], T. J. M. Ruers[3,4], F. J. Siepel[2], M. B. Karakullukcu[1]

[1]Netherlands Cancer Institute, Verwelius 3D Lab, Head and Neck Oncology and Surgery, Amsterdam, Netherlands [2]University of Twente, Robotics and Mechatronics, Enschede, Netherlands [3]Netherlands Cancer Institute, Department of Surgery, Amsterdam, Netherlands [4]University of Twente, Faculty of Science and Technology, Enschede, Netherlands

Keywords surgical navigation, mandible reconstruction, fibula free flap, electromagnetic tracking.

Purpose

The golden standard to reconstruct the original contour of the mandible after oncologic resection is with a fibula free flap (FFF). Virtual surgical planning (VSP) is performed prior to surgery to determine the locations and orientations of the osteotomy planes on the mandible and fibula. Currently, patient-specific cutting guides are designed and 3D printed to translate the VSP to the patient in the operating room. However, these cutting guides have shortcomings; they lack adaptability when the intraoperative situation is different than expected, e.g. due to tumor progression. Also, designing and obtaining the guides is a time-consuming and costly procedure. In the meantime, the tumor can grow outdating the virtually planned margins and increasing the patient's risk of death by 2.2% per week. Universal cutting guides that can be positioned on the bone using surgical navigation could overcome these problems. Previous research has already shown the feasibility of navigated mandibular osteotomies [1]. The aim of this study is to assess the feasibility of

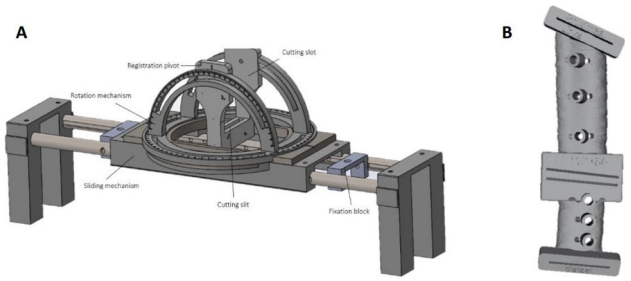

Fig. 1 Illustration of the surgical cutting guides used in this study: **A** EM navigated cutting guide, **B** 3D printed patient-specific cutting guide

performing fibular osteotomies using a novel electromagnetic (EM) navigated surgical cutting guide.

Methods

Fibular 3D models were constructed from computed tomography angiography (CTA) scans from five patients who received mandible reconstruction surgery in our institute. To simulate surgery, the fibular head and lateral malleolus were cut off and only the fibular shafts were 3D printed. All models were 3D printed twice. For every phantom, a CT scan was obtained with 0.5 mm slice thickness, the phantom was segmented, and a 3D model was constructed in 3D Slicer software. The 3D models were registered on the actual patient CTA scans so that the original VSP, including four osteotomy planes, could be used. Five phantoms were sawn with the navigated cutting guide (see Fig. 1): an EM sensor was attached to the proximal end of the fibula phantom to track its movements during the navigation procedure. An NDI Aurora field generator was positioned at 12 cm from the phantom, similar to intraoperative use. The cutting guide was registered to the phantom by pinpointing eight widely spaced pivots on the cutting slot with a tracked probe. Next, the tracked probe was used to register the VSP to the phantom by performing a hybrid registration, consisting of a point registration for initialization and a surface registration for optimization [2]. Then, the phantom was fixated into the cutting guide with screws and the cutting slot was (manually) moved over the phantom to navigate to the planned osteotomy planes using 3D Slicer software. For every osteotomy plane, the following procedure was performed: when the actual osteotomy plane was superimposed on the planned (virtual) osteotomy plane, the cutting slot was fixated and the osteotomy was performed with a handheld saw. To compare the outcomes of the navigated procedure with the current state-of-the-art, five phantoms were also sawn with patient-specific cutting guides printed in Polyamid12 (see Fig. 1). For all sawn fibula segments, a CT scan was obtained (0.5 mm slice thickness) and a 3D model was constructed.

Table 1 The length, yaw, and roll errors of the fibula segments obtained with the navigated cutting guide and the patient-specific cutting guides (mean ± SD)

	Navigated cutting guide n = 10	Patient-specific cutting guide n = 10
Length error (mm)	2.19 ± 2.19	0.66 ± 0.41
Yaw (°)	2.18 ± 1.64	1.96 ± 2.35
Roll (°)	1.64 ± 1.40	5.19 ± 3.37

For every segment, length errors and angular deviations (yaw and roll) were calculated using MATLAB R2022a.

Results

In total, five fibula phantoms were sawn with a first prototype of a novel EM navigated cutting guide (resulting in 10 segments) and five fibula phantoms were sawn with the originally designed patient-specific cutting guides (resulting in 10 segments). For the navigated procedure, the average fiducial registration errors (FRE) after cutting guide registration and hybrid registration were 1.28 ± 0.05 mm and 0.32 ± 0.03 mm respectively. The mean length, yaw, and roll errors of the obtained fibula segments using the navigated cutting guide and patient-specific cutting guides can be found in Table 1.

Conclusion

The preliminary results of this study show that a novel EM navigated surgical cutting guide could be a feasible method to perform fibular osteotomies during mandible reconstruction surgery. However, there is still a large variation in segment length. Therefore, the design of the EM navigated cutting guide and navigation procedure requires further optimization. Currently, the experiments are repeated with an improved version of the navigated cutting guide in order to reduce the length errors.

References

[1] Ter Braak TP, Brouwer de Koning SG, van Alphen MJA, van der Heijden F, Schreuder WH, van Veen RLP, Karakullukcu MB (2020) A surgical navigated cutting guide for mandibular osteotomies: accuracy and reproducibility of an image-guided mandibular osteotomy. Int J CARS 15:1719–1725

[2] de Geer AF, van Alphen MJA, Zuur CL, Loeve AJ, van Veen RLP, Karakullukcu MB (2022) A hybrid registration method using the mandibular bone surface for electromagnetic navigation in mandibular surgery. Int J CARS 17:1343–1353

Role of preoperative navigated transcranial magnetic stimulation for surgical management of gliomas located near motor function area

M. Tamura[1,2], A. Kuwano[1,2], T. Takakura[1], J. F. Mangin[3], I. Sato[4], M. Nitta[1,2], T. Saito[1,2], T. Kawamata[2], K. Masamune[1], Y. Muragaki[1,2]

[1]Tokyo Women's Medical University, Institute of Advanced Biomedical Engineering and Science (TWIns), Tokyo, Japan [2]Tokyo Women's Medial University, Department of Neurosurgery, Tokyo, Japan [3]Neurospin, Biomedical Imaging Institute, CEA, The Computer Assisted Neuroimaging Labratory, Gif/Yvette, France [4]Future University Hakodate, Faculty of System Information Science Engineering, Hakodate, Japan

Keywords nTMS, MEP, brain mapping, hot spot.

Purpose

We have clinically operated an intraoperative MRI-guided navigation system since 2000 in more than 2200 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 500 cases. To avoid post-operative complications concerning motor function in glioma patients, we provided pre-operative evaluation of Diffusion Tensor Imaging (DTI) tractography datasets for brain mapping of white

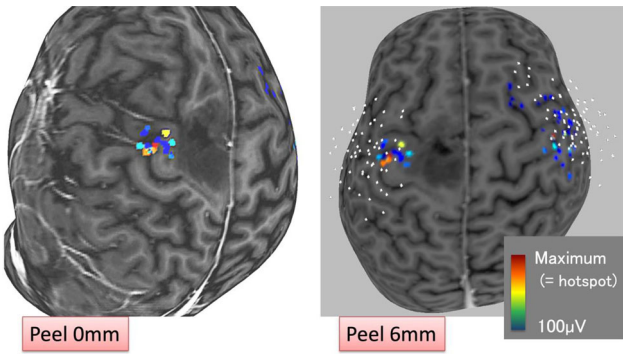

Fig. 1 Preoperative navigated TMS revealed monitoring APB muscle contraction as a pre-operative motor function area as upper limb motor area. We could confirm real motor function with direct response by electrical stimulation on cortex and direct muscle contraction

matter pathways during glioma resection. In pre-operative DTI tracking, only CST near lower limb was described and not visualized for upper limb fibers, which means false negative. We perform the pre-operative navigated transcranial magnetic stimulation (nTMS) to evaluate upper and lower limb motor function comparing with the DTI tractography, and finally confirmed with MEP monitoring and direct muscle test in OR. As we reported the usefulness of nTMS [1] in glioma resection near the motor area, including the supplementary motor area (SMA), we used the results of pre- and post-operative nTMS to predict postoperative motor function complications. In this study, we compared mapping findings with pre-operative nTMS during awake craniotomy and verified its usefulness.

Methods

A prospective observational study of 17 consecutive cases of glioma in the vicinity of the pre-central gyrus with preoperative nTMS consent was obtained. Sixteen cases in awake craniotomy was included (40 ± 12 years old, 10 males, 10 inital cases, 8 on the left side, 5 Grade II, 10 Grade III and one Grade IV in WHO classification). Preoperative nTMS mapping performedmotor evoked potential (MEP) measurements with resting stimulation threshold (RMT) as the contralateral abductor pollicis brevis (APB) target to identify the site of maximum amplitude recording (hot spot), and we added to targets as the orbicularis oris and anterior tibialis muscles based on tumor localization. In cases where resting MEP induction was difficult, hot spots were identified with activation. Awake electrical stimulation mapping (DES) was performed on the intraoperative brain surface, and the results were combined with the subcortical results during resection and compared with preoperative nTMS.

Results

In 16 cases where brain surface mapping corresponding to target muscles including APB was performed by nTMS before surgery, hot spots were identified in all cases including 5 cases with activation, Fig. 1. Motor areas were identified in 11 cases by awake DES on the brain surface, of which 8 (72.7%) were consistent with preoperative nTMS mapping sites. In addition, 8 of the 11 cases of nTMS performed with RMT without activation matched the actual brain surface DES location, while all 5 cases with activation did not identify the motor cortex wiht the area of interest around the tumor at the awake brain surface DES ($P = 0.026$, Fisher's exact test). It is possible that mapping around the tumor was difficult due to the elevated threshold, but DES was performed even during resection, and the site of interest was identified in 14 out of 16 awake resections. Intraoperative motor paralysis was observed in 13 cases, all of which were accompanied by findings of decreased MEP.

Conclusion

In resection of glioma near awake motor cortex combined with intraoperative MEP, preoperative nTMS provides accurate suggestions for intraoperative DES findings and resection planning and is useful.

Acknowledgement

This study was obtained from the commissioned research by Grant-in-Aid for Scientific Research (B, 22H03443), and National Institute of Information and Communications Technology (NICT), Japan.

References

[1] Takakura T, Muragaki Y, Tamura M, Maruyama T, Nitta M, Niki C, Kawamata T (2017) Navigated transcranial magnetic stimulation for glioma removal: prognostic value in motor function recovery from postsurgical neurological deficits. J Neurosurg:1–15. https://doi.org/10.3171/2016.8.JNS16442

Augmented Reality guidance for the surgical localization of pediatric chest wall tumors

R. van der Woude[1,2], J. van der Zee[1,2], M. Fitski[1], C. van de Ven[1], M. Wijnen[1], F. Siepel[3], A. van der Steeg[1]

[1]Princess Máxima Center, Pediatric Surgery, Utrecht, Netherlands
[2]University of Twente, Technical Medicine, Enschede, Netherlands
[3]University of Twente, Robotics and Mechatronics, Enschede, Netherlands

Keywords Augmented reality, Pediatric surgical oncology, tumor localization, chest wall.

Purpose

Surgical treatment of pediatric chest wall tumors is complex. Due to the highly malignant nature of these tumors, complete tumor resection with a wide tumor margin is the main goal of surgery. However, the inevitable removal of multiple ribs can lead to significant chest wall deformities that negatively affect respiration, mobilization and aesthetics. Consequently, to achieve radical resection while still sparing as much healthy tissue as possible, pediatric chest wall resections require meticulous surgical planning and accurate tumor localization. However, tumor localization can be difficult as these tumors have often become invisible and non-palpable due to neoadjuvant chemotherapy. Surgical decision making is currently based on multiple two-dimensional (2D) imaging modalities, palpation and thoracoscopy prior to resection. Subsequently, by using solely 2D imaging, the three-dimensional (3D) perception and anatomical relation of the tumor greatly depend on the surgeon's spatial interpretation.

The use of intraoperative 3D guidance, e.g. Augmented Reality (AR), could overcome these challenges and significantly improve surgical planning and anatomical understanding, thereby facilitating surgical decision making. After a first introduction of an AR system by Spijkerboer et al. [1], we have used the HoloLens 2 (Microsoft Corporation, Redmond, WA, USA) to intraoperatively localize chest wall tumors of five patients treated in our center. We present our early experience with this system and discuss the feasibility of AR guidance for the surgical localization of pediatric chest wall tumors.

Methods

From the first of January 2021 to the end of 2022, a total of five pediatric patients underwent surgical resection of a chest wall tumor

Table 1 Patient characteristics

Patient	Sex	Age (y) Median = 7	Tumor	Resected ribs	Neoadjuvant chemotherapy	Diagnostic tumor measurements (AP × RL × CC) (cm)	Preoperative tumor measurements (AP × RL × CC) (cm)
1	F	6	Ewing sarcoma	1 (7th)	Yes	11.2 × 14.9 × 19.9	4.6 × 2.5 × 5.9
2	M	12	Ewing sarcoma	3 (8th–10th)	Yes	5.9 × 5.4 × 7.4	5.8 × 1.3 × 4.5
3	M	7	Ewing sarcoma	3 (8th–10th)	Yes	4.6 × 6.4 × 5.6	4.4 × 3.3 × 1.6
4	M	2	Ewing sarcoma	1 (7th)	Yes	5.0 × 3.0 × 5.0	2.6 × 1.8 × 1.1
5	M	13	Mesenchymal chondrosarcoma	3 (5th–7th)	No	9.1 × 7.1 × 7.6	9.1 × 7.1 × 7.4

Fig. 1 Holographic overlay of the 3D model including the tumor, ribs, lung and anatomical landmarks (purple spheres). The reference QR code is fixated on the patient's hip. Note that there can be an apparent misalignment between patient and hologram due to the displacement between the camera and surgeon's line of sight

in the Princess Máxima Center (Utrecht, The Netherlands) (Table 1). For all patients, a patient-specific 3D model was created by segmenting the tumor and relevant anatomy from preoperative computed tomography (CT) images. Segmentation was done by technical physicians specifically trained in Ewing sarcoma delineation together with the performing surgeon (CvdV). Finally, the model was projected onto the patient in the operating room (OR) by a five-point registration method based on anatomical landmarks.

The preoperative CT scan was performed with the patients lying in the surgical lateral decubitus position. During this scan, five 1.5 mm radiopaque skin markers (Suremark®, Mesa, AZ, USA) were attached to recognizable landmarks such as scars, birthmarks, or the nipple. The tumor, adjacent ribs, five landmarks and involved anatomical structures were segmented from the CT images and the 3D model was integrated into the HoloLens 2 application in Unity (Unity Technologies, San Francisco, CA, USA).

Registration was performed pre-incisional by using a 3D printed pointer with a quick response (QR) code recognized by the HoloLens 2. A reference QR code was attached onto the patient to enable a stable visualization and adjust for respiratory movement. During registration, the surgeon pinpointed the anatomical landmarks and voice commands were used to save the real 3D position of each point. Subsequently, a Procrustes algorithm computed the most accurate transformation of the virtual 3D model onto the patient and the holographic overlay was realized. A more extensive explanation of the used method is described by Spijkerboer et al. [1].

Results

Registration and holographic overlay was achieved in all five patients. Figure 1 shows the results of the AR-guided tumor localization in one case. The pre-incisional holographic overlay appeared to be accurate for most patients, though it is difficult to quantify misalignments of the holographic overlay since it is unclear which localization is to be kept as ground truth. In some cases, minor disagreements between the overlay and expected tumor location based on the conventional localization methods were found. Moreover, the system was unable to maintain a correct projection of the 3D model once the skin had been opened. Lastly, as the chest wall generally lacks the presence of distinguishable and rigid landmarks, the five-point registration method proved to be prone to inaccuracies and user-dependent errors. The holographic overlay seemed most accurate when the five-landmarks were positioned in a non-symmetric configuration in proximity to the tumor.

Conclusion

Our results prove the further applicability of AR guidance for the pre-incisional localization of pediatric chest wall tumors during surgery. The system has the potential to enable intraoperative 3D visualization, hereby facilitating surgical planning and management of chest wall resections. Misalignments of the holographic overlay may be due to insufficient tracking of the reference QR code or inaccurate landmark selection during registration. Therefore, we are currently exploring the feasibility of different registration methods, such as surface matching. Moreover, as it remains difficult to quantify misalignments of the hologram within our current workflow, we are working on methods to measure our system's accuracy and to validate its performance.

References

[1] Spijkerboer KGB, Fitski M, Siepel FJ, van de Ven CP, van der Steeg AFW (2022) Augmented reality-guided localization of a chest wall tumor in a pediatric patient. European Journal of Cancer 170: 103–105

Augmenting instrument segmentation in video sequences of minimally invasive surgery by synthetic smoky frames

T. Rückert[1], M. Rieder[1], D. Rauber[1,2], M. Xiao[3], E. Humolli[3], H. Feussner[3,4], D. Wilhelm[3,4], C. Palm[1,2]

[1]Ostbayerische Technische Hochschule Regensburg (OTH Regensburg), Regensburg Medical Image Computing (ReMIC),

Regensburg, Germany [2]Ostbayerische Technische Hochschule Regensburg (OTH Regensburg), Regensburg Center of Health Sciences and Technology (RCHST), Regensburg, Germany [3]Klinikum rechts der Isar, Technical University of Munich, MITI Research Group (Minimally Invasive Interdisciplinary Therapeutical Interventions), Munich, Germany [4]Klinikum rechts der Isar, Technical University of Munich, Department of Surgery, Faculty of Medicine, Munich, Germany

Keywords Surgical instrument segmentati, smoke simulation, unpaired image-to-image transl, robot-assisted surgery.

Purpose

The precise segmentation of surgical instruments in endoscopic video images based on purely visual information is a central component for the development of robot-assisted procedures in minimally invasive surgical interventions. Recently, considerable progress has been made in this area of research, particularly through the use of deep neural networks. The main difficulty in the development of new approaches is to overcome issues encountered in real-world minimally invasive surgery, such as heavy smoke, unexpected bleeding, or instrument motion blur due to fast movements. In this work, we present a method that focuses on improving the segmentation quality of surgical tools in scenarios where visibility is limited by heavy smoke. To this end, we generate artificially smoked images from unsmoked video frames using an unpaired image-to-image [2] translation approach. Then, the resulting synthetic images augment the training process of a deep neural network for semantic image segmentation. We validated our method using a six-fold cross-validation approach, for which six recordings of real operations are provided, and compare our results to a baseline setup that involves no synthetic smoke images at all.

Methods

Overview

For the development of our approach, six videos were collected showing endoscopic images of laparoscopic surgeries performed in a minimally invasive manner. The videos were captured at a frame rate of 25 frames per second, and the duration of the recordings ranges from 29 to 52 min. In each video, one frame per second was annotated manually. Twelve different surgical tools are used in the surgical procedures, six of which can be divided into shaft and tip, providing a total of 18 classes as well as the background for annotation. Our approach is composed of two steps, as described in the following.

Stage 1: Synthetic video frame generation

In order to perform unpaired I2I translation, the images of the laparoscopic recordings were first manually divided into three modalities, namely non smoky, slightly smoky, and strongly smoky. Subsequently, using the images of the non smoky and strongly smoky categories, synthetic images are generated by a CycleGAN [2] architecture, using only every fifth frame for efficiency and resizing all input images to 912×513. As a result, for each unsmoked image there then exists a generated synthetically smoked version and vice versa.

Stage 2: Synthetic smoke image augmentation

A semantic segmentation model on basis of a DeepLabV3 + [1] architecture and a Lovász-Softmax loss is trained. In each training epoch a random number of non smoky images is selected and the respective strongly smoky counterpart is added to the training set. While the number of slightly smoky images remain untouched, as many artificially generated strongly smoky images are added that the sets of non smoky and strongly smoky images are the same size.

Experimental evaluation

To quantitatively compare the proposed approach using generated synthetic (GS) images, we first train baseline (BL) models using only the existing video images and the corresponding annotations. We evaluate both the baseline method as well as our proposed approach using six-fold cross-validation, where in each run the video frames of one of the six videos are considered as the validation dataset, and the images of the remaining five videos are used for training. The resizing of the input images is identical to that applied during I2I translation. To determine the impact of the approach on segmentation quality with respect to images with different smoke levels, the trained models are validated on the individual smoke modalities separately.

Results

Table 1 shows the quantitative validation results of our experiments for the BL and for GS. For each modality, the mean intersection over union (mIoU) and mean dice value (mDice) averaged over all folds and all instrument classes, except background, are reported as percentages. As can be seen, our method outperforms the baseline method with respect to both metrics, regardless of the considered smoke modality. Regarding the non smoky images, an improvement of 1.5% mIoU and 1.1% mDice was achieved, for the slightly smoky images the quality of the segmentations increased by 4.3% mIoU and 3.2% mDice, and for the strongly smoky images an increase of 1.2% mIoU and 1.1% mDice was observed.

A qualitative result is shown in Fig. 1. The left image shows a heavily smoked input of the network, in the center the segmentation prediction by applying BL is presented, and the right figure shows the segmentation output with GS. For a better visualization, the categories are encoded with different colors. It can be seen that the method still recognizes the shaft and parts of the tip of the center instrument quite well despite heavy smoke and poor visibility.

Conclusion

The presented approach offers promising results in terms of improving segmentation quality under challenging conditions, especially in the presence of heavy smoke during a minimally invasive surgical procedure. The results show that the use of generated synthetic training images using an unpaired I2I approach can improve segmentation results regardless of the severity of the smoke present. One component of future work is to further investigate the impact of the synthetic images on the segmentation results and the extent to which these segmentation results can be improved by more realistic smoke simulations.

Table 1 Results for BL and GS

	Non smoky		Slightly smoky		Strongly smoky	
	mIoU	mDice	mIoU	mDice	mIoU	mDice
BL	61.7	73.6	53.2	65.4	51.8	64.8
GS	62.6	74.4	55.5	67.5	52.4	65.5
	+ 1.5	+ 1.1	+ 4.3	+ 3.2	+ 1.2	+ 1.1

All results are averaged over all classes and folds and expressed in percent. The last row shows the improvement of GS with respect to BL

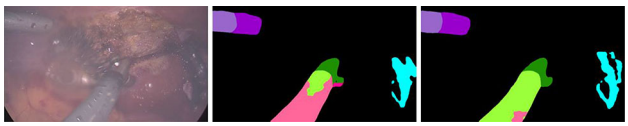

Fig. 1 Example of a smoky input image, the segmentation output using BL, and the prediction obtained by GS. The different classes are color coded for better illustration

References

[1] Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Computer Vision—ECCV 2018—15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII, volume 11,211 of Lecture Notes in Computer Science, 833–851. Springer.

[2] Zhu, J.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22–29, 2017, 2242–2251. IEEE Computer Society.

Deep learning-based soft-tissue-driven craniomaxillofacial surgical planning

X. Fang[1], D. Kim[2], X. Xu[1], T. Kuang[2], N. Lampen[1], J. Lee[1], H. H. Deng[2], J. Gateno[2,3], M. A. K. Liebschner[4], J. J. Xia[2,3], P. Yan[1]

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

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

Purpose

Craniomaxillofacial (CMF) surgery corrects deformities of the skull, the jaws, and the face. In these operations, surgeons cut the bones into multiple segments and rigidly move them into normal positions. The facial soft tissues change accordingly.

Current CMF surgical planning focuses on correcting the bony deformities, expecting a normal facial appearance to follow. However, bone-driven surgical planning is limited because correcting the bony deformity may not completely fix the distortion of the overlying soft tissues. The abovementioned problem occurs because the relationship between bones and the covering soft tissue is complex and nonlinear [1, 2].

Moreover, planning surgery with a bone-driven approach is time-consuming, even when one can predict the soft tissue changes that result from the bony movements. The bony movements need to be iteratively revised until an ideal facial appearance is achieved [1, 2]. Iterative plan revision is, thus, clinically impractical.

In this study, we propose a novel *soft-tissue-driven* surgical planning method that overcomes the limitations of conventional bone-driven surgical planning. Our deep learning-based planning method directly estimates the ideal bony movements (surgical plan) to achieve a desired facial appearance without iterative plan revisions.

Methods

We develop a deep learning-based framework that estimates the rigid movement of bony segments (LeFort 1, distal, right, and left proximal segments) to achieve desired facial soft tissue appearance. The proposed framework takes three inputs: pre-operative bony points, pre-operative facial points, and desired post-operative facial points. Each set of points is the point cloud sampled from the corresponding surface model. An overview of the proposed method is shown in Fig. 1.

First, using the three inputs, we predict the point-wise displacement of the bony points necessary to achieve the desired facial appearance. For the prediction, we used the Attentive Correspondence assisted Movement Transformation network (ACMT-Net) [2]. Second, the point-wise displacement is added to their corresponding bony points to estimate a non-rigidly transformed bony surface. Third, the transformed bony points are divided into four sets representing each bony segment. Fourth, four rigid transformation matrices are sequentially computed from the paired preoperative and transformed bony points using the Kabsch algorithm. Finally, the rigid transformation matrices are applied to their corresponding bone segment.

The performance of the proposed network is evaluated using 13 (11 for training and 2 for testing) sets of data randomly selected from our digital archive. The dataset composes of retrospectively analyzed data (surgical plan) of patients who underwent double-jaw orthognathic surgery. For evaluation, we assume the postoperative face is the desired facial appearance.

Our network was trained by minimizing the mean square error (MSE) loss between the predicted and postoperative bony points (ground-truth). Two different losses were considered during the training. 1. Non-rigid loss: MSE between the ground truth and non-rigidly transformed bony surface and 2. Rigid loss: MSE between the ground truth and bony surface rigidly transformed by the acquired transformation.

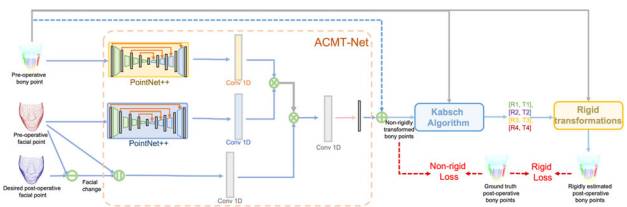

Fig. 1 Overview of the proposed soft-tissue-driven framework for bony movement prediction

Table 1 Results of the network supervised by non-rigid and rigid loss

Methods	Mean absolute error (mm)				
	LeFort 1	Distal	Left proximal	Right proximal	Entire bone
Non-rigid loss	2.12	2.16	1.47	1.4	1.82
Rigid loss	1.37	2.4	1.14	1.26	1.62

For efficient training, 4096 facial points were downsampled from each preoperative and postoperative facial model, respectively. 2048 bony points were downsampled from the preoperative bony model (512 points for each bony segment).

Results

The prediction accuracy of the proposed framework was evaluated using mean absolute error (MAE) between the predicted and ground-truth post-operative bones. Besides "Rigid loss" which is directly utilized to supervise the rigid estimation, we also report the performance of "Non-rigid loss", which is used to supervise the intermediate non-rigid estimation. "Non-rigid loss" has been commonly used in movement prediction tasks, e.g. facial change prediction. For detailed evaluation, MAE was also separately calculated for each bony segment, including LeFort 1, Distal, Left Proximal, and Right Proximal. The results of the network supervised by non-rigid and rigid loss are reported below. As shown in Table 1, compared to "Non-rigid loss", directly minimizing "Rigid loss" improves performance on three bony segments and the entire bone.

Conclusion

We developed a soft-tissue-driven method that directly predicts the rigid movement of four bony segments for CMF surgical planning. Instead of simulating facial tissue from bony movement, this is the first method that estimates bony movement to achieve a desired facial appearance. By removing the guess work of adjusting bony movement out of the process, the proposed method carries great promise for simplifying the surgical planning procedure.

References

[1] Kim, D., Kuang, T., Rodrigues, Y. L., Gateno, J., Shen, S.G., Wang, X., Stein, K., Deng, H.H., Liebschner, M.A., Xia, J.J., A novel incremental simulation of facial changes following orthognathic surgery using FEM with realistic lip sliding effect. Medical Image Analysis 72, 102,095 (2021).

[2] Fang, X., Kim, D., Xu, X., Kuang, T., Deng, H.H., Barber, J.C., Lampen, N., Gateno, J., Liebschner, M.A., Xia, J.J. and Yan, P., 2022. Deep Learning-Based Facial Appearance Simulation Driven by Surgically Planned Craniomaxillofacial Bony Movement. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 565–574). Springer, Cham.

25th International Workshop on Computer-Aided Diagnosis and Artificial Intelligence

Chair: Hiroyuki Yoshida, PhD (US)

Creating Pseudo Datasets from Disparate Datasets for Domain Adaptation for Brain Metastases Detection

A. Liew[1], C. C. Lee[2], B. L. Lan[1,3], M. Tan[1,4]

[1]Monash University Malaysia, School of Engineering, Selangor, Malaysia [2]Sunway Medical Centre, Radiology Department, Selangor, Malaysia [3]Advanced Engineering Platform, Monash University Malaysia, School of Engineering, Selangor, Malaysia [4]The University of Oklahoma, School of Electrical and Computer Engineering, Oklahoma, United States

Keywords Brain Metastases Detection, Domain Adaptation, Iterative Learning, Self-training.

Purpose

Brain metastases (BM) are the most common brain tumours in adults—typically 10–26% of patients with primary cancers will develop secondary BM. These lesions are small, have poor contrast and erroneously mistaken as blood vessels. Our work aims to apply deep learning methods to assist the neuroradiologist in identifying BM. For training and validation, we have obtained a dataset consisting of 121 patients with BM from a private medical facility in Malaysia (known as the SUNMED dataset from hereon). Due to the limited amount of datapoints in the SUNMED dataset, our proposed scheme jointly trains on two publicly available datasets, BrainMetShare [1] and BrATS [2]. However, it is widely established that training all 3 datasets together will not improve model performance. Here, we introduce the concept of pseudo datasets, made up of intermediary datasets for each learning iteration, based on a selection criterion. We will show that training the model with a diverse set of data improves the performance of the BM detection.

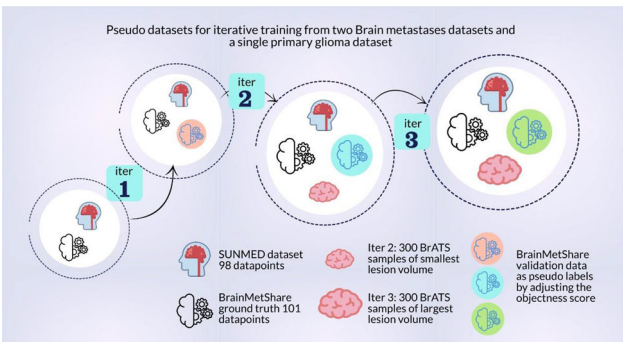

Fig. 1 Flowchart for the pseudo dataset selection criteria for iterative training

Methods

We propose a method to jointly train the SUNMED metastatic lesion dataset consisting of 121 head MRIs with two publicly available datasets, BrainMetShare [1] and BrATS [2]. While these three datasets consist of head MRIs, they represent 3 levels of domain shifts from one another—as they consist of different MRI sequences. Despite BrainMetShare being a brain metastases dataset, the origin of the primary cancer influences the appearance of the metastatic lesions. As there is no way to ensure that the distribution of patients with primary cancers are the same in both the SUNMED and BrainMetShare dataset, we assume a small domain shift between them. A larger domain shift occurs between the primary glioma dataset BrATS with metastases dataset BrainMetShare and SUNMED. The model is trained for 3 iterations with pseudo datasets consisting of pseudo labels from the validation dataset from BrainMetShare and the ground truth from BrATS. At every iteration, a new pseudo dataset is curated based on increasing the objectness scores of the pseudo labels and increasing the number of data points within a chosen volume (see Fig. 1). To evaluate the improvements to the SUNMED dataset, we calculate the sensitivity and false positive rate (FPR) in comparison with other methods.

Results

We compare a few methods of creating these pseudo datasets in Table 1: Observe that the sensitivity increases by 13 percentage points, while false positive rate drops drastically from 3.952 to 3.130 when iterative training is applied. Three versions of the pseudo dataset selections are investigated:

1. Limited to 375 of data points of comparable lesion volume
2. Using the next 300 datapoints with increasing lesion volume

•Using the 300 largest lesion volume datapoints

Our experimental results show that the largest 300 lesion volume datapoints to construct the pseudo dataset performs better than using only small and medium sized lesions. This could be attributed to the fact that the model trained on small and medium sized lesions have a smaller data distribution compared to having a dataset of large lesions. The model trained on the dataset with large lesions could have some important learned representations, which are missed in the pseudo datasets with less variance in lesion size.

Conclusion

Our study shows that for tasks with scarce datasets, performance improvements can be achieved by incorporating dissimilar datasets in a pseudo dataset approach. The pseudo datasets driven by the lesion volume selection helps bridge the large domain gap.

References

[1] Grøvik E, Yi D, Iv M, Tong E, Rubin D, Zaharchuk G (2020) Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI. J. Magn. Reson.

Table 1 Sensitivity and false positive (FP) metastases detection on SUNMED dataset

Description	True Positive (TP)	False positive (FP)	False positive rate (FPR)	Sensitivity
Non local block without iterative training	47	83	3.952	0.610
Non local block 3 iterations with pseudo datasets (small volume)	57	72	3.130	0.740
Non local block 3 iterations with pseudo datasets (300 next data)	58	69	3	0.753
Non local block 3 iterations with pseudo datasets (300 largest volume)	**58**	**52**	**2.261**	**0.753**

Imaging, vol. 51, no. 1, pp. 175–182, Jan. 2020, https://doi.org/10.1002/jmri.26766.

[2] Baid U, Ghodasara S, Mohan S, Bilello M, Calabrese E, Colak E, Farahani K et al. (2021). The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. https://doi.org/10.48550/arxiv.2107.02314.

Machine Learning Based Computer-Aided Simple Triage (CAST): Capability for COVID-19 Pneumonia Triage in Multicenter and Multi-Reader Study

Y. Ohno[1,2], T. Aoki[3], M. Endo[4], H. Koyama[5], H. Moriya[6], F. Okada[7], T. Higashino[8], H. Sato[7], N. Manabe[9], J. Matsumoto[10], K. Arakita[11], K. Aoyagi[11], Y. Ikeda[11], S. Kaminaga[11], A. Taniguchi[11], N. Sugihara[11]

[1]Fujita Health University School of Medicine, Radiology, Toyoake, Japan [2]Fujita Health University School of Medicine, Joint Research Laboratory of Advanced Medical Imaging, Toyoake, Japan [3]University of Occupational and Environmental Health School of Medicine, Radiology, Kitakyusyu, Japan [4]Chiba University Hospital, Comprehensive Radiology Center, Chiba, Japan [5]Osaka Police Hospital, Radiology, Osaka, Japan [6]Ohara General Hospital, Radiology, Fukushima, Japan [7]Oita Prefectural Hospital, Radiology, Oita, Japan [8]National Hospital Organization Himeji Medical Center,, Radiology, Himeji, Japan [9]Jichi Medical University Saitama Medical Center, Radiology, Saitama, Japan [10]St. Marianna University School of Medicine, Emergency and Critical Care Medicine, Kawasaki, Japan [11]Canon Medical Systems Corporation, Otawara, Japan

Keywords Computer-aided diagnosis, Computer-aided triage, Machine learning, CT.

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. Although reverse transcription polymerase chain reaction (RT-PCR) is the gold standard for diagnosing COVID-19, some studies also suggest that chest CT in particular may be positive in the setting of a negative RT-PCR test. Therefore, several reporting systems such as Reporting systems include the RSNA Expert Consensus Statement system, etc. were proposed for providing standardized language and diagnostic categories aiming to convey the likelihood that lung abnormalities on CT images represent COVID-19 [1,2]. Under the above-mentioned circumstances, we developed a machine learning (ML)-based CT texture analysis software for simple triage of COVID-19 based on the RSNA Expert Consensus Statement system. The purpose of this study was to determine the capability of ML-based computer-aided simple triage software (CAST) based on RSNA expert consensus statement for diagnosis of COVID-19 pneumonia in multicenter and multi-reader study.

Methods

174 cases underwent CT and polymerase chain reaction (PCR) test for COVID-19 (87 PCR positive and 87 PCR negative cases) were retrospectively included in this multicenter study. Then, CT data were assessed by CAST and consensus from three independent board-certified chest radiologists with equal to or more than 20-year experiences. Then, all cases were divided into two groups as follows: positive (i.e. typical and indeterminant appearances) and negative (i.e. atypical appearance and negative for pneumonia). To determine radiological finding evaluation capability on CAST, three other board-certified chest radiologists with equal to or more than 8-year experiences also assessed CAST results of radiological finding including pulmonary emphysema, nodular lesion, consolidation, ground-glass opacity (GGO), reticulation and honeycombing within 156 slices selected by central reading team and classified into five criteria as follows: agree, acceptable, disagree, true negative and false positive. To determine the agreements between CAST software and each investigator or among all investigators, inter-observer agreements between CAST software and each investigator was determined by Cohen's kappa statistics with $\chi 2$ test, and inter-rater agreement among all investigators were evaluated by Fleiss" kappa statistics. For comparison of diagnosis for COVID-19 pneumonia based on RT-PCR between ML-based CAST software and consensus evaluation in all cases as well as cases with COVID-19 pneumonia finding on CT, sensitivity, specificity and accuracy were compared by McNemar's test. To determine agreement for each radiological finding evaluation between ML-based CT texture analysis on CAST software and each investigator on all evaluated slices, inter-rater agreement among all investigators were evaluated by Fleiss" kappa statistics. Then, accuracies for all radiological evaluations between the ML-based CAST software and each investigator were compared each other by McNemar's test.

Results

Agreements between CAST software and consensus evaluation and each investigator were determined as "moderate" ($0.41 < \kappa < 0.56$, $p < 0.0001$). Inter-rater agreements of diagnosis for COVID-19 pneumonia among three investigators was determined as "moderate" (Fleiss's kappa value = 0.57, $p < 0.0001$). Although there was no significant difference of diagnostic performance between ML-based CAST and consensus evaluation in cases with COVID-19 pneumonia finding on CT ($p > 0.05$), specificity (SP) and accuracy (AC) in all cases on ML-based CAST software were significantly lower than those on consensus evaluation (SP: $p < 0.0001$, AC: $p < 0.0001$). Inter-rater agreement for each radiological finding among all investigators were determined as "moderate" to "substantial" ($0.54 <$ Fleiss's $\kappa < 0.81$, $p = < 0.0001$). On comparison of agreement accuracies for all radiological finding evaluations emphysema evaluation accuracy on investigator A (AC = 91.7%) was significantly lower than that of investigator B (100%, $p = 0.0009$) and C (100%, $p = 0.0009$).

On thin-section CT, ground-glass opacities and reticulation with crazy-paving were observed at peripheral lung in the both lungs. The CAST demonstrate ground-glass opacities as green and reticulation with crazy-paving as yellow in the both lungs. PCR test also showed this case as "positive". All chest radiologists and the CAST software were accurately evaluated this case as "positive case", and this case was determined as true-positive case in this study (Fig. 1).

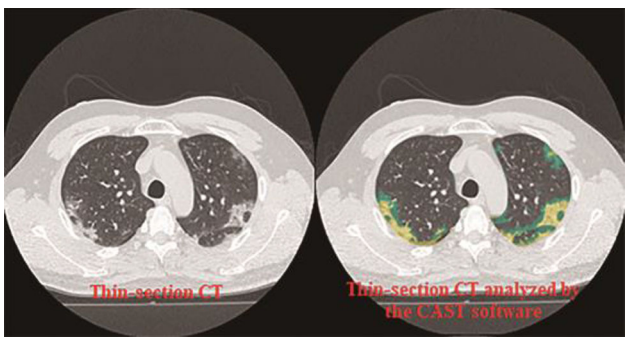

Fig. 1 COVID-19 pneumonia patient with typical appearance

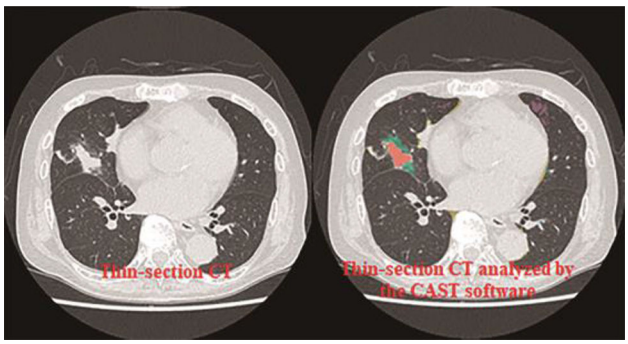

Fig. 2 Suspected COVID-19 pneumonia patient with atypical appearance

On thin-section CT, consolidation with ground-glass opacities was shown in the right middle lobe, and emphysema was also observed in the lingula segment. The CAST demonstrate consolidation and ground-glass opacities as pale beige and green in the right middle lobe, and emphysema as purple. PCR test also showed this case as "negative". All chest radiologists and the CAST software were evaluated this case as "atypical appearance" and "negative case". Therefore, this case was determined as true-negative case in this study (Fig. 2).

Conclusion

This multicenter study shows CAST is considered at least as valuable as chest expert radiologists for COVID-19 pneumonia triage with accurate radiological finding evaluations. This software has a potential to play as complementary role for management of suspected COVID-19 pneumonia patients as well as RT-PCR test in routine clinical practice.

References

[1] Simpson S, Kay FU, Abbara S, et al. Radiological Society of North America Expert Consensus Document on Reporting Chest CT Findings Related to COVID-19: Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. Radiol Cardiothorac Imaging. 2020 Mar;2(2):e200152.

[2] Simpson S, Kay FU, Abbara S, et al. Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA—Secondary Publication. J Thorac Imaging. 2020 Jul;35(4):219–227.

3-D SEH-YOLO: A YOLO-based Computer-aided Detection model for Lung Nodule Detection on Low-dose Computed Tomography

Y. S. Huang[1], T. A. Chen[2], R. F. Chang[2]

[1]National Changhua University of Education, Department of Computer Science and Information Engineering, Changhua, Taiwan [2]National Taiwan University, Department of Computer Science and Information Engineering, Taipei, Taiwan

Keywords Computer-aided detection, YOLO, Squeeze-and-Excitation, Hyper receptive field.

Purpose

Lung cancer is the most deadly cancer in the world. Early detection and treatment is the best way to reduce the mortality. Low-dose computed tomography (LDCT) screening is a widely used examination for lung cancer detection. Lung nodule is a symptom and may evolve into lung cancer. Therefore, it is important to discover suspicious nodules from LDCT screening. But, reviewing LDCT screening is a time-consuming process for the radiologist because of hundreds of slices and the small size of the nodule in each scan. Hence, the computer-aided detection (CADe) system has become a second reviewer to assist the radiologist. In recent, the convolution neural network-based (CNN-based) CADe system is proven an outstanding architecture in the medical imaing tasks due to its powerful performance. Many literatures present that CADe system with CNN actually could discover more ambiguous nodules. Therefore, in this research, a 3-D SEH-YOLO is proposed for the lung nodule detection.

Methods

In this research, our 3-D SEH-YOLO for lung nodule detection consists of the volumes of the data preprocessing and a 3-D SEH-YOLO detection model. In the data preprocessing, the voxel intensity normalization is performed to conver the intensity of each VOI into the range from 0 to 255. Then, the volumes of interest (VOIs) extraction is performed to separate the whole screening into VOIs with $80 \times 80 \times 80$ pixels. After the data preprocessing, the normalized VOIs are deliverd to the 3-D SHE-YOLOv4 model to detect suspicious nodules. The proposed detection model is constructed based on the YOLOv4 architecture and the feature extraction block, cross stage partial block, is modified by embedding the squeeze-and-excitation (SE) and the hyper receptive field block (RFB) modules for performance improvement.

Results

In experiments, the used dataset was the Lung Nodule Analysis 2016 (LUNA16) collected from the Netherlands and Italy in 2016. In this dataset, the slice thickness of scan greater than 2.5 mm and nodule size smaller than 3 mm are exculuded. Hence, there are 888 LDCT scans including 1186 nodules in total and all nodules are annotated by at least three experienced thoracic radiologists. Moreover, the LDCT scan is composed of a variable number of slices and the resolution of each slice is 512×512 pixels. In the system validation, the competition performance metric (CPM), the average sensitivity at seven pre-defined false positives (FPs) per scan, is employed as the evaluation metric. A higher CPM score means better performance. In the results, the CPM of proposed detection system is 0.898 and the sensitivity at eight FP is 0.96. The results are higher than some state-of-the-art detection models. In conclusion, the proposed 3-D SEH-YOLO modified by embedding SE and RFB modules into YOLOv4 architecture could improve the nodule detection capability.

Conclusion

In this study, a YOLO-based CADe system including the data preprocessing and the 3-D SEH-YOLO detection model is proposed for lung nodule detection from LDCT scans. As the results presentation, the proposed detection system in this study could achieve a good detection performance. In future, the system would be further improved by using other architectures or learning strategies, such as the knowledge distillation or self-supervised learning.

Virtual Radiomics Biopsy for the Histological Diagnosis of Pulmonary Nodules

G. Torres[1], D. Gil[1], S. Baeza Mena[2,3], A. Rosell Gratacós[2,3], C. Sanchez[1]

[1]Computer Vision Center (CVC), Universidad Autónoma de Barcelona (UAB), Computer Science Department (UAB), Barcelona,

Spain [2]Hospital Universitari Germans Trias i Pujol, Respiratory Medicine Department, Barcelona, Spain [3]Germans Trias i Pujol Research Institute (IGTP), Barcelona, Spain

Keywords radiomics, optimization, lung cancer, histology.

Purpose

Several clinical studies have demonstrated that Lung Cancer Screening (LCS) with low dose chest-CT (Computed Tomography) reduces the mortality rate by more than 20%, but there are challenges in its implementation, such as reducing the false positive rate. Besides the classification of the histological types is crucial to make targeted personalized treatments decisions timely, which have a positive impact on patients" comfort and survival rate. Currently, a definitive diagnosis of pulmonary nodules, including their histological type, can only be achieved by histopathological analysis of a tissue sample obtained by a biopsy surgical procedure.

Artificial intelligence methods applied to radiomics could represent a critical shift in the reduction of the false positive rate and an improvement of early diagnosis of lung cancer histological type providing clinicians with a detailed image-based diagnosis. Existing methods [2] either address the diagnosis of nodules (i.e. detection of whether a lesion is benign or malignant) or the determination of the histological type of nodules previously diagnosed as malignant. Although, many methods for malignancy detection can achieve an accuracy over 90%, the detection of the histological type still has room for improvement.

The goal of this work is to develop a 1-shot algorithm based on analysis of CT scans for the join detection of malignancy and the histological diagnosis of nodules for the two most frequent malign types, adenocarcinoma (ADC) and squamous cell carcinoma (SCC). In particular, we use radiomic features as input to a fully connected network and present a strategy for the optimization of its hyperparameters based on a multi-objective genetic algorithm combined with a nested cross-validation to optimize statistical metrics of the performance of networks.

Methods

This is a prospective cohort study initiated in December 2019 that includes patients who underwent surgery for PN in routine practice and from LCS, demographic and pathological data, and low-dose unenhanced chest CT scan.

An overview of our workflow is shown in the Fig. 1. In each CT a ROI volume enclosing the nodule is marked and segmented using Otsu thresholding and morphological operations to extract the nodule and mask. Then both are used to extract 24 Gray Level Co-occurrence Matrix GLCM features in axial 2D slices (PyRadiomics v3.01). A t-test is then used to select 19 GLCM features with maximum correlation with benign (B), ADC and SCC. These features are the input to training an optimized neural network and make the nodule's prediction based on max-voting (the most frequent classification) of the nodule's 2D slices. To compensate the imblalance between the classes a weighted cross-entropy is used.

For the network optimization, we formulate hyperparameters as a multi-objective optimization problem on the space of network architecture. The search strategy for solving the multi-objective problem is the Non-Dominated Sorting Genetic Algorithm (Optuna v2.10). The search space are the different models parameterized by the hyperparameters that define the network architecture: number of layers, number of neurons, activation function, weight initialization, optimizer and its weight decay. Besides architecture, the learning rate is included since it has an impact on performance. The performance of the different hyperparameters configuration is assessed by a nested cross-validation splitting of the train set. The μ (mean) and σ (standard deviation) of the loss function evaluated on the test folds define our multi-objective problem.

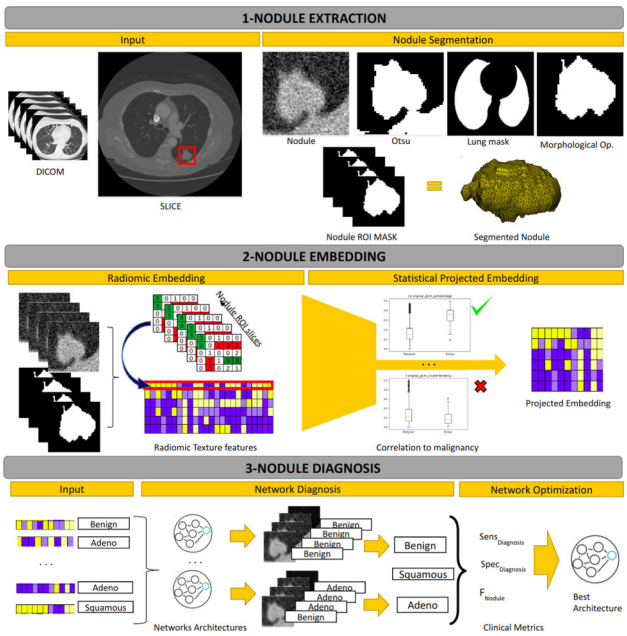

Fig. 1 An overview of our workflow

Results

84 patients have been recruited, dividing 59 pulmonary nodules (PN) (ADC: 42, SCC: 9, B: 8) into the training group and 25 PN (ADC: 15, SCC: 6, B: 4) into the test group. The dataset is publicly available at http://iam.cvc.uab.es/portfolio/radiolung-database.

The training group was used to optimize the internal parameters of the neural network. The optimal network consists of an input layer (19 neurons), a hidden layer (9 neurons), and an output layer (3 neurons), a hyperbolic tangent of activation function before and after the hidden layer and an orthogonal weight initialization of the layers, a Stochastic Gradient Descent of optimizer with 0.036 of weight decay and 0.015 of learning rate.

The optimal network has been evaluated on the independent test set to verify the reproducibility of the system. For the sake of comparison to existing methods, we assessed our method in terms of detection of malignancy (classification between malign and benign nodules) and the 2 malignant histological types (ADC, SCC) among the nodules detected as malignant. For each binary problem, the metrics used were the accuracy, specificity, sensitivity, F1-score and AUC, considering as positive class, malignant nodules and adenocarcinoma, respectively.

Table 1 reports the comparison to SoA methods with best performers in bold. For malignancy detection we achieve 100% of sensitivity like model3 proposed in [1], highest F1-score (0.98) and has top accuracy (96%). The sensitivity and AUC are significantly decreased because it misclassifies a benign nodule and the small number of such samples. Regarding the ADC/SCC classification, we have the highest F1-score (0.86) and are very close to the best performers in the other metrics (except AUC, which is close to the second best).

Conclusion

We have presented a hybrid method based on classic radiomic features combined with a network with an architecture optimized for the histological diagnosis of a PN using a novel strategy based on multi-objective optimization. Our optimized approach achieves competitive (being best for some metrics) results for identification of, both,

Table 1 Comparison to SoA methods with best performers

Study	Approach	Classes	Accuracy (%)	Specificity (%)	Sensitivity (%)	F1-Score	AUC
Peikert et al. [1]	Radiomics	Malignant/bening	–	85.5	90.4	–	0.939
Zhang et al. [1]	Machine learning		96.09	**95.34**	96.84	–	**0.979**
Multicrop [1]	Deep CNN		87.14	93	77	–	0.93
Nodule-level 2D [1]	Deep CNN		87.30	86	88.5	0.872	0.937
Vanilla 3D [1]	Deep CNN (3D)		87.40	85.2	89.4	0.873	0.947
DeepLung [1]	Deep CNN		90.44	–	81.42	–	–
AE-DPN [1]	Deep CNN		90.24	88.94	92.04	0.905	0.933
NASLung [1]	Deep CNN		90.77	95.04	85.37	0.89	–
model3 [1]	Hybrid		**96.30**	83.33	**100**	0.977	0.94
Our	Hybrid		96	75	**100**	**0.98**	0.8
Chaunzwa et al. [2]	VGG16 + CNN	ADC/SCC	68.6	**82.9**	37.5	–	0.709
Liu et al. [2]	CapsNet (3D)		**81.3**	80.7	**82.2**	0.796	**0.852**
Marentakis et al. [2]	LSTM + Inceptionv3		74	67	81	0.76	0.78
Our	Hybrid		80	80	80	**0.86**	0.77

malignancy and histological type using a highly unbalanced small size number of cases. These intermediate results show that radiomics are able to approximate the histological diagnosis of a pulmonary nodule and encourage further research including a higher number of cases and the optimization of convolutional architectures (like CapsNet [2]).

References

[1] Torres G, Baeza S, Sanchez C, Guasch I, Rosell A, Gil D (2022) An Intelligent Radiomic Approach for Lung Cancer Screening. Appl. Sci. 2022, 12, 1568. https://doi.org/10.3390/app12031568.

[2] Tomassini S, Falcionelli N, Sernani P, Burattini L, Dragoni A (2022) Lung nodule diagnosis and cancer histology classification from computed tomography data by convolutional neural networks: A survey. Computers in Biology and Medicine, 2022, 146, 105,691.

Investigation of federated learning for automated cerebral aneurysm detection in head MR angiography images

A. Yamada[1], S. Hanaoka[2], H. Shibata[2], T. Takenaga[2], T. Yoshikawa[3], Y. Nomura[3,4]

[1]Chiba University, Department of Medical Engineering, Faculty of Engineering, Chiba, Japan [2]The University of Tokyo Hospital, Department of Radiology, Tokyo, Japan [3]The University of Tokyo Hospital, Department of Computational Diagnostic Radiology and Preventive Medicine, Tokyo, Japan [4]Chiba University, Center for Frontier Medical Engineering, Chiba, Japan

Keywords computer-aided detection, federated learning, magnetic resonance angiography, cerebral aneurysm.

Purpose

In the development of computer-aided detection (CAD) software, it is desirable to use data collected from multiple sites to ensure stable performance across various data. However, collecting data from multiple sites requires resolving ethical and data management issues. Recently, distributed learning, which allows training a model on distributed datasets while keeping data locally, has been attracting attention. In particular, federated learning (FL)[1] has the advantage that the model can be updated using information from all sites participating in the training. While several research groups have reported the application of FL to medical image processing tasks, to the best of our knowledge, there has been no report on the application of FL to the automated detection of cerebral aneurysms in head magnetic resonance angiography (MRA) images. In this study, we investigated the application of FL to develop automated detection of cerebral aneurysm detection in head MRA images using data collected from multiple sites, scanner vendors, and magnetic field strengths.

Methods

(1) Dataset

This study was approved by the ethical review boards of our institutions. We collected a total of 315 MRA images from three in-hospital datasets. The details of the datasets are shown in Table 1. Each case included at least one aneurysm of 2 mm or more in diameter, which at least two experienced radiologists determined. Among the 105 cases for each dataset, 75 cases were used for training, while the remaining cases were equally assigned to validation and test sets.

(2) Detection of cerebral aneurysms in MRA images

The input MRA images are first resampled to a 0.3 mm isotropic voxel size using tricubic interpolation, and the signal intensity distributions of images were standardized by the global piecewise linear mapping. After that, we extracted arterial voxels from the MRA images by the region-growing-based method. We trained the 3D U-Net model, using each pair of the original MRA image and the area of aneurysms. For each aneurysm, a 48 × 48 × 48 cubic volume of interest (VOI) around the center of gravity of the aneurysm region is extracted. In addition, four augmented VOIs are generated by random shifts within ± 24 voxels on the x-, y-, and z-axes and random rotation (0°/90°/180°/270°) in each of the axial, coronal, and sagittal planes. The augmented VOIs are changed for each epoch. In the detection phase, 48 × 48 × 48 cubic VOIs are extracted at lattice points (whose intervals are 24 voxels in each direction) in the bounding box of the extracted arterial volume, and all VOIs are fed into the trained 3D U-Net model. The outputs of the U-Net model are combined by taking the average of overlapped voxels. After that, the

Table 1 Specification of the MRA datasets

Dataset ID	No. of cases	MR scanners	Magnetic field strength (Tesla)
A	105	Two Signa HDxt and one Discovery MR750 (GE Healthcare, Waukesha, WI, USA)	3
B	105	Skyra (Siemens Healthcare, Erlangen, Germany)	3
C	105	MRT200PP2 and Titan (Canon Medical Systems Co., Otawara, Japan)	1.5

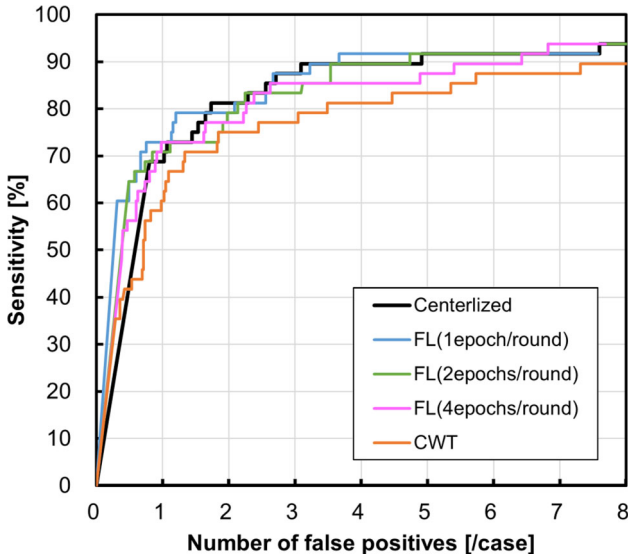

Fig. 1 FROC curves

lesion candidates are extracted by binarization of the combined output and the following connected component analysis.

(3) Federated Learning (FL)

The framework of FL is a server-client architecture, which consists of one server and k clients. In this study, $k = 3$ was set to match the number of datasets in Table 1. In FL, a global model is sent to each client to train using its local data. The model, rather than the data, is moved around the network. The server combines model updates to generate a new global model that is sent back to the local client for further training. This flow is defined as one round. We used the Federated Averaging (FedAvg) [1] generalized approach to update the global model.

We used a conventional strategy that trains the model using data collected from each site (hereafter, Centerlized) as a baseline. We employed 20 trials of hyper-parameter tuning with random search and utilized the area under the curve (AUC) value of the free-response receiver operating characteristic (FROC) curve, with the upper limit of 2 false positives (FPs) per case, as an evaluation criterion. The tuned hyper-parameters were: the depth of 3D U-Net (3/4/5), the number of filters of the first convolution layer (4–32, step: 2), the batch size (2–32, step: 2), the learning rate of the SGD (10^{-4}–10^{-2}). The number of maximum epochs and the patience of early stopping were set to 500 and 50, respectively. We compared cyclical weight transfer (CWT) [2] as another distributed learning method. The number of epochs per round for FL was set to 1, 2, and 4, and the number of epochs per site for CWT was set to 1. The other parameters for FL and CWT were set the same as those of Centerlized.

Results

Figure 1 shows the FROC curves for each method. The sensitivity at 2.0 FPs per case was 81.3% for Centerlized, 79.2% for FL with one epoch per round, and 75.0% for CWT. Among FL, the number of epochs per round of four was inferior in performance compared to that of one and two.

Conclusion

We investigated the application of FL to develop automated detection of cerebral aneurysms in MRA images. The results showed that federated learning is feasible for training the model using data collected from multiple sites, scanner vendors, and magnetic field strengths. Our future works include validating federated learning using other types of automated lesion detection and introducing differential privacy to enhance data security.

References

[1] McMahan B, Moore E, Ramage D, Hampson S, Arcas B (2017) Communication-efficient learning of deep networks from decentralized data. Proceedings of Machine Learning Research 54:1273–1282

[2] Chang K, Balachandar N, Lam C, Yi D, Brown J, Beers A, Rosen B, Rubin D, Kalpathy-Cramer J (2018) Distributed deep learning networks among institutions for medical imaging. Journal of the American Medical Informatics Association 25(8):945–954

Detection and Segmentation of Pancreatic IPMN Cysts in Multi-Sequence MRI by Deep Learning

N. Mazor[1], G. Dar[2], R. Lederman[1], N. Lev-Cohain[2], J. Sosna[2], L. Joskowicz[1]

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

Keywords Pancreatic cysts, Detection and segmentation, Multisequence MRI, Deep learning.

Purpose

Patients with pancreatic cancer have the lowest 5-year survival rate, which is often diagnosed late. In 80–85% of new diagnoses, metastases are already present, indicating an advanced disease stage [1]. Pancreatic cancer may originate from pancreatic cysts that become malignant. The most common cysts that have malignant potential are Intraductal Papillary Mucinous Neoplasms (IPMN) [2]. They are best detected and followed over time on MRI studies including TSE (Turbo Spin Echo) and MRCP (Magnetic Resonance Cholangiopancreatography) sequences. The detection and volumetric measurement

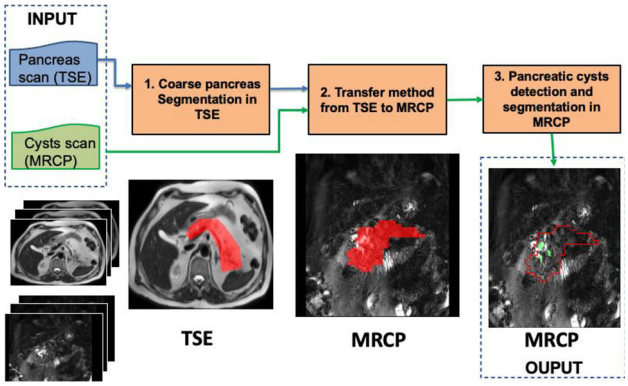

Fig. 1 Flow diagram of the multisequence cascaded 3D U-Net pipeline. The inputs are the TSE scans and MRCP scans of a patient. The outputs are the cysts segmentations in the MRCP scan. The method steps are: (1) coarse pancreas segmentation in TSE; (2) transfer of the coarse segmentation mask (red, lower left) in TSE to the MRCP sequence (red, lower middle); (3) detection and segmentation of pancreatic cysts (green, lower right; coarse pancreas boundaries shown in red)

of IPMN in MRCP scans remains challenging, time-consuming and requires expertise.

Methods

We developed a novel method for pancreatic cyst detection and segmentation in multisequence MRI scans. The method, called multisequence cascaded 3D U-Net, is a pipeline that consists of three steps: (1) coarse pancreas segmentation with a 3D U-Net in the axial TSE scan; (2) transfer of the coarse pancreas segmentation to the coronal MRCP MRI scan; (3) detection and segmentation of pancreatic cysts with a 3D U-Net in the MRCP scan, see Fig. 1. The method, which follows the current clinical guidelines [2], is the first to automatically detect and compute volumetric measurements for pancreatic cysts of all sizes in MRI. The multisequence cascaded 3D U-Net pipeline leverages the information of the MRI sequences with large resolution differences and different acquisition axes.

The method utilizes a new patch selection technique for deep learning network training with few annotated datasets based on hard negative mining. It starts by performing inference on a pre-trained 3D U-Net in the MRCP scans on the negative patches group that were not part of the training. It then constructs a False Positive (FP) Negative patches dataset consisting of only patches from the Negative patches group with misclassified voxels. The 3D U-Net model is then fine-tuned on a dataset consisting of a balanced patches dataset of Positive patches and FP Negative dataset patches.

To test our method, we retrospectively collected 158 MRI studies with TSE and MRCP sequences of patients with IPMN undergoing follow-up. The dataset was split into training/validation/test sets of 118/17/23 scans. The dataset contains a total of 840 cysts, of which 619 are > 5 mm and 221 are > 10 mm. The mean number of cysts per scan is 5.3 (2.6) with a mean cyst diameter (volume) of 7.4 mm (0.91 cc). The computed test set results were then compared to their respective manual ground truth delineations.

Results

For cysts of diameter > 5 mm, the method achieves a mean recall of 0.80 ± 0.19, a mean precision of 0.75 ± 0.26, a mean Dice score of 0.80 ± 0.19 and a mean Average Symmetric Surface Distance (ASSD) of 0.60 ± 0.53 mm. The patch selection method with hard negative mining outperforms the precision and recall of existing patch selection methods by 5.5% and 16.9% (F1 scores of 0.77 vs. 0.73 and 0.97 vs. 0.84) for cysts > 5 mm and > 10 mm, respectively. The cyst inclusion recall in the coarse pancreas segmentation is 0.94 (0.22). These results outperform the detection and Dice scores for pancreatic cysts in CT scans reported by others, Table 1.

Conclusion

Automatic IPMN cyst detection and segmentation in MRI may provide an accurate and reliable method for precise disease evaluation that that allows radiologists to reliably assess cyst size and longitudinal change.

Table 1 Results on three sets of pancreatic cysts (cysts with diameter > 10 mm, cysts with diameter > 5 mm, all cysts) of two patch selection methods: balanced positive and hard negative (*Bal-Pos-H-Neg*) and all patches (*All*)

Cysts diameter	Patch selection type	cyst detection		Cyst segmentation	
		Precision	Recall	Dice	ASSD
> 10 mm	*Bal-Pos-H-Neg*	0.95 (0.16)	**0.99** (0.05)	**0.81** (0.11)	**0.64** (0.50)
	All	**0.98** (0.11)	0.73 (0.30)	0.80 (0.17)	1.08 (0.98)
> 5 mm	*Bal-Pos-H-Neg*	0.75 (0.26)	**0.80** (0.19)	**0.80** (0.08)	**0.60** (0.53)
	All	**0.81** (0.27)	0.67 (0.27)	0.74 (0.20)	1.13 (0.98)
All cysts	*Bal-Pos-H-Neg*	0.61 (0.27)	**0.80** (0.22)	**0.80** (0.09)	**0.73** (0.63)
	All	**0.78** (0.27)	0.62 (0.27)	0.72 (0.20)	L17 (1.15)

Columns 3–4 list the mean (std) cysts detection precision and recall. Columns 5–8 list the mean (std) lesion segmentation Dice and Average Symmetric Surface Distance (ASSD). Boldface numbers indicate the best overall per-class results. The test set consists of 23 scans with a total of 110 IPNM cysts

References

[1] Mizrahi JD, Surana R, Valle JW, Shroff RT (2020) Pancreatic cancer. *The Lancet* 395(10,242):2008–20

[2] Tanaka M, Fernández del Castillo C, Kamisawa T, Jang JY, Levy P, Ohtsuka T, Wolfgang CL(2017) Revisions of the international consensus Fukuoka guidelines for the management of IPMN of the pancreas. *Pancreatology* 17(5):738–753,

3-D ResNeSt tumor diagnosis for automated breast ultrasound image

R. F. Chang[1,2], Y. S. Huang[3], C. Dai[1]

[1]National Taiwan University, Department of Computer Science and Information Engineering, Taipei, Taiwan [2]National Taiwan University, Graduate Institute of Biomedical Electronics and Bioinformatics, Taipei, Taiwan [3]National Changhua University of Education, Department of Computer Science and Information Engineering, Changhua City, Taiwan

Keywords automated breast ultrasound, computer-aided diagnosis, 3-D ResNeSt, breast cancer.

Purpose

The automated breast ultrasound (ABUS) was one of the most widely-used examinations for breast cancer. However, physicians were susceptible to the shape and texture of tumors and led to large variations in image interpretation. The computer-aided diagnosis (CADx) system was developed to offer a second opinion to deal with this problem. In recent years, applying convolution neural network (CNN) on CADx system became a tendency and achieved great success on medical images. Hence, a CNN-based CADx system was proposed for breast tumor diagnosis in this study.

Methods

Our CADx system was a 3-D model and consisted of data preprocessing, 3-D tumor segmentation, and 3-D tumor classification. In the data preprocessing, the volume of interest (VOI) was firstly extracted by the experienced physicians, and the tumor VOI would be resized and conducted histogram equalization. After that, the resized VOI would be sent into the segmentation model using U-Net + + to obtain the corresponding tumor mask. Lastly, the resized VOI, the equalized VOI, and the tumor mask were fed into our classification model, 3-D MASP-ResNeSt which was built based on the ResNeSt, the SPP block, and the MA block, to determine the tumor was benign or malignant.

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. In total, there were 396 patients with 444 pathology-proven lesions in our dataset. In 444 pathology-proven lesions, there are 226 malignant tumors and 218 benign lesions, respectively. In our experiment, the proposed CADx system achieved 89.6% accuracy, 88.9% sensitivity, and 90.4% specificity. The results indicated the proposed CADx system had a good performance and might be a second opinion for physicians to make the decision.

Conclusion

In this study, the CADx system, which was based on the tumor segmentation model and the tumor classification model, was proposed for tumor diagnosis. In the tumor segmentation, the tumor masks were generated by 3-D U-Net + + , and then the tumor mask would be sent into the classification model as a part of the input. In the tumor classification, the 3-D MASP-ResNeSt was constructed by the modified 3-D ResNeSt with the 3-D MA block and the 3-D SPP block for tumor diagnosis. The 3-D ResNeSt, which utilized the channel attention and group convolution in the architecture, was adopted as our backbone model. In addition, the 3-D MA block modified the same radix to different radix for learning the variety of features. The 3-D SPP block pooling the feature maps with different kernels was also embedded in our model to obtain a more robust combination of features.

Computer-aided Diagnosis of Colorectal Polyps in Photon-counting CT Colonography: Phantom Study

J. Näppi[1,2], T. Hironaka[1], D. Wu[1,2], R. Gupta[1,2], R. Tachibana[1,2], K. Taguchi[3], H. Yoshida[1,2]

[1]Massachusetts General Hospital, Radiology, Boston, United States [2]Harvard Medical School, Boston, United States [3]The John Hopkins University, Radiology, Baltimore, United States

Keywords Computer-aided detection, Photon-counting CT, CT colonography, Colon cancer.

Purpose

Colorectal cancer (CRC) is the third most common cancer and the second most common cause of cancer deaths worldwide. Colorectal screening for large polyps and early-stage cancers would largely avoid the morbidity, mortality, and high treatment costs of advanced CRC. Computed tomographic (CT) colonography provides a nearly ideal primary screening strategy for addressing the issues and problems inherent with other CRC screening tests. However, the ionizing radiation of CT technology has been cited as a concern for using CT colonography in population screening. Photon-counting CT (PCCT) can be used to address that concern. A PCCT imaging system is designed to count the number of individual x-ray photons that exceed a specified energy level. By setting the photon-counting energy threshold slightly above the level of electronic noise from the electronic circuits of the underlying x-ray detection system, the electronic noise of conventional CT can be excluded from the measured photon count data [1]. This rejection of electronic noise enables PCCT to be used at a much lower dose than conventional CT while yielding the same if not better overall CT image quality. Preliminary studies have shown that PCCT can yield up to a 30% reduction in the radiation dose of CT examinations without compromising image quality. This pilot study aimed to investigate the feasibility of performing computer-aided diagnosis of clinically significant polyps in PCCT colonography.

Methods

An anthropomorphic phantom (Phantom Laboratory, Salem, NY, USA) that had been designed to resemble human anatomy on clinical 120 kVp CT colonography scan images was filled partially with a 300 mL mixture of saline, a non-ionic iodinated contrast agent (300 mg/ml Omnipaque, GE Healthcare, Chicago, IL) in 40 mg/ml concentration, aqueous fiber (30 g of psyllium), and ground foodstuff (10 g of cereal) to simulate a laxative-free CT colonography examination. The phantom contained 23 simulated pedunculated, sessile, and flat polyps 6–15 mm in their largest diameter in combination with

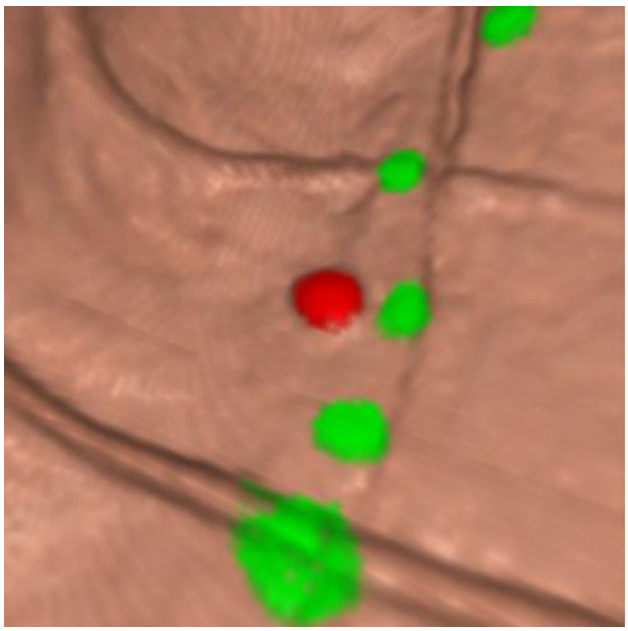

Fig. 1 In this example, the true-positive polyp candidate detected by the CADe system is indicated by red color, whereas the false-positive polyp candidates that were excluded correctly by the SVM as false positives are indicated by green color

simulated haustral folds. The phantom was scanned by use of a 16-slice PCCT scanner (OmniTom, NeuroLogica, MA, USA) at 120 kVp voltage, 40 mAs current, 0.71-mm slice thickness, and a 50 keV x-ray photon-counting threshold. The PCCT images were reconstructed at a 0.6-mm spatial resolution and 0.6-mm reconstruction interval by use of a filtered back-projection algorithm. After the CT image acquisition, polyps were detected automatically from the volumetric PCCT datasets by use of a fully automated computer-aided detection (CADe) system [2]. The CADe system performs a fully automated 3D segmentation of the colon region from the input dataset, followed by volumetric shape-based detection of polyp candidates along the extracted region of the colon wall. False-positive (FP) detections were reduced based on a radiomic analysis of the detected polyp candidates by use of a support vector machine (SVM). The polyp detection performance was evaluated by use of a tenfold cross-validation method.

Results

The CADe system detected all of the 23 clinically significant simulated polyps from the PCCT datasets (100% detection sensitivity). After the FP reduction step, no FP detections remained (0 FP detections per patient). Figure 1 illustrates the detection of a simulated polyp from the phantom PCCT dataset.

Conclusion

A PCCT colonography examination can be performed at a much lower radiation dose than what is required by conventional CT colonography. In this pilot study, we showed that PCCT colonography has the potential to yield a high accuracy in the automated detection of colorectal polyps. The significance of the result is that an effective ultra-low-dose PCCT colonography examination would largely eliminate the remaining concerns about the theoretical radiation risk of CT technology which has been cited as a major concern regarding the use of CT colonography in population screening.

References

[1] Rajendran K, Petersilka M, Henning A, Shanblatt E, Schmidt B, Flohr T, Ferrero A, Baffour F, Diehn F, Yu L, Rajiah P, Fletcher J, Leng S, McCollough C (2022). First clinical photon-counting detector CT system: technical evaluation. Radiology 303:130–8

[2] Näppi J, Yoshida H. Fully automated three-dimensional detection of polyps in fecal-tagging CT colonography (2007). Academic Radiology 14(3):287–300

Performance evaluation of a weakly unsupervised GAN-based survival prediction model on progressive fibrosing interstitial lung disease

M. Okamoto[1], J. Näppi[1], C. Watari[1], H. Yoshida[1]

[1]Harvard Medical School/Massachusetts General Hospital, Radiology, Boston, United States

Keywords computer-aided risk assessment, generative adversarial network, prognostic imaging biomarker, Progressive interstitial lung disease.

Purpose

Progressive fibrosing interstitial lung disease (PF-ILD) is a group of diseases characterized by increasing self-sustaining fibrosis, progressive worsening of dyspnea, progressive decline in lung function, limited response to immunomodulatory therapies, and high mortality [1]. Due to the highly variable rates of decline and poor prognosis of PF-ILD, accurate individualized prediction of mortality is crucial for therapeutic decision-making and management of the patients. However, no formal staging system based on prognosis has been established for PF-ILD, and none of the developed existing prognostic biomarkers, including quantitative CT imaging biomarkers and radiomics, have been considered accurate enough for establishing such a system.

Recently, we developed and showed that a survival analysis method based on a conditional generative adversarial network (cGAN), called pix2surv, can provide an effective image-based prognostic predictor for COVID-19 [2]. The purpose of this study was to evaluate the performance of the pix2surv model on the prediction of the survival of patients with PF-ILD compared with existing clinical biomarkers of the gender, age, and physiology (GAP) index and the composite physiologic index (CPI).

Methods

Idiopathic pulmonary fibrosis (IPF) is the most common form of PF-ILD. Thus, we retrospectively identified 75 patients with a diagnosis of IPF who had high-resolution chest CT scans and pulmonary function tests performed at the time of the diagnosis. The CT scans of the patients were performed with full inspiration from the lung apices to their bases. CT images were obtained from the CT scans performed at inspiration, and the supine position was included. The slice thickness was 0.6–1.5 mm, and the detector configuration was 16–64 × 0.625. The tube voltage was 120 kVp or 140 kVp with automatic tube current modulation.

Figure 1 shows the schematic architecture of our pix2surv model, where the time generator G is used to generate a "survival-time image" from the input CT image to predict the survival of the patient. In this study, we defined the survival time of a patient as the number of days from the patient's chest CT scan to death. The survival-time image has a single survival-time value at each voxel, and the discriminator D attempts to differentiate the "predicted pairs" of input CT images and their predicted survival-time images, from the

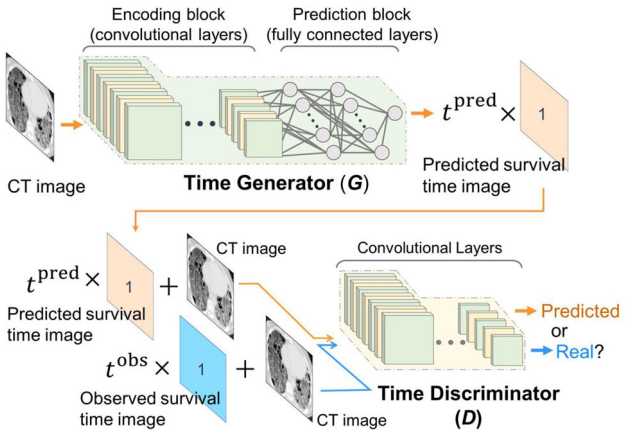

Fig. 1 The schematic architecture of our weakly unsupervised cGAN-based survival prediction model, pix2surv

"observed pairs" of the input CT images and their observed (actual) survival-time images of the patients. The training of vox2surv involves the optimization of G and D through a modified min–max objective function so that G can learn to generate a survival time that is close to the observed survival time. The predicted survival time of a patient is calculated as the median of the predicted survival times of the patient's CT images.

We used the concordance index (C-index) as the metric to evaluate the performance of the prognostic prediction. A bootstrap method with 50 replications was used to estimate the C-index. The prognostic prediction performance of pix2surv was compared with those of the GAP index and the CPI using a two-sided unpaired t-test. We also compared the equivalence of Kaplan–Meier (KM) survival curves generated by pix2surv, GAP, and CPI to the actual survival curve of the patient cohort by using a non-parametric equivalence test.

Results

Table 1 shows the median C-index values calculated from the bootstrap evaluation, which shows that the performance of the mortality prediction by the pix2surv model (C-index: 82.6% [95% CI: 80.8, 83.8]) was statistically significantly higher (bootstrap t-test, $p < 0.0001$) than those of the CPI (63.0% [61.7, 64.3]) and the GAP index (65.9% [64.7, 67.1]).

Figure 2 shows the KM survival curves predicted by the CPI, GAP index, and pix2surv for the mortality of the IPF patients in magenta, red, and green colors, respectively, where, in comparison, the actual survival curve of the patient cohort is shown in blue. Non-parametric equivalence test showed that the survival curves predicted by pix2surv were statistically equivalent to the actual survival curves over the period of 0 to 5000 days, whereas those predicted by the CPI and

Table 1 The C-index values estimated by the bootstrap evaluation of the CPI, GAP, and vox2pred. 95% CI = 95% bootstrap confidence interval. *Two-tailed t-test

Clinical/Radiomic Features	C-index (%) [95% CI]	P-value*
composite physiologic index (CPI)	63.0 [61.7, 64.3]	P < 0.0001
Gender, age, and physiology (GAP) index	65.9 [64.7, 67.1]	P < 0.0001
pix2surv	82.6 [80.8, 83.8]	

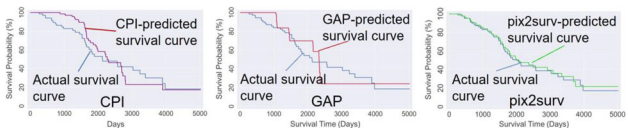

Fig. 2 Predicted KM survival curves by the CPI, GAP index, and pix2surv in comparison with the actual survival curves for the patient cohort

GAP were not. Also, visual assessment indicates that pix2surv approximates the actual survival curve substantially better than do CPI and GAP.

Conclusion

We evaluated the performance of our image-based weakly supervised survival prediction model, pix2surv, which can directly predict the survival of patients from their chest CT images, on the survival prediction of patients with PF-ILD. We showed that pix2surv outperforms the current standards of GAP and CPI in predicting the survival of patients with PF-ILD, indicating that pix2surv can be an effective prognostic biomarker for PF-ILD.

Acknowledgments

This study was supported partly by the NIH/NIBIB/NCI grants of R01HL164697 (PI: Yoshida).

References

[1] Fischer A, Antoniou KM, Brown KK, Cadranel J, Corte TJ, Du Bois RM, Lee JS, Leslie KO, Lynch DA, Matteson EL, Mosca M, Noth I, Richeldi L, Strek ME, Swigris JJ, Wells AU, West SG, Collard HR, Cottin V (2015) An official European Respiratory Society/American Thoracic Society research statement: Interstitial pneumonia with autoimmune features. European Respiratory Journal 46:976–987.

[2] Uemura T, Näppi JJ, Watari C, Hironaka T, Kamiya T, Yoshida H (2021) Weakly unsupervised conditional generative adversarial network for image-based prognostic prediction for COVID-19 patients based on chest CT. Medical Image Analysis 73:1–14.

Application of Convolutional Neural Network with Transfer learning to pattern recognition of tuberculosis in chest X-ray images

L. Lins de Lima[1,2], R. Rodrigues[1], J. Saito[3], O. Yussuf[1], L. Peron[1], K. Lopes[1], M. Koenigkam-Santos[1,4], P. Azevedo-Marques[1]

[1]University of Sao Paulo—USP, Ribeirao Preto Medical School—Medical Imaging, Hematology and Oncology, Ribeirao Preto, Brazil [2]University of Sao Paulo, Interunit graduate program in Bioengineering, Sao Carlos, Brazil [3]University of Sao Paulo, Medical Physics, Ribeirao Preto, Brazil [4]University of Sao Paulo, Bauru Medical Course, Bauru, Brazil

Keywords Tuberculosis, X-ray images, Convolutional Neural Network, Transfer Learning.

Purpose

Tuberculosis is the disease responsible for most deaths provoked by an infectious agent, and it caused around 2.5% of deaths worldwide in 2004 [1]. Tuberculosis can also increase the risk of lung malignancy, chronic pulmonary aspergillosis, and septic shock, even in patients that survived the primary infection caused by the "Mycobacterium tuberculosis," responsible for this illness [1]. The treatment is successful at about 85%, and the mortality rate is about 15% [1].

In diagnosing pulmonary diseases, including tuberculosis, the initially requested imaging exam is the chest X-ray. Although it is considered a simple exam, its evaluation can sometimes be complex. To support radiologists" decision-making, computer-aided diagnosis (CAD) systems have been developed to act as a second opinion through a computer-supplied suggestion.

The use of artificial intelligence (AI) to support diagnostic decision-making in radiology has grown exponentially in recent years. Machine learning (ML) has been the basis for CAD systems. ML is a way of "training" an algorithm so that it can learn. "Training" involves providing large amounts of data to the algorithm, allowing it to adjust and improve its performance. ML algorithms look for patterns within a dataset.

From 2010 to 2012, an approach called Deep Learning (DL) was increasingly adopted for solving ML problems. One approach based on DL is the Convolutional Neural Network (CNN), focusing on classifying images. Modeling the best CNN architecture for any situation by hand can be exhausting, time-consuming, and expensive. Another difficulty is getting 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, for example, the ImageNet [2].

This work aimed to develop a CNN model using transfer learning to support the diagnosis of tuberculosis in chest radiographic images in frontal projection.

Methods

Our institutional review board approved this retrospective study with a waiver of the patient's informed consent (CAAE: 25,762,319.7.0000.5440). For the application of CNN models, a database of chest X-ray images was used with a total of 547 images, with 382 images presenting normal cases and 165 images from patients diagnosed with tuberculosis at the Ribeirao Preto Clinics Hospital. The database was structured with radiographic exams saved in.png format and with three channels (RGB).

The processing environment consists of a server that was accessed remotely; this server has an NVIDIA® TeslaTM T4 16 GB GPU and operates with the Linux 18.04 LTS system. For the execution of the codes, the Jupyter Notebook software was used, which is an interface to browse the server files and execute the codes in the Python language (version 3.9) using the TensorFlow framework (version 2.8).

Five types of transfer models based on CNN were tested. The selected architectures were: VGG19, InceptionV3, DesNet201, ResNet125v2, and XceptionV3. For the use of the models, a global average pooling layer was added to the output of the last layer in each architecture. The original fully connected layers were removed and replaced by only one fully connected layer with two output neurons using the activation function Softmax. All the weights of the models were retrained. These architectures were trained individually, and a standard preprocessing approach with the function "preprocess_input" was performed on the original images to evaluate each model. The number of epochs used was 100. The callback function ReduceLROnPlateau was applied to find the best learning rate value on the validation dataset during the training with values of 0.5, 5, 'max,' and 0.00001 for factor, patience, mode, and min_lr, respectively. The batch size used was 30, and the optimizer function used was Adam to minimize the categorical cross-entropy.

To evaluate the models' performance, the samples were shuffled. The Holdout method was applied to train the networks and classify the images, where 481 images were used for training, 10 percent were separated for validation, and 66 were used for testing the architectures. The models' performance was evaluated based on specificity, sensitivity, and accuracy.

Table 1 Results obtained for each tested model

Model	Specificity (%)	Sensitivity (%)	Accuracy (%)
VGG19	90	90	90
Inception V3	100	81	90
Desnet201	96	81	89
Xception	57	96	77
Resnetl25v2	33	96	65

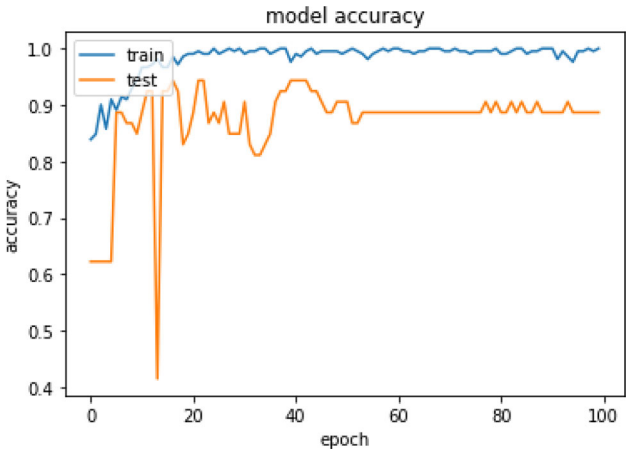

Fig. 1 Accuracy values obtained during the training and testing phases of the VGG19 architecture

Results

At the end of each architecture's training and testing, results presented in Table 1 were generated. During the training and testing phases, the VGG19 model gave the best and most stable results, with specificity, sensitivity, and accuracy of 90%.

Figure 1 presents the accuracy values during the training and testing phases of the VGG19 model.

Conclusion

Based on the results obtained, it can be stated that the networks performed very well when classifying positive images for tuberculosis. It can be seen that the Resnet125v2 and Xception networks had an excellent performance when evaluating the normal cases; however, when considering the images with a positive result for tuberculosis, the VGG19, InceptionV3, and Desnet201 networks presented better results.

In general, the networks presented a satisfactory performance for classifying tasks, and other tests will be carried out using public databases to simulate situations of exams from different sources to verify their generalization power.

Acknowledgments

This study was financed in part by the Coordenacção 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] Alzayer Z, Al Nasser Y. (2022). Primary Lung Tuberculosis. Treasure Island (FL): StatPearls Publishing. PMID: 33,620,814.

[2] Lima L L, Koenigkam-Santos M, Peron L, Fontes Lopes K L, Lourenço Repolês, M, Koenigkam-Santos M, Azevedo-Marques, P. (2022). Classifying pulmonary interstitial opacities in chest X-ray images using convolutional neural network and transfer learning. In: CARS 2022—Computer Assisted Radiology and Surgery Proceedings of the 36th International Congress and Exhibition Tokyo, Japan, June 7–11, 2022. Int J CARS (2022) 17 (Suppl 1): S93-S94. https://doi.org/10.1007/s11548-022-02635-x

Distal radius fracture classification on dual-view radiography using ensemble deep learning framework

H. Min[1,2], A. Wadhawan[3], Y. Rabi[4], P. Bourgeat[1], J. Dowling[1,2,5,6,7], J. White[3,8], A. Tchernegovski[9], B. Formanek[10], M. Schuetz[4,11,12,13], G. Mitchell[3,8,11], F. Williamson[3,8,11], C. Hacking[3,8], K. Tetsworth[3], B. Schmutz[4,11,12,13]

[1]CSIRO, Australian e-Health Research Centre, Herston, Australia [2]University of New South Wales, South Western Clinical School, Sydney, Australia [3]Royal Brisbane and Women's Hospital, Herston, Australia [4]Queensland University of Technology, School of Mechanical, Medical and Process Engineering, Brisbane, Australia [5]University of Wollongong, Centre for Medical Radiation Physics, Wollongong, Australia [6]The University of Sydney, Institute of Medical Physics, Sydney, Australia [7]University of Newcastle, School of Mathematical and Physical Sciences, Newcastle, Australia [8]University of Queensland, Medical School, Brisbane, Australia [9]Monash Medical Centre, Clayton, Australia [10]University of Queensland, St Lucia, Australia [11]Jamieson Trauma Institute, Brisbane, Australia [12]Queensland University of Technology, ARC Training Centre for Multiscale 3D Imaging, Modelling, and Manufacturing, Brisbane, Australia [13]Queensland University of Technology, Centre of Biomedical Technologies, Brisbane, Australia

Keywords Distal radius fracture, X-ray, Deep learning, Ensemble learning.

Purpose

Distal radius fractures (DRFs) are one of the most common fractures treated surgically. To examine the injured wrists, standard radiographs, including posteroanterior (PA) and lateral views, are often taken in the emergency department. DRFs can be classified into intra- and extra-articular fractures. In extra-articular fractures, the fracture line does not extend to the joint, while the intra-articular fractures involve the articular surface, which may require further evaluation and more complex treatments. Identifying DRFs as intra- or extra-articular can be useful for guiding further treatment. However, radiographic classification of DRFs is challenging due to the extreme variability of fracture patterns, complex anatomy of the wrist and variability in imaging quality of radiography. The aim of this study is to propose a deep learning (DL) framework incorporating both PA and lateral view X-rays for automatic DRF classification and evaluate the framework on clinically acquired wrist X-ray dataset.

Methods

The proposed framework consists of a distal radius region of interest (ROI) detection stage and a DRF classification stage as shown in Fig. 1. The distal radius ROI detection stage used an ensemble model of 10 YOLOv5 [1] base networks which is a recent release of the YOLO object detection network. This step allows the framework to zoom in on the relevant regions on PA and lateral view X-rays for fracture pattern analysis. Following the ROI extraction, an ensemble model of 10 dual-branch EfficientNet (DB-EffiNet) was applied to classify the DRFs into intra- or extra-articular fracture. The DB-EffiNet is a novel adaptation of the EfficientNet [2] constructed in this study, which consists of two EfficientNet-b0 branches taking PA and lateral view X-rays as input respectively. The two branches were fused at the last linear layer via summation, followed by an additional linear layer to generate the final classification output.

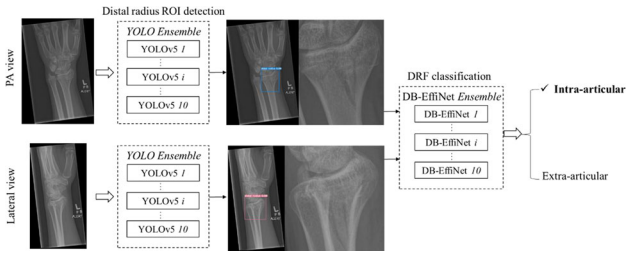

Fig. 1 Diagram of the DL-based DRF classification framework using both PA and lateral view X-rays. DB-EffiNet stands for dual-branch EfficientNet

The dataset used for evaluating the DL framework contains 302 cases of clinically retrieved wrist X-rays. The dataset was randomly split into a training set of 251 cases with 257 fractures and a testing set of 51 cases with 52 fractures. There are 193 and 38 intra-articular DRFs in training and testing set respectively. The training set was randomly partitioned into 10 folds for cross-validation.

For distal radius ROI detection, the YOLOv5s variant was trained on the PA and lateral view X-rays separately for 100 epochs within tenfold cross-validation, generating 10 YOLO base models for each view. The batch size was set as 8 and image size as 1280 × 1280, with stochastic gradient descent (SGD) used as the optimizer. Translation, scaling, horizontal flip and mosaic augmentations were adopted during training. For DRF classification, the DB-EffiNet was trained on the PA-lateral ROI pairs for 50 epochs within tenfold cross-validation. Each EfficientNet-b0 branch was pre-loaded with ImageNet pretrained weights. The ROI images were resized to 256 × 256 and normalized to the mean and standard deviation of ImageNet. The Adam optimizer was used with a learning rate of 0.0001 and the batch size was set as 16. Horizontal flip, rotation and brightness adjustment based on Power–Law transformations were used for augmentation during training. The model with the best area under the receiver operating characteristic curve (AUROC) on the validation set was saved as the base model in each cross-validation iteration.

Given an unseen testing instance, the distal radius ROI was detected on each view through merging the 10 YOLO base models by enabling the model ensemble feature of YOLOv5. The ROIs on the PA and lateral views were then passed into the 10 DB-EffiNet base models. The ensemble probability was computed by averaging the probabilities across all base models.

Results

When evaluated on the testing data, the YOLO ensemble model successfully detected all distal radius ROIs on PA and lateral view X-rays with no false positives. As for differentiating intra- from extra-articular DRFs, the DB-EffiNet ensemble model achieved an AUROC of 0.90, an accuracy of 0.87, a sensitivity of 0.87 and a specificity of 0.86.

Conclusion

This study proposed a dual-view DL framework for automatic DRF classification on wrist radiography. Evaluated on a clinically acquired wrist X-ray dataset, this framework attained promising performance in identifying intra- and extra-articular DRFs using both PA and lateral view X-rays, which demonstrated its potential to assist clinicians in fracture pattern analysis and further treatment planning.

References

[1] https://github.com/ultralytics/yolov5
[2] Tan M, Le Q. (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning. pp. 6105–6114.

Searching for infiltrative brain tumor using qMRI and deep learning model explainability

I. E. Tampu[1,2], I. Blystad[2,3], N. Haj-Hosseini[1,2], A. Eklund[1,2,4]

[1]Linköping University, Department of Biomedical Engineering, Linköping, Sweden [2]Linköping University, Center for Medical Image Science and Visualization (CMIV), Linköping, Sweden [3]Linköping University, Department of Radiology and Department of Health, Medicine and Caring Sciences, Linköping, Sweden [4]Linköping University, Division of Statistics & Machine Learning, Department of Computer and Information Science, Linköping, Sweden

Keywords quantitative MRI, deep learning, model explainability, brain tumor.

Purpose

Gliomas are malignant brain tumors known for their infiltrative nature. Identification and subsequent targeting of all active tumor regions plays an important role in improving patient survival rate. However, the active infiltrative tumor region is invisible to the eyes of radiologists in the conventional magnetic resonance images (cMRI) [1]. This work aims at investigating the possibility of using deep learning model explainability methods to identify infiltrative tumor regions from models trained on quantitative MRI (qMRI) data.

Methods

Dataset

qMRI data [2], measuring R1 and R2 relaxation and proton density (PD) of brain tissue, from 23 subjects diagnosed with glioma was included in this ethically approved study. Pre- and post-contrast gadolinium data was available for each subject along with the manual annotation of the tumor structure obtained from cMRI images. Preprocessing of the qMRI data included skull stripping, per-subject registration to the conventional T1w-post-gadolinium volume, and per-modality normalization using min–max normalization. Registration to the T1w post-gadolinium volume was done to match the qMRI data to the tumor annotation. The final volume resolution was $512 \times 512 \times [24, 26]$ voxels in the x, y and z directions, respectively. Transversal slices (n = 528 of which 136 showing tumor) were used for model training instead of volumetric data given the small number of subjects and the higher spatial resolution in the x and y directions.

Deep learning model

Transversal slices from the pre- and post-contrast qMRI volumes were used to train a shallow multi-channel 2D convolutional neural network (CNN) model for the task of tumor detection (presence or not of tumor in the slice). The input to the model were six-channel 512×512 images (R1, R2 and PD, pre- and post-contrast). The model was composed of a CNN image encoder with 4 convolutional blocks and a fully connected network, with two dense layers, as classifier. Each convolutional block was composed of two consecutive 2D convolutions (kernel_size = 3×3) each followed by batch normalization and ReLU activation. The number of filters for the 2D convolutions was constant and set for each block. Here, 64, 128, 256 and 512 were used. Between each convolutional block, a 2D maxPool operation was used (pool_size = 2, stride = 2). The output of the last convolutional block was flattened using a GlobalAveragePooling and fed to the classifier part of the model. The first fully connected layer contained 128 nodes and used the ReLU activation function. The output layer had two nodes and used the softmax activation function, returning the probability of the input belonging to each of the classes.The model was trained using binary cross entropy loss using the LookAhead optimizer with Adam inner optimizer. Learning rate was set to $1e-6$ and kept constant during training which was carried out for 300 epochs without early stopping. Models were trained using a ten-times repeated five-fold cross validation scheme and evaluated using Matthew correlation coefficient (MCC), which has been shown to be a better metric for unbalanced datasets. The benefit of using a repeated cross validation scheme is twofold: (1) it reduces the effect of sampling bias of the test set given the small dataset, and (2) it allows a relatively large number of models which can be used to obtain reliable model explainability results.

Model explainability

Occlusion mapping was used as post-hoc model explainability method, where patches of the image size 5×5 were iteratively set to background (value = -1) and the difference in model predicted probability between the un-occluded image and the occluded image was computed and used to obtain a relevance map. Considering the tumor presence class, high values in the relevance map indicate regions in the image important for the detection of the tumor. The relevance maps obtained from each of the models trained through the repeated cross validation scheme were averaged and thresholded (th = 0.03) to obtain a model explainability-derived mask of the brain tissue relevant for tumor detection. Using the manual annotation of the tumor, the relevant regions within and outside the annotation could be identified and the change in relaxation values after contrast injection could be investigated. In particular, the contrast agent deposits in those brain regions where active tumor is present, increasing the tissue relaxation values.

Results

Tumor detection could be achieved with an MCC of $0.72 +\ -0.13$ (accuracy $0.92 +\ -0.06$). From the model explainability analysis, the occlusion relevance maps highlight regions important for the tumor detection that match well with the annotation of the tumor structure. A representative example is shown in Fig. 1. Analysis of the relaxation values pre- and post-contrast for the relevant regions inside and outside the manual annotation show that the positive shift in relaxation values seen for regions outside the annotation is comparable to that of the regions inside the annotation and which are visible as tumors in the conventional images. The distribution of R1 and R2 relaxation inside and outside the annotation, and the pixel-wise difference pre and post contrast are shown as violin plots in Fig. 1.

Conclusion

Model explainability analysis of models trained on qMRI data highlight brain regions relevant for the detection of the tumor which are adjacent to the manual annotation of the tumor visible in the conventional images. The shift in relaxation values after contrast injection of these regions, similar to that of regions inside the manual

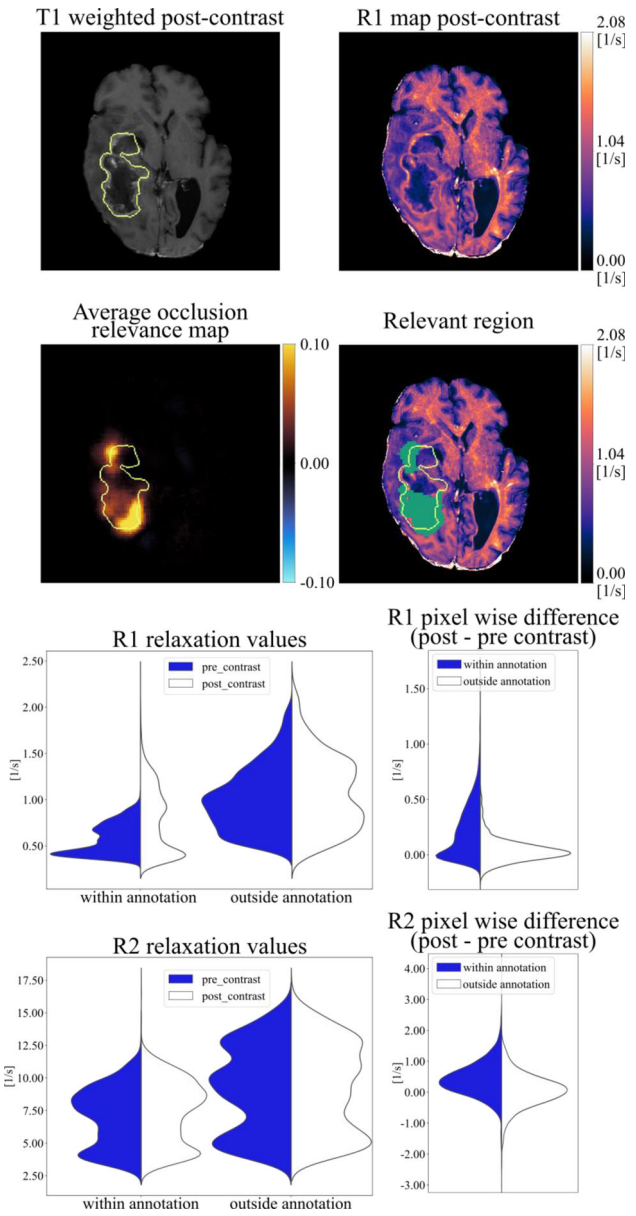

Fig. 1 Post-contrast T1w and quantitative R1 of representative transversal with corresponding raw and thresholded occlusion relevance map. Tumor annotation is shown as a yellow contour. The distribution of R1 and R2 quantitative values inside and outside the tumor annotation, and pre- and post-contrast are shown as violin plots. The shift in quantitative values after contrast injection for the relevant regions inside and outside the annotation is also presented

annotation, suggest the presence of infiltrative tumor tissue in the immediate region outside the visible tumor. Further analysis is needed, however, to include a larger number of subjects and validate the identified regions with image guided biopsies.

References

[1] Konukoglu, E., Clatz, O., Bondiau, P. Y., Delingette, H., & Ayache, N. (2010). Extrapolating glioma invasion margin in brain magnetic resonance images: Suggesting new irradiation margins. Medical image analysis, 14(2), 111–125.

[2] Gurney-Champion, O. J., Mahmood, F., van Schie, M., Julian, R., George, B., Philippens, M. E., & Redalen, K. R. (2020). Quantitative imaging for radiotherapy purposes. Radiotherapy and Oncology, 146, 66–75.

Performance evaluation of self-supervised 3D GAN for electronic cleansing in photon-counting CT colonography

R. Tachibana[1,2], J. Näppi[2], T. Hironaka[3], S. Yoshida[3], D. Wu[2], R. Gupta[2], K. Taguchi[4], H. Yoshida[2]

[1]National Institute of Technology, Oshima College, Oshima, Japan [2]Harvard Medical School/Massachusetts General Hospital, Boston, United States [3]Massachusetts General Hospital, Boston, United States [4]John Hopkins University, Baltimore, United States

Keywords CT colonography, photon-counting CT, electronic cleansing, self-supervised learning.

Purpose

Colorectal cancer is one of the leading causes of cancer deaths worldwide, although it would be preventable by early detection and removal of the benign precursor polyps 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. Because some of the polyps could be missed due to being covered by residual materials in the colon, electronic cleansing (EC) is used to perform virtual subtraction of the residual materials from the CTC images to enable radiologists and computer-aided detection (CADe) systems to detect all polyps. Previously, we developed an EC scheme based on self-supervised artificial intelligence (AI), where some of the training samples were self-generated by use of a self-supervised 3D generative adversarial network (GAN) [1]. The 3D-GAN EC was trained to directly transform an uncleansed CTC volume into the corresponding virtually cleansed image volume.

However, CTC screening has been criticized because the examination involves a small radiation dose. Photon-counting CT is an emerging technology that has the potential to minimize the radiation exposure of CT while also being able to reconstruct images at a higher spatial resolution than current CT technology. In this study, we evaluated the effect of the use of photon-counting CTC on the performance of our self-supervised 3D-GAN EC scheme based on a phantom study and a simulated clinical CTC study.

Methods

For the phantom study, an anthropomorphic phantom (Phantom Laboratory, Salem, NY) designed to imitate the appearance of the human colon in CT scans was filled with 300 ml 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 photon-counting CT scanner (OmniTom PCD, NeuroLogica, Boston, MA, USA) in a single-energy mode with 120 kVp, 40 mAs, and 0.707-mm slice thickness, as well as by use of a conventional dual-energy CT scanner (SOMATOM Definition Flash, Siemens Healthcare, Erlangen, Germany) in a single-energy mode with 120 kVp, 41mAs, 0.6-mm collimation, and 0.6-mm

reconstruction interval. After the CT scans, the CTC image volumes were registered to match spatially at each voxel by use of a rigid 3D registration method (VersorRigid3DTransform) in the Insight Toolkit.

For the simulated clinical CTC study, we first acquired a clinical CTC patient case with reduced cathartic preparation, where the CT image volumes were scanned by use of the conventional CT scanner and reconstructed at 120 kVp, 1.0-mm slice thickness, 0.627 mm pixel spacing, and 0.7-mm reconstruction interval. The corresponding photon-counting CTC was then simulated based on the acquired CT image volumes by use of the photon counting toolkit [2].

The 3D GAN of the EC scheme has a 3D generator (3D U-Net) and a 3D discriminator network. The 3D generator has six down-/up-convolution layers. 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, where the CT dataset of the corresponding native phantom acted as the ground truth. The statistical significances of the differences of the PSNRs between the supervised and self-supervised versions of the 3D-GAN EC images on single-energy CTC and photon-counting CTC were assessed by use of the t-test with Bonferroni correction. The image quality of EC on the simulated clinical photon-counting CTC study was evaluated by visual comparison to the corresponding conventional CTC case.

Results

Figure 1 shows the boxplots of the PSNRs of the phantom study over the 100 test VOIs from the single-energy and photon-counting CTC cases. The differences between the supervised and self-supervised versions of 3D-GAN EC between the photon-counting CTC and single-energy CTC were all statistically significant ($p < 1e-6$).

Figure 2 shows a visual comparison of the EC results on a 2D image and the corresponding virtual endoscopic image of the clinical CTC study. As shown in these images, the use of simulated photon-counting CTC with the 3D-GAN EC scheme generated higher-quality EC images than those obtained by use of conventional CTC.

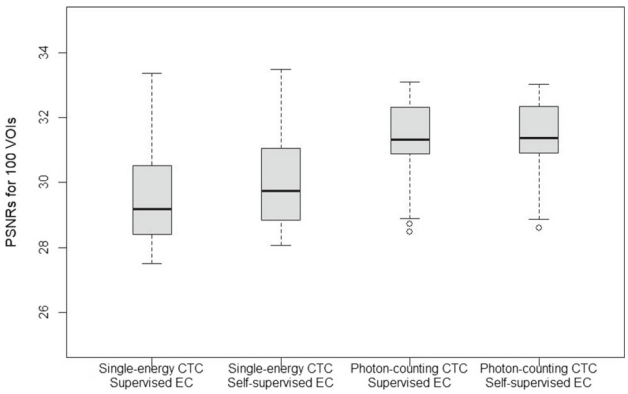

Fig. 1 Comparison of the PSNRs of the 100 electronically cleansed VOIs by the supervised and self-supervised 3D-GAN EC schemes in conventional single-energy and photon-counting CTC

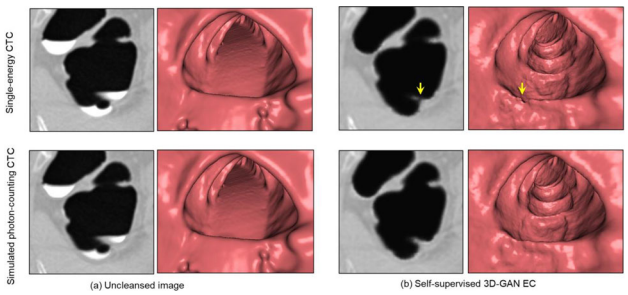

Fig. 2 An example of the performance of EC in conventional CTC (top row) and the simulated photon-counting CTC (bottom row) in a clinical case. **a** Comparison of the original uncleansed axial 2D CT and the corresponding virtual endoscopic images. **b** Comparison of the corresponding 2D and virtual endoscopic images cleansed by use of the self-supervised 3D-GAN EC. The yellow arrow indicates the location of an EC artifact in conventional CTC

Conclusion

We evaluated the effect of the use of photon-counting CTC on the performance of our self-supervised 3D-GAN EC scheme. Our preliminary results indicate that photon-counting CTC can be employed to generate higher-quality EC images than those obtained by use of conventional CTC at a comparable radiation dose. Thus, photon-counting CTC in combination with our self-supervised 3D-GAN EC scheme is expected to provide EC images of the highest quality in low-dose fecal-tagging CTC.

References

[1] Tachibana R, Näppi JJ, Hironaka T, Yoshida H (2022) Self-supervised adversarial learning with a limited dataset for electronic cleansing in computed tomographic colonography: a preliminary feasibility study. Cancers (Basel) 14(17):4125

[2] Taguchi K, Polster C, Segars WP, Aygun Nafi, Stierstorfer K (2022) Model-based pulse pileup and charge sharing compensation for photon counting detectors: A simulation study. Medical Physics 49(8):5038–5051

28th Computed Maxillofacial Imaging Congress

Chairman: Christos Angelopoulos, DDS (US),
Co-Chair: Yoshihiko Hayakawa, PhD (JP)

Orbital reconstruction with Gaussian processes

P. Vanslambrouck[1,2], R. Willaert[1,3], M. Bila[1,3], C. Politis[1,3], P. Claes[4,5], J. Van Dessel[1,3], Y. Sun[1,3]

[1]University Hospitals Leuven, Campus St Rafael, oral and maxillofacial surgery, Leuven, Belgium [2]KU Leuven, Department of Computer Science, Leuven, Belgium [3]KU Leuven, Department of Imaging & Pathology, Leuven, Belgium [4]KU Leuven, Department of Electrical Engineering (ESAT), Leuven, Belgium [5]KU Leuven, Department of Human Genetics, Leuven, Belgium

Keywords statistic shape modelling, Gaussian processes, orbital reconstruction, outlier detection.

Purpose

Orbital defects make up a substantial part of maxillofacial trauma cases. To achieve an optimal functional and esthetic outcome, accurate reconstruction of the original bone level is required. The standard method for reconstruction is mirroring the healthy orbital to the defective side. However, this method disregards the natural asymmetry of the human skull and cannot be applied to bilateral defects. It is impossible to build a robust workflow for orbital reconstruction via mirroring. The purpose of this study is to apply statistical shape modeling to reconstruct defects in the orbital region.

Methods

In our study, 50 healthy (bilateral) orbital regions were first segmented using the BrianLab iPlan software. The results are visually inspected by an experienced engineer and the results are saved in STL format as training shapes.

In the first step, a reference shape is selected randomly from the 50 healthy shapes. By applying kernel functions that enforce smooth and predominantly symmetric deformations on the reference shape, a prior model that favors realistic deformations is obtained. An Iterative Closest Points (ICP) algorithm based on Gaussian Process Morphable Models (GPMMs) [1] is applied to register the reference shape to the other 49 healthy orbital shapes. In this step, non-rigid registration is performed to find the correspondence between each point on the reference shape and the other 49 training shapes. Validation of the parameters of this prior model is done by estimating the generalization and specificity metrics using the training dataset, which is traditionally used for the validation of SSMs. The trade-off between these metrics means a compromise between flexibility of the model and plausibility of the deformations is necessary. Each iteration step of the ICP algorithm consists of a first step of finding corresponding points for each point of the model and a second step of computing a regularized transformation corresponding to the observed correspondences. To find the regularized transformation, the prior model is fitted to the identified point correspondences by computing the mean of the regularized Maximum A Posteriori estimate [2].

In the second step, after building up the point correspondence within the 50 training models, Principal Component Analysis (PCA) is applied to create a statistical model of the deformations. However, because of the small number of degrees of freedom of this PCA-model, the model is augmented with a custom kernel of Gaussian process. This augmentation provides more flexibility to the SSM and improves generalization to new shapes. However, it also introduces deformations that are not present in the training set, i.e. the specificity is higher. Hence, augmentation of the PCA-model is subject to the same trade-off as the prior model used in step 1 for non-rigid registration and the validation of the parameters of the augmentation is performed similarly.

In the third step, the resulting SSM model created in step 2 is used to reconstruct the defective orbital shape by applying an ICP algorithm that iteratively fits the model to a defective shape. However, in this case some of the points from SSM can not have a valid projection on the defective orbital model. Therefore, an outlier detection algorithm is developed to automatically identify those points on the SSM model during the ICP registration procedure.

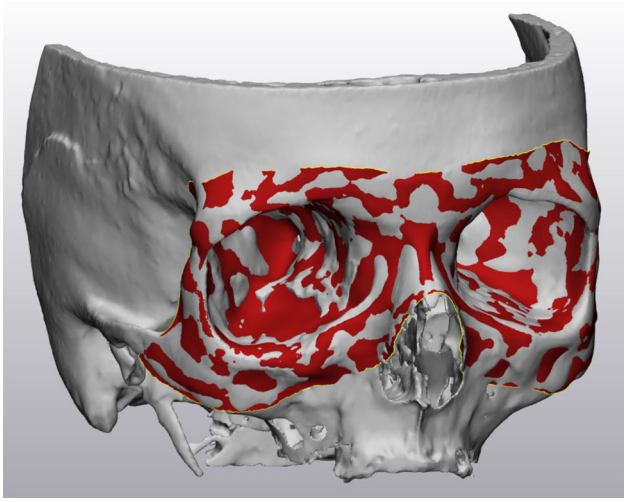

Fig. 1 Reconstruction of a defect in the orbital floor using a PCA-based SSM augmented with additional modes from a Gaussian process with a Gaussian kernel. The reconstruction is computed iteratively by means of an Iterative Closest Points algorithm with outlier detection

The outlier detection is performed based on the distance between the two points, the surface orientation of both points and the orientation of the deformation vector with respect to the surface normal of the mesh of the model. The outlier score is computed by applying robust statistics on these metrics to identify anomalous point pairs. Concretely, an inlier score is attached to each pair of corresponding points and points with a low score are downweighted in the fitting step by increasing the regularization noise parameter [2].

Results

The registration algorithm based on Gaussian Process Morphable Models achieves a Root Mean Square Error (RMSE) of 0.2 mm on average within the 50 training models, but it should be noted that a lower number is not always desired in this case. The RMSE can be made arbitrarily small by choosing different parameters for the prior model, but this disregards the specificity and would consequently ruin the resulting SSM. Additionally, an accuracy of 0.2 mm is sufficiently low compared to the resolution of the training data.

An SSM is computed based on the meshes in point correspondence. By augmenting the PCA-model, generalization RMSE is reduced from 0.8 mm to 0.3 mm without significant degradation of the specificity of the model.

Finally, reconstruction is performed by using the created SSM with outlier detection. An example result is shown in Fig. 1. Depending on the size of the defect, the maximal reconstruction error (the surface distance) ranges from 1.0 to 0.3 mm.

Conclusion

In summary, a registration algorithm based on Gaussian Process Morphable Models (GPMMs) was developed. Instead of using a traditional PCA model, the model is augmented with additional GPMM modes to obtain better generalization. The resulting SSM is used for reconstruction by applying an ICP algorithm with integrated outlier detection to identify the defective regions. Quantitative results prove that this pipeline is a promising alternative to current reconstruction approaches. Further study is required to analyze the factors that influence the accuracy of reconstruction results.

References

[1] Lüthi, M., Gerig, T., Jud, C., & Vetter, T. (2017). Gaussian process morphable models. IEEE transactions on pattern analysis and machine intelligence, 40(8):1860–1873.

[2] Albrecht, T., Lüthi, M., Gerig, T., & Vetter, T. (2013). Posterior shape models. Medical image analysis, 17(8):959–973.

AI-based characterization of partially edentulous jaws in panoramic x-rays

H. Meine[1], L. Brandenburg[2], M. Detering[1,3], P. Weingart[2], M. C. Metzger[2], J. Georgii[1]

[1]Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany [2]Medical Center – University of Freiburg, Center for Dental Medicine, Department of Oral and Maxillofacial Surgery, Freiburg, Germany [3]Fraunhofer MEVIS, Bremen, CÔte D'ivoire

Keywords detection, tooth, classification, deep learning.

Purpose

Panoramic x-rays are used as a basic diagnostic tool for initial assessment of patients in dental and maxillofacial medicine as they depict the full set of teeth and its adjacent bones in one image. Especially for prosthodontic planning purposes the detection and location of abutment teeth, implants and pre-existing restorations is vital. The goal of this work is to provide methods that characterize the prosthodontic value of the jaws. At its core, it requires detection and segmentation of regions with teeth or implants. Challenges in the automatization are missing teeth or pontics. We propose an approach based on Mask R-CNN (regions with convolutional neural networks) that detects regions with healthy teeth, teeth with fixed dental prostheses and implants. The approach is designed to predict further parameters following the classification of Kennedy in the future.

Methods

For the study, 209 panoramic x-ray images from 206 patients from the Clinic for Oral and Maxillofacial Surgery of the University of Freiburg, Germany have been employed. Using the MeVisLab-based software platform SATORI [1], the images were manually annotated, contouring the available tooth regions and labeling them with the respective type (tooth vs. implant). The dataset has been randomly split into five equally sized folds for cross validation. We trained five Mask R-CNN models on three folds, leaving out one fold each as validation set (for convergence detection) and one as test set. Learning rate was initially 10–3 and was halved after 25 epochs without improvement on the validation set. Training was stopped after five such reductions (on average, about two hours). In order to prevent overfitting, we applied L2 regularization (with a weight of 5·10–5) and selected the model with the best validation score for application. For validation and testing, we assessed detection quality via the standard AP_{50} score, which gives the average precision of detection (0 to 1, corresponding to the area under the precision-recall curve), where an object has to have at least 50% geometric overlap (IoU) with the reference.

Results

The above training setup led to good convergence, and we found the Mask R-CNN architecture to be able to robustly detect teeth and implants, with 78% of the test cases getting an AP_{50} score of 90% or higher (median 93.8%, cf. Figure 1). Contrary to shape model-based approaches [2], our deep learning approach is able to cope well with missing teeth, pontics, or implants. Accordingly, even the segmentation quality is relatively high (median Dice similarity coefficient (DSC) of 93.5%,, 91% of cases with a Dice above 90%), although our dataset showed severe pathologies. There are a few cases with overlapping duplicate detections, in which case the correct detection usually has the highest score. Other, more seldom failures include incorrect teeth detected far from the jawline (note, however, that the model had to and did correctly learn to detect unerupted wisdom teeth within the jaw).

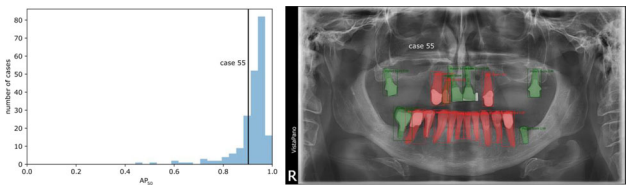

Fig. 1 Distribution of AP_{50} detection performance on all 219 cases (computed with five distinct models in a cross-validation manner, not used during the respective training), with an example case highlighted (*red:* detected teeth, *green:* detected implants)

Conclusion

We presented a method for assessment and prosthodontic characterization of the jaws. The method shows good performance on the data from one site. We plan to incorporate further parameters like fillings/inlays/onlays, root canal treatment, root resection, caries or bone loss into the approach, thereby enabling an automatic classification of Kennedy. A dedicated preprocessing step should be able to address many of the observed errors. Additionally, data from multiple sites has to be included to increase the model robustness.

References

[1] J. Klein, M. Wenzel, D. Romberg, A. Köhn, P. Kohlmann, F. Link, A. Hänsch, V. Dicken, R. Stein, J. Haase, A. Schreiber, R. Kasan, H. Hahn, and H. Meine, "QuantMed: Component-based deep learning platform for translational research," in Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications, vol. 11,318, pp. 229–236, International Society for Optics and Photonics, SPIE, 2020.

[2] Wirtz, A., Mirashi, S.G., Wesarg, S. (2018). Automatic Teeth Segmentation in Panoramic X-Ray Images Using a Coupled Shape Model in Combination with a Neural Network. Proc. MICCAI 2018. LNCS 11,073. Springer, Cham.

Incremental Biomechanical Deep Learning Modeling of Facial Tissue Deformation

N. Lampen[1], D. Kim[2], X. Xu[1], X. Fang[1], T. Kuang[2], H. H. Deng[2], M. A. K. Liebschner[3], J. J. Xia[2,4], J. Gateno[2,4], P. Yan[1]

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

Keywords Deep learning, Finite element method, Orthognathic surgery, Soft-tissue change prediction.

Purpose

Biomechanical modeling of facial tissue deformation is the most popular and accurate facial change prediction method that is essential in surgical planning for orthognathic surgical procedures [1]. Currently, most biomechanical simulations are performed using the Finite Element Method (FEM) [1], which uses numerical methods to approximate a solution to a physical problem described by a boundary condition using discretization. To further improve the accuracy of FEM, previous studies divided FEM simulations into smaller incremental steps [1]. While breaking the simulation into incremental steps increased accuracy, it also greatly increased computation time, with some simulations taking up to 30 min [1]. Recently, deep learning (DL) methods have been proposed as an alternative to FEM for simulating many physical phenomena, including tissue deformation. While DL methods require large amounts of time and data to train, once a network is trained, simulations can be performed extremely quickly, often within seconds. However, previous DL-based methods have been trained only using a non-incremental approach, i.e., all input movement is applied in a single step rather than incrementally, thus limiting accuracy. We argue that such non-incremental training limits the deep neural network by (1) forcing the network to be adaptable to inconsistent magnitudes of deformation and (2) preventing the network from learning from temporal trends in the incremental simulations. To address these problems, we propose a method for incremental biomechanical modeling, where multiple deep neural networks are trained to perform incremental simulations for different magnitudes of maximum deformation to create the optimal balance between the speed of DL and the accuracy of using an incremental approach. Each deep neural network is trained using incremental FEM data to be optimized for a specific magnitude of maximum deformation.

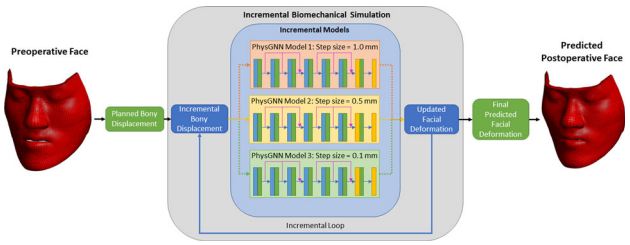

Fig. 1 Incremental biomechanical modeling of facial deformation using several deep neural networks trained for specific magnitudes of maximum deformation

Methods

A dataset of FEM simulations was generated by collecting 19 subjects who underwent double-jaw orthognathic surgery from our digital archive. The FEM simulations were performed using the standard clinical approach, where the movement of bony segments is retrospectively planned and then used as input deformation in the FEM simulation of the soft tissues of the face [1]. The FEM method discretizes the total input deformation into many increments adaptively, choosing the step size of each increment adaptively to arrive at a stable solution. Because the increments are chosen adaptively, the maximum deformation of each increment may be unique. From the FEM simulation for a given subject, a dataset of many "sub-simulations" was created. Each sub-simulation can be seen as a subset of the original simulation increments with a maximum deformation threshold T between the neighboring increments.

All FEM simulations were performed on a mesh with 3960 nodes. A PhysGNN [2] network was utilized to predict the mesh deformation based on the input node features, which consisted of the planned bony displacement (surgical plan) at the interface between the moving bony surface and the soft tissue. Nodes at the interface of a fixed bony segment were fed a feature vector of zeros, as were all other nodes where the input displacement was unknown. Adjacency matrices were generated for each subject and the edge weights were computed to be the distance between neighboring nodes. The planned bony displacement was broken into incremental bony displacements based on the maximum deformation threshold for the given network. The output of the network was the updated facial deformation, which was then fed back to the network in an incremental feedback loop, where another incremental bony displacement step was applied. In this way, the final predicted facial deformation was simulated by applying many incremental steps, Fig. 1.

We separately trained three networks using sub-simulations with maximum deformation threshold $T\{0.1, 0.5, 1.0\}mm$, respectively. Each network was trained in a leave-one-out cross-validation on a total of 19 subjects (Table 1). The networks were trained for 100 epochs, then tested on the left-out subject. Each network trained on a different deformation threshold was evaluated on the test independently.

Results

The incremental simulation approach was extremely accurate, with all three incremental models achieving a mean error of less than 1 mm on most subjects. The prediction accuracy was mostly consistent between the different maximum deformation thresholds. However, the best performing model changed from subject to subject.

Table 1 The mean Euclidean distance error [mm] on incremental models across subjects

Subject	1	2	3	4	5	6	7	8	9	10
Model										
1 mm	0.486	0.514	0.604	0.632	0.574	0.272	0.490	0.925	0.610	0.521
0.5 mm	0.537	0.488	0.346	0.508	0.731	0.308	0.410	0.970	0.476	0.423
0.1 mm	0.364	0.558	0.316	0.514	0.653	0.376	0.434	1.105	0.488	0.602
Subject	11	12	13	14	15	16	17	18	19	
Model										
1 mm	0.303	0.602	0.471	0.413	0.594	0.633	0.629	0.506	0.817	
0.5 mm	0.308	0.578	0.461	0.369	1.100	0.438	0.490	0.414	0.905	
0.1 mm	0.329	0.642	0.650	0.419	1.078	0.491	0.433	0.595	1.103	

Conclusion

The proposed incremental simulation approach can be used to simulate facial tissue deformation accurately and efficiently. Interestingly, the 0.1 mm deformation model was not always the most accurate, suggesting that there may be a trade-off between increment size and the size of the training dataset. Our proposed method of having multiple incremental models trained for several deformation magnitudes allows for flexibility when optimizing a simulation for facial deformation prediction.

References

[1] D. Kim, Kuang T, Rodrigues YL, Gateno J, Shen SGF, Wang X, Stein K, Deng HH, Liebschner AAK, Xia JJ (2021) A novel incremental simulation of facial changes following orthognathic surgery using FEM with realistic lip sliding effect," *Med. Image Anal.*, vol. 72, p. 102,095, Aug. 2021, https://doi.org/10.1016/J.MEDIA.2021.102095.

[2] Salehi Y, Giannacopoulos D (2021) PhysGNN: A Physics-Driven Graph Neural Network Based Model for Predicting Soft Tissue Deformation in Image-Guided Neurosurgery. Sep. 2021, https://doi.org/10.48550/arxiv.2109.04352.

Mandible landmark detection for craniomaxillofacial surgical planning using graph convolutional network

X. Xu[1], T. Kuang[2], A. Chen[2], N. Lampen[1], X. Fang[1], D. Kim[2], M. A. K. Liebschner[3], J. J. Xia[2,4], J. Gateno[2,4], H. H. Deng[2], P. Yan[1]

[1]Rensselaer Polytechnic Institute, Department of Biomedical Engineering, Troy, United States [2]Houston Methodist Research Institute, Department of Oral and Maxillofacial Surgery, Houston, United States [3]Baylor College of Medicine, Department of Neurosurgery, Houston, United States [4]Cornell University, Department of Surgery (Oral and Maxillofacial Surgery), New York, United States

Keywords Mandible landmark detection, Craniomaxillofacial surgery, Graph convolutional network, Deep learning.

Purpose

Mandible landmark detection (digitization) is a critical prerequisite for craniomaxillofacial (CMF) surgical planning. By locating a set of landmarks on the (pre-segmented) 3D mandible models, the bone can be cut into several individual pieces for planning and adjustment. In clinical practice, mandible landmark detection is manually accomplished by physicians, which is time-consuming and labor-intensive. Therefore, automatic landmark detection methods are highly desired. Benefitting from the development of deep learning (DL) in recent years, convolutional neural networks (CNNs) are widely used for various object detection tasks. However, when it comes to the mandible landmark detection, this methodology exhibits two limitations. First, since conventional CNN can only process grid-structured data (e.g., matrix and image), they can only take input from the CBCT images. This makes it hard to utilize the geometric information from the 3D mandible mesh, which critically affected the position of the landmarks. Second, CNN has no explicit constraint on the output to ensure the detected landmarks are attached on the surface of the 3D mandible mesh. This often leads to suboptimal results requiring further refinement. To address the above two issues and achieve better performance in mandible landmark detection, in this study, we propose a deep learning-based method using graph convolutional network (GCN) [1] combined with CNN.

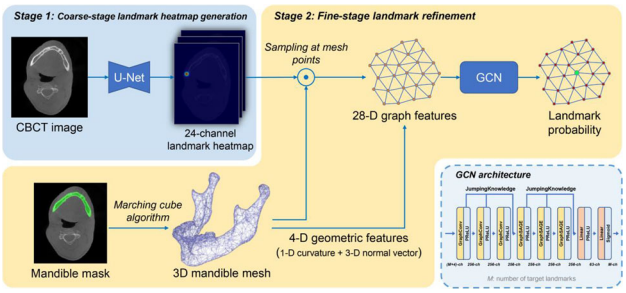

Fig. 1 Overview of the proposed mandible landmark detection method for craniomaxillofacial surgical planning

Methods

Our method has two stages in a coarse-to-fine fashion shown in Fig. 1. The core idea is to utilize a GCN to extract the geometric information from the 3D mandible mesh to complement the CBCT image information for landmark detection.

1) U-Net for coarse-stage landmark heatmap generation: We first employ a U-Net [2] to roughly locate the landmarks. The input and output channels of the U-Net are 1 and M, respectively, indicating the single-channel CBCT image and the M-channel heatmap for M mandible landmarks. Each channel of the heatmap has the same size as the CBCT image, and a gaussian-shaped kernel is located at the pixel nearest to the landmark. The U-Net is trained to predict such an M-channel heatmap for the landmarks given an CBCT image.

2) GCN for fine-stage landmark refinement: Although the landmarks can be roughly located by finding the maximums in above predicted heatmaps, they are not necessarily to be attached to the surface of the 3D mandible model, which is undesired for CMF surgical planning. Therefore, we incorporate a GCN in our method to refine the landmark position. Specifically, we first convert the pixel-wise mandible segmentation mask to a 3D triangle mesh via the marching cube algorithm. Since the 3D mesh can be seen as a graph with the mesh points treated as graph vertices, the problem of locating a landmark on a mesh surface is equivalent to finding out the most likely vertex from a given graph to be a landmark. This process can be naturally implemented by GCN, which is specialized in processing graph data. In this study, we build a 6-layer GCN to achieve the purpose. The input of this GCN is a graph containing N vertices as well as an $N \times N$ adjacent matrix to describe the edges. Each vertex has an $(M + 4)$-dimensional feature vector to represent the information associated with that mesh point (M is the number of landmarks). The first M-channel of this feature vector comes from the M-channel heatmap yielded from the first stage, in which we took samples at the mesh point. The last 4-channel is the geometric features of that mesh point, including a 1D curvature and 3D normal vector. The output graph of the GCN has the same number of vertices but each vertex is associated with an M-dimensional vector, which indicates the probabilities of each mesh point being one of the M mandible landmarks.

Results

We conducted experiments using a clinical dataset containing 56 subjects. Each subject included a whole head CBCT image, a mandible segmentation mask, and $M = 24$ mandible landmarks (represented as 3D coordinates in the same space as the CBCT image). We evenly divided the dataset into four parts and performed fourfold cross-validation (three/one folds for training/testing respectively) for evaluation. Distance error from the predicted landmark to the ground-truth landmark was used as the metric to quantitatively assess the model performance.

Table 1 Distance error of detected landmarks generated by different methods

Models	Average distance error [mean ± SD mm]		
	18 bilateral landmarks	6 non-bilateral landmarks	All 24 landmarks
U-Net	2.62 ± 1.73	2.16 ± 1.59	2.50 ± 1.71
U-Net refined by bone mesh attaching	2.47 ± 1.84	**1.88 ± 1.77**	2.32 ± 1.84
GCN	**2.36 ± 1.81**	2.25 ± 1.62	2.33 ± 1.77
GCN + U-Net (proposed)	**2.36 ± 1.81**	**1.88 ± 1.77**	**2.24 ± 1.81**

We compared landmark detection results of four methods in Table 1. The first method showed the baseline result of a U-Net trained to predict the landmark heatmap merely using CBCT image as input. The average error over 24 landmarks is 2.50 ± 1.71 mm. However, this model was applied on CBCT and thus did not have constraints on detected landmarks to ensure they are attached to the bone surface. Therefore, in the second method, we refined the result of U-Net by directly attaching the predicted landmarks to the 3D bone mesh surface. This refinement yielded a slightly lower distance error of 2.32 ± 1.84 mm, indicating a positive effect of incorporating the 3D bone mesh into the detection framework. The third method showed the result of GCN, which can explicitly leverage the geometric information from bone mesh to detect landmark. Although the global average error of 2.33 ± 1.77 mm is close to that of the refined U-Net, we argue that the GCN did a better detection on the 18 bilateral landmarks with more discriminative geometric features (e.g., distributed on the ridge of the mesh. Therefore, in the fourth method, we merged the GCN's prediction on the bilateral landmarks with the U-Net's prediction on the rest landmarks, which yielded the lowest error shown in the last row of Table 1.

Conclusion

In this study, we proposed a mandible landmark detection method for CMF surgical planning using GCN combined with CNN. By incorporating a GCN into the landmark detection pipeline, our method can explicitly utilize the geometric information of the 3D bone mesh to enforce the detected landmarks to be attached to the bone surface, resulting in lower detection error compared with the conventional CNN-based methods. Moreover, by merging the prediction of GCN with that of the refined CNN, the landmark detection accuracy can be further improved, demonstrating the effectiveness of the proposed design.

References

[1] Kipf T, Welling M. (2017) Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations.

[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.

Mouth and rip motion capture during speech and the development of rip-reading AI

T. Hirabayashi[1], R. Hosoya[1], Y. Hayakawa[1]

[1]KItami institute of Technology, Dept. of Engineering on Intelligent Machines & Biomechanics, Kitami, Japan

Keywords Rip-Reading, Face recognition, Machine learning, Recognition matching.

Purpose

Several kinds of man–machine interface are used at various situations, such as keyboard, mouse, game-controller, gesture capture device, speech and natural language recognition, hand-written character recognition, etc. We explored the another possibility, namely, automatic mouth and rip reading. As the application of the general object detection and motion capture and tracking, the mouth and rip motion were captured during speech and the development of rip-reading AI was tried in the study. In particular, morphological features and changes in oral region (rip and mouth) during vowel sounds and consonants were extracted. Such various images of oral cropped face (oral and surrounding tissue) region were collected as the learning data and a rip-reading AI system was tried to develop.

Methods

Our hardware and software environments for the object detection and tracking tasks are as follows; PC: MacBook Air (2018, Intel Core i5-8120Y, 8 GB RAM, mac OS Venture, in-camera 1280 × 720), Code Editor: Visual Studio Code, Language: Python 3.8.13, and Code Libraries: Media Pipe 0.8.10, Open CV 4.5.5 and NumPy 1.20.3. Also the environment for the AI development are as follows; Google Colaboratory, Language: Python 3.8.16, and Code Libraries: Tensor Flow 2.9.2 and Keras 2.2.5. Media Pipe is image processing and recognition (computer vision) library and Tensor Flow is also machine learning library, provided by Google, Inc. The face recognition worked to extract 478 feature point on face, and the oral region were cropped under a constant rule (for example, zygomatic bones on both sides are included in cropped images.) and mesh images were created for the learning data and test data as shown in Fig. 1. At first, long tone voices of three vowel sounds "a," "i" and "u" were recorded as shown in Fig. 2. Furthermore, the data of other vowel sounds "e" and "o" (There are only and simply five vowel sounds in Japanese language.) and consonants such as "ka" and "sa," etc., were added. In order to distinguish among these sounds, the learning

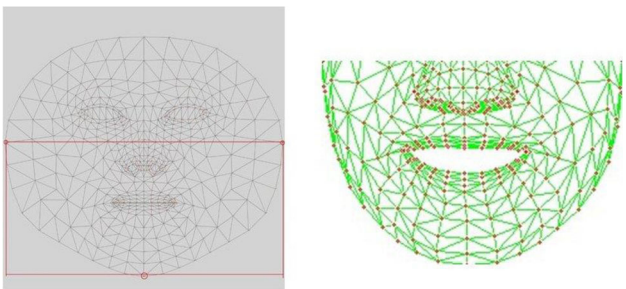

Fig. 1 Face recognition worked to extract 478 feature point, and the oral region were cropped and mesh images were created

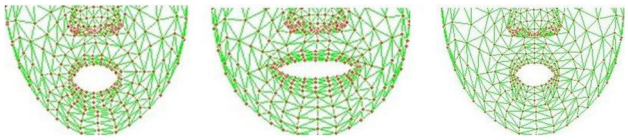

Fig. 2 Three samples images. Long tone voices of three vowel sounds "a," "i" and "u" were recorded

system was worked. The number of epochs was primarily set at 40. The classification performance for five vowel sounds and limited consonants ("ka" and "sa") was evaluated using the test data.

Results

We tried to develop three kinds of automatic rip-reading AI systems. The classifications of three classes of three vowel sounds "a," "i" and "u," showed the good performance. The accuracy is over 0.9. The classifications of five classes of five vowel sounds "a," "i," "u," "e," and "o" and the classifications of seven classes of five vowel sounds "a," "i," "u," "e," and "o," in addition, consonants such as "ka" and "sa" showed relatively lower accuracy.

Conclusion

The three-classes classification showed the satisfied performance but both five classes and seven classes classifications did not show such results. We try to optimize the number of epochs, to improve the volume and quality of the learning data. We also explore the other possibility, such as the application of the 3D projection for mesh-image creation and cropping procedures. This is thought to be so beneficial for the distinguish between "a" and "u" as observed in Fig. 2.

A feasibility study on estimating desired postoperative face using deep learning for patients with craniomaxillofacial deformities

J. Lee[1], D. Kim[2], X. Xu[1], X. Fang[1], T. Kuang[2], N. Lampen[1], H. H. Deng[2], M. A. K. Liebschner[3], J. J. Xia[2,4], J. Gateno[2,4], P. Yan[1]

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

Keywords Orthognathic surgery, Surgical Simulation, Deep Learning, Geometric analysis.

Purpose

Orthognathic surgery is a surgical procedure to correct jaw deformities by repositioning bony segments to normal positions, improving facial aesthetics. In current clinical practice, surgical planning is a bone-driven method. Surgeons plan the movement of each bony segment to achieve a normal skeleton, assuming the postoperative facial appearance is normalized following the underlying bony skeleton. However, facial deformity may remain after the surgery due to the complex and nonlinear relationship between the bone and soft tissue. In addition, bone-driven planning is clinically impractical because it requires time-consuming plan revisions [1]. The surgical plan (i.e., bony movement) needs to be iteratively revised until an ideal facial appearance is achieved by facial change simulation [2].

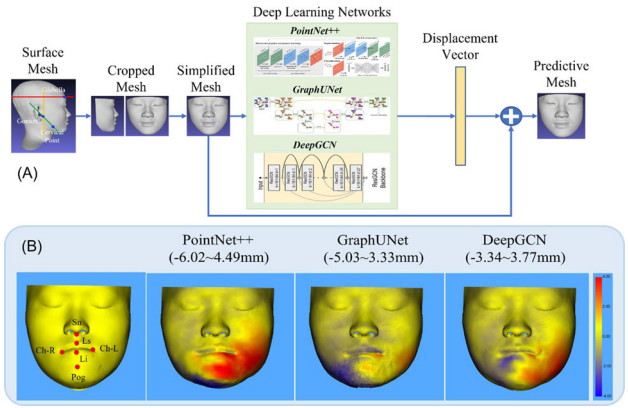

Fig. 1 **A** Overall diagram of deep learning architecture, **B** surface deviation errors between the predicted results and ground-truth

A soft-tissue-driven planning method can overcome this problem by directly estimating the surgical plan (bony movement) from the given ideal facial appearance. However, accurate estimation of the ideal appearance based on a deformed preoperative face is challenging. This is because the ideal postoperative appearance is patient-specific and must satisfy both quantitative and qualitative clinical criteria.

We assume deep learning techniques are ideal methods to tackle this problem because information about facial deformities can be effectively extracted through deep learning based on datasets. Therefore, in this study, we investigate the feasibility of deep learning methods to predict a desired postoperative facial appearance based on a patient's preoperative facial appearance. Three different deep learning methods are chosen for performance comparison.

Methods

We evaluated the performance of three deep learning models (PointNet ++ , GraphUNet, and DeepGCN) to predict an ideal postoperative facial appearance (Fig. 1A). The prediction accuracy of the three models is quantitatively and qualitatively compared.

We first select PointNet ++ because it is reported to successfully predict facial change following orthognathic surgery [2]. However, PointNet ++ is limited because it is designed for unstructured point cloud data, while facial surface data is often represented with a structural three-dimensional (3D) mesh surface. Therefore, graph convolutional network (GCN), which is specialized in 3D mesh data, is also investigated. GCN leverages the structural information using an adjacency matrix that converts 3D mesh data into graph data. Two major variants of GCN, GraphUNet and DeepGCN, are selected for investigation.

We hypothesize that not only local but also global structural information is important for the accurate estimation of the ideal facial appearance. PointNet + + architecture inherently can handle both local and global information successfully. However, this is not the case for GCN-based methods. To achieve a balance between local and global information, various receptive fields are implemented for both GraphUNet and DeepGCN. Specifically, in GraphUNet, Top-K-Pooling is replaced with distance-based pooling. This can prevent trained model to be biased toward the lower part of the facial mesh where most deformities exist. In distance-based pooling, entire facial mesh is divided into K-clusters using K-means clustering. Then, the most significant points are selected from each cluster, resulting in evenly distributed downsampled points over the entire facial mesh. DeepGCN leverages jumping knowledge (or residual learning) to prevent over-smoothing problems caused by the deep structure. However, this may take too many irrelevant points into account while most deformities exist in the lower face. Therefore, we replace its

Table 1 Quantitative prediction accuracy [The average landmark-based error (mean ± standard deviation) between the predicted and ground-truth of test set (mm)]

Models	Sn	Ls	Li	Ch-L	Ch-R	Pog	Average
PointNet + +	1.21 ± 0.97	1.09 ± 1.00	1.63 ± 1.23	1.75 ± 1.27	2.14 ± 1.49	2.53 ± 1.68	1.72 ± 0.28
GraphUNet	1.11 ± 1.00	0.97 ± 0.81	1.73 ± 1.08	1.54 ± 1.07	1.96 ± 1.46	2.16 ± 1.61	1.58 ± 0.30
DeepGCN	**1.03 ± 0.96**	**0.79 ± 0.80**	**1.50 ± 0.98**	**1.25 ± 0.99**	**1.93 ± 1.41**	**1.79 ± 1.71**	**1.38 ± 0.34**

GCN layer with Graph Attention Layer (GAT), which can adaptively adjust the importance of each point by assigning different weights.

The evaluation is conducted on a dataset consisting of 65 paired preoperative and postoperative facial surface meshes of patients who underwent orthognathic surgery. We assume that the patient's postoperative facial mesh is an ideal facial appearance to achieve and thus used as the ground truth. We split the dataset into two subsets, 52 for training and 13 for testing. To avoid overfitting, we applied fivefold cross-validation on the training set to train five models.

For training efficiency, each facial mesh was simplified into a mesh with 3072 points. The prediction accuracy was quantitatively evaluated using the average distance error of clinically important six landmarks: Subnasale (Sn), Labiale Superius (Ls), Labiale Inferius (Li), Left Cheilion (Ch-L), Right Cheilion (Ch-R), and Soft Tissue Pogonion (Pog) (Fig. 1B) and qualitatively evaluated by surface deviation errors.

Results

Table 1 shows the quantitative prediction accuracy of three deep learning methods. DeepGCN showed the smallest error, especially at the Pog landmark where largest facial deformity exists. Figure 1B shows a randomly selected example of the estimated facial mesh with color-coded surface deviation error. PointNet ++ was the worst, showing clinically unacceptable large error. GraphUNet showed an improved prediction accuracy especially in the region of large deformity. However, results of GraphUNet suffered from unnatural artifacts especially in the chin. DeepGCN showed the best overall performance in terms of surface deviation error, surface quality, and landmark-based error.

Conclusion

This study investigated the feasibility of deep learning methods to estimate desired postoperative facial appearance for patients with deformities, which is a critical step towards the soft-tissue-driven surgical planning. Three deep learning methods, PointNet + +, GraphUNet, and DeepGCN, with different mechanisms were evaluated on a clinical dataset.

The experimental results showed that the two GCN-based methods generally exhibited better performance than the point cloud-based method (i.e., PointNet + +) because they considered structural information including mesh connectivity. Among GCN-based methods, DeepGCN performed better than GraphUNet because pooling process of GraphUNet may lose topological information during the training. On the other hand, DeepGCN can effectively extract the information of deformity by concentrating only on important regions based on attention and preserving the existing graph without pooling. In the future, we will develop methods that can incorporate the characteristics of the normal subjects in the predictions to improve the estimation.

References

[1] 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

[2] Fang, X., Kim, D., Xu, X., Kuang, T., Deng, H.H., Barber, J.C., Lampen, N., Gateno, J., Liebschner, M.A., Xia, J.J. and Yan, P., (2022). Deep Learning-Based Facial Appearance Simulation Driven by Surgically Planned Craniomaxillofacial Bony Movement. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 565–574). Springer, Cham.

Poster Session

.

Including planning information into image fusion during left atrial appendage closure

I. Vernikouskaya[1], D. Bertsche[1], P. Metze[1], L. M. Schneider[1], V. Rasche[1]

[1]Ulm University Medical Center, Department of Internal Medicine II, Ulm, Germany

Keywords image fusion, pre-procedural planning, left atrial appendage, multiplanar reconstruction.

Purpose

Left atrial appendage (LAA) closure with catheter-based devices is a minimally invasive cardiac intervention for preventing embolization of thrombus from LAA. LAA is a highly variable structure consisting of lobes and trabeculations. Thus, imaging is essential to pre-plan equipment selection and implantation strategy, to guide procedural device implantation, and also for device surveillance post implantation [1]. Whereas X-ray fluoroscopy (XR) and TEE are the preferred modalities for intra-procedural imaging for LAA closure, a careful pre-procedural planning based on CTA is crucial to minimize procedural or device related complications and to improve patient's outcome. Several small studies have compared LAA measurements from CTA to TEE and fluoroscopy. Overall, CTA produces the largest LAA measurements and more accurately predicted device sizes [2], due to superior multiplanar 3D imaging. In contrast to CTA, cardiac MRI (CMR) provides excellent assessment of the intracardiac and vascular anatomy without any radiation exposure. Thus, the objective of this study was to implement an open-source tool to integrate CMR-based pre-procedural planning of LAA landing zone into image fusion independently of underlying software solutions and to evaluate its application for device selection.

Methods

Ten LAA occlusion cases with WATCHMAN FLX (Boston Scientific, Marlborough, MA, USA) device, for which a 3D CMR with Dixon sequence at the cardiac phase with the largest LAA dimension (atrial diastole) corresponding to 30–40% of the RR interval has been acquired, were chosen retrospectively. Segmentation of the left atrium with LAA was performed from 3D volume using 3DSlicer software. To integrate pre-procedural information obtained from CMR images into image fusion, we implemented a new module for oblique multiplanar reconstruction (MPR) within open-source XR guidance framework 3D-XGuide [3], which allows reconstruction of oblique views of the LAA from 3D volume-rendered images (Fig. 1). Measurement of the LAA was performed according to expert recommendations for the WATCHMAN device [4] with device landing zone (LZ) defined as a line connecting the circumflex coronary artery and a point 1–2 cm inside the LAA, measured from the left upper pulmonary ridge tip. Automatically calculated area-derived diameter of the LZ on "en face" double-oblique view was then compared with the measurements performed intra-procedurally on 2D XR fluoroscopy for device size selection (Fig. 2a). The LZ as defined in the MPR was extracted as triangle mesh and together with the segmented model of the left atrium and LAA (Fig. 2b) registered to XR fluoroscopy, and finally overlaid on the XR frame in which the XR measurement took place for verification (Fig. 2c). If the measured

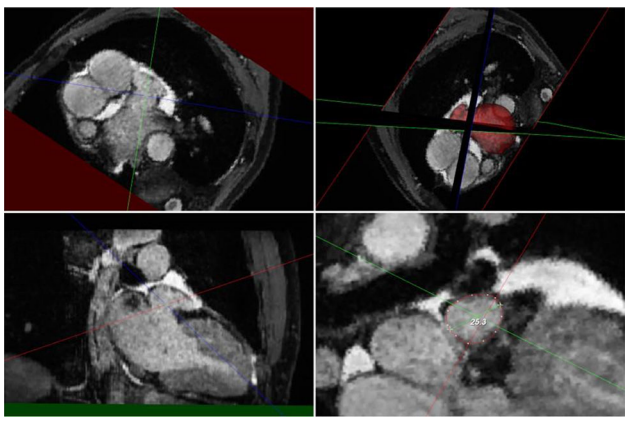

Fig. 1 Double-oblique multiplanar views of the LAA with blue line indicating "en face" plane where the measurement of LAA landing zone is performed. 3D scene allows visualization of three oblique planes within 3D volume with overlaid segmentation model of the left atrium and LAA

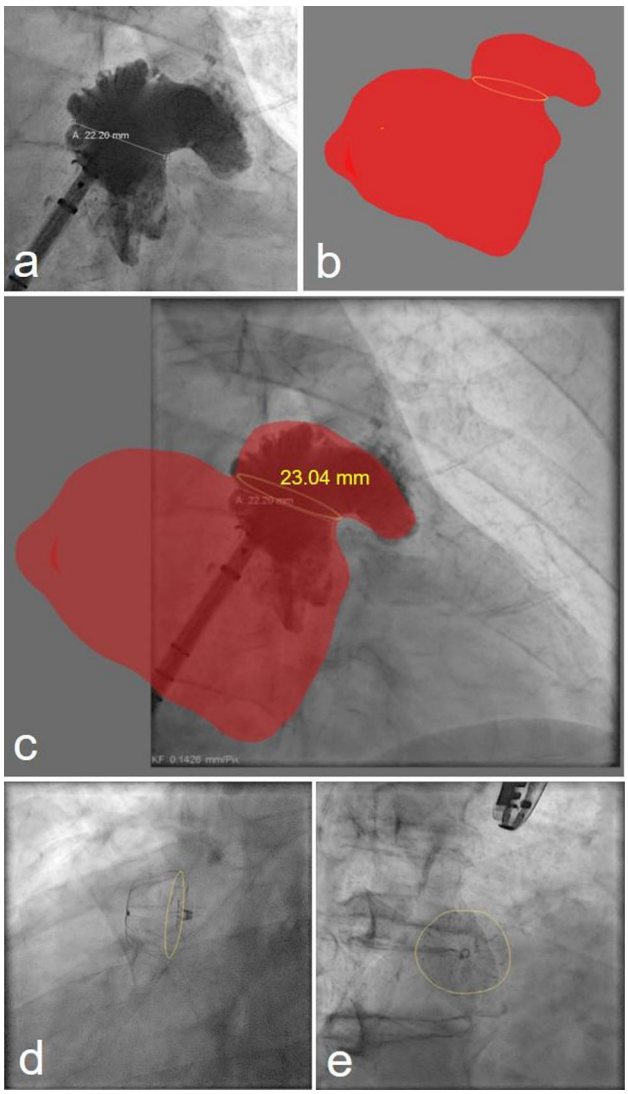

Fig. 2 **a** Measurement of the LAA landing zone (LZ) on XR fluoroscopy. **b** Segmented model of the left atrium incl. LAA with the LZ contour defined in MPR. **c** Overlay of the registered model of the left atrium onto XR fluoroscopy for a frame in which XR measurement was taken. **d** and **e** Overlay of the defined LZ contour onto XR fluoroscopy in two different angulations after device deployment

Table 1 Comparison of measurements of the LAA landing zone performed on CMR with MPR viewer versus XR fluoroscopy with the corresponding selected device size based on manufacturer recommendations

XR measurement		MPR measurement	
Selected device size	Landing zone (mm)	Landing zone (mm)	Selected device size
FLX24	21.66	19.83	FLX24
FLX24	16.39	17.21	FLX24
FLX31	24.45	24.02	FLX31
FLX27	20.17	21.02	FLX27
FLX24	17.66	19.45	FLX24
FLX27	22.2	23.04	FLX27
FLX31	26.79	26.09	FLX31
FLX20	15.89	15.86	FLX20
FLX27	19.2	22.0524	FLX27
FLX31	22.77	23.3041	FLX31

LZ did not coincide with the in XR defined LZ, the planes in MPR were adjusted and LZ was measured again. Doing so, we could make sure that the in CMR measured LZ is defined at the same location at which XR measurement were taken in 2D.

Results

Figure 2d, e visualize position of the released device in two different XR projections with respect to pre-defined LZ contour.

Table 1 summarizes the results of the LAA LZ measurements performed in MPR viewer vs. XR fluoroscopy for the chosen 10 cases. Average LZ diameter measured on XR fluoroscopy was 20.72 ± 3.35 mm, whereas derived from the area of the defined in MPR contour 21.19 ± 3.00 mm ($p = 0.288$). Very good correlation with the Pearson coefficient of $r = 0.93$ ($p = 0.0001$) was achieved among two groups of measurements. Bland–Altman plot demonstrates good agreement between CMR-derived and XR-derived measurements with a mean difference of 0.47 mm (95% limits of agreement − 1.97, 2.91). Device size could be predicted correctly based on the manufacturer sizing chart for the WATCHMAN FLX device in 100% of cases.

Conclusion

CMR is a good alternative to the traditional gold-standard imaging methods for LAA occlusion having several advantages such as detailed 3D characterization of the LAA anatomy, accurate sizing, radiation-free and non-invasive acquisition. MPR viewer implemented in 3D-XGuide as add-on module allows accurate pre-planning of the suitable device size and direct incorporation of the planning information (*e.g.* LAA ostium, landing zone) into the image fusion for control of the device position during implantation without the use of external planning software.

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] Glikson M, Wolff R, Hindricks G, Mandrola J, Camm A, Lip G, Fauchier L, Betts T, Lewalter T, Saw J, Tzikas A, Sternik L, Nietlispach F, Berti S, Sievert H, Bertog S, Meier B (2020) EHRA/EAPCI expert consensus statement on catheter-based left atrial appendage occlusion—an update. EuroIntervention 15(3).

[2] Goitein O, Fink N, Hay I, Di Segni E, Guetta V, Goitein D, Brodov Y, Konen E, Glikson M (2017) Cardiac CT Angiography (CCTA) predicts left atrial appendage occluder device size and procedure outcome. Int J Cardiovasc Imaging 33(5):739–747.

[3] Vernikouskaya I, Bertsche D, Rottbauer W, Rasche V. (2021) 3D-XGuide: open-source X-ray navigation guidance system. Int J Comput Assist Radiol Surg 16(1):53–63.

[4] Korsholm K, Berti S, Iriart X, Saw J, Wang DD, Cochet H, Chow D, Clemente A, De Backer O, Møller Jensen J, Nielsen-Kudsk JE (2020) Expert Recommendations on Cardiac Computed Tomography for Planning Transcatheter Left Atrial Appendage Occlusion. JACC Cardiovasc Interv 13(3):277–292.

Temporal subtraction technique for phalanges CR images using U-Net and Geometric-matching CNN

H. Ono[1], T. Kamiya[1], T. Aoki[2]

[1]Kyushu Institute of Technology, Department of Mechanical and Control Engineering, Kitakyushu, Japan [2]University of Occupational and Environmental Health School of Medicine, Department of Radiology, Kitakyushu, Japan

Keywords Computer Aided Diagnosis, Image Registration, Temporal Subtraction, Geometric Matching.

Purpose

X-ray examination is widely used to diagnose the condition of rheumatoid arthritis and to monitor its activity. However, its diagnosis depends on the subjective evaluation of the physician. In addition, the need to evaluate many joints can be time consuming and increase the workload to the physician on their visual screening. Therefore, the development of a CAD (Computer Aided Diagnosis) system for rheumatoid arthritis is needed.

The temporal subtraction technique enhances the temporal changes by subtracting two successive images. Visualization of temporal changes is expected to improve diagnostic accuracy and shorten reading time. Therefore, we propose a temporal subtraction technique for phalanges CR images, which consists of three steps: automatic extraction of phalangeal regions, image registration, and generation of the subtraction image.

Methods

From CR images taken of both hands, corresponding phalangeal regions between past and current images should be extracted. We use U-Net, which is a CNN (Convolutional Neural Network) model for semantic segmentation, to extract phalangeal regions. Extraction classes are proximal and media phalanges. The distal phalanges are not included in this study because bone deformities due to rheumatoid arthritis are rarely seen. The images used for the training were manually annotated under the guidance of a radiologist. The hand CR images are input into U-Net for segmentation of the proximal and middle phalanges regions. Then, phalangeal region images are generated by extracting only the region of each phalanges using the mask image obtained by U-Net.

Pairs of corresponding phalangeal region images between the past and the current are aligned. We use Geometric-matching CNN [1] algorithm to estimate geometric transformation parameters. The images used in this study were taken with the hand and imaging device in the same position. Therefore, the scale of the phalangeal region on the image is considered unchanged and alignment is performed by rigid transformation. The output layer of the Geometric-matching CNN is three-dimensional because the rigid transformation model needs to estimate three parameters: the rotation angle, and the translation components and.

The geometric transformation parameters estimated by the CNN can be interpreted as simply an approximation or initialization to the optimal deformation [2]. Therefore, after initial alignment by Geometric-matching CNN, the weight parameters of the CNN model can be improved by using instance-specific optimization [2]. We input the initially aligned image pairs again to the Geometric-matching CNN and independently fine-tune the parameters using gradient descent. The fine-tuning of the parameters is accomplished by minimizing the MSE (Mean Squared Error) between the past and current phalangeal region images.

Results

We experimented with segmentation and registration. In the experiment of segmentation, we used 202 hand CR images divided into right and left hands in 101 cases and evaluated by fivefold cross validation. Figure 1 shows an experimental result. In Fig. 1a, b show original, deformed phalangeal region and (c) shows a fusion image which is obtained temporal subtraction image by image registration, respectively. IoU (Intersection over Union) and mIoU (mean Intersection over Union) were used as evaluation metrics. Segmentation of the phalangeal region by U-Net resulted in IoU of 0.949 for the middle phalanges, IoU of 0.957 for the proximal phalanges, and mIoU of 0.953.

In the experiment of registration, we used 560 phalangeal region images generated by U-Net. However, it is difficult to evaluate accurate registration with real data consisting of pairs of past and current images due to bone deformation and other factors. Therefore, we evaluated the synthetic data. The synthetic data was generated by randomly applying rotation of − 15 to 15 degrees and translation of − 10 to 10 pixels to the phalangeal region image. The Dice score was used as the evaluation metric. The proposed method used gradient descent for 50 iterations on each test pair. The proposed method was compared with the previous registration methods, Geometric-matching CNN without instance-specific optimization, SIFT (Scale-Invariant Feature Transform) and GA (Genetic Algorithm). Experimental results show that the proposed method achieves a Dice score of 99.0%, outperforming Geometric-matching CNN without instance-specific optimization, SIFT and GA by 5.5%, 3.6% and 3.2%, respectively. In addition, the computational time of the proposed method is fast, less than 1 s using a GPU.

Fig. 1 Experimental Result: **a** original, **b** deformed phalangeal region, and **c** fusion image

Conclusion

We propose a temporal subtraction method for phalangeal CR images using U-Net and Geometric-matching CNN with instance-specific optimization. The proposed registration method outperforms the accuracy of previous methods. The temporal subtraction images created by the proposed method may be useful for radiologists in detecting interval changes on CR images.

References

[1] Rocco I, Arandjelovic R, Sivic J (2017): Convolutional neural network architecture for geometric matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition:6148–6157.

[2] Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV (2019): VoxelMorph: A learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging: 1788–1800.

Deep learning-based mediastinum suppression technique to evaluate pleural invasion in lung cancer on dynamic chest radiography

R. Nagatani[1], R. Tanaka[2], F. Goshima[1], W. Segars[3], E. Abadi[4], E. Samei[5]

[1]Kanazawa University graduate school of Medical Sciences, Health Sciences, Kodatsuno Kanazawa city, Japan [2]Kanazawa University, AI Hospital/Macro Signal Dynamics Research and Development Center, Kanazawa city, Japan [3]Duke University, School of medicine, Durham, United States [4]Duke University, Clinical science, Durham, United States [5]Duke University, Radiology, Physics, BME, ECE, and Medical Physics, Durham, United States

Keywords dynamic chest radiography, deep learning, 4D extended cardiac-torso phan, mediastinum.

Purpose

Preoperative assessment of lung cancer invasion and adhesions to the mediastinum is crucial for determining the appropriate surgical procedure. Although previous studies have reported that four-dimensional (4D) computed tomography and magnetic resonance imaging can effectively assess the invasion and adhesions in lung cancer, neither can be implemented as a routine protocol because of their high cost and limited availability [1]. Dynamic chest radiography (DCR) is a recently developed, low-cost, low-dose functional X-ray imaging method that uses a flat-panel detector (FPD) to evaluate the pleural invasion and adhesions in lung cancer based on motion information during respiration [2]. Several approaches, such as dual energy subtraction and deep learning-based bone suppression image processing techniques, have been implemented in clinical practice to improve the detection of lung nodules by reducing bone

shadows. However, deep learning-based mediastinum suppression (MS) technique has not been developed, because it is not technically possible to collect large datasets of paired images with and without a mediastinum in clinical images. Therefore, we focused on a 4D extended cardiac-torso (XCAT) phantom. The XCAT phantom flexibly generates realistic virtual patients using defined structures and parameterized models for cardiac and respiratory motions. We believe that two-dimensional projection images of the XCAT phantom with and without a mediastinum can be used to train and validate a network model that enables the MS technique. This study aimed to develop a deep learning-based MS technique using a training dataset created from an XCAT phantom with and without a mediastinum as well as assess its usefulness in evaluating pleural invasion in lung cancer on DCR.

Methods

Phantom preparation: A total of 22 XCAT phantoms with heart rate of 60 beats/min; forced breathing, 10 breaths/min; diaphragm motion, 4 cm; and body mass index, 22–30 kg/m2 were generated using the XCAT program (male:female ratio, 16:6). The XCAT phantoms consisted of 150 respiratory phases in 10 s to simulate 15 frames per second (fps) imaging of a real DCR. A tumor sphere of 30 mm in diameter was randomly inserted around the mediastinum in the lungs. The XCAT phantoms without the mediastinum structures, such as the pericardium, aorta, and esophagus, were also used to produce a perfect projection image without the mediastinum.

Creation of projection images: The XCAT phantoms were projected using the X-ray simulator set to model a dynamic FPD system (PaxScan, 4343CB, Varian Medical Systems, Inc.) having the same imaging geometry and conditions as those of actual DCR (100 kV; 0.2 mAs/frame; 15 fps; and source-to-detector distance, 2.0 m). Virtual imaging was performed in the posteroanterior direction, and 150 respiratory phases of projection images were obtained for each XCAT phantom. In total, 3300 paired projection images with and without a mediastinum were obtained. The matrix size, pixel size, and grayscale range were 1024×1024, 417×417 μm2, and 32 bits Real, respectively. The image histogram was matched with that of actual DCR images, and the resulting images were then rescaled into 8-bit grayscale images in the PNG format.

Network training and creation of MS images: In this study, 2400 paired projection images with and without a mediastinum created from 16 XCAT phantoms were used for training, while 900 projected images with a mediastinum created from 6 XCAT phantoms were used for testing. The generative adversarial network's pix2pix model was trained to estimate synthetic images consisting only of mediastinum structures from the original image. The test dataset was input to the trained generator to obtain a mediastinum image, which was subtracted from the original images to create an MS image (Fig. 1). The resulting MS images were evaluated based on the structural similarity index measure (SSIM) to assess their similarity to the corresponding projection images of the XCAT phantom without mediastinum, i.e., GS images, and on the peak signal-to-noise ratio (PSNR) to assess the reproducibility of the obtained MS images relative to the GS images.

Motion tracking of lung tumor: The kernelized correlation filter tracking function of OpenCV (version 3.4.2) was used to track the simulated tumor automatically in both the original and MS images. The trajectories were compared to the trajectories tracked on the projected images of the tumor without a background, which was the ground truth (GT) in this study. The root mean squared error (RMSE) was calculated to assess the accuracy of tumor tracking on the MS images.

Results

The MS technique could selectively suppress the mediastinum shadow in any test dataset. The MS images of the XCAT phantoms showed a low PSNR (20.4) and SSIM (0.837). The means ± standard deviations of RMSE in the horizontal, vertical, and oblique directions were 0.195 ± 0.436, 16.2 ± 11.5, and 16.3 ± 11.4 for the original images and 1.12 ± 0.920, 7.38 ± 3.85, and 7.51 ± 3.87 for the MS images, respectively. In the original images, the mediastinum shadow caused a tracking error. However, no significant errors were observed in any frame of the MS images. The trajectory of the motion target on the MS images was almost the same as the GT, while that on the original images deviated from the GT because of the mediastinum shadows. These results indicate the feasibility of deep learning-based MS in DCR to allow precise motion tracking of lung cancer and evaluate pleural invasion. Further studies are required to verify the usefulness of the proposed method for application in clinical cases.

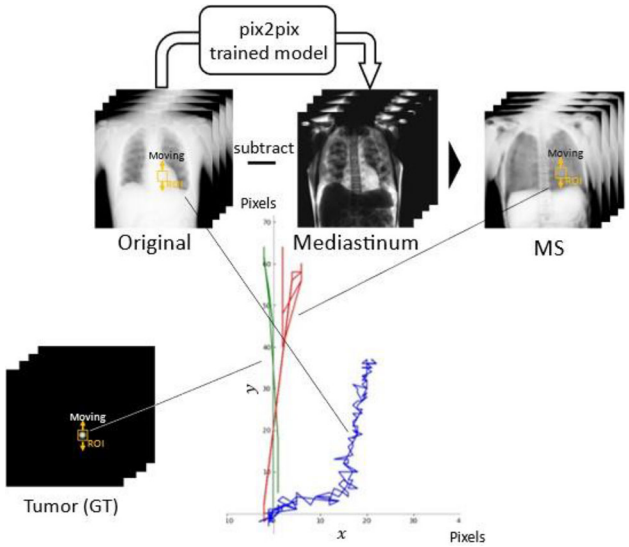

Fig. 1 Improved accuracy of tumor tracking on generated mediastinum suppression (MS) images compared to original images (*ROI* region of interest, *MS* mediastinum suppression, *GT* ground truth)

Conclusion

We developed a deep learning–based MS technique to evaluate the pleural invasion in lung cancer on DCR. The proposed MS technique improves the tracking performance of lung tumors around and/or behind the mediastinum on DCR. Furthermore, this study demonstrates that simulated images work effectively in deep learning models, which would facilitate the research and development of deep learning-based image processing.

References

[1] Tanaka R (2022) Preoperative evaluation of pleural adhesions with dynamic chest radiography: a retrospective study of 146 patients with lung cancer. Clinical Radiology 77(9):689–696

[2] Tanaka R (2016) Dynamic chest radiography: flat-panel detector (FPD) based functional X-ray imaging. Radiol Phys Technol 9(2):139–153

Edge Reinforced Co-training Network based Bimodal Segmentation of Abnormal Nerve and Vascular for Microvascular Decompression

R. Tu[1], L. Yu[2], W. Si[1]

[1]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Institute of Advanced Integration Technology, Shenzhen, China [2]The University of Hong Kong, Department of Statistics & Actuarial Science, Hong Kong, China

Keywords Co-training, 3D segmentation, vessel-like structure, Attention.

Purpose

Microvascular decompression (MVD) is a surgery to relieve abnormal compression of a cranial nerve causing trigeminal neuralgia, glossopharyngeal neuralgia, or hemifacial spasm. In this regards, precise preoperative segmentation of abnormal nerve together with surrounding vessels in multi-modal images is crucial for surgical planning. Abnormal nerve and vascular are typical vessel-like structures, which occupy a wide area and require segmentation of sharp edges, thus it presents a great challenge in segmentation. In three-dimensional (3D) medical imaging, the distribution of the edge and non-edge voxels of the vessel-like structure is typically very unbalanced since the majority of the voxels are non-edge, making it difficult to identify sharp edges. Consequently, the extraction of edge part should be particularly emphasized. On the other hand, existing methods rarely combine different modalities in abnormal nerve and vessel segmentation, while they can provide more detailed features of different structures in different modalities required by preoperative spatial relationship analysis of nerve and vessel.

Methods

In our setting, two MRI modalities (T2 and TOF) are available and provide different contrast levels. Our proposed method consists of a bimodal path and a unimodal path with only T2 modality to acquire the general features and the most relevant features, respectively. We introduce an attention-based strategy to discover and better preserve the spatial edge information within each modality and a mutual information-based co-training strategy to retain the rich mutual information between two paths. In order to achieve satisfying segmentation performance, the skeleton and the edge parts of vessel-like structures should be emphasized equally. Inspired by the reverse edge attention module in [2], we introduce an attention-based strategy to strengthen the ability of context perceiving as well as edge information discovering. By deleting the estimated object areas from the high-level side-output characteristics, it is inserted between adjacent layers of the encoder to uncover the edge information, progressively exploiting complimentary regions and details. Co-training approach has been applied to tasks with multiple views, where each view is independently trained and provides complementary information. In our setting, two MRI modalities (T2 and TOF) are available and provide different contrast levels. Here, we construct a bimodal path and a unimodal path with only T2 modality to acquire the general features and the most relevant features, respectively. We introduce modality-mutual knowledge transfer learning to allow efficient mutual information interaction between two paths to promote the unimodal path to learn from the bimodal path and adds necessary guidance and regularization to bimodal path.

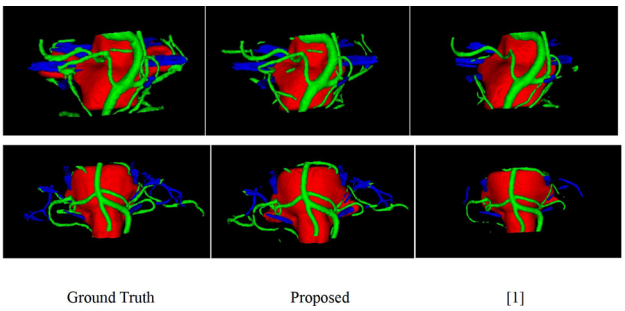

Fig. 1 Visualization of the segmentation results. The two rows represent the 3D rendering of two segmentation volumes. From left to right in each column: ground truths, segmentation results of the proposed method and [1]

Results

We perform experiments on our in-house dataset consisting of 100 MRI scans with two modalities: (1) T2-weighted and (2) Time-of-Flight (TOF) MRA volumes. Each of these modalities capture three main structures: brainstem, abnormal vascular and nerve. We compared our method with current SOTA method for MVD-related structure segmentation [1], both are (re)implemented on the nnUNet framework. To the best of our knowledge, this is the first work that employ bimodal MRI to achieve brainstem, abnormal vascular and nerve segmentation. Experimental results presented in Table 1 demonstrate that our method can achieve better performance in all three targets, and improves significantly in abnorml nerve which is the hardest to segment. Segmentation results are shown in Fig. 1. Both methods can detect brainstem and most major vessels, but [1] performed poorly in identifying the continuity of thin vessels and the presence of irregular nerves. In contrast, our proposed method achieved better segmentation results in maintaining the vessel continuity and reduce false-positive results. For quantitative results, our method achieves better scores in DSC and Jaccard, indicating the superior performance in the overall segmentation results, lower AHD indicates our proposed method has better classification ability of edge part voxels.

Conclusion

In this work, we presented two effective modules, namely a reverse edge attention module to discover edge part and preserve the spatial edge information and a co-training strategy between unimodal path and bimodal path to align feature distributions between modalities. Experimental results demonstrate the proposed method can outperform the SOTA. However, considering that each modality can provide information for each target of varying uncertainty, our immediate

Table 1 Segmentation performance of comparison methods

Methods	DSC			HD			Jaccard		
	BS	CV	CN	BS	CV	CN	BS	CV	CN
[1]	0.909	0.689	0.542	**1.414**	5.817	8.819	**0.907**	0.601	0.271
Proposed	**0.912**	**0.722**	**0.616**	2.738	**5.000**	**7.382**	0.860	**0.677**	**0.607**

BS brainstem, *CV* cerebrovascular, *CN* cranial nerve

plan is to focus on utilizing the uncertainty when fusing information from different modalities.

Acknowledgement
This work was supported in part by Shenzhen Fundamental Research Program (JCYJ20200109110420626, JCYJ20200109110208764), in part by National Natural Science Foundation of China (U22A2034) and in part by Guangdong Basic and Applied Basic Research. Foundation (2023A1515030268, 2021A1515012604).

References

[1] Lin, J., Mou, L., Yan, Q., Ma, S., Yue, X., Zhou, S., … & Zhao, Y. (2021). Automated Segmentation of Trigeminal Nerve and Cerebrovasculature in MR-Angiography Images by Deep Learning. Frontiers in Neuroscience, 1684.

[2] Xia, L., Zhang, H., Wu, Y., Song, R., Ma, Y., Mou, L., … & Zhao, Y. (2022). 3D vessel-like structure segmentation in medical images by an edge-reinforced network. Medical Image Analysis, 82, 102,581.

Automatic measurement of the muscle cross-sectional area on the 1st and 3rd lumbar vertebra levels by two U-nets

S. Hanaoka[1], W. Gonoi[1], S. Inui[1], N. Akamatsu[2], Y. Nomura[3,4], T. Takenaga[1], S. Miki[4], T. Yoshikawa[4], N. Hayashi[4], K. Sugawara[5,6], S. Taguchi[7], K. Kishitani[7], H. Kume[7], T. Kawai[8], T. Nakagawa[8], O. Abe[1]

[1]The University of Tokyo Hospital, Department of Radiology, Tokyo, Japan [2]Nerima Hikarigaoka Hospital, Tokyo, Japan [3]Chiba university, Chiba, Japan [4]The University of Tokyo Hospital, Department of Computational Diagnostic Radiology and Preventive Medicine, Tokyo, Japan [5]Saitama Cancer Center Hospital, Department of Gastroenterological Surgery, Saitama, Japan [6]The University of Tokyo Hospital, Department of Gastrointestinal Surgery, Tokyo, Japan [7]The University of Tokyo Hospital, Department of Urology, Tokyo, Japan [8]Teikyo University School of Medicine, Department of Urology, Tokyo, Japan

Keywords Sarcopenia, Segmentation, Unet, Abdominal muscle.

Purpose

Accurate evaluation of sarcopenia, or decrease of the body muscle volume, is important in the assessing and predicting the health status of patients. In this study, a fully automated segmentation method for abdominal wall muscles in the cross-section at the 3rd lumbar vertebra (L3) or the 1st lumbar vertebra (L1) levels in CT images.

Methods

Two U-nets with residual convolutional network components were used in this study. The two U-nets share the network architecture (Fig. 1).

First, frontal and lateral maximum intensity projection images and also frontal and lateral mean intensity projection images of the input CT volume are created. There four images are inputted to the first U-net, and the cross-sectional line of the body on L3 (or L1) level is detected. In the training, the inputted ground-truth label was the corresponding (i.e., frontal or lateral) projection of the final ground-truth segmentation label of the abdominal muscle in the target (e.g., L3) level slice. In the test, the outputted likelihood images (the frontal and lateral images) were summed up along x- or y-axis, and the slice with the maximal sum likelihood was selected as the output.

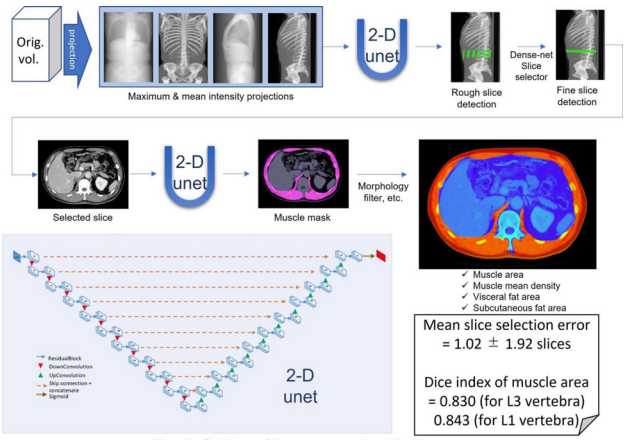

Fig. 1. Outline of the proposed method.

Fig. 1 Outline of the proposed method

The L3 (or L1) level detection are slightly improved using another DenseNet-based slice-wise detector. In detail, the detected L3 slice and ± 4 slices above and below the detected slice are inputted to the DenseNet-based slice-wise detector and the slice with the highest output likelihood value is selected. This two-stage vertebral cross-sectional slice detection scheme imitates the detection manner of radiologists who firstly detect the target vertebra roughly in the (for example) scout image or the mid-sagittal cross-section, then refine the level by checking neighbor slices in axial images.

After the level determination, the detected cross-sectional CT slice is inputted to the second U-net (for segmentation) and the final segmentation result is outputted. Additionally, the fat region areas of the outer (subcutaneous) and the inner (visceral) areas were automatically calculated.

The training was performed with 127 CT volumes which has various sizes of field of view and imaging range. Although most of cases were with intravenous contrast agent, some cases were plain CT. Board-certificated radiologists manually inputted the ground truth muscle area in the L3 and L1 level cross-section. The loss function used was Dice loss for both U-nets and binary labels for the DenseNet. The training for L1 and L3 was performed separately, thus finally we trained six models (three for each vertebra).

The test was performed with the other 100 CT volumes. The evaluation was performed for both the slice detection and the Dice score in the selected slice (between the automatically segmented muscle mask and the manually imputed true label binary image). Note that the calculation of Dice score was performed even when the slice detection was failed.

Results

The slice determination accuracy (mean absolute difference of slice position) was 1.02 ± 1.92 slices (mean ± standard deviation), and the mean similarity (Dice coefficient) between the hand-made and automated region of interest was 0.830 ± 0.108 for the L3 vertebra level. For L1 vertebra level, the slice determination accuracy was 0.92 ± 2.32 slices the Dice score was 0.843 ± 0.080. These results are comparable to those of Pickhardt et. al. [1], in which slice detection error was 1.41 ± 5.92 slices and Dice score was 0.96 ± 0.02. Another previous study [2] reported slice selection error was 1.41 ± 5.02 and Dice score was 0.96 ± 0.02. In comparison, the stability and robustness of our system for detecting the L3 vertebra level were proven. On the other hand, the Dice score was inferior to other two studies.

Conclusion

A fast, accurate and robust sarcopenia evaluation tool was presented. Our automatic system has a sufficient accuracy, and it is ready to be used for sarcopenia analysis with a large cohort.

References

[1] Pickhardt PJ, Perez AA, Garrett JW, Graffy PM, Zea R, Summers RM (2022) Fully Automated Deep Learning Tool for Sarcopenia Assessment on CT: L1 Versus L3 Vertebral Level Muscle Measurements for Opportunistic Prediction of Adverse Clinical Outcomes. AJR Am J Roentgenol. 2022;218: 124–131.

[2] Islam S, Kanavati F, Arain Z, Da Costa OF, Crum W, Aboagye EO, Rockall AG (2022) Fully automated deep-learning section-based muscle segmentation from CT images for sarcopenia assessment. Clin Radiol. 2022;77: e363–e371.

Automatic measurement system for the diameter of the inferior vena cava using deep neural network

H. Noro[1], N. Koizumi[1], Y. Nishiyama[1], J. ZHOU[1], T. Ishikawa[1], R. Tsumura[1,2], K. Yoshinaka[2], H. Tsukihara[3], N. Matsumoto[4], R. Masuzaki[4], M. Ogawa[4], K. Numata[5]

[1]The University of Electro Comunications, Graduate School of infomatics and Engineering, Chofu, Japan [2]National Institute of Advanced Industral Science and Technology, Health and Medical Research Institute, Tsukuba, Japan [3]The University of Tokyo, School oh Engineering, Bunkyo-ku, Japan [4]Nihon University Surugadai Hospital, Department of Gastroenterology,and Hepatology, Chiyoda-ku, Japan [5]Yokohama City University Medical Center, Gastroenterological Center, Yokohama, Japan

Keywords IVC, Deep Learning, Robotic ultrasound diagnosis, Diameter measurement of vessel.

Purpose

Compared to MRI and CT, ultrasound has the advantage of being able to visualize the internal structures of the body noninvasively and in real time, regardless of the location. On the other hand, ultrasound diagnosis requires operator skill in acquiring diagnostic images and is prone to variations in the measurement of feature values due to operator habits. In particular, it is extremely difficult for unskilled operators to distinguish the inferior vena cava (IVC) from the abdominal aorta, and even skilled operators are prone to variations in the measurement of the internal diameter in the same cross section due to their own habits.

Based on the above, the purpose of this paper is to propose a new image processing method to easily identify the IVC from the abdominal aorta and to measure the time series of the inner diameter in the same cross section[1] without operator's habitual errors by using deep learning technology.

Methods

Figure 1 shows an overview of the newly proposed method to identify the IVC from the abdominal aorta and to measure its inner diameter in the same cross-section. Specifically, first, the IVC is identified from the abdominal aorta with the aid of YOLACT[2], and segmentation is performed together with the liver. Since the IVC and abdominal aorta are parallel, the above two vessel images can be alternately switched and displayed by manipulating the probe in the elevation direction. Using this, the probe is operated to switch the above vessel images until the displayed vessel is identified from the abdominal aorta and identified as the IVC.

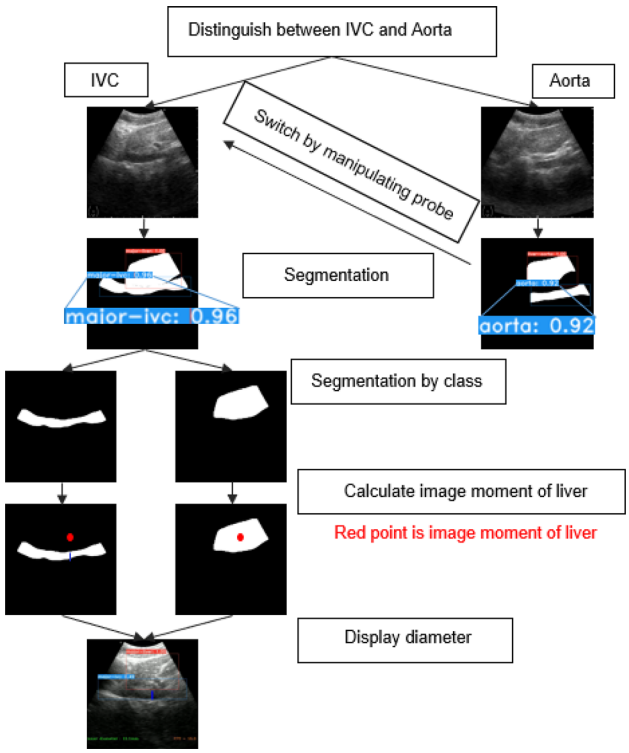

Fig. 1 Proposed method

Next, the inner diameter of the IVC is measured in a straight line approximately perpendicular to the IVC, passing through the center of gravity of the liver identified from the segmentation results. Here, we take advantage of the fact that the liver and IVC operate in a similar manner.

Results

The results of the IoU and Dice coefficients are shown in Table 1. The goal of this study was to measure the internal diameter of the IVC in real time, and we believe that YOLACT is effective because it can detect the IVC in more than 60% of cases and because YOLACT is excellent for real-time segmentation. Some of the images show proper internal diameter measurement, while others do not. The accuracy was 3.30 (mm). The images that failed were those in which the segmentation failed. We believe that the number of failures will decrease as the accuracy of the segmentation is improved.

One point for improvement is that the blood vessels are measured perpendicular to the image. In the actual inspection, the measurement is made perpendicular to the blood vessels, so improvement is necessary in the future. Fps was about 18. Since the fps of the probe is about 15, there is no problem with real-time performance.

Conclusion

In this study, we evaluated the accuracy of the segmentation of the target organ and the accuracy of the measurement of the internal diameter. In the future, we would like to construct a system that measures perpendicularly to the internal diameter of the IVC and improve the accuracy of segmentation and internal diameter measurement.

Table 1 Segmentation performance results

	Major-ivc	Aorta	Major-liver	Liver-aorta
IoU	0.641	0.066	0.88	0.781
Dice係数	0.727	0.08	0.936	0.87

References

[1] Mesin L, Pasquero P, Albani S, Porta M, Roatta S (2015) Semi-automated tracking and continuous diameter monitoring of inferior vena cava in simulated and experimental ultrasound imaging. Ultrasound Med Biol, 2015 Mar;41(3):845–57.
[2] Bolya D, Zhou C, Xiao F, Jae Lee Y (2019) YOLACT Real-time Instance Segmentation, IEEE Trans Pattern Anal Mach Intell, Feb;44(2):1108–1121.

Blood Vessel Segmentation via Cascaded Enhance-and-Segment Framework in Dissecting Microscopic Images of Cranial Windows

Y. Wu[1], Y. Hayashi[1], M. Oda[1,2], T. Zheng[1], S. Kawamura[3], T. Takebe[3,4,5], K. Mori[1,6,7]

[1]Nagoya University, Department of Intelligent Systems, Nagoya, Japan [2]Nagoya University, Information and Communication, Nagoya, Japan [3]Tokyo Medical and Dental University, Institute of Research, Tokyo, Japan [4]Cincinnati Children's Hospital Medical Center, Division of Gastroenterology, Hepatology and Nutrition, Developmental Biology, and Center for Stem Cell and Organoid Medicine (CuSTOM), Cincinnati, United States [5]University of Cincinnati, Department of Pediatrics, College of Medicine, Cincinnati, United States [6]Nagoya University, Information Technology Center, Nagoya, Japan [7]National Institute of Informatics, Research Center for Medical Bigdata, Hitotsubashi, Japan

Keywords Dissecting Microscope Images, Image enhancement, Image Denoising, Blood Vessel Segmentation.

Purpose

This paper proposes an end-to-end cascaded enhance-and-segment framework to improve blood vessel segmentation result from dissectingmicroscopic images of cranial windows. Cranial window allows the cortical vasculature to be easily visualized by dissecting microscope. Dissecting microscope, also known as a stereomicroscope, allows long-term and real-time imaging and observation of non-sectioned samples. Many disease experiments are performed in the cortical vessels of mice's cranial windows. The changes in the blood vessels and the manipulation of the vessels play a key role in these experiments.

With the development of medical image processing, it is expected to utilize dissecting microscopic images to develop an automated system that can automatically perform automatic manipulation and structure extraction of blood vessels in the cranial window. However, the problems of low contrast and noise in dissecting microscope images make blood vessel segmentation from dissecting microscope images become difficult.

Therefore, in this paper, we propose a cascaded enhance-and-segment framework to improve blood vessel segmentation result in low-quality (LQ) dissecting microscopic images of cranial windows by using synthesized high-quality (HQ) images. Compared with LQ images, HQ images have higher contrast and contain less noise. In addition, compared with LQ images, blood vessels are easier to be observed and extracted from HQ images. The implementation of segmentation of blood vessels in dissecting microscope images can help the robot and researcher can easier localize the position of blood vessels in cranial windows.

Methods

Overview

Our proposed cascaded enhance-and-segment framework includes two parts: (1) HQ image generation network and (2) Blood vessel segmentation network as shown in Fig. 1. The input of the cascaded framework is noisy, low-contrast LQ microscope images amd the output is a segmentation result. In the following, we describe our framework in detail.

Image Enhancement and Denoising Based Image Generation Method (U-Net 1)

We follow the U-Net [1] generator of the Pixel2pixel [2] as our image enhancement and denoising network to synthesize HQ images from LQ images. Compared to the Pixel2pixel, we don't use the discriminator in our framework, because the subsequent segmentation network (U-Net 2) provides feedback to the U-Net 1 and make synthesized HQ images more beneficial for vessel segmentation. In the training phase of the U-Net 1, to enhance contrast and denoise dissecting microscopic images, we use LQ images as input and HQ images as the ground truth. We use L1 loss as the loss function to train the enhancement and denoising network. In the testing phase, 3-channel LQ images are input U-Net 1 and its outputs is 3-channel synthesized HQ images.

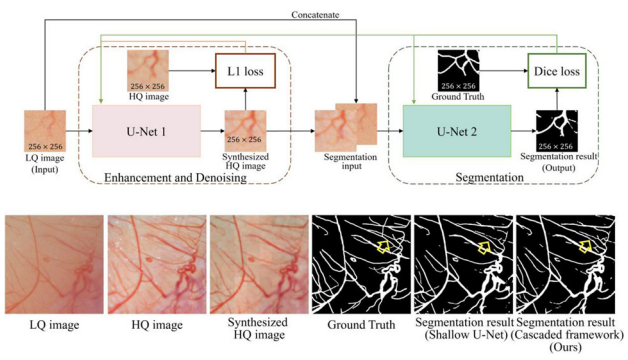

Fig. 1 Upper half: proposed end-to-end cascaded enhance-and-segment framework. Lower half: qualitative results of image enhancement and segmentation

Table 1 Cascaded framework had better quantitative results than the shallow U-Net in microscope image data

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	Dice (%)	clDice (%)
Shallow U-Net	92.55 ± 0.54	76.38 ± 2.45	95.76 ± 1.02	77.29 ± 1.93	76.92 ± 2.09
Cascaded framework (ours)	92.67 ± 0.53	76.60 ± 2.37	95.88 ± 1.10	77.64 ± 1.80	77.51 ± 1.71

Blood Vessel Segmentation in Dissecting Microscopic Images Based U-Net (U-Net 2)

In our work, we used a shallow U-Net (some layers are removed from the original U-Net [1]) as our blood vessel segmentation network. The schematic illustration is shown in Fig. 1. The input of the U-Net is a 6-channels image which is obtained by combining a 3-channels LQ image and a 3-channels synthesized HQ image. The output of the U-Net is a binary image. In the training phase of the U-Net 2, we use the Dice as a loss function to train the segmentation network. In the testing, 3-channel LQ images and 3-channel synthesize HQ images are combined and input to the shallow U-Net (U-Net 2) to segment blood vessels in microscopic images.

Results

Dataset

69 raw LQ-HQ stereomicroscopic image pairs of pixels were taken by an experienced researcher using a camera on the single eyepiece of a Leica S9D dissecting microscope (stereomicroscope). HQ images are taken at × 5 magnification, and LQ images are taken at × 1 magnification. For training, we registered HQ and LQ images using affine transformation with the SURF feature. Then we cropped LQ region corresponding to the field of view of HQ images and enlarge cropped LQ images to the same size as HQ images which means cropped LQ image is enlarged by 5 times. Therefore, we obtained 65 HQ-LQ paired images with the size of 1792 × 1792 pixels as our dataset for generator (U-Net 1). For segmentation, the ground truths were made manually by annotating at the pixel-level.

Because of computation resource limitations, we resized the images to size of 1024 × 1024 pixels and cropped patches with a size of 256 × 256 as input of our cascaded network for training and testing. Additionally, we conducted a five-fold cross-validation experiment. Each fold contained 13 images, four were used for training and one for testing.

Quantitative and qualitative results

As shown in the lower half of Fig. 1, our framework could synthesize HQ images. In addition, to verify that the synthesize HQ image can improve the segmentation result as part of the input, we compared results of our cascaded framework and the shallow U-Net (only performsegmentation) in our microscopic dataset. The segmentation results had better performance and especially had better connectivity for some blood vessels, compared to using only the shallow U-Net as shown in Fig. 1. We used the accuracy, sensitivity, specificity, Dice, and clDice measures to quantitatively evaluate the segmentation results. As shown in Table 1, our cascaded framework had better quantitative results than the shallow U-Net in our microscope image data. Compared with the shallow U-Net, Dice and clDice score of our framework were improved by 0.35% and 0.59% respectively.

Conclusion

We proposed an end-to-end cascaded enhance-and-segment framework to improve blood vessel segmentation result in dissecting microscopic images of cranial window. The experimental results showed that our proposed method could obtain better vascular segmentation results than training the segmentation network using only low-quality data as input. In the future, we will further improve the vascular segmentation results by improving the network structure and the cascade training loss function.

References

[1] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-assisted Intervention. LNCS 9351 (pp. 234–241). Springer, Cham.

[2] Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1125–1134).

A robust respiratory gating method for rodent micro-CT imaging

M. Fukushima[1], T. Okamoto[2], H. Haneishi[2]

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

Keywords micro-ct, respiratory gating, small animal imaging, retrospective.

Purpose

Micro-computed tomography (micro-CT) is an X-ray-based imaging modality that obtains detailed three-dimensional information on the internal structure of a target object. This technology has recently been applied to in-vivo small animal imaging in preclinical studies because micro-CT enables us to perform noninvasive imaging. However, respiratory motion for living small animals causes motion artifacts (blurring) on reconstructed images.

To cope with the severe artifacts, some researchers have proposed retrospective respiratory gating methods that collect the subset of intensity data for each respiratory phase. Hu et al. [1] discriminated respiratory phases using the total intensity value of the region of interest (ROI) set on all intensity data. In addition, a retrospective respiratory gating system using the average intensity value of ROI has been implemented in a commercial micro-CT scanner [2] (CosmoScan GX (Rigaku Corp., Tokyo, Japan)). This system is effective for the stable respiratory cycle; however, this system is less effective in shallow and abnormal breathing cases. In this study, we proposed a robust respiratory gating method that is applicable to various respiratory cycles and abnormal breathing cases.

Methods

The proposed method consists of three steps: (1) motion gating signal extraction, (2) determination of projection angles to be used, and (3)

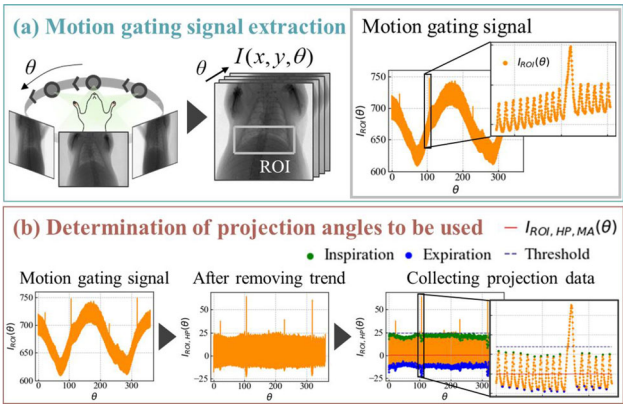

Fig. 1 Outline of the proposed method. **a** Extraction of motion gating signal from all intensity data. **b** Determination of projection angles to be used

reconstruction. Figure 1 shows an outline of the proposed method. A detailed description of each method is presented below.

Motion gating signal extraction: Let us define all intensity data as $I(x,y,\theta)$. Here denotes spatial coordinate of image and denotes the projection angle. We manually set a rectangle ROI around the diaphragm at the same location for all intensity data and calculate the average intensity value, $I_{ROI}(\theta)$. This intensity waveform includes the respiratory and heart movement. For example, when the average intensity value decreases, the proportion of air decreases, indicating an expiration process; conversely, when the average intensity value increases, the proportion of air increases, indicating an inspiration process. This intensity waveform is used as a motion gating signal.

Determination of projection angles to be used: Inspiration and expiration phases to be used for reconstruction are determined from $I_{ROI}(\theta)$. First, as $I_{ROI}(\theta)$ includes a projection angle-oriented sine wave trend as shown in Fig. 1a, it is removed by applying a high-pass filter. We call the filtered signal $I_{ROI,HP}(\theta)$. Then, its smoothed version, $I_{ROI,HP,MA}(\theta)$, is also calculated by moving average technique with a proper width over one respiration cycle. In the proposed method, as inspiration phase we extract projection angles satisfying the following conditions: $I_{ROI,HP}(\theta)$ has the local maximum and is greater than $I_{ROI,HP,MA}(\theta)$. Similarly, as expiration we extract projection angles satisfying the following conditions: $I_{ROI,HP}(\theta)$ has the local minimum and is smaller than $I_{ROI,HP,MA}(\theta)$. Finally, outliers in $I_{ROI,HP}(\theta)$, which refer to abnormal breathing, are removed by a threshold value computed by considering the overall waveform.

We compared the performance of the proposed method with the previous method [2] implemented in commercial micro-CT (CosmoScan GX). To verify the applicability of the proposed method for various respiratory cycles, we applied the proposed and the previous methods to two respiratory cycle data: approximately 1 s (1 s respiration case) and approximately 0.4 s (0.4 s respiration case), which is shorter than usual. The specimens used for the experiments were two 10-week-old adults Wistar rats. We qualitatively evaluated the adequateness of projection angle selection as well as reconstruction image quality of 1 s respiratory case.

Results

For 1 s respiration case, both methods collected projection data for inspiration and expiration phases adequately. For 0.4 s respiration case, on the other hand, the proposed method succeeded in collecting the subset of projection data for inspiration and expiration phases stably, whereas the previous method failed. The previous method used a band-pass filter in the frequency domain to extract respiratory signal from the motion gating signal. However, since the range setting of the band-pass filter is fixed, the respiratory signal is out of range when the case of a rodent whose respiratory cycle is shorter than usual. The experimental results demonstrate that the proposed method is more robust for various respiratory cycles and abnormal breathing cases.

In terms of the reconstructed image, there is little difference between the proposed and previous methods in the inspiration phase. However, in the expiration phase, the proposed method reconstructed CT images with less blur compared to the previous method.

Conclusion

We proposed a respiratory gating method that collects the subset of projection data for inspiration and expiration phases using the motion gating signal. Experimental results show that the proposed method could collect the subset of projection data more stable than the previous method.

References

[1] Hu J, Haworth ST, Molthen RC, Dawson CA (2004) Dynamic small animal lung imaging via a post-acquisition respiratory gating technique using micro-cone beam computed tomography. Academic radiology 11(9):961–970

[2] Koike T, Maesawa M, Hara Y (2018) CT-IMAGE PROCESSING APPARATUS AND METHOD. US Patent 9,895,129 B2, 20 February 2018

Correction software of ADC bias induced by gradient non-linearity in DWI: breast phantom study

T. Yoshida[1], A. Urikura[2], M. Endo[3], T. Aramaki[1]

[1]Shizuoka Cancer Center, Diagnostic Radiology, Nagaizumi, Japan [2]National Cancer Center Hospital, Diagnostic Radiology, Tokyo, Japan [3]Chiba University Hospital, Division of Comprehensive Radiology Center, Chiba, Japan

Keywords Magnetic Resonance imaging, Diffusion-weighted imaging, Apparent diffusion coefficient, Post-processing filter.

Purpose

Gradient non-linearity is known to be one of the causes of the apparent diffusion coefficient (ADC) bias in diffusion-weighted magnetic resonance imaging (DWI) [1]. A correction software is recently available in a commercially available scanner. This study aimed to evaluate the correction software for the ADC bias induced by gradient non-linearity using an ice water phantom in breast DWI.

Methods

An ice water phantom shows a known ADC value of 1.1×10^{-3} mm^2/s at 0 °C [2]. The phantoms simulated bilateral breasts constructed from 1.5-L cylindrical plastic bottles with a 120-mm diameter. Five plastic tubes with a 100-mm long and 15 mm diameter filled with distilled water were set at the center, upper, right, bottom, and left positions into the plastic bottles filled with an ice-water bath. The longitudinal axis of the tube was aligned perpendicular to the main field and separated by about 30 mm from each other.

The magnetic resonance scanner used for data collection was a 3.0 T whole-body scanner (Ingenia 3.0 T CX, Philips Healthcare, Best, the Netherlands) with a dedicated 16-channel bilateral breast coil (dStream Breast 16ch coil, Philips Healthcare, Best, the Netherlands). The scanner has a maximum gradient amplitude of 45 mT/m and a slew rate of 200 mT/m/ms. Echo-planar imaging sequence was used for DWI acquisition and the following parameters were used: repetition time, 10,000 ms; echo time, shortest; flip angle,

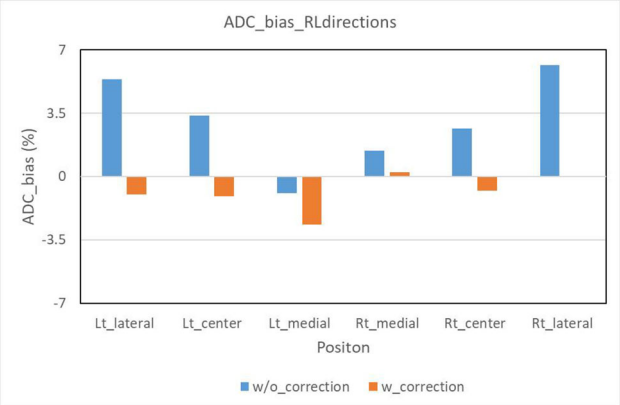

Fig. 1 The ADC bias in the RL direction. The blue and orange bar indicate ADC bias without and with correction, respectively. Our correction software improves the ADC bias, especially at the off-center positions

90 degrees; b-value, 0 and 800 mm2/s; bandwidth, 2025 Hz/pixel; field of view 350 mm, EPI factor, 65; scan time, 120 s; number of excitations, 2; slice orientation, axial; motion probing gradient, three orthogonal directions. Correction software named "ADC correction" can be selected in the mode, on/off in DWI acquisition. Five data were acquired for each correction mode.

The ADC maps were generated with mono-exponential fitting using b = 0 and b = 800 images on the scanner console and transferred to and analyzed with the SYNAPSE VINCENT (Fujifilm Medical Co., Ltd., Tokyo, Japan). The ADC analysis was performed at three slices (center slice, 30-mm superior slice, and 30-mm inferior slice) for each ADC map. The 4 × 4 square region of interest (ROI) was manually set at the center of each tube to prevent the image artifact. The mean ADC in the ROI was recorded. ADC bias was defined as the relative percentage difference of ADC from that of 1.1×10^{-3} mm^2/s.

ADC bias was evaluated in the anterior–posterior (AP), right-left (RL), and superior-inferior (SI) directions. Two ROIs were set at ± 25-mm away from the center ROI in the central tube to evaluate the ADC bias in the AP direction. ADC bias in the RL direction and SI direction were evaluated using ADC of the center, right, and left tubes and the center, upper, and bottom tubes, respectively.

Results

ADC bias was observed in the RL directions. ADCs at the outermost tubes for each phantom were about 5–6% higher than that of 1.1×10^{-3} mm^2/s. The ADC correction software improves the ADC bias at the off-center position; however, Fig. 1, ADCs after correction were slightly lower than that of 1.1×10^{-3} mm^2/s. ADC bias was also observed in the SI direction; however, the correction software did not improve the ADC bias and caused about a 2–3% decrease on average. In the AP direction, the evident ADC bias was not observed and ADCs after the correction decreased by about 3% as well as in the other directions.

Conclusion

Our correction software improved the ADC bias in the RL directions; however, the systematical decrease of ADC was observed after the correction.

References

[1] Malyarenko D, Galban CJ, Londy FJ, Meyer CR, Johnson TD, Rehemtulla A, Ross BD, Chenevert TL. Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. J Magn Reson Imaging. 2013;37:1238–46.

[2] Holz M, Stefan RH, Antonio S. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. Phys Chem Chem Phys. 2000;20:4740–2.

Patch-based respiratory phase matching method for digital subtraction angiography under natural respiration

Y. Sekiguchi[1], T. Okamoto[2], K. Fujiwara[3], T. Kondo[3], J. Koizumi[3], H. Haneishi[2]

[1]Chiba university, Graduate School of Science and Engineering, Chiba, Japan [2]Chiba university, Center for Frontier Medical Engineering, Chiba, Japan [3]Chiba university, Graduate School of Medicine, Chiba, Japan

Keywords DSA, TACE, Respiratory phase matching, Patch-based.

Purpose

Transcatheter arterial chemoembolization (TACE) is a widely accepted locoregional therapy for patients with unresectable hepatocellular carcinoma. Intraoperatively, the surgeon uses digital subtraction angiography (DSA) to visualize blood vessel structure, blood flow, and tumor location. DSA images are generated by subtracting the images before and after contrast agent injection (we call them mask image and contrast image in this summary). To obtain clear DSA images, patients need to hold their breathing during DSA acquisition to avoid artifacts due to respiratory motion. However, most patients for TACE are elderly and have difficulty in holding their breathing. Therefore, the acquisition of DSA images under natural respiration is required.

Previously, our group proposed a respiratory phase matching method for respiratory-synchronized DSA acquisition under natural respiration [1]. In that method, a set of contrast images covering a respiration cycle are captured preoperatively and then subtraction is performed between arbitrary target intraoperative image without contrast agent and the contrast image which is most similar to the target image. In this technique, however, the artifacts tend to appear due to the insufficient respiratory phase matching between contrast and mask images or the organs' movement unrelated to the respiration.

In this paper, we propose a different subtraction scheme and also modify our previous method. First, we assume that a set of mask images covering at least one respiration cycle are captured prior to live imaging with contrast agent. Next, as modification of the method, the image matching between for each live image (contrast image) and a set of mask images is performed on the basis of patch or sub-block.

Methods

The proposed method consists of the following three steps. Let $(\boldsymbol{M}^1, \boldsymbol{M}^2, \ldots, \boldsymbol{M}^S) \in \mathbf{R}^{S\times H\times W}$ represent a set of mask images covering one respiratory cycle which are acquired preoperatively. Here S represents the number of images in the set, and H and W represent the height and width of each image, respectively. Furthermore, let $\boldsymbol{C} \in \mathbf{R}^{H\times W}$ represent a certain contrast image during live image acquisition.

Step 1. For the image $\boldsymbol{C}$ and all images $\boldsymbol{M}$, we split each image into sub-blocks with the size of $H/N \times W/N$ pixels, generating sub-blocks $\{\boldsymbol{c}_{1\times1}, \ldots, \boldsymbol{c}_{i\times j}, \ldots, \boldsymbol{c}_{N\times N}\}$and $\{(\boldsymbol{m}^1{}_{1\times1}, \ldots, \boldsymbol{m}^1{}_{i\times j}, \ldots, \boldsymbol{m}^1{}_{N\times N}), (\boldsymbol{m}^2{}_{1\times1}, \ldots, \boldsymbol{m}^2{}_{i\times j}, \ldots, \boldsymbol{m}^2{}_{N\times N}), \ldots, (\boldsymbol{m}^S{}_{1\times1}, \ldots, \boldsymbol{m}^S{}_{i\times j}, \ldots, \boldsymbol{m}^S{}_{N\times N})\}$. Parameter

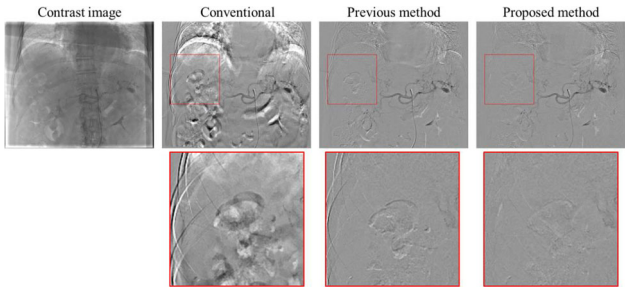

Fig. 1 DSA images and zoomed images generated by the conventional method, the previous method, and the proposed method

N is the number of patch splits in each direction, and (i, j) represents the location of the sub-block ($i, j = [1, N]$). We select the $\boldsymbol{m}^{k}_{i\times j}$ with the highest similarity to each $\boldsymbol{c}_{i\times j}$ from $\{\boldsymbol{m}^{1}_{i\times j}, \boldsymbol{m}^{2}_{i\times j}, \ldots, \boldsymbol{m}^{S}_{i\times j}\}$ using a pattern matching technique. Zero normalized cross-correlation (ZNCC) is used as a similarity measure.

Step 2. Fine tuning is performed in this step. Let $\boldsymbol{m}^{k}_{i\times j}$ be the sub-block selected as the most similar to $\boldsymbol{c}_{i\times j}$ in *Step 1*. In step 2, we search for a more similar sub-block than $\boldsymbol{m}^{k}_{i\times j}$ by shifting two dimensionally in the sub-block area in $\boldsymbol{M}^{k}$. The downhill simplex method is used for this search.

Step 3. We repeat the above process for all sub-blocks and generate whole DSA images. In addition, we perform brightness correction for all sub-blocks when a sub-block has extremely different brightness compared to the surrounding sub-blocks.

We conducted comparative experiments using contrast and mask images obtained from 11 patients in Chiba University Hospital. We generated DSA images by the simple subtraction with a fixed mask image (conventional method), the previous method (Ohnishi et al. [1]), and the proposed method. The image size, $H \times W$, was 512 × 512 pixels, with 8-bit pixel depth. The frame rate of both contrast and mask images was 5.0 frames per second. The number of a set of mask images, S, was approximately 30. We set N be 10.

Results

Figure 1 shows DSA images and zoomed images generated by each method. The previous method visualized blood structure clearly, but some artifacts remained in bone and intestinal tract regions. The proposed method, on the other hand, suppressed artifacts around them. We confirmed that the proposed method achieved superior performance than the previous method for all cases.

Conclusion

We proposed a patch-based respiratory phase matching method for DSA under natural respiration. The experimental results demonstrated that the proposed method effectively suppressed artifacts compared with the previous method. For improvement of image quality, the choice of similarity measure and the way of brightness correction will be further investigated.

References

[1] Ohnishi T, Takano Y, Kato H, Ooka Y, Haneishi H (2018) Respiratory-synchronized digital subtraction angiography based on a respiratory phase matching method. Signal, Image and Video Processing, 12(3): 539–547

Automated detection of knee joint from X-ray fluoroscopic images based on deep learning and accuracy validation of 2D/3D registration

T. Yamazaki[1], H. Hayashida[1], F. Itami[1], T. Tomita[2], K. Sugamoto[2]

[1]Saitama Institute of Technology, Information Systems, Fukaya, Japan [2]Osaka University Graduate School of Medicine, Orthopaedic Biomaterial Science, Suita, Japan

Keywords Automated detection, Knee joint, Deep learning, 2D/3D registration.

Purpose

Accurate and quantitative assessment of 3D dynamic knee kinematics under in vivo conditions is very important for diagnosis, treatment, and surgical planning of various joint diseases and dysfunction. For artificial knee implants, to achieve 3D measurement of the dynamic kinematics, 2D/3D registration techniques which use X-ray fluoroscopic images and computer-aided design model of the implants have been applied to clinical cases. These fluoroscopy-based techniques have also been applied for motion measurement in knee joints without implants, where 3D bone models created from CT or MRI images are utilized. In most conventional techniques, however, the 3D knee kinematic measurement process requires some manual operations (for example, setting the region of interest in knee joint images, and setting initial pose for 2D/3D registration, etc.) and is still labor-intensive and time-consuming work.

In labor-intensive manual operations, to automatically detect and classify the knee joint (femur, tibia/fibura and patella) from X-ray fluoroscopic image is very important and meaningful. Such an automatic detection of the knee joint is thought to be also useful for improving the accuracy of 2D/3D registration. In this study, therefore, we present automated detection methods of knee joint from X-ray fluoroscopic images using deep learning. In addition, we validate the accuracy of 2D/3D registration using the detected each knee image under in vivo conditions.

Methods

In this study, for detection and classification of knee joint (femur, tibia/fibura and patella) from X-ray fluoroscopic images based on deep learning, two object detection methods, single shot multibox

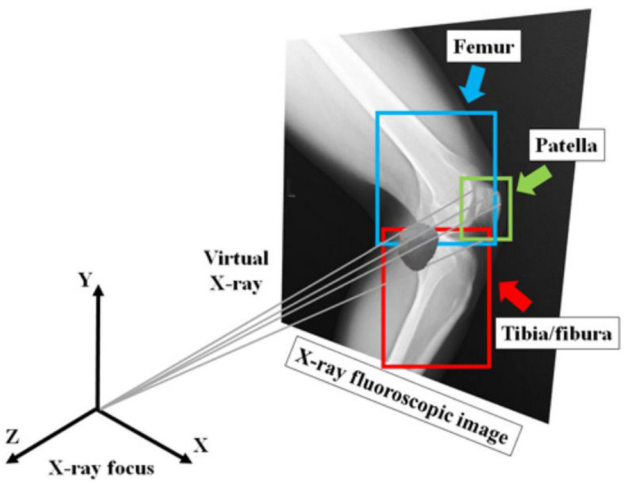

Fig. 1 Pose estimation of each knee bone model using intensity-based 2D/3D registration

detector (SSD) and you only look once (YOLO) were utilized. In addition, for pose estimation of femur, tibia/fibura and patella from the detected each knee image, an intensity-based 2D/3D registration technique [1] was utilized (Fig. 1).

In order to investigate the performance of detection and classification of each knee joint (femur, tibia/fibura and patella) from X-ray fluoroscopic images based on deep learning (SSD and YOLO), a total of 1222 X-ray fluoroscopic images of actual knee flexion movements (14 cases) were used. All images used were labeled (annotated) with femur, tibia/fibula, and patella, and the performance was evaluated with a tenfold cross validation test. The detection accuracy for each knee joint was determined by calculating the Intersection over Union (IoU) based on the position estimated by each method (SSD and YOLO) and the position of the correct value determined in advance on the X-ray image. In this study, the performance was evaluated using the detection rate when the IoU threshold was 0.8.

In the accuracy validation of the pose estimation of femur, tibia/fibura and patella using the intensity-based 2D/3D registration, experiments using actual X-ray fluoroscopic images (one case of knee flexion movement) were conducted. For the correct pose (reference data), we used 3D pose data obtained by giving appropriate initial pose through manual operation and carefully applying the intensity-based 2D/3D registration. Initial guess poses of each knee bone model were randomly given from within $\pm$ 3 mm and $\pm$ 3 degree of the correct pose. Errors in the 3D pose of the model were determined by comparing the estimated pose to the known pose (the correct pose). The translation error in the Z direction (perpendicular to the image plane) was not evaluated in this experiment because this parameter is not used in the actual knee motion analysis. In this study, the accuracy of the intensity-based 2D/3D registration was verified in three cases, case using the image without preprocessing, case using the image of regions detected by SSD, case using the image of regions detected by YOLO.

Results

As results of the detection rate for each knee joint (femur, tibia/fibura and patella) using each method (SSD and YOLO), the SSD method showed an extremely high detection rate (almost 100%) for all knee joints, while the YOLO method showed a slight decrease in detection rate for the patella (98%).

The results of the accuracy of the intensity-based 2D/3D registration are summarized in Table 1. The root-mean-square errors are given for each image case and for each knee bone model. In all cases, the pose estimation accuracy of the patella was particularly poor.

Conclusion

In this study, with the aim of reducing the labor-intensive and time-consuming work for the 3D knee kinematic measurement process based on 2D/3D registration using X-ray fluoroscopic images, automated detection methods of knee joint (femur, tib-ia/fibura and patella) from X-ray fluoroscopic images using deep learning (SSD and YOLO) were presented. In addition, the accuracy of the intensity-based 2D/3D registration using the detected each knee image under in vivo conditions was validated.

In the result of experiment for the performance of detection and classification of each knee joint, the SSD method showed a very high detection rate (almost 100%) for all knee joints, while the YOLO method showed a slight decrease in detection rate for the patella. The reason for this is that YOLO method uses Darknet-53, a network with a large number of layers, and it is thought that patella with relatively small shapes lose features during feature extraction, leading to poor detection.

In the result of experiment for accuracy validation in the pose estimation of each knee joint using the intensity-based 2D/3D registration, in particular, the patella was found to be inaccurate and unstable (see Table 1). While, for the femur and tibia/fibura, the accuracy was high enough to allow 3D kinematic analysis using actual X-ray fluoroscopic images, especially when using images of regions detected by SSD. As a future issue, it is necessary to investigate various image conditions for the purpose of fully automated 3D knee kinematic measurement.

Table 1 Root-mean-square errors of pose estimation for each image case and for each knee bone model

Image used	Bone model	Rotation (°)			Translation (mm)		
		X	Y	Z	X	Y	Z
Image without preprocessing	Femur	1.10	3.94	1.11	0.67	1.14	–
	Tibia/fibura	1.92	1.52	1.37	0.48	0.73	–
	Patella	5.87	3.67	4.85	1.02	1.82	–
Image of regions detected by SSD	Femur	1.12	2.03	1.54	0.80	1.10	–
	Tibia/fibura	1.27	0.98	1.17	0.48	0.46	–
	Patella	4.20	13.44	9.66	7.18	10.98	–
Image of regions detected by YOLO	Femur	1.78	2.26	3.05	0.87	2.52	–
	Tibia/fibura	1.25	0.94	1.03	0.42	0.46	–
	Patella	4.16	5.55	8.87	1.91	4.58	–

References

[1] Penney GP, Weese J, Little JA, Desmedt P, Hill DLG, Hawkes DJ (1998) A comparison of similarity measures for use in 2-D-3-D medical image registration. IEEE Trans Meg Imag 17(4):586–595

3d-Printed Patient-specific Perfused Aortic Dissection Model for Training of Endovascular Interventions

L. Mohl[1], R. F. Karl[1], A. Runz[2], M. Hagedorn[3], M. Tölle[1], J. Hatzl[3], K. Meisenbacher[3], C. Uhl[3], D. Böckler[3], S. Engelhardt[3]

[1]Heidelberg University Hospital, Working Group Artificial Intelligence in Cardiovascular Medicine, Heidelberg, Germany [2]German Cancer Research Center (DKFZ), Division of Medical Physics in Radiation Oncology, Heidelberg, Germany [3]Heidelberg University Hospital, Department of Vascular and Endovascular Surgery, Heidelberg, Germany

Keywords 3D-Printing, Patient-specific model, Aortic dissection, Endovasculular Training.

Purpose

Aortic dissections are the most common cause for emergency operations at the aorta. Affecting 7 in 100.000 inhabitants they are more than twice as common as ruptured aortic aneurysms. An aortic dissection Stanford Type B is characterized by a proximal entry distal to the orifice of the left subclavian artery in the descending part of the aorta. Blood flows through this entry tear into the wall thereby creating a true and a false lumen separated by the dissection membrane. The false lumen can grow and eventually compress the true lumen causing organ and leg malperfusion. The treatment of aortic dissections requires a lot of experience in distinguishing the true and false lumen when steering the catheter, since a misplaced stentgraft can be fatal for the patient.

The goal of our research is to develop a simulator for intervention training and treatment planning on aortic dissections to provide surgeons with patient-specific training opportunities. However, due to the fragile and complex anatomy, aortic dissection models have been challenging to manufacture. Previously, only few models were ever made, mostly using complex silicone injection processes not easily accessible [1]. Zimmermann et al. [2] showed the possibility to 3d print a model of an aortic dissection, though their model does not incorporate abdominal outflows. In this abstract we present a 3d-printed patient-specific anatomical aortic model including abdominal branching vessels and a thin, flexible dissection flap. The main goals were as follows: First, the models should work in a hemodynamic environment in a physiological pressure range. Secondly, the movement of the dissection flap should be observable via ultrasound. Thirdly, the model should be suitable for stentgraft placement.

Methods

A CT angiography dataset of a patient with an acute Stanford type B dissection was selected, it featured one main entry and one main reentry. The dissection ended shortly above the bifurcation. The aorta was segmented including the dissection flap and major aortic outflows (brachiocephalic trunk, left common carotid artery, left subclavian artery, celiac trunk, superior mesenteric artery, renal arteries, common iliac arteries). The voxel-based segmentation was then converted into a 3d-surface-model. After cleanup the patient-specific aortic model was printed on an industrial Stratasys Objet 500 Connex 3 3D printer (Stratasys Inc., Rechovot, Israel) using a flexible rubber material (Tangoplus) which is known to reflect the elastic wall properties of the aorta. To provide realistic extravascular pressure onto the model during flow, it got embedded in food grade gelatin which also provided ultrasound visibility. The model was then connected to a pulse duplicator (Vivitro Labs Inc., Victoria, Canada). In this first setup the flow and pressure in the system were measured and the movement of the dissection flap was recorded on five previously defined positions along the dissected aorta. In a second step an already deployed stentgraft with a length of 10 cm (CTAG, W.L. Gore & Associates, Flagstaff, AZ, USA) was manually repackaged using an inhouse developed, 3d-printed applicator and released inside the model to cover the main entry. Stentgraft position, flow and

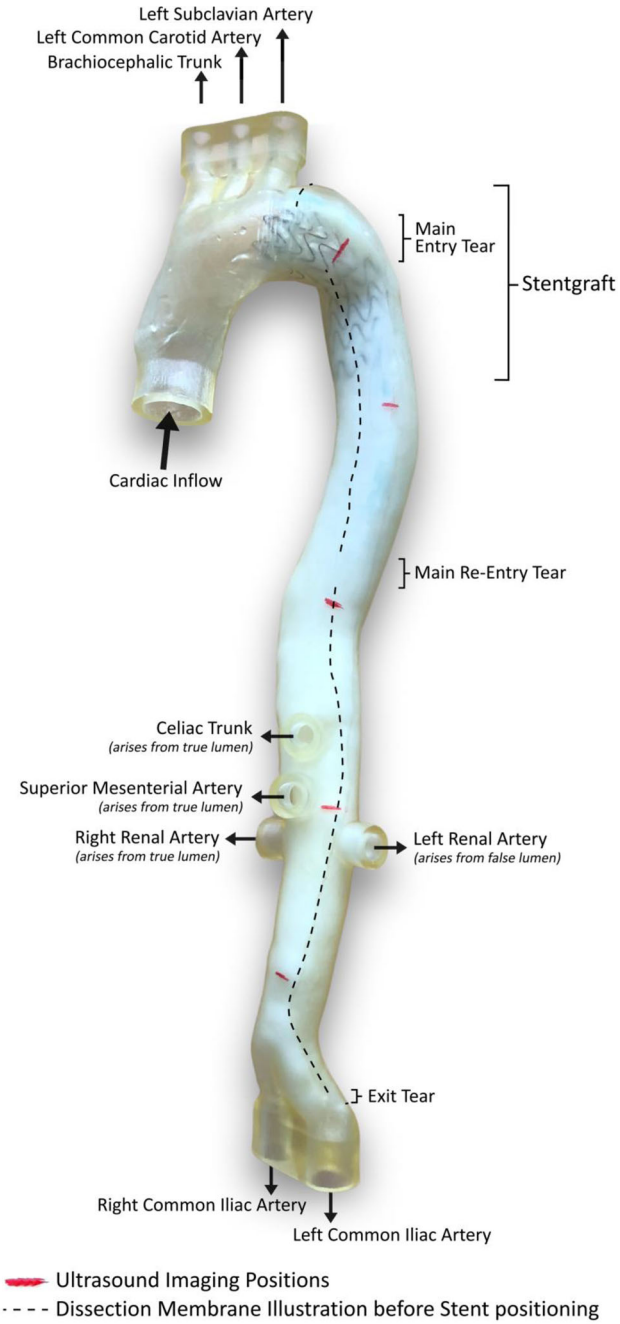

Fig. 1 3d-printed model with released stent

movement changes of the dissection flap were observed on the previously defined positions.

Results

With the model connected to the pump via a fluid circuit a systolic blood pressure between 100 and 110 mmHg was applied. The measured total average flow was around 5.9 l/min. With a heart rate of 60 bpm all three parameters are well within a physiological range. On all vessel outlets the individual flow rates were recorded as a baseline for later comparison.

In the ultrasound the dissection flap was clearly visible and moving in all five defined positions. True and false lumen were clearly distinguishable.

For this proof-of-concept study, the model was removed from the circuit and the stentgraft was successfully released in the true lumen covering the primary entry and extending close to the left subclavian artery (see Fig. 1). The model was then reconnected to the pump.

After the intervention the total average flow had slightly dropped to 5.1 l/min. The ultrasound showed that the stentgraft had displaced the dissection flap to the aortic wall, thereby compressing the false lumen and re-extending the true lumen to nearly original size. Right below the stentgraft, the dissection flap curved back to its former position. Comparing the flap movement in two measuring positions below the stentgraft revealed a noticeably more pronounced flap movement after the intervention, indicating a change in pressures between the true and false lumen.

Comparing the different outflow volumes was especially interesting for the renal arteries since the left renal artery arose from the false lumen. Its average flow dropped from 0.61 l/min down to 0.35 l/min while the right renal artery arising from the true lumen jumped from 0.29 l/min to 0.7 l/min, further indicating an increased true lumen and decreased false lumen pressure after the intervention.

Conclusion

Our patient specific model incorporates major outlets of the aorta, combined with image-based entry and reentry tears and a thin, flexible dissection flap. It works in a hemodynamic environment under physiologic pressures. A stentgraft was successfully deployed in the model and changes in dissection flap movement and flow differences in aortic outlets were observed, reflecting shifts in pressure between true and false lumen. All in all, our models might enable future patient-specific intervention training and pre-procedural planning.

References

[1] Morris L, Tierney P, Hynes N, Sultan S (2022) An in vitro Assessment of the Haemodynamic Features Occurring Within the True and False Lumens Separated by a Dissection Flap for a Patient-Specific Type B Aortic Dissection. Frontiers in cardiovascular medicine 9: 797,829. https://doi.org/10.3389/fcvm.2022.797829

[2] Zimmermann J, Bäumler K, Loecher M, Cork TE, Kolawole FO, Gifford K, Marsden AL, Fleischmann D, Ennis DB (2021) Quantitative Hemodynamics in Aortic Dissection: Comparing in Vitro MRI with FSI Simulation in a Compliant Model. In: Ennis DB, Perotti LE, Wang VY (eds) Functional Imaging and Modeling of the Heart. Springer International Publishing, Cham, pp 575–586

Unsupervised Segmentation of Fetal Brain MR Images Using Multi-Atlas Segmentation and Cascaded Registration

V. Comte[1], M. Alenya[1], A. Urru[1], A. Nakaki[2], F. Crovetto[2], O. Camara[1], E. Eixarch[2], F. Crispi[2], G. Piella[1], M. Ceresa[1], M. A. González Ballester[1]

[1]Universitat Pompeu Fabra, Department of Information and Communication Technologies, Barcelona, Spain [2]Maternal Fetal Medicine, BCNatal, Center for Maternal and Neonatal Medicine, Barcelona, Spain

Keywords Fetal brain, MRI, Registration, Segmentation.

Purpose

Deep Learning is now well established as the most efficient method for medical image segmentation. Yet, it requires large training sets and ground-truth labels, annotated by clinicians in a time-consuming process. We propose an unsupervised segmentation method using multi-atlas registration. The architecture of our registration model is composed of cascaded networks that produce small amounts of displacement to warp progressively the moving image towards the fixed image. Once trained, this model can be used to register multiple annotated magnetic resonance (MR) fetal brain images with a new target image, combining the propagated labels to form a refined segmentation. Our experiments show that our segmentation method produces results similar to one of the most robust state-of-the-art segmentation methods, without the need for labeled training data.

Accurate tissue segmentation is a crucial first step in the analysis of fetal brain morphology, as it allows for the identification of specific regions of interest that may be affected in cases of abnormal brain development. Numerous deep learning approaches for medical image segmentation have been proposed, many of which utilize convolutional neural networks (CNNs) with an encoder-decoder structure, such as U-Net [2]. Before deep learning techniques became the dominant method for image segmentation, multi-atlas segmentation (MAS) was often considered the most accurate approach. MAS involves the use of multiple atlases, or reference images, which are registered to the target image and used to guide the segmentation process. One of the key advantages of MAS is that it can be used in an unsupervised manner. This is particularly useful in cases where ground-truth labels are difficult to obtain or where the number of available training examples is limited. In general, the quality of the segmentation results is directly related to the accuracy of the registration. In recent years, deep learning DL-based registration methods have been proposed as a faster alternative to classical methods, achieving similar results in a shorter amount of time. These (DL)-based approaches are typically based on UNet-like networks and use the intensity-based similarity metrics as loss function. In this work, we propose a multi-atlas segmentation method based on DL cascaded registration, which includes:

- A novel deep learning-based image registration method that utilizes cascaded networks to predict a series of deformation fields. These deformation fields are then combined to align the moving image with the fixed image.
- A contracted architecture of VoxelMorph [1] for the cascaded networks, which reduces the memory cost while achieving similar results, allowing to use more cascades.
- A multi-atlas segmentation (MAS) method using our registration network, which obtains similar results than nnU-Net [2], without using labeled data for training.

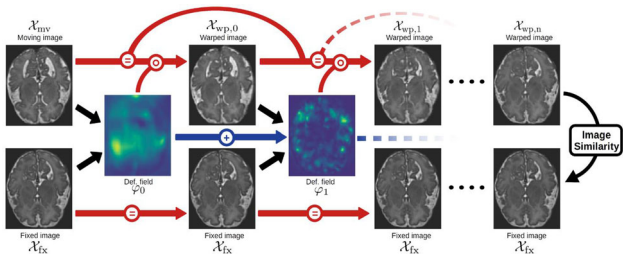

Fig. 1 Overview of the cascaded registration model, /home/valentin/Pictures/model.png

Methods

Cascaded registration

We present a cascaded registration model that utilizes multiple networks to generate successive deformation fields, which are then combined to warp the atlases into alignment with the target image. This cascaded approach allows us to divide the registration process into smaller, simpler transformations that can be processed at various spatial levels. The general architecture of our model is shown in Fig. 1, the first network takes the moving and fixed images, Xmv and Xfx, as input and produces a dense deformation field φ0 that partially aligns the moving image with the fixed image, forming Xwp,0. The second network then inputs Xwp,0 and Xfx, and generates φ1, which is summed with φ0 to warp Xmv into Xwp,1. This process is repeated with successive networks in a recursive manner.

Multi-atlas segmentation

Following the cascaded registration method presented in the previous section, we propose a segmentation method based on multi-atlas registration. This approach involves using a series of atlas images, each with its own associated ground-truth labels, and registering them to the target image in order to create an accurate segmentation. Once the atlas images have been registered to the target, we select the ones that align most closely based on the average local cross-correlation between the atlas and target images. Finally, we combine the labels from the selected atlas images using a local weighted voting strategy, which consists of propagating the labels of the warped images based on a weighting strategy, giving more weight to the labels corresponding to the highest local similarity.

Dataset

We used a dataset of 170 fetal brain Magnetic Resonance Images between 32 and 37 gestational weeks. The scans were aquired in Hospital San Joan de Déu and Hospital Clínic of Barcelona. The dataset was split into 140–10-20 for train, validation and test sets, respectively. As a preprocessing step, the images were cropped to remove the non brain regions, resized to 128 × 128 × 128 voxels, and normalized between 0 and 1.

Results

We evaluated the performance of our cascaded registration model on the IMPACT dataset. The average Dice score obtained by our model was 0.86 ± 0.02. For comparison, the average Dice score obtained by VoxelMorph [1] was 0.81 ± 0.02, when trained and tested on the same dataset. The table presents the results of our multi-atlas segmentation method. In order to compare the performance of our method to a state-of-the-art deep learning-based approach, we also trained nnU-Net [2] on the same training set and included the resulting Dice scores in the table. The results demonstrate that our method is able to achieve results that are comparable to nnU-Net in terms of segmentation accuracy.

Label	Ours	nnU-Net
CSF	**0.923 ± 0.006**	0.897 ± 0.011
Grey matter	**0.877 ± 0.006**	0.871 ± 0.012
White matter	0.915 ± 0.006	**0.919 ± 0.006**
Ventricules	**0.902 ± 0.005**	0.879 ± 0.011
Cerebellum	0.964 ± 0.004	**0.965 ± 0.004**
Thalamus	0.950 ± 0.002	**0.957 ± 0.004**
Brain stem	**0.955 ± 0.003**	0.953 ± 0.006
Average	**0.926 ± 0.012**	0.920 ± 0.014

Conclusion

In this work, we introduce a registration model based on cascaded networks for multi-atlas segmentation in medical imaging. Our model decomposes the registration field into a series of simpler transformations that operate at different spatial scales, allowing for more efficient and accurate alignment of the atlases with the target image. Our results show that our approach outperforms state-of-the-art (DL)-based registration method [1]. In addition, the resulting multi-atlas segmentation method achieves performance similar to one of the most robust state-of-the-art segmentation method [2], despite not requiring any labeled training data. This is a significant advantage, as annotated fetal MR images can be difficult to obtain.

Acknowledgment

This publication is part of the project PCI2021-122044-2A, funded by the project ERA-NET NEURON Cofund2, by MCIN/AEI/10.13039/501100011033/ and by the European Union "NextGenerationEU"/PRTR. G. Piella is supported by ICREA under the ICREA Academia programme.

References

[1] Balakrishnan G, Zhao A, Sabuncu MR, Guttag JV, Dalca AV (2018) Voxelmorph: A learning framework for deformable medical image registration. CoRR, abs/1809.05231

[2] Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, Wasserthal J, Koehler G, Norajitra T, Wirkert S, Maier-Hein KH (2018) nnu-net: Self-adapting framework for u-net-based medical image segmentation, arXiv:1809.10486

Construction of organ rotation estimation system using deep learning

M. Sano[1], N. Koizumi[1], Y. Nishiyama[1], J. Zhou[1], T. Fujibayashi[1], M. Matsuyama[1], M. Yamada[1], T. Ishikawa[1], A. Katsuragi[1], S. Monma[1]

[1]The University of Electro Communications, Graduate School of Informatics and Engineering, Chofu-Shi, Japan

Keywords HIFU therapy, Organ movement estimation, Segmentation, Robot diagnosis and treatment.

Purpose

In recent years, many people around the world have died from malignant tumors. HIFU (High Intensity Focused Ultrasound) therapy, a non-invasive treatment method, has been attracting attention as a treatment method for this problem. In HIFU therapy, ultrasound waves must be focused and continuously irradiated onto the affected area for a certain period of time, whereas the abdominal organs move relatively because the abdomen moves in accordance with respiration (hereinafter referred to as "body movements"). This makes incineration of the affected area difficult, and there is a risk of damaging surrounding healthy tissues around the affected area. Therefore, in order to realize safe and effective HIFU therapy of abdominal organs, it is necessary to track and follow the movement of organs due to body movements, and methods should be investigated [1]. In this report, we aim to track organs by estimating 3D organ movements caused by respiration in advance and utilizing a ultrasound diagnostic robot for HIFU therapy to cope with the abovementioned problem.

Methods

We conducted experiments using the Robotics Ultrasound Diagnostic System (RUDS) developed in our laboratory for image acquisition and robotic manipulation, using Kyoto Chemical Corporation's ultrasound phantom (ABDFAN) as the imaging target. In the proposed method, one cycle of echograms is acquired with the probe fixed in RUDS. The motion of organs is estimated by performing this operation on multiple cross sections. The motion of organs due to body movement can be divided into translation and rotation. We have calculated the amount of translational movement in three dimensions by calculating and combining the optical flow from the echograms in two cross sections. In this report, the amount of rotational translation was calculated by combining the proposed method with orientation estimation by AEMAD + +, which has been performed in our laboratory [2].

Specifically, we acquired echograms for 90 degrees with breathing stopped, and extracted the area, diagonal length, vertical length, and horizontal length of the segmented organs as feature values. This was also conducted for orthogonal cross sections, and the two were combined and stored as a database. Afterwards, body movements for one cycle were imaged in the same cross section as when the database was constructed, and the orientation were estimated by referring to the database using approximate nearest neighbor search.

We constructed the system based on U-Net for organ segmentation. 469 images were used for training, 151 images for validation, and 60 images for the test data set. ResNet50 was utilized as the encoder. IoU and Dice Loss were used as evaluation metrics, and Adaptive Moment Estimation (Adam) was applied for optimization.

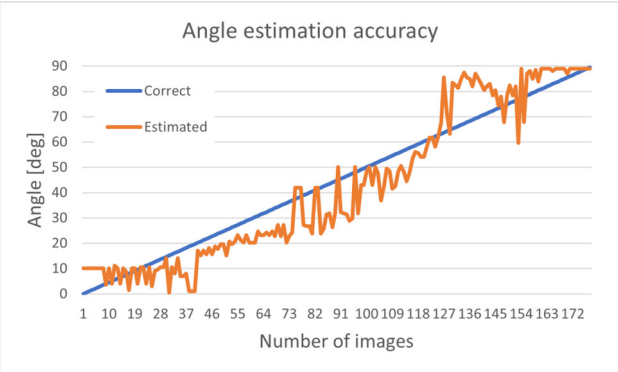

Fig. 1 Experimental result of angle estimation

Table 1 Angle estimation accuracy at each angle

Allowable error	± 0 deg (%)	± 2.5 deg (%)	± 5 deg (%)
0–10	0	25	40
10–20	5	20	45
20–30	0	0	30
30–40	0	0	15
40–50	0	16	16
50–60	0	15	40
60–70	5	20	35
70–80	0	5	20
80–90	5	37	63

Results

First, the segmentation results were 0.9737 for IoU and 0.0144 for Dice Loss. Second, the results of angle estimation are shown in Fig. 1. Table 1 shows the accuracy when the allowable error was set to ± 0 deg, ± 2.5 deg, and ± 5 deg. Although the accuracy at each angle was lower than in the previous study [2], it was confirmed that the accuracy in angle estimation at the short-axis image (50 ∼ 90 deg), which had been a weak point, was improved. One of the reasons for the low accuracy is that the size of the organs differs between the database and the test images. Therefore, we plan to investigate the accuracy when applying relative values based on one of the features.

Conclusion

The aim of this study is to estimate 3D organ motion due to breathing using an ultrasound robot. We acquired ultrasound image with body movements in two cross sections, and the rotational movement of organs was estimated by modifying the method of a previous study. In our method, the two cross sections are not necessarily orthogonal. The results showed decrease in accuracy in the long-axis image compared to the previous study. However. it shows improvement in accuracy in the short-axis image. We will continue to improve the performance of our method so as to be applied in the clinical use.

References

[1] Arnold P, Preiswerk F, Fasel B, Salomir R, Scheffler K, Cattin PC (2011) 3d organ motion prediction for mr-guided high intensity focused ultrasound. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 623–630. Springer, 2011.

[2] Fujibayashi T, Koizumi N, Nishiyama Y, Watanabe Y, Zhou J, Matsuyama M, Yamada M, Tsumura R, Yoshinaka K, Matsumoto N, Tsukihara H, Numata K (2022) A Study on Deep Learning for Automatic Diagnosis and Treatment Robots Automatic Diagnosis and Treatment Robot, CARS 2022—Computer Assisted Radiology and Surgery Proceedings of the 36th International Congress and Exhibition Tokyo, Japan, June 7–11, 2022. Int J CARS 17 (Suppl 1), 1–147 (2022). https://doi.org/10.1007/s11548-022-02635-x

Patient positioning assistance applicable to existing computed tomography systems

A. Urikura[1,2], Y. Miyauchi[2], T. Yoshida[2], Y. Ishita[2], K. Takiguchi[2], T. Aramaki[2]

[1]National Cancer Center Hospital, Radiological Technology, Tokyo, Japan [2]Shizuoka Cancer Center, Diagnostic Radiology, Shizuoka, Japan

Keywords Computed tomography, Localizer radiograph, ATCM, 3D camera.

Purpose

Patient positioning in computed tomography (CT) examinations is performed manually by the operator. Adjusting of the table height correctly is important for optimizing patient dose because inaccurate positioning in the height direction affects patient exposure dose. Some modern CT scanners can use automated positioning techniques with a three-dimensional camera (3D camera) system; assisted positioning techniques with 3D camera systems are useful to increase the throughput of CT examinations and to optimize patient exposure dose. However, automatic height adjustment capability using a 3D camera system has only been realized in a few modern CT scanners. The purpose of this study is to develop a patient positioning assistance technique using CT localizer radiographs instead of a 3D camera system.

Methods

This retrospective study was approved by the Institute's Review Board. The study included a total of 210 patients, divided into 127 regression studies and 83 validation studies.

Regression analysis between localizer image and ideal table height

A linear regression analysis was performed to determine the R^2 coefficient and the linear equation relating the mean pixel value of localizer radiograph image and ideal table height. To measure the mean pixel value corresponds to patient's anteroposterior (A–P) diameter, a 50 × 250 mm of rectangular region of interest was placed on the localizer image and measured pixel value. A–P diameter from axial section images were measured and averaged. A total of 51 axial images per patient were scanned automatically along the vertical direction, and upper- and lower-coordinates was recorded in each image; provided that the lower coordinate was defined as the lowest value of the coordinates recorded. The A–P diameter was obtained by multiplying the difference of the coordinate values by the pixel size. A–P diameter was defined by an averaged it from 51 images. Measurement of A–P diameter were parsed automatically using a homegrown Python script. The ideal table height (TH_{ideal}) was calculated using axial images.

Estimation of ideal table height

To estimate the ideal table height, the average pixel values of the localizer adiograph image were substituted into the regression equation obtained in the previous section.

Results

Linear regression analysis indicated high coefficient of determination (R^2 = 0.91) between mean pixel value of the localizer radiograph image and TH_{ideal}. The correlation coefficient between TH_{ideal} and the table height determined by the regression equation (TH_{est}) was 0.95, 95% confidence interval 0.92–0.97 ($P < 0.0001$), systematic bias was 0.2 mm, and limits of agreement were − 5.4–5.9 (P = 0.78). The offset of the table height with TH_{est} indicated 2.8 ± 2.1 mm, Fig. 1.

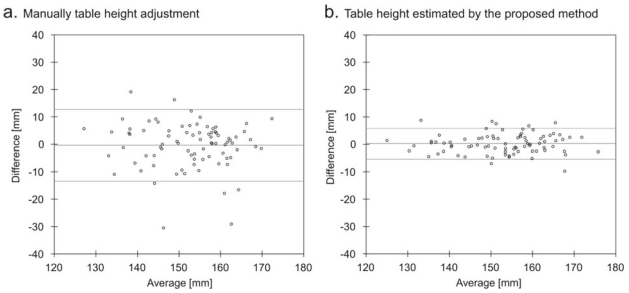

Fig. 1 Bland–Altman plot of the table height for manually adjusted by the operator (**a**) and estimated by our method (**b**)

Conclusion

This study showed important results for a technique that assists accurate table height setting for CT examinations. Notably, our method requires no additional hardware such as 3D camera system, to optimize the table height. Our results surpassed the positioning accuracy of CT systems equipped with a 3D camera. The proposed estimation method using localizer radiograph images could improve the automatic optimization of table height in CT and does not require additional hardware and can be a general-purpose automatic positioning technique.

Evaluation of hand tracking accuracy of HoloLens 2 for palpation during closure of ventricular septal defects

A. Bordin[1], G. Tibamoso-Pedraza[1], J. Miró[2], L. Duong[1]

[1]Ecole de technologie superieure, Department of software and IT engineering, Montreal, Canada [2]CHU Sainte-Justine, Department of Pediatrics, Montreal, Canada

Keywords Ventricular septal defects, Augmented reality, Hololens 2, Hand tracking.

Purpose

3-D printed virtual heart models can facilitate navigation guidance during a hybrid procedure to close ventricular septal defects (VSD) [1]. However, interventional cardiologists lose direct observation of the patient"s heart when they focus on the virtual models to manipulate their surgical instruments. HoloLens 2 and other augmented or mixed reality technology has been proposed for image-guided interventions, mainly in interventional workflow to superimpose pre or intra operative image directly onto the patient, to see the target. Recent advances in computer vision allowed significant improvements of the tracking of the eyes and the tracking of the hand for such application. This study evaluates the hand tracking accuracy of the augmented reality helmet HoloLens 2, for augmented reality in interventional cardiology specifically for manual palpation of the heart. The purpose of this project was to determine the accuracy of HoloLens2's detection system, first in terms of movement detection, but also in terms of object tracking and detection, in the context of a surgical intervention.

Methods

For movement detection, we focused on hand tracking, specifically the thumb and the index, which are the two mainly used fingers in our context. For object detection, we want to determine if the HoloLens 2 can be used as a real-time tool tracking device. Ground truth measurements were obtained using an electromagnetic measurement

system (EMS) (NDI Aurora, Waterloo, Canada). This device, used in the medical field, allows to track any sensor linked to the device in real-time with an accuracy between 0.4 mm and 1.4 mm. The EMS' coordinate system's position in the real world is known, which will be useful for calibration.

A custom software was developed to record the fingers position in 3D of the HoloLens2. This application was developed using the Unity software with the Microsoft Mixed Reality Toolkit (MRTK) extension and the VisionLib library was chosen to test object tracking. To ensure that the HoloLens2 and the ground truth are measuring into the same 3D coordinate system, we chose to implement a QR code scanning system into the HoloLens2: by placing the QR code physically at the centre of the electromagnetic emitter, which corresponds to the origin of the EMS' coordinate system, and then scanning the code with the helmet, we can place a virtual 3D object oriented so that its x (horizontal), y (vertical) and z (depth) axes are aligned with the EMS ones, and its origin at the same global position as the EMS origin. We then use this object's local coordinate system to measure our coordinates. To acquire the data, users were asked to wear an electromagnetic sensor on the index and thumb and to wear the Hololens2 while our application is running. Simple tasks were defined and organized in different categories: (1) no hand movement, (2) simple hand movements and (3) using a custom 3-D printed object. Timestamps were used to synchronize both data stream in Python, and outliers were filtered out using a distance-between-points threshold.

The accuracy was computed using the Absolute Mean Error (AME) between the electromagnetic sensor and the HoloLens 2.

Results

Figure 1 shows the thumb's position in the x-axis over time during the first experiment (static hands with head movements), measured by both HoloLens 2 and the EMS. The time gap between data points is used to distinguish every gesture of our experiment. While there is a visible difference between the HoloLens 2's measures and the ground truth, the graphs' tendencies remain very close, showing that the helmet is robust in its detection. To calculate the accuracy, an error is defined as significant only above 1.40 mm, that being the minimum accuracy of our reference.

The overall accuracy ranged between 16.3 mm and 19.7 mm, which is slightly better than a similar study [2], which also measured HoloLens2's hand tracking accuracy (between 18.3 and 26.3 mm). First in terms of axis, the measurements in X (horizontal) axis often show poor accuracy versus the Y and Z axis. Between the thumb and the index, there is no notable difference in accuracy. Accuracy decreased when the hands were moving (− 2 mm), and even more when an object masked part of the fingers (− 20 mm).

We obtained better results than prior studies [2], but, as it is, an accuracy of nearly 2 cm is good for most usage but can't be accepted in the medical field for high precision interventions. Nonetheless, the experiment was done with only one test subject. If the error appears to be systematic with multiple users, we could consider calibrating the system and counterbalance that error. As it can be seen in Fig. 1, it looks like the error isn't random, so finding the source of error will hopefully allow to virtually manipulate the surgical site with precision. We could also use the EMS to assist the helmet and increase its accuracy.

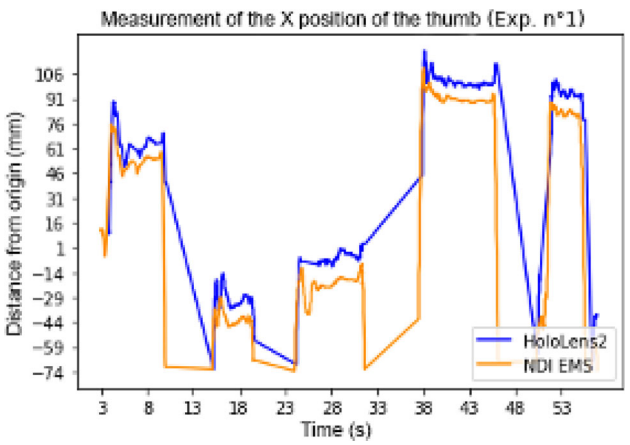

Fig. 1 Position measurement (x-axis) of the thumb during experience n°1

Conclusion

Mixed reality is still very promising as new technologies could emerge in the years to come and become a reliable asset in the medical field. While the accuracy is not suitable directly for direct navigation guidance, with proper calibration and with EMS guidance in the surgical tool and in the index of the cardiologist, the HoloLens 2 could provide some guidance to the target. As of future work, we will evaluate the HoloLens 2 with a larger sample of users. We will also integrate this system with the 3D printed heart models [1], and with ultrasound guidance in real-time, with gesture and 3D models of the heart in the same reference frame.

References

[1] Tibamoso-Pedraza G, Amouri S, Molina V, Navarro I, Raboisson M, Miró J., Lapiere C, Ratte S, Duong L (2022). Navigation guidance for ventricular septal defect closure in heart phantoms. International Journal of Computer Assisted Radiology and Surgery,17(10): 1947–1956

[2] Soares I, Sousa R, Petry M, Moreira, AP (2021). Accuracy and Repeatability Tests on HoloLens 2 and HTC Vive. Multimodal Technologies and Interaction, 5(8):47.

Post-hoc aleatoric uncertainty estimation through generalized Gaussian distribution for semantic segmentation in laparoscopic images

J. Qiu[1], Y. Hayashi[1], M. Oda[1,2], T. Kitasaka[3], K. Mori[1,4]

[1]Nagoya University, Informatics, Nagoya, Japan [2]Nagoya University, Information and Communications, Nagoya, Japan [3]Aichi Institute of Technology, Graduate School of Information Science, Toyota, Japan [4]National Institute of Informatics, Research Center for Medical Bigdata, Tokyo, Japan

Keywords post-hoc, uncertainty, segmentation, laparoscopic.

Purpose

This paper proposes a post-hoc aleatoric uncertainty estimation method for trained semantic segmentation networks. Our approach is an incremental work based on BayesCap [1], which estimates aleatoric uncertainty for trained neural networks in regression tasks (super-resolution, deblurring, etc.).

Semantic segmentation in the laparoscopic video is an important task to provide information for further computer-assisted surgeries (CAS) systems and it has been advanced by the power of neural networks in recent years. However, the predictions from the segmentation network could be inconsistent and unreliable due to some problems such as insufficient training or noise in training data.

Therefore, uncertainty estimation for network's outputs is crucial regarding the practical usage in CAS. There are some methods to estimate uncertainty such as minimizing Gaussian Negative Likelihood (GNLL) [2]. However, these methods require training networks from scratch. Besides, in most cases, since different institutes have developed their own segmentation network, retraining these networks is resource insufficient. Even though we have enough time and data to retrain these networks, it doesn't guarantee new networks with the same segmentation performance. These problems turn us to seek for post-hoc uncertainty estimation methods for segmentation networks.

One of the important previous works for post-hoc uncertainty estimation is called BayesCap [1]. BayesCap estimates aleatoric uncertainty by constructing a twin network that is identical to the trained network but with two extra branches and minimizes Generalized Gaussian Negative Log-likelihood (GGNLL) regarding their outputs. However, it is designed for regression tasks such as super-resolution. The post-hoc manner and strong performance of BayesCap make it a potential solution to previous problems. However, it is not trivial to extend BayesCap to segmentation networks due to the following reasons: (1) regression task requires minimizing GGNLL over networks" output logits while segmentation network outputs probability maps; (2) Segmentation networks usually have outputs with more than one channel (each channel per semantic class) while regression task generally has one-channel outputs.

To solve these problems, we propose a new post-hoc aleatoric uncertainty estimation framework for semantic segmentation in laparoscopic images with (1) computing GGNLL across all probability maps; (2) a fusion scheme utilizing multi-scale representations from the frozen network to maintain segmentation performance; (3) a light-decoder that can replace twin network in BayesCap to reduce parameters.

Methods

The proposed framework is based on BayesCap with some modifications. BayesCap first trained a regression network for a certain task, then freeze that network and build up a twin network (with a similar structure to the frozen network) at the end. By training the newly added parts with GGNLL, BayesCap can estimate uncertainty and a refined output. Like BayesCap, we first train our segmentation networks (segformer, b0 in our case, where its decoder utilizes multi-scale representations from the encoder.) with the training dataset of EndoVis18 (https://endovissub2018-roboticscenesegmentation.grand-challenge.org/) and save the trained network. Then the trained net work (hereinafter referred to as T_{fix}) is plugged into our framework. There is a light decoder on the top of T_{fix} whose inputs are multi-scale representations of T_{fix} and outputs are probability maps, uncertainty alpha (scale), and uncertainty beta (shape) respectively (we encourage readers to check the definition of GGNLL in [1] for details). Finally, we train other parts of our framework except T_{fix} with GGNLL to each channel of outputs. The followings are details of our framework:

· (1) Light-weight decoder with uncertainty estimation branches

Compared to BayeCap which uses a twin network as an uncertainty estimation branch, instead, we leverage a lightweight decoder on the top of T_{fix}. Its inputs are the outputs and multi-scale representations from T_{fix}. Two more decoder paths built with two convolution blocks are used for estimating uncertainty alpha and uncertainty beta. The reduction of parameters could speed up training and reduce hardware requirements.

· (2) Multi-scale Fusion

As mentioned above, the lightweight decoder utilizes multi-scale representations, and therefore a fusion module to mix these representations to the same scale is necessary. Hence, we added a fusion module by a sequence of several convolution blocks and downsampling to fuse representations from T_{fix} to fully utilize its information.

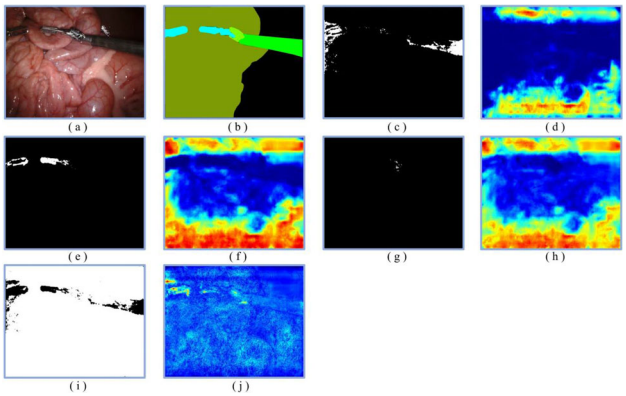

Fig. 1 Visualizations. **a** Laparoscopic image; **b** GT label images; **c** ∼ **d**, **e** ∼ **f**, **g** ∼ **h**, and **i** ∼ **j** are prediction labels and alpha uncertainty maps for the shaft (green), clasper (blue), wrist (light green) and small intestine (dark green). We use JET colormap to visualize alpha uncertainty and red means higher while blue means lower uncertainty

· (3) GGNLL over Prediction Probability Maps

In regression tasks, the outputs are usually one-channel therefore GGNLL can easily be applied. However, segmentation tasks contain multi-channel outputs. Therefore, we straightforwardly apply GGNLL to all output channels respectively.

Results

We followed the official EndoVis18, which is a laparoscopic image dataset, for training and testing. 10 semantic classes including shaft, clasper, wrist, parenchyma, covered kidney, thread, clamps, needle, intestine, and probe were used for evaluation. The segmentation performance (IoU, %) for the trained network was 63.02, 29.55, 31.66, 74.65, 44.66, 6.87, 25.58, 0, 71.12, and 4.51, respectively. After adding the uncertainty estimation branch, the segmentation performance became 53.04, 32.71, 25.86, 67.41, 51.48, 5.54, 33.36, 0, 67.74, and 5.87, which was a comparable performance to the frozen network. The visualization of uncertainty (alpha uncertainty) was shown in Fig. 1. From Fig. 1 we observed the prediction mask for each class was highly related to the low uncertainty region (e.g.: the prediction maps of each class have the lowest uncertainty in their corresponding uncertainty maps).

Conclusion

We investigated an approach for estimating aleatoric uncertainty based on BayesCap in the laparoscopic video. The visualization results showed the relation between segmentation regions and corresponding uncertainty. Further investigation will be on how to handle the temporal consistency within a laparoscopic video clip.

References

[1] Uddeshya U, Karthik S, Chen Y, Mancini M, Akata Z (2022) BayesCap: Bayesian Identity Cap for Calibrated Uncertainty in Frozen Neural Networks. European Conference on Computer Vision.

[2] Kendall A, Gal Y (2017) What uncertainties do we need in bayesian deep learning for computer vision? Advances in neural information processing systems 30.

Cold Ablation Robot-Guided Laser Osteotomy in Hand, Wrist and Forearm Surgery – a feasibility study

E. Coppo[1,2], M. Hofer[2], M. Morawska [3], M. Müller-Gerbl[4], P. Honigmann[1,2]

[1]Kantonsspital Baselland, Hand and peripheral Nerve Surgery, Liestal, Switzerland [2]Universität Basel, Medical Additive Manufacturing, Biomedical Eingineering, Basel, Switzerland [3]Advanced Osteotomy Tools AG, Basel, Switzerland [4]Universität Basel, Medizinische Fakultät, Basel, Switzerland

Keywords Robotic assisted surgery, Ablation laser, Forearm osteotomy, Hand surgery.

Purpose

Bone surgery is associated with direct contact of instruments with the bone causing friction, heat and pressure and hence, damaging the bone and surrounding soft tissues. We introduce the technology of navigated cold ablation robot-guided laser osteotomy, present potential applications, and preliminary pre-clinical cadaver test results in the field of hand-, wrist- and forearm surgery [1,2].

Methods

Technique

CARLO® is a miniaturized ablation laser with an optical system controlled by a navigation system. The system uses Yttrium Aluminum Garnet doped with Erbium (Er:YAG), with a wavelength of 2943 nm, which corresponds to peak absorption coefficient for water and hydroxyapatite. The energy of a laser pulse hitting the bone tissue heats up the water content of the bone and vaporizes it. The increase in local pressure causes "micro-explosions", breaking up the bone structure. The debris is being expelled immediately and at high velocity, providing a clean-cut line with preservation of the bone structure. CARLO® also has an Optical Coherence Tomography (OCT)-based depth control system to visualize the current cutting level and avoid soft-tissue damage.

Cadaver tests

We first evaluated if laser-osteotomies in long bones were possible. CARLO® was then used for corrective osteotomies of the distal metaphyseal radius, ulna as well of the metacarpal one (Wilson's Osteotomy). Standard surgical approaches were carried out and the navigation device mounted at the end of the surgical field. Different cutting patterns were applied and tested for precision and primary stability. The osteotomies were stabilized using lag screws only.

Results

Laser osteotomies in long bones were feasible using the OCT. Best cutting patterns were sine and sawtooth in terms of primary stability and precision, Fig. 1. Multiple plane cuts were performed to allow for corrections in the x, y, z plane. The bone cuts in dia-, meta and epiphyseal regions did not show any carbonisation. The lag screws provided good compression and stability.

Conclusion

First cadaveric results are promising. New bone cutting patterns are feasible which are not possible to carry out manually. CARLO® allows to perform corrective osteotomies without the use of patient-specific guides with high precision and probably less hardware. Future steps are stability testing of the osteosynthesis and application in smaller bones (carpal, metacarpal and digital bones) followed by certification and first use in patients.

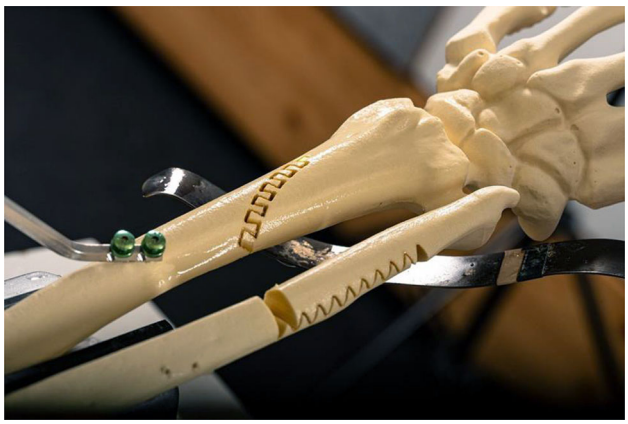

Fig. 1 Possible cut geometries on a 3D printed model

References

[1] Baek, K.; Deibel, W.; Marinov, D.; Griessen, M.; Dard, M.; Bruno, A.; Zeilhofer, H.; Cattin, P.; Juergens, P. A Comparative Investigation of Bone Surface after Cutting with Mechanical Tools and Er:YAG Laser. *Laser Surg Med* 2015, *47*, 426–432, https://doi.org/10.1002/lsm.22352.

[2] Matys, J.; Hadzik, J.; Dominiak, M. Schneiderian Membrane Perforation Rate and Increase in Bone Temperature During Maxillary Sinus Floor Elevation by Means of Er. *Implant Dent* 2017, *26*, 238–244, https://doi.org/10.1097/id.0000000000000520.

Effects of patient demographics on anatomical fitting of a distal femur plate: A 3D modelling study

B. Schmutz[1,2,3], M. T. Phan[1], J. Pople[2,4], B. C. W. Lam[2,4], E. Schoofs[2,4], J. Warren[2,5], H. Minehara[6,7], K. Tetsworth[2,8], M. Schuetz[2,8,5]

[1]Queensland University of Technology, School of Mechanical, Medical & Process Engineering, Brisbane, Australia [2]Jamieson Trauma Institute, Brisbane, Australia [3]Queensland University of Technology, ARC Training Centre for M3D, Brisbane, Australia [4]University of Queensland, Faculty of Medicine, Brisbane, Australia [5]Queensland University of Technology, School of Clinical Sciences, Brisbane, Australia [6]Fukushima Medical University, Department of Traumatology, Kawasaki-city, Japan [7]Shin-yurigaoka General Hospital, Trauma and Reconstruction Center, Kawasaki-city, Japan [8]Royal Brisbane and Women's Hospital, Department of Orthopaedics and Trauma Service, Brisbane, Australia

Keywords Anatomical Fit, Fracture Fixation Plates, 3D Models, Distal Femur.

Purpose

Lateral locking-plate fixation is commonly used for distal femur fractures, and pre-contoured plates attempt to match the bony anatomy of the target patient population. However, plate fit is highly variable due to inter-subject morphological differences. Average proximal plate misfit of 11.4 mm has been reported for Korean cadaver femora [1]. Digital plate templating on pre-operative total knee arthroplasty (TKA) x-rays demonstrated an average metaphyseal misfit of 6.6 mm [2]. The plate is often used as a reduction template;

if plate misfit is not recognised and addressed, this can result in axial malalignment. This study evaluates the effects of age, sex, height, and ethnicity on the anatomical fit of the LCP Distal Femur plate, to better inform the clinical community and aid future implant shape optimisation.

Methods

Unilateral 3D bone models of 70 (35 male, 35 female) Caucasian and 60 (30 male, 30 female) Vietnamese femora were utilised from earlier studies. Mean age and height were 58 years (range 22–96) and 166 cm (range 150–193) for Caucasians, and 58 years (range 31–84) and 160 cm (range 144–176) for Vietnamese. Both cohorts were separated into subgroups of young (< 65 years) and old (≥ 65 years) subjects, with groups of 33 young and 37 old Caucasian, and 30 young and 30 old Vietnamese.

Plate undersurface and screw trajectories were obtained from reverse engineered 3D scans of an 11-hole and a 13-hole LCP Distal Femur plate (DePuy Synthes, Switzerland). Plate length was selected such that the proximal tip was ≥ 2 cm away from the lesser trochanter. Through an iterative process using reverse engineering software (RapidForm2006, Inus Technology, Korea), the undersurface of the plate was positioned correctly on the bone model (Fig. 1).

The undersurface was positioned to achieve plate-bone contact at three locations: distal anterior, distal posterior, and shaft [1]. It was ensured that there was no screw penetration in the intercondylar fossa, and the most proximal screw was located within the intramedullary canal. The plate fit criteria (Fig. 1) were defined based on clinical requirements:

Criteria 1: Plate to bone distance ≤ 1 mm at most anterior distal location;
Criteria 2: Plate to bone distance ≤ 2 mm at nine points over the metaphyseal region, including distal tip;
Criteria 3: Plate to bone distance ≤ 3 mm for points (anterior, centre, posterior) on the shaft spaced at ∼ 4 cm intervals, including proximal tip.

A customised batch-process module (MATLAB, MathWorks, USA) measured/recorded plate-bone distances on the plate-bone assemblies. Anatomical plate fit was defined as satisfying all three criteria.

Femoral radius of curvature (ROC) was measured to assess the effect of mismatch of the plate bow radius on plate fit.

Results

Satisfactory plate conformity was achieved from the plate head up to screw hole 6, with most measurement locations fitting between 52–100% of bones for both ethnicities. The exception were locations (3rd to 5th point) along the posterior plate edge in the metaphyseal to distal shaft region, with plate fit of 7–40%. The most anterior point on the plate head also fit poorer ($p = 0.005$) for Caucasians (43%) compared to Vietnamese (63%), although the posterior point fit 100% for both.

Other ethnicity specific differences were: a larger ($p < 0.001$) mean plate-bone distance at the plate tip of 1.8 mm (0.7–5.9 mm) for Caucasian vs 1.3 mm (0.4–2.5 mm) for Vietnamese; and smaller ($p < 0.032$) mean plate-bone distances along the 2nd and 3rd rows of distal measurement points ranging from 1.1–3.4 mm for Caucasian vs 1.3–4.1 mm for Vietnamese. For both ethnicities, there was plate misfit from level of proximal 5–6 holes, with plate fit of 0–44% in that region, and a mean distance of 11.7 mm (0–23 mm) at the proximal plate tip for Caucasians and 15.4 mm (4.8–31.2 mm) for Vietnamese, respectively. Plate misfit was worse ($p \leq 0.01$) for Vietnamese in that region.

For the entire dataset, and for the ethnic subgroups, age (young vs old) had limited impact on the plate-bone distances with no significant differences for height and ROC, except for the younger Vietnamese who had a larger ($p = 0.022$) ROC compared to the older.

Sex affected plate fit as females had larger ($p < 0.038$) plate-bone distances at some distal locations and for most of the shaft; they were also shorter ($p < 0.001$) and had a smaller ($p = 0.030$) ROC. Female Caucasians were on average shorter ($p < 0.001$) while their ROC was not significantly different. However, their plate-bone distances at level of proximal 8–10 holes were larger ($p < 0.045$). Vietnamese female had larger ($p < 0.039$) plate-bone distances for the majority of measurements points, were shorter ($p < 0.001$), and had a smaller ($p = 0.012$) ROC.

Based on the three defined criteria, anatomical fitting was not achieved for any bones in our dataset. A plate ROC = 940 mm was obtained from the reverse engineered plate. The mean ROC of 1008 mm (615–1739 mm) for Caucasians was larger ($p < 0.001$) vs 866 mm (471–1505 mm) for Vietnamese. Caucasians were also taller ($p = 0.002$) on average, and ROC positively correlated ($r_s = 0.40$; $p < 0.001$) with height. Height and ROC positively correlated ($r_s = 0.43, 0.33$; $p < 0.001$) with the distal plate tip-bone distance, while most plate-bone distances in the proximal metaphysis and in the shaft showed negative correlations ($rs = -0.20$ to -0.55; $p < 0.023$). Age only showed positive associations ($r_s = 0.24$ to 0.25; $p < 0.008$) with plate-bone distances along the 1st row of distal measurement points.

Conclusion

This study demonstrated satisfactory fit from plate head to midshaft for both ethnicities. To our knowledge this is the first report of proximal shaft plate misfit for Caucasians, which was of similar magnitude to Koreans [1]. The metaphyseal misfit in our study was below reported values [2], perhaps reflecting differences between 3D modelling and 2D templating, and/or normal anatomy vs Caucasian TKA patients.

Combining 3D modelling with patient demographics revealed both patient height and ROC are negatively associated with plate fit, particularly in the shaft. This, together with the positive correlation of ROC and height, helps to explain the poorer plate fit observed clinically for Asians and females, who are generally shorter compared to Caucasians and males.

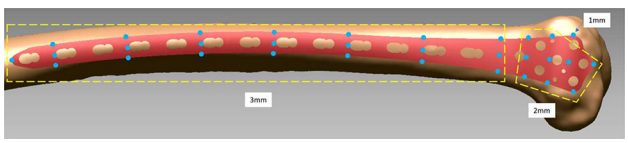

Fig. 1 13-Hole plate undersurface positioned on lateral distal femur with fit criteria for shaft and plate head. Light blue dots show locations of the plate-bone distance measurements, along with maximal acceptable distance thresholds

References

[1] Hwang JH, Oh JK, Oh CW, Yoon YC, Choi H (2012) Mismatch of anatomically pre-shaped locking plate on Asian femurs could lead to malalignment in the minimally invasive plating of distal femoral fractures: a cadaveric study. Archives of Orthopaedic and Trauma Surgery 132:51–56.

[2] Campbell ST, Bosch LC, Swinford S, Amanatullah DF, Bishop JA, Gardner MJ (2019) Distal femur locking plates fit poorly before and after total knee arthroplasty. Journal of Orthopaedic Trauma 33(5):239–243.

Navigation makes more accurate femur resection compared to conventional TKA in lateral femoral bowing greater than 5 degrees

W. K. Choi[1]

[1]Daugu Catholic University Medical Center, Orthopaedics, Dae gu, South Korea

Keywords Navigation, lateral femoral bowing, total knee arthroplasty, outlier.

Purpose

There are many disagreements about the merits of navigation in total knee arthroplasty (TKA). We compared and analyzed the difference in the accuracy of femoral resection according to the degree of lateral femoral bowing in two surgical methods (conventional and navigation assisted TKA) [1,2].

Methods

A total of 238 cases participated, with 100 cases underwent navigation TKA and 138 cases underwent conventional TKA from 2016 to 2021. The surgeon first performed TKA as main operator since 2016. Most of the TKA was navigation assisted between 2016 and 2018, and most of the TKA was conventional between 2019 and 2021. For more accurate distal femoral cutting in patients with lateral femoral bowing, preoperative scanogram was used to set up the insertion point of IM rod in conventional TKA, see also Figs. 1 and 2. Femoral lateral bowing was divided into three groups based on the angle. Group 1 is divided into groups with negative values, group 2 is divided into mild bowing of 0 to 5 degrees, and group 3 is divided into groups with moderate bowing of 5 degrees or more. Postoperative mechanical hip knee ankle (mHKA) angle and mechanical lateral distal femoral angle (mLDFA) were aimed to be 0° and 90°. The allowable range of each value was set as 0° ± 3° and 90° ± 3°

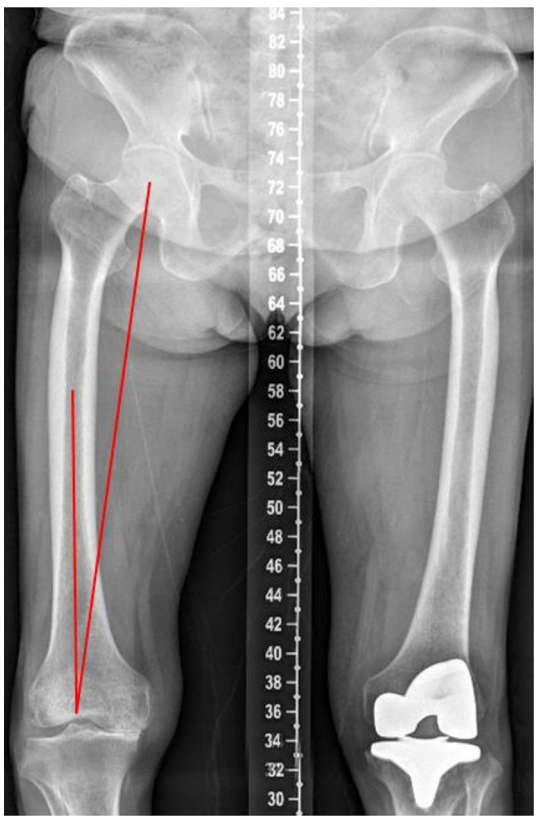

Fig. 1 For more accurate distal femoral cutting in patients with lateral femoral bowing, preoperative scanogram was used to set up the insertion point of IM rod in conventional TKA (**A**). The more severe the lateral femoral bowing, the lateralization of the IM rod (**B**)

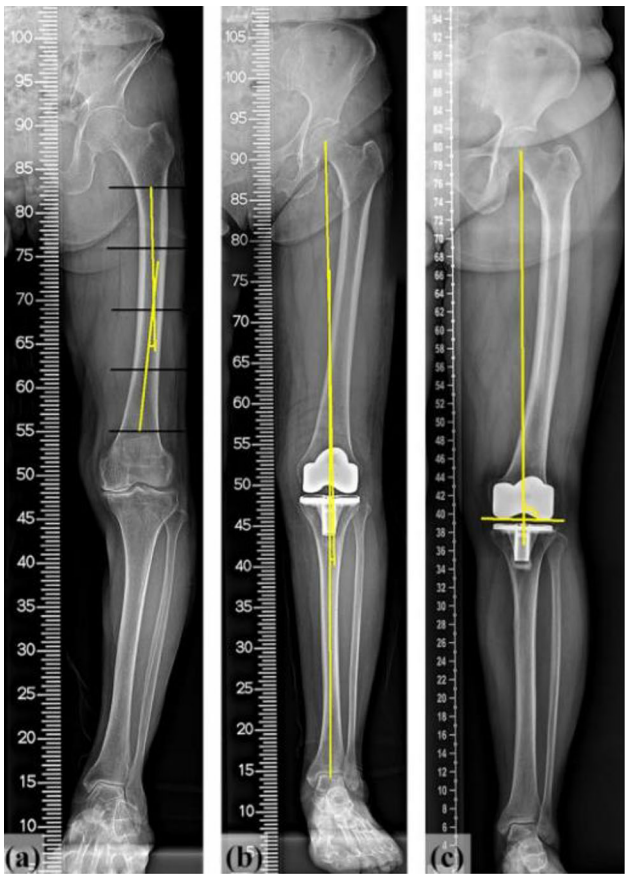

Fig. 2 Lateral femoral bowing, mechanical hip knee ankle angle, and mechanical lateral distal femoral angle were measured using scanogram

Results

The distribution of outliers of mLDFA in the three groups divided according to lateral femoral bowing in the navigation-assisted group was not statistically significant (P = 0.59). In the other hand, the distribution of outliers of mLDFA was statistically significant in conventional method group (p = 0.01). The odds ratio of outlier occurrence of mLDFA in the conventional method was 2.50, which was statistically significant (P = 0.03). And, when the lateral femoral bowing value was moderate, having 5 degrees or more, the odds ratio was 4.20, which was statistically significant (P = 0.003).

Conclusion

For patients with lateral femoral bowing greater than 5 degrees, navigation helps more accurate femur resection compared to conventional TKA.

References

[1] Lasam MP, Lee KJ, Chang CB, Kang YG, Kim TK. Femoral lateral bowing and varus condylar orientation are prevalent and

affect axial alignment of TKA in Koreans. Clinical orthopaedics and related research. 2013;471(5):1472–83
[2] Yau WP, Chiu KY, Tang WM, Ng TP. Coronal bowing of the femur and tibia in Chinese: its incidence and effects on total knee arthroplasty planning. Journal of orthopaedic surgery. 2007;15(1):32–6.

Towards Development of a Mixed-Reality Surgical Simulator

A. W. Palliyali[1], S. Paul[1], J. Abinahed[1], A. Mohamed[2], E. Yaacoub[2], N. Navkar[1]

[1]Hamad Medical Corporation, Department of Surgery, Doha, Qatar
[2]Qatar University, Computer Science and Engineering, Doha, Qatar

Keywords Surgical Training, Surgical Simulators, Box Trainers, Mixed Reality.

Purpose

Minimally invasive laparoscopic surgery requires specialized skills, including hand–eye coordination (to operate on the tissue) and depth perception (to view and navigate the operative field displayed on a flat two-dimensional screen). This presents unique challenges to surgical training. The conventional surgical training apprenticeship model (where a mentor teaches a mentee) is difficult to execute due to the work-hour constraints, demand to increase operating room throughput (as operative time is extended), and ethical concerns about novice surgeons operating on live patients. As a result, surgical simulators are used for training of minimally invasive surgeries. These simulators can be broadly classified into: virtual reality (VR) based simulators (e.g. LapVR—CAE and Lap Mentor—Simbionix) and box-trainers (e.g. Simsei and Laparo). Realism of tool-tissue interaction is essential to simulate realistic tissue behavior and properties when operated by a surgical instrument. However, VR simulators lack this feature as virtual models of the tissue and surgical instrument are used in computer-generated environment. The tool-tissue interactions (such as during cutting, suturing, and cauterization) cannot be simulated with high fidelity (by a computer) as compared to the real-world. Also, the force-feedback on the simulator's interfaces mimicking surgical instruments is absent or relatively poor. On the other hand, while the box trainers overcome the limitation of VR simulators, they are not able to render realistic operative field inside the insufflated cavity of a surgical scenario. This gap between the training on real surgical environment and on box trainers usually leads to a suboptimal training experience for surgeons [1], and are thus limited to FLS training curriculums (which mostly include peg transfer, pattern cutting, ligating loop, and intra-corporeal and extra-corporeal knot tying). This work proposes a mixed reality surgical simulator that merges the advantages of both box trainers (i.e. realistic tool-tissue interactions) and virtual reality simulators (i.e. true-to-life operative field). The simulator enables an operator to view the surgical site from different perspectives using a scope and perform the tool-tissue interaction required to operate at the surgical site.

Methods

The proposed mixed reality surgical simulator setup (as shown in Fig. 1) includes a workstation, visualization screen, an optical tracking system, scope, and a simulation box. The simulation box consists of a box tracking frame, a Chroma background, and holes (mimicking incisions) to insert the minimally invasive surgical instruments. The video stream (acquired from the scope) and the tracking data (acquired from the optical tracking system) are fed to the simulation workstation. The simulation workstation processes the

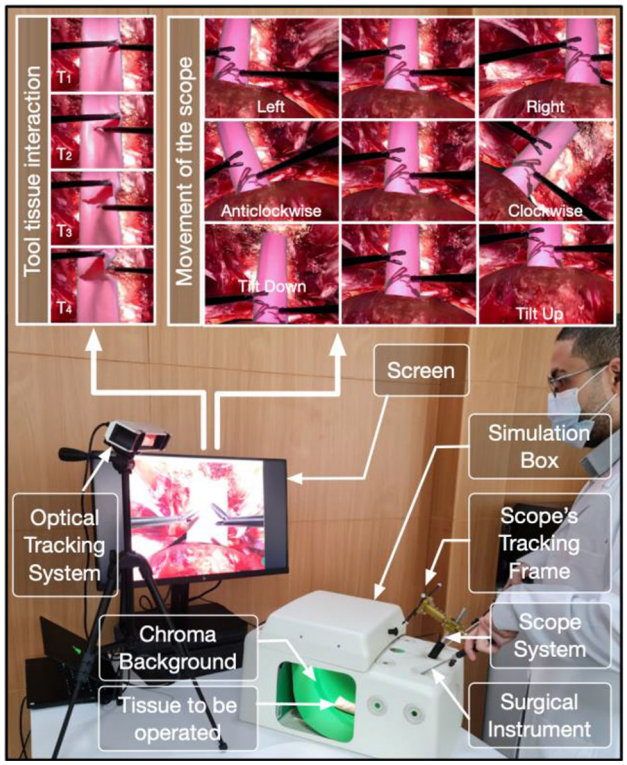

Fig. 1 Setup of the mixed-reality surgical simulation system. The sequential time-stamped images (T_i denotes ith timestamp) of the simulated operative field depict the dissection of a tubular soft tissue performed by the operator. During the scope movment (panning left and right, rotating anticlockwise and clockwise, and tilting down and up), the relative pose of the 3D virtual meshes (constituting the foreground and background of operative field) and physical models (comprising of the tissue to be operated) remains same creating an immersive experience

information and renders a mixed reality scene. The software modules were implemented based on an augmented reality platform [2]. The simulator can be configured with different procedure specific operative fields. A surgical scene involves multiple tissues. If the tissue is operated by the surgical instrument (such as during cutting, cauterization, suturing, grasping), a deformable soft tissue physical model is used. The model is placed at a specific pose inside the simulation box with a Chroma background. On other hand, if the tissue in the surgical scene does not interact with surgical instrument, a virtual mesh model of the surrounding tissue structures is generated and is registered with respect to the simulation box to be rendered as a part of the background/foreground scene of the operative field. Both the virtual and physical tissue models are registered with respect to the simulation box. An optical tracking system is used to continuously track the pose of the scope and simulation box during simulation. The video frames acquired from the scope are processed using imaging filters and Chroma key technique to remove the background. As the pose of simulation box and scope camera is tracked in real-time using optical tracking system, a mixed reality scene of the operative field is generated by merging the virtual mesh models and the viewing plane of the filtered video stream depicting the interaction of the surgical instrument with deformable soft tissues.

Results

A mixed-reality scene of the operative field is generated by the surgical simulator (Fig. 1). The rendering produced by the simulator

demonstrate the relative poses of the virtual operative field, soft tissue models, and surgical instruments are synchronized during the movement of the scope. The developed mixed reality surgical simulator was evaluated for accuracy in aligning the physical world with the virtual environment during scope movements (panning, tilting, and zooming). A rectangular strip (4 cm × 3 cm) of a silicon tissue composing the operative field was placed in front of the scope at a distance of 10 cm. The scope was actuated with different angular speeds (varying from 5, 10, and 15 degrees per seconds at remote centre of motion) using a UR5e robot. The pose of the scope tip and tissue were tracked using an optical tracking system. This enabled rendering of a 2D virtual plane at the position of the tissue in the simulator. The video feed acquired from the scope was processed to extract the corners of the real tissue and were compared with the virtual ones projected on the scope's viewing plane. For a video feed of 640 × 480 pixels, the average alignment error for 5, 10, and 15 degree/second were [4.2, 3.2, 2.7], [6.9, 4.5, 2.8], and [8.4, 5.1, 2.8] pixels in case of panning, tilting, and zooming, respectively. The alignment was accurate (with errors less than 1.5%) to create a seamless and immersive mixed reality experience. It enabled the operator to perceive the virtual elements as if they were physically present and facilitated the physical elements to behave realistically within the virtual environment.

Conclusion

The proposed mixed-reality surgical simulator tends to provide true-to-life tool-tissue interactions while rendering a realistic operative field during surgical training. The simulator can also be configured for robot-assisted minimally invasive surgeries as well. As a part of future work, we aim to conduct user-studies with surgeons to assess the mixed reality surgical simulator for procedure as well as patient specific surgical scenarios.

Acknowledgment

This work was supported by NPRP award (NPRP12S-0119-190006) from the Qatar National Research Fund (a member of The Qatar Foundation). All opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of our sponsors.

References

[1] Jin C, Dai L, Wang T (2021) "The application of virtual reality in the training of laparoscopic surgery: a systematic review and meta-analysis." International Journal of Surgery 87: 105,859.

[2] D Shabir, Abdurahiman N, Padhan J, Trinh M, Balakrishnan S, Kurer M, Ali O, Al-Ansari A, Yaacoub E, Deng Z, Erbad A, Mohammed A, Navkar N (2022) Towards development of a tele-mentoring framework for minimally invasive surgeries. The international Journal of Medical Robotics and Computer Assisted Surgery 18(5).

Surgical workflow recognition for effective use of plastic surgery videos

T. Kobayashi[1], H. Kajita[2], Y. Takatsume[3], Y. Aoki[4]

[1]Keio University, Integrated design engineering, Yokohama, Japan [2]Keio University School of Medicine, Department of Plastic and Reconstructive Surgery, Tokyo, Japan [3]Keio University School of Medicine, Department of Anatomy, Tokyo, Japan [4]Keio University, Department of Electrical Engineering, Faculty of Science and Technology, Yokohama, Japan

Keywords Surgical Workflow, Surgical Phase Recognition, Plastic Surgery, Transformer.

Purpose

The automatic classification of surgical processes can make the surgical workflow more distinct, making it an integral aspect of the efficient use of long videos. This task can be applied to video captioning and summarization for the training of physicians. In addition, real-time process classification can be utilized for the decision support of physicians. Given the foregoing, it is evident that surgical workflow recognition has a wide range of applications. Many studies have been conducted in the domain of laparoscopic surgery because of the original use of a camera and the large amount of data that has been collected in this area. On the other hand, these studies are limited to endoscopic surgeries, and only a few studies have been conducted on other surgeries, including open surgery. There are two reasons for this:

1. In the case of a non-endoscopic surgery, it is necessary to film the surgery from the outside, and the quality of the video is low due to the occlusion caused by the hand and head.
2. The variety of body parts and the types of surgeries require a versatile model.

The purpose of this study is to create a dataset of first-person viewpoint videos using a loupe-mounted camera in plastic surgery. Thereafter, a surgical workflow recognition model will be proposed for utilizing surgical videos in a wider area.

Methods

The flow of the surgical process classification is shown in Fig. 1. The inputs to our model are the visual features obtained from the feature extractor and the surgical tool presence classification results. They are combined and input into the ASFormer [1] to obtain the final classification result. ResNet50 is used as the visual feature extractor. Since our datasets have strong data imbalance, Importance Balanced Loss and Focal Loss(IBFloss) are introduced. In addition, for tool identification, there are frames wherein some annotated tools do not appear. Therefore, they are trained by binary cross-entropy after Sigmoid activation. In addition, the ResNet50 model is a multitask learning process that classifies workflow and surgical tools, so we introduce automatic weighted loss to adjust these weights.

Datasets

We captured 15 plastic surgery videos using the NanoCamHDi. In the first-person video by NanoCam, there are frames in which the surgeon is not looking at the surgical field. These frames are irrelevant to the surgical process and need to be removed. We created a video with

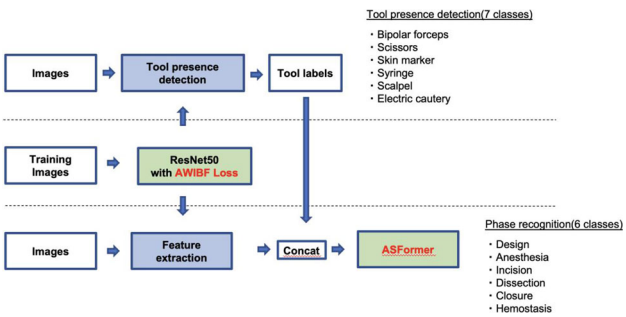

Fig. 1 Proposed process flow

Table 1 Results of feature extraction

	Acc	F1	Acc(VA)	F1(VA)	Jacc	Acc(Tool)	F1(Tool)
ResNet + AW	0.617	0.547	–	–	–	0.934	0.572
ResNet + AWIBF	0.649	0.597	–	–	–	0.938	0.621
TeCNO	0.633	0.574	0.684	0.582	–	–	–
TeCNO + ours	0.664	0.62	0.715	0.628	–	–	–
ASFormer	0.732	0.694	0.733	0.679	0.55	–	–
Ours	0.776	0.753	0.786	0.733	0.611	–	–
Ours w/o IBF loss	0.75	0.679	0.747	0.718	0.602	–	–
Ours w/o Tool label	0.742	0.712	0.754	0.718	0.602	–	–

these frames removed using optical flow and semi-supervised learning. The algorithm consists of three steps:

1. The inference results are obtained in a ResNet50 model that has been trained to determine whether a surgical field frame or an irrelevant frame is used in five surgical videos.
2. We divide the video into five-second segments and use the following as a criterion of confidence: whether or not at least one optical flow point could be traced in each segment. Specifically, if no optical flow is obtained, it is likely to be an irrelevant frame; if it is obtained, it is likely to be a surgical field frame.
3. Frames with high confidence are used as input to the ResNet50 model as a new dataset.

To build a dataset, we annotate 15 recreated videos with the surgical process and surgical instruments. The videos are labeled at 30 fps with phase and tool presence annotations. Since there are only a few irrelevant frames in this video, no annotations are made in these. Annotation is also not done if the phase recognition is ambiguous or if the person wearing the NanoCam is not the main surgeon. Since tools may be only partially visible and difficult to recognize visually, a tool is defined as present in the image if at least half its tip is visible. The tool is annotated only with the surgical tools in the hand of the surgeon performing the main operation, so a maximum of one tool is annotated per frame.

Experiments

Our dataset is divided into 12 videos for training and three videos for testing. The dataset was subsampled to 5fps and used as input for the model. The visual feature extractor is trained using frames segmented at 1 fps, and approximately 45,000 frames are used for training. The models are evaluated using accuracy, F1 score, and Jaccard index. The accuracy and F1 score are evaluated using the overall test data and the average of the per-video values (VA). We also perform ASFormer, Tool Presence, and IBFloss ablation experiments to validate each element of our model. TeCNO[2] is used for comparison.

Results

As shown in Table 1, our method achieves high accuracy (78%) on difficult datasets and an F1 score of 73%. This shows that universal features can be extracted and taken into account in different surgeries. Furthermore, the use of the ASFormer as a temporal model greatly improves the accuracy compared with existing surgical workflow recognition models. This indicates that the model works efficiently in general action segmentation tasks because our plastic surgery videos show hand movements and their relationship with the surgical tools.

Feature Extractor

As seen in Table 1, the feature extractor improves the F1 score by 5% using our proposed loss. This is because the impact of class imbalance in surgical process classification is significant, and our approach eliminates this problem.

Tool Presence

We input the tool classification results into the temporal model as new features. This way, we can increase the accuracy by about 2%. Since the F1 score of the surgical tool classification is 62%, further improvement of the accuracy is expected to make a further contribution.

Conclusion

In this study, we were able to achieve high accuracy for the difficult task of process classification in plastic surgery. The findings showed the possibility of utilizing surgical videos over a wider area by performing process classification in plastic surgery. Furthermore, this method has the potential to be applied to existing datasets.

References

[1] Yi F, Wen H, Jiang T (2021). ASFormer: Transformer for action segmentation. In: BMVC.

[2] Czempiel T, Paschali M, Keicher M, Simson W, Feussner H, Kim ST, Navab N (2020). Tecno: Surgical phase recognition with multi-stage temporal convolutional networks. In: Medical Image Computing and Computer Assisted Intervention. Vol. 12,263, pp. 343–352.

Low Cost On-chip Magnetic Sensors for Electromagnetic Navigation Bronchoscopy

K. O'Donoghue[1], M. Srivastava[1], A. Sidun[1], H. A. Jaeger[2], D. O'Hare[1], P. Cantillon-Murphy[1,2]

[1]Tyndall National Institute, MCCI, Cork, Ireland [2]University College Cork, Electrical and Electronic Engineering, Cork, Ireland

Keywords electromagnetic tracking, magnetic sensors, surgical navigation, on-chip sensor.

Purpose

In this work we present a low cost device for lung navigation utilising a novel magnetic sensor technology. Typical sensors used in most electromagnetically tracked devices are simple wire-wound coil based sensors, but by moving to an on-chip solution, the cost of disposable medical devices can be considerably reduced. We demonstrate the operation of this device in a lung phantom and report accurate navigation to within 2 mm.

Electromagnetic Tracking Systems (EMTS) are a commonly used method of determining the location of surgical tools and instruments without direct line of sight [1]. EMTS generally comprises a source of

magnetic fields and a miniature sensing element that can be integrated into surgical devices. By measuring the magnetic field seen by the sensor, its position and orientation can be calculated [2].

Typical EMT sensors used in various image guided interventions can be an expensive component, particularly in low margin disposable procedures. The coil based sensors can often cost between $10–200 depending on the size and complexity of the sensor, and this can be a significant portion of the overall materials cost for a device. Coil-based sensors are also susceptible to cable noise and electromagnetic interference (EMI) effects. By migrating to sensors that can be mass produced using standard silicon fabrication techniques and processes, these magnetic sensors can be produced for substantially less than $1 when manufacturing at large scale.

Recent chip-based solutions yield only 3D position coordinates without angular information. Pitch, yaw, and roll angles are critical for clinical applications [1], such as robotic bronchoscopy, which uses angular information to visualize lung airways, and electrophysiology (EP) mapping catheters and biopsy devices, which require extremely small sensors, typically less than 0.5 mm in diameter. Previous chip-based solutions have failed to integrate on-chip sensing coils resulting in large footprints offering little advantage in form-factor and scaling compared to discrete solutions.

Methods

AC electromagnetic tracking systems are typically based on Faraday's law to induce a voltage in a search coil [1], [2]. EM tracking generally operates with magnetic fields from multiple sources that are spatially varying functions given. Once the flux has been measured, the position and orientation of the sensor can be determined by using a non-linear least squares algorithm such as the Levenberg–Marquardt (LM) method [2].

The sensor proposed in this work is created on a standard silicon process in the form of an air-core, multilayer spiral coil as seen in Fig. 1b. The sensor is a rectangular coil (1200 µm × 450 µm) and this form factor is used to minimise the width of the device. The on-chip sensor achieves improved sensitivity by the usage of thick metals at top and thin metals at bottom with varied width and number of turns on each layer. The wide and thick metals are used at top to minimise resistance, but the number of turns is high enough to obtain good sensitivity to the magnetic field. The multilayer stacking improves sensor coil sensitivity by increasing the area of metal traces as well as the amount of field captured by increasing the effective cross-sectional area within a given region. The on-chip sensor uses Metal Layers M8, M9 and M10 of a 65 nm silicon process arranged in stacked fashion, implemented to maximize sensitivity and minimize resistance in a given area. This method allows to capture more flux by increasing the effective cross-sectional area without increasing the area of the inductor on a die. The weak sensor signal is amplified by an on-board low noise, high gain amplifier, the output of the chip sensor is then digitised and processed in Matlab to calculate its position and orientation.

The resulting sensor is wire-bonded onto a flex PCB substrate and inserted into flexible catheter tube for evaluation. Flex PCB carrier measures 1.7 mm wide, 0.4 mm thick and 1 m long. This is shown Fig. 1a. Accuracy of the system is verified using a rigid mechanical grid and pre-clinical validation was demonstrated in a 3D printed lung phantom.

Results

To demonstrate the tracking performance of the sensor, full EMT system measurements were carried out using the Anser open-hardware EMT system [2]. The sensor outputs were sampled at 200 kHz and 10,000 samples were used per pose estimate, resulting in a 20 Hz update rate (substantially faster than 1–2 Hz) The accuracy was determined by recording a grid of measurements and comparing to the true mechanical positions of the sensor holder. 243 test points were recorded in a 9 × 9 × 3 grid at x–y intervals of 15.9 mm on each layer and z = 38.4 mm between layers. The tests were repeated with two different sensor orientations (z-directed and y-directed) giving a total of 486 test points. Position error is defined as the error between the measured grid of points and its ideal grid when both grids are aligned using Horn's Absolute Orientation algorithm. Each test point is averaged for 2 s with the sensors position sampled at 20 Hz. The measured results of the accuracy experiments are shown in Fig. 1d. The mean measured results show an accuracy of 2 mm position error with standard deviation of 1.4 mm and 1° errors for orientation. Table 1 compares this sensor solution with a commercially available magnetic sensor by Northern Digital (Waterloo, Canada). A critical

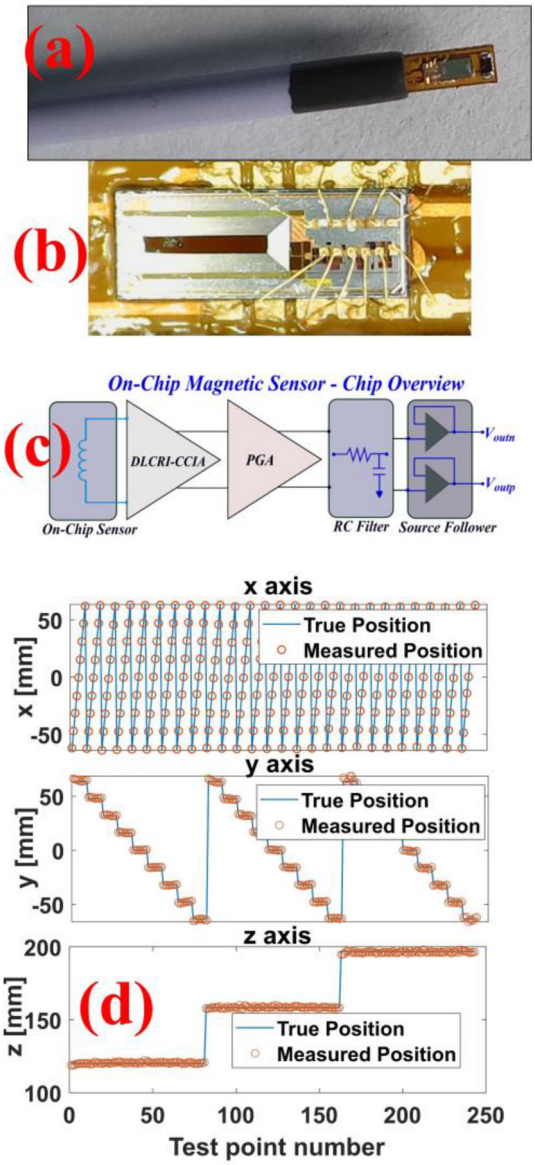

Fig. 1 **a** Wire bonded sensor attached to a flex PCB substrate and inserted into flexible catheter tube for evaluation, **b** close up of wire bonded chip to carrier, **c** schematic representation of the on-chip coil sensor with front-end low noise amplifiers, **d** accuracy results showing the true position of the sensor compared to the predicted position based on the sensors magnetic field measurements, a mean accuracy of 2 mm is observed during this test

Table 1 Comparison between the performance and parameters of the sensor described in this work and a traditional, off the shelf wound coil sensor

Metric	This work	Open source EMTS with NDI Aurora 610,099
Localization accuracy (mm)	2	1.1
Angular accuracy (°)	1	0.07
Sensor type	On-chip	Discrete
Sensor volume (mm^3)	0.18	1.27
Sensitivity (V/Hz/T)	0.0002	0.12
Sensor resistance (Ω)	1300	68
Sensor inductance (mH)	0.012	4.9
Sensor coil thermal noise density (nV/sqrt(Hz))	4.6	1

advantage of this design is a significant reduction in overall sensor volume, crucial in minimally invasive image-guided interventions.

Conclusion

The integrated EMT sensor described in this work is the first of its kind designed for clinical EMT of image guided interventions. On-chip EMT was possible due to the on-board low noise, high gain amplifier even with induced voltages on the order of μV, resulted in an overall accuracy of system accuracy of 2 mm. The solution can dramatically cut sensor cost, increase scalability and reduce reliance on harmful x-ray radiation guided imaging.

References

[1] Sorriento A, Porfido MB, Mazzoleni S, Valvosa G, Tenucci M, Ciuti G, Dario P (2020) Optical and Electromagnetic Tracking Systems for Biomedical Applications: A Critical Review on Potentialities and Limitations. IEEE Reviews in Biomedical Engineering, vol. 13. Institute of Electrical and Electronics Engineers, pp. 212–232, 2020, https://doi.org/10.1109/RBME.2019.2939091.

[2] Jaeger HA, Franz AM, O'Donoghue K, Seitel A, Trauzettel F, Maier-Hein L, Cantillon-Murphy P (2017) Anser EMT: the first open-source electromagnetic tracking platform for image-guided interventions, Int. J. Comput. Assist. Radiol. Surg., vol. 12, no. 6, pp. 1059–1067, Jun. 2017, https://doi.org/10.1007/s11548-017-1568-7.

A bendable burr of robot for minimally invasive bone tumor surgery and its preclinical evaluation using MR-scannable bone tumor models

S. Kim[1], D. Shin[1], C. Lee[1], D. Yu[1], J. Cho[1], H. Bang[2], H. J. Lee[2], D. H. Kim[2], I. Park[2], J. Hong[3], S. Joung[1]

[1]AIRS, Department of R&D, Daegu, South Korea [2]Kyungpook National University Hospital, Department of Orthopedic surgery, Daegu, South Korea [3]DIGST, Department of Robotics Engineering, Daegu, South Korea

Keywords MR-scannable bone tumor, Robotic surgery, Bendable drilling end-effector, Preclinical study.

Purpose

Bone tumors are accidentally diagnosed and primarily treated by traditional open surgery. This surgical treatment not only accompanies a wide incision and a large bone window into the patient but also increases the risk of infection, a lengthy recovery period, and secondary fracture. Although minimally invasive procedures for bone tumor surgery have been reported, it is not widely spread because of a lack of dedicated surgical tools Thus, this study aimed to evaluate a proposed bone tumor surgery robotic system in a series of experiments on phantoms and a tibia of animal studies.

Methods

The Bone tumor surgical robotic system has been designed and manufactured under IEC-60601–1(International Product Safety Standards for Medical Devices) and IEC 80,601–2-77 (Particular requirements for the basic safety and essential performance of robotically assisted surgical equipment). The robotic system has a bendable drilling module and a remote controller. The bendable drill can be inserted and debrided the tumor legion inside the bone via a 12 mm bone hole. Performance tests and IEC regulation tests evaluated the safety and validation. The mimicking materials for bone tumors were fabricated with 2.0 wt% agarose solution and 0.1 wt% MRI contrast media. Three series of phantom and animal models were manufactured by inserting a 25(L) mm × 25(W) mm × 15(H) mm volume of a solid block into a sawbones femur and the left tibia of a dead cow. The robotic system debrided the bone tumor models and evaluated the operation. We compared the remaining bone tumor model using MR images after the removal procedures.

Results

The safety and design regulation of the robotic surgery system was proved by compliance with (IEC 6060–1-1-2nd and IEC-60601–1-2-3rd) at a qualified testing agency. The validation was verified by providing the repeatability of the Standard deviation for degree travel is less than 1 degree and the travel range is less than 0.3 mm. The removal rate for bone tumor-mimicking materials in the phantom model was 88.5%, 100%, and 100%. In the preclinical studies, the removal rate was 54.87%, 98.08%, and 75.07%. The three surgeries were performed within 50 min with a bendable endoscope. An incision size was 45 mm and the bone hole size was 12 mm.

Conclusion

In this paper, we have developed a robotic system for assisting bone tumor surgery and designed it according to safety standards. A 5 mm Outer diameter bendable burr can be used for minimally invasive bone tumor removal. We tested the safety of the robotic system according to safety standards and evaluated the robot system using MR-scannable bone tumor models. We proved the preclinical validation. The average removal rate for the MR-scannable bone tumor model was 88.5% to 100% in the phantom studies and 54.87% to 98.08% in the preclinical studies. This bone tumor robot surgery

approach method will apply to future studies regarding human cadavers and clinical trials.

Validation of an augmented reality based functional method to determine and render the hip rotation center during total hip arthroplasty

Q. Neuville[1,2], T. Frantz[1,2], F. Van Gestel[1,2], B. Janssen[3,4], J. Vandemeulebroucke[3,4], J. Duerinck[1,2], T. Scheerlinck[5,6]

[1]Vrije Universiteit Brussel, Universitair Ziekenhuis Brussel, Department of Neurosurgery, Brussels, Belgium [2]Vrije Universiteit Brussel, Research Group Center For Neurosciences (C4N-NEUR), Brussels, Belgium [3]Vrije Universiteit Brussel, Department of Electronics and Informatics (ETRO), Brussels, Belgium [4]imec, Leuven [5]Vrije Universiteit Brussel, Universitair Ziekenhuis Brussel, Department of Orthopedic Surgery and Traumatology, Brussels, Belgium [6]Vrije Universiteit Brussel, Research Group Beeldvorming en Fysische wetenschappen (BEFY-ORTHO), Brussels, Belgium

Keywords augmented reality, navigation, total hip arthoplasty, head mounted display.

Purpose

We present a method to determine and visualize the functional center of rotation (FCOR) during total hip arthroplasty (THA) using off-the-shelf augmented reality (AR) hardware. It is demonstrated through this phantom study that current AR technology, together with internally developed image processing techniques, are sufficient for interventional use. It eliminates the need for traditional computer-aided navigational (CAN) hardware outside the AR head mounted device (HMD) itself.

Methods

An AR-HMD (Microsoft Hololens II) was adapted to provide inside-out 3D tracking of existing commercial passive infrared (IR) trackers. The tracked instrumentation being a generic tracker rigidly affixed to cadaver femurs and a reference tracker fixed to 3D printed cups to define a cartesian coordinate system. 20 different cadaver femurs were rotated into 3D printed acetabulum analogues with respect to the physiological range of motion. Both observers rotated each femur twice in its matching cup, producing 80 measurements. Through a pivot-fitting algorithm, the FCOR was determined based on the point cloud generated from displacement of the femoral tracker to the acetabular tracker. This FCOR was compared to the ground truth femoral and acetabular COR determined on CT.

Results

Determination of FCOR through the proposed functional, AR-based method resulted in an absolute error of 2.9 ± 1.4 mm and 2.9 ± 1.1 mm for the acetabular cups and femoral heads respectively. The 95th percentiles were below 5.6 mm and 4.7 mm respectively.

Conclusion

Our AR-HMD imageless navigation system offers an accurate and easy way to determine and render the femoral and acetabular FCOR in an experimental setting. After validation in a cadaver setting, this system could be used to navigate the insertion of both, the femoral and acetabular components of a total hip arthroplasty.

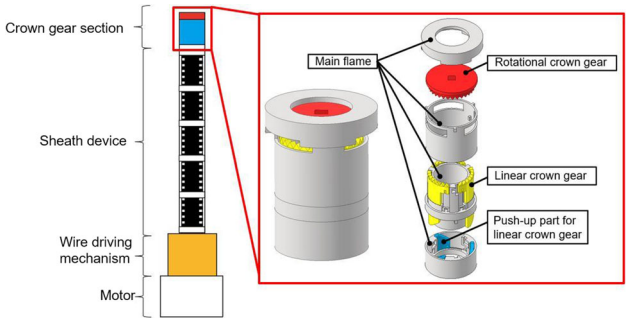

Fig. 1 Schematic image of the proposed crown gear

Rotational performance evaluation of a rotation mechanism including crown gear for controlling the posture of a flexible surgical tool

T. Okada[1], T. Dohi[2,3], R. Doine[2,3], K. Kuwana[1,4]

[1]Tokyo Denki University, Graduate School of Engineering, Tokyo, Japan [2]Tokyo Denki University, Research Institute for Science and Technology, Center for Research and Collaboration, Tokyo, Japan [3]Tohto University, Faculty of Human Care at Makuhari, Chiba, Japan [4]Tokyo Denki University, School of Engineering, Tokyo, Japan

Keywords Rotating mechanism, Crown gear, outer sheath, flexible surgical tools.

Purpose

Recently, laparoscopic surgery is widely employed in various kinds of treatment. The surgical tools used in laparoscopic surgery are straight in shape, making it difficult to operate deep inside the body. To solve this problem, we proposed a method in which a flexible surgical tool is used after being secured the path of the tool with an outer sheath that can be changed the rigidity of the main body [1, 2]. Although a rigidity variable outer sheath that can be bent deep inside the body has been developed, it lacks the degree of freedom for rotating the flexible surgical tool which is inside the sheath. In this paper, the rotational performance of the crown gear rotation mechanism, which we proposed for the outer sheath, was evaluated for controlling the rotation posture of the flexible surgical tools around the axis.

Methods

We proposed a method to rotate a flexible surgical tool around the axis of the tool by attaching a rotation mechanism to the tip of the outer sheath. The rotation mechanism is based on crown gears. For controlling the posture of the flexible tool, the rotational angle is required more than 360° around the axis of the tool. Figure 1 shows a schematic image of the proposed crown gear. The rotation mechanism consists of a rotating crown gear, three linear crown gears, three push-up parts for linear crown gear and main frame. Each linear crown gear is placed with the gap of one third pitch of the crown gear each other. This structure enables to generate a force to rotate the rotational crown gear when the linear crown gear is pushed up by the push-up part. It is possible to rotate the crown gear by pushing up the three linear crown gears with shifted cycle by one third cycle each other. The three push-up parts are driven by wire. The wires for driving the push-up parts are connected to the wire driving mechanism based on cylindrical cam to realize wire traction for three linear crown gears with shifted cycle by one third cycle each other. The rotation angle, rotation speed, and rotation torque were evaluated as the rotational

performance of the crown gear rotation mechanism. The angle and the speed of the rotational crown gear were measured using potentiometers. The rotational torque was defined as the required torque to stop the rotation during the crown gear rotating mechanism rotates steadily. The rotational torque was calculated from the output of a torque transducer coupled to the center of the crown gear.

Results

The crown gear rotation mechanism rotates 500° clockwise with a maximum error of 1.8° and 500° counterclockwise with a maximum error of 11.2°. A possible cause of the larger counterclockwise rotation error is the tilting of the linear crown gear. The measured rotational speeds were 0.6, 1.2, 2.4 rpm compared to the designed values of 0.6, 1.2, 2.4 rpm, respectively. The rotational torques were 48, 40, 36 mN·m for rotational speeds of 0.6, 1.2, 2.4 rpm, respectively. The torque varied linear with the reciprocal of the rotational speed.

Conclusion

In this paper, the angle, the speed and the torque were evaluated as the rotational performance of the crown gear rotation mechanism for controlling the rotational posture of a flexible surgical tool around the axis. Though the rotation angle had a maximum error of 1.8° clockwise and 11.2° counterclockwise, more than 360° rotational angle was achieved. The rotational mechanism rotated the designed rotational speed. These results indicate that the rotational mechanism is applicable to the rotation mechanism of the outer sheath for controlling the posture of a flexible surgical tool.

References

[1] Zuo S, Masamune K, Kenta M, Tomikawa M, Ieiri S, Ohdaira T, Hashizume M, Dohi T (211) Nonmetallic Rigid–Flexible Outer Sheath with Pneumatic Shapelocking Mechanism and Double Curvature Structure, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011, pp.169–177

[2] Sajima, H. Kamiuchi, K. Kuwana, T. Dohi, K. Masamune (2012) MR-Safe Pneumatic Rotation Stepping Actuator, Journal of Robotics and Mechatronics Vol.24 No.5, 2012,pp.280–827

A method for analyzing position of polar body in oocyte from microscopic images for assisting intracytoplasmic sperm injection

Y. Hayashi[1], T. Aoyama[2], G. Koutaki[3], K. Mori[1,4]

[1]Nagoya University, Graduate School of Informatics, Nagoya, Japan [2]Nagoya University, Graduate School of Engineering, Nagoya, Japan [3]Kumamoto University, Faculty of Advanced Science and Technology, Kumamoto, Japan [4]National Institute of Informatics, Research Center for Medical Bigdata, Tokyo, Japan

Keywords ICSI, Microscope, Deep learning, Micromanipulation.

Purpose

Intracytoplasmic sperm injection (ICSI) is performed as one of in vitro fertilization (IVF) procedures. In ICSI, an operator (embryologist) injects the sperm into the ooplasm in the oocyte by micromanipulations using micropipettes while observing magnified views from a microscope. Since these operations require highly skilled operators, an assistance system for the micromanipulations in ICSI is desired. Therefore, we have been developing a micromanipulation assistance system in ICSI [1]. In ICSI, it is vital to inject the sperm into the ooplasm without damaging the spindle. However, the spindle cannot be observed using a conventional microscope. Since the spindle is generally located near the polar body, the operator adjusts the oocyte orientation using a micropipette while observing the polar body in the oocyte before performing the microinjection. Therefore, it is crucial to recognize the polar body position in the oocyte during ICSI. A micromanipulation assistance system for ICSI would be able to provide appropriate support if it could automatically recognize the position of the polar body in the oocyte from microscopic images. The purpose of this study is to develop a method for analyzing the position of the polar body in the oocyte from microscopic images for assisting ICSI.

Methods

The proposed method recognizes the position of the polar body in the oocyte from microscopic images. The proposed method consists of two parts, (1) segmentation of structures in the oocyte in the image and (2) classification of the status of the oocytes in the image based on the position of the polar body. In the first part, we extract the polar body, the transparent body, and the ooplasm in the microscopic image using 2D U-Net [2]. We use Dice loss as a loss function and Adam as an optimizer for training the 2D U-Net. The segmentation results from 2D U-Net have some false negative and false positive regions. We modify the segmentation results to reduce these regions by applying morphological opening and closing operators, and connected component analysis. In the second part, we classify the statuses of the oocyte in the image based on the positions of the polar body in the oocyte. After the segmentation, the proposed method calculates the center of gravity in each segmented region. Positional relationship between the polar body and the ooplasm regions are analyzed using their centers. Then, we compute the clock position of the polar body in the oocyte. The proposed method classifies the images into five statuses based on the position of the polar body. Five statuses are (Status 1) no polar body locates in the oocyte, (Status 2) the polar body locates around the 3 o'clock position in the oocyte, (Status 3) the polar body locates around the 6 o'clock position in the oocyte, (Status 4) the polar body locates around the 9 o'clock position in the oocyte, and (Status 5) the polar body locates around the 12 o'clock position in the oocyte.

Results

We applied the proposed method to microscopic images from five porcine oocytes. We used about 200 images from four porcine oocytes for training and about 50 images from one porcine oocyte for evaluation. Mean Dice coefficients of the segmentation results using 2D U-Net for the polar body, the transparent body, and the ooplasm were 0.38, 0.86, and 0.97, respectively. The classification accuracy of five statuses based on the polar body positions was 70.0%. An example of a correctly classified result is shown in Fig. 1. In center image of Fig. 1, green, blue, and purple regions indicate the polar body, the transparent body, and the ooplasm regions extracted using 2D U-Net. In the right image of Fig. 1, green and blue regions show the centers of gravity in the polar body and ooplasm regions to

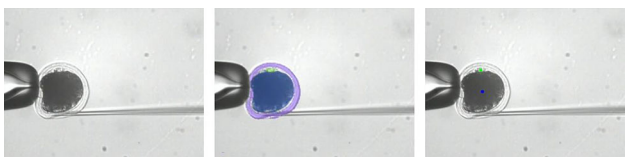

Fig. 1 Example of correctly classified result. This image is classified as polar body located 12 o'clock position in oocyte. Left, center, right images indicate microscopic image, segmentation result of polar body (green), transparent body (blue), and ooplasm (purple), and centers of polar body region (green) and ooplasm region (blue), respectively

compute the clock positions. This image was correctly classified as (Status 5) the polar body located the 12 o'clock position in the oocyte. As shown in this figure, the proposed method could classify microscopic images into five statuses using the segmentation of structures in the oocyte. Since the proposed method obtains the position of the polar body in the oocyte from a microscopic image, it will provide useful information to a micromanipulation assistance system in ICSI.

Conclusion

This paper reported a method for analyzing the position of the polar body in the oocyte from microscopic images. We extracted structures in the oocyte from microscopic images using 2D U-Net and analyzed the position of the polar body in the oocyte based on the segmentation results. The experimental results showed that the proposed method could obtain the position of the polar body in the oocyte from microscopic images. Future works include evaluation using a large amount of data and application of the proposed method to a micromanipulation assistance system.

References

[1] Fujishiro T, Aoyama T, Hano K, Takasu M, Takeuchi M, Hasegawa Y (2021) Microinjection system to enable real-time 3D image presentation through focal position adjustment. IEEE Robotics and Automation Letters 6(2): 4025–4031

[2] Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional networks for biomedical image segmentation. MICCAI 2015, LNCS 9351: 234–241

Identification of Diabetic Retinopathy from Fundus Images Using Deep Auto Encoder

K. Abe[1], H. Takeo[1], Y. Takahashi[2], Y. Nagai[3]

[1]Kanagawa Institute of Technology, Electrical and Electronic Engineering, Atsugi, Kanagawa, Japan [2]Takahashi Eye Clinic, Hospital director, Isehara, Kanagawa, Japan [3]National Cancer Center Hospital East, Radiology, Kashiwa, Chiba, Japan

Keywords Diabetic retinopathy, Deep Auto Encoder, Fundus image, Abnormality classification.

Purpose

Research and development activities aimed at image-based computer aided diagnosis (CAD) are being conducted in various medical fields. Additionally, active studies on organ and lesion detection using AI technologies are also underway [1]. Diabetic retinopathy, one of the three major complications of diabetes, is regarded as typical cause of acquired blindness. Conventional diagnosis of diabetic retinopathy is accomplished by an ophthalmologist comparing a series of fundus images taken over time. However, the diagnostic performance of this method has considerable dependence on ophthalmologist skill, presenting a degree of difficulty in the objective and accurate identification of lesions. This study considers a method of identifying diabetic retinopathy that relies on abnormality detection using an autoencoder that is trained using deep learning technology. The autoencoder is prepared by training with only normal fundus images. When a diseased fundus image is input into the trained autoencoder, abnormal regions are not reconstructed. This enables the detection of abnormal shadows by obtaining the difference with the original image.

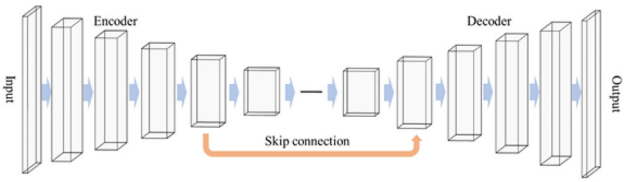

Fig. 1 Diagram of the auto-encoder used in this method

Methods

(1) Preprocessing

First, the color fundus image is converted to an image containing only the green component. The extracted green component has higher contrast with blood vessels and surrounding tissues compared with the color fundus image. At the same time, image size homogenization is also performed.

(2) Skip-GANomaly-based abnormality detection

The autoencoder configured in this study is based on Skip-GANomaly by Akçay et al.[2] Both encoder and decoder were extended to five-layer architectures. In the original study, skip connections were used for all layers. For this study, this was changed to just the intermediate 4th layer. Figure 1 shows the configuration of the autoencoder used in this study. This change was determined by conducting a preliminary study of the optimal skip connection configuration. The all-layer skip connection configuration used in the original study resulted in the incorporation of diseased region parameters for which learning had not been conducted, impeding abnormality detection. Integrations of the skip connection into various layers, such as the 5th layer, the 4th and 5th layers, and the three layers from the 3rd to the 5th, were similarly evaluated. These showed that integration of the skip connection in the 4th layer was optimal for fundus images.

(3) Method of diabetic retinopathy identification

In diabetic retinopathy identification, judgement is made by calculating agreement using the root mean squared error (RMSE) of input images and of reconstructed images obtained from abnormality detection. The higher the agreement with reconstructed images, the lower the RMSE. Conversely, the higher the number of abnormal regions, the less reconstruction, resulting in a higher RMSE. This makes possible the use abnormality detection to identify diabetic retinopathy.

Results

(1) Autoencoder training

Autoencoder training was conducted using fundus images from 476 healthy subjects. Training parameters were: No. of epochs 5000; Batch size 64; Latent dimensions 100.

(2) Evaluation using test data

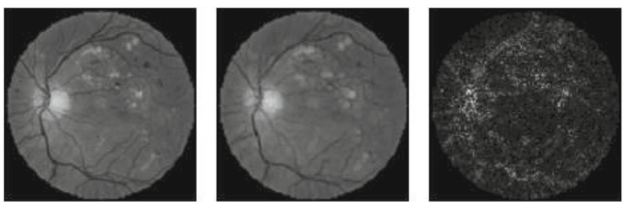

Fig. 2 Reconstructed image of diabetic retinopathy

The image data used in this experiment as test data comprised 52 healthy subject images and 98 diabetic retinopathy patient images for a total of 150 images.

Figure 2 shows an example of a reconstructed image from the autoencoder and an input/output difference image. The difference is expressed as the absolute value difference. Contrast is set to five times to facilitate viewing the displayed image. The average RMSE for healthy subjects was 9.71, while the average for diabetic retinopathy patients was 13.40. A t-test was performed to verify the significance of the difference between RMSE values. A p-value of 0.000005 (< 0.05) was obtained, confirming significance.

Conclusion

This study considered RMSE-based diabetic retinopathy identification in which the detection of diabetic retinopathy abnormalities was conducted using a method that extends Skip-GANomaly by adjusting it for medical images. Reconstruction effected from abnormality detection and differences with the original image enabled detection of leukoderma and hemorrhage areas. T-tests based on the RMSE values showed that the differences were a significant. These results suggest the effectiveness of this method in the identification of diabetic retinopathy.

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] Akçay S, Atapour-Abarghouei A, Breckon TP (2019) Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8

Development of an Interpretation System for Diagnostic Imaging Methods in CAD Combined Reading of Breast Cancer

A. Sugawara[1], R. Ohsaka[1], K. Abe[1], H. Takeo[1], S. Nawano[2], Y. Nagai[3]

[1]Kanagawa Institute of technology, Electrical and Electronic Engineering, Atsugi, Kanagawa, Japan [2]Shinmatsudo Central General Hospital, radiology, Matsudo,Chiba, Japan [3]National Cancer Center Hospital East, radiology, Kashiwa, Chiba, Japan

Keywords CAD, Interpretation, Breast Cancer, ROC.

Purpose

In general, medical students, residents, and inexperienced physicians learn how to read diagnostic imaging using so-called teaching files, which are collections of typical and characteristic diagnostic imaging cases [1]. However, with the spread of Computer-Aided Diagnosis (CAD), the way of diagnostic imaging is changing. That is, the final diagnosis is now made by referring not only to the images but also to the results of abnormal shadow detection by CAD. Currently, there is no learning system for this type of diagnosis method.

Therefore, this study proposes a learning system for diagnostic imaging methods in a reading method based on the combined use of CAD. This also allows users to learn about CAD characteristics. In addition, this study proposes a CAD system installed in a particular hospital that should evolve by learning additional methods from the hospital's images, and this diagnostic imaging method learning system is also enhanced as the CAD system evolves. As CAD evolves, the number of teaching files in the diagnostic imaging method learning system increases.

Methods

(1) CAD system

The CAD system used in this study was based on breast cancer mass shadows. CAD discriminator consisting of 350 training images, 44 validation images, and 216 test images were developed using 550 images collected from five institutions. Yolov5, an AI-based automatic segmentation process, was used for training. In 50 unknown cases (4 images in 2 directions per person, left and right, for a total of 200 images), the CAD detection performance was 86.0% (%) for TP and 0.504 (points/image) for FP.

(2) Imaging Diagnostics Learning System

The 4 K LCD monitor displays four images in two directions, left and right, for each patient. At the click of a button, the results of CAD detection are displayed as ROIs, with TP indicated by a white border, FP by a yellow border, and FN (lesions not detected by CAD) by a blue border. This database of 54 examples is saved and can be viewed over and over again to learn how to read breast cancer mass shadows as well as learn CAD characteristics.

(3) Evolutionary image processing method learning system

The CAD system was trained with image data collected from several hospitals, but the CAD system was also designed to enhance the learning system for image reading methods when the CAD system was installed at a particular facility. With CAD installed in the hospital, the physician's accurate diagnosis results are fed back into this diagnostic imaging method learning system and added to the case database. The CAD system is updated and evolves by regularly adding and retraining the database itself as learning data. In other words, CAD is customized to the images taken at the hospital, which are added as a teaching file to this diagnostic imaging learning system.

(4) Reading evaluation experiment

To confirm the effectiveness of the proposed system, reading experiments were conducted with the following three systems. System A was a case where reading is performed in a test case with no learning at all; System B assumes the conventional learning method and performs a reading test by learning with the location of the lesion indicated on the images; System C assumes the proposed system, where the original images were displayed, and displays the CAD results, and correct answers were showed in colored frames when

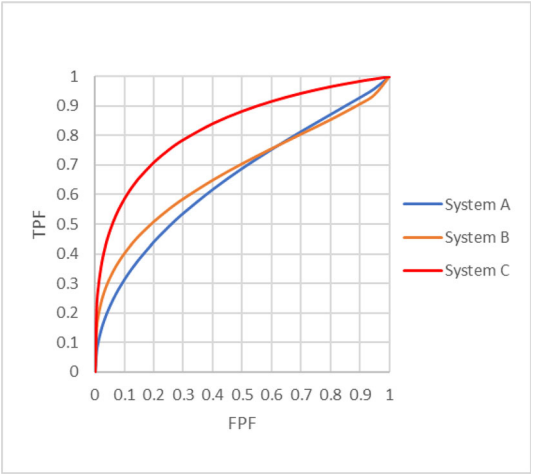

Fig. 1 ROC curves with three system learning methods

clicked. The System C displays the original image and, at the click of a button, displays the CAD results and correct answers. Furthermore, in System C, the number of cases for learning was increased, assuming that a certain period of time has passed since the introduction of the system to the hospital.

(5) Experimental conditions and evaluation methods

Interpretations experiments with systems A, B, and C were conducted on 14 medical students. After learning each system, the students were asked to read 200 images from a separately prepared test case, and the diagnostic performance of each system was compared using the area under the ROC curve Az based on the 3-step confidence method ROC evaluation [2]. First, all patients were evaluated with System A. After an interval of about 2 months, half of the patients were evaluated with System B and the other half with System C. ROC evaluation results were compared using System B and System C, respectively. After a further interval of approximately 2 months, the remaining systems were evaluated.

Results

Figure 1 shows the ROC curves summarizing the results of the reading experiment with the test cases of the three systems. The Az for System A in the case of no learning at all is 0.6524. The Az of System B is 0.6771 when learning with the conventional teaching file, and the Az of System C is 0.8288 when learning with the proposed evolutionary image diagnostics learning system.

Conclusion

The results of the reading experiment with the test cases confirmed that the interpretation method using the proposed system significantly improves interpretations ability, suggesting its effectiveness.

References

[1] Advanced Search | Eurorad

[2] https://www.ahajournals.org/doi/10.1161/circulationaha.105.594929

Classification of Pancreatic Cystic Lesions with multiple annotators merging techniques

M. Riera-Marín[1,2], B. Sastre-García[1,2], J. Rodríguez-Comas[3], J. García-López[1], M. Á. González-Ballester[2,4]

[1]Sycai Medical, Technical department, Barcelona, Spain [2]Universitat Pompeu Fabra, BCN—MedTech, Barcelona, Spain [3]Sycai Medical, Scientific department, Barcelona, Spain [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Keywords pancreatic cystic lesions, diagnosis, CT, machine learning.

Purpose

Medical imaging plays a crucial role in the diagnostic process, as it allows physicians to visualize and analyze internal organs and tissues for signs of disease or abnormalities. However, the low resolution of such images can compromise their interpretability, leading to uncertainty in the labeling of structures or abnormalities and posing challenges in the training and evaluation of machine learning (ML) models. In particular, the labels provided by different experts may vary, resulting in uncertainty in the training process.

In addressing this clinical challenge, this study proposes a ML approach for the classification of pancreatic cystic lesions (PCLs) as potentially malignant or non-potentially malignant. PCLs are a common type of lesion that can be difficult to detect on Computed Tomography (CT) scans and have the potential to progress to pancreatic cancer, a highly lethal form of cancer with a poor prognosis.

$$\gamma_p(x,y,z) = \begin{cases} 1, & if\ \frac{1}{N}A(x,y,z) \geq p \\ 0, & otherwise \end{cases}$$

Fig. 1 Equation used to obtain the final merged ground truths. A(x, y, z) is the number of annotators that labeled voxel (x, y, z) as containing the structure of interest; N the total number of annotators; and, p the threshold applied to generate the final ground truth

Furthermore, the study aims to address the issue of having multiple labels for the same CT image. To do this, various techniques for merging these labels are evaluated, and their influence on the performance of the classification of PCLs is studied in order to identify the most effective method. Through this approach, the ultimate goal of the study is to develop a reliable ML approach for accurately classifying PCLs, thereby improving the classification and management of pancreatic cancer.

Methods

The four types of PCLs that are commonly encountered in clinical practice are: intraductal papillary mucinous neoplasm (IPMN), mucinous cyst neoplasm (MCN), serous cystadenoma (SCA), and pseudocysts. These PCLs are divided into two main groups based on their likelihood of progressing to pancreatic cancer: high probability to malignize (IPMN and MCN) and low probability to malignize (SCA and pseudocysts).

Several factors that may influence the likelihood of progression to pancreatic cancer of the PCLs were considered: the location of the cyst (inside, partially inside, or outside of the pancreatic organ), the position with respect to the pancreas (head, body or tail), the presence of calcifications in the cyst or pancreas, and the ratio of the biggest to smallest axis of the volume of the PCL.

The training dataset for this study consisted of 154 studies with 81 non-mucinous and 73 mucinous, corresponding approximately to the 50% each. The distribution of sex among the subjects was 95 male subjects (60%) and 59 female subjects (40%). The mean age was 69 years, with a standard deviation of 12 years, meaning that the main subjects are found within 6912 years. The range of ages in the training dataset was from 33 to 91 years, with a median of 62 years.

The testing dataset consisted of 95 studies with 53 not mucinous and 42 mucinous, which approximately corresponds to 55% and 45% respectively. The distribution of sex among the subjects was also roughly equal, with 60 male subjects (60%) and 35 female subjects (40%). The mean age was 70 years, with a standard deviation of 12 years and the range of ages was from 38 to 87 years, with a median of 62.5 years.

Overall, both the training and testing datasets were well-balanced with respect to the likelihood of malignancy and sex, and the age distributions were consistent with similar statistical values.

The ML method used was the Gradient Boosting classifier. It consists of an ensemble of weighted decision trees which each of them outputs a prediction that is then averaged with the other's applying their corresponding relevance and merged into one final prediction. The Gradient Boosting classifier was trained with 10,000 estimators, which are the boosting stages to perform; a maximum depth of 1000, which limits the nodes of each tree; and, an initial learning rate of 0.0001.

This training was repeated per each dataset generated from the merging methods stated below and used to study their performance and influence in the PCLs classification.

Table 1 PCLs classification results for the different label merging techniques

Training dataset	Accuracy (%)	Accuracy test (%)	Sensitivity test (%)	Specificity test (%)
No merging	88	79	82	76
Intersection	72	75	77	71
Union	72	75	79	69
Majority voting	81	73	81	62

The first column indicates which case we are studying: no merging (all labels used for training as independent studies), intersection, union and majority voting merging techniques. The second column shows the accuracy obtained in the training and the third, fourth and fifth columns are metrics obtained from inferring the testing dataset

Three main techniques used for merging labels were studied [1]. They are carried out through probability maps, as depicted in the equation in Fig. 1, assigning each voxel to the final corresponding class depending on the applied threshold p:

- Intersection (p = 1), voxels marked by all annotators.
- Union (p = 1/n), voxels marked by at least one annotator.
- Majority voting (p = 0.5), voxels marked by at least half of the annotators after smoothing the probability map with a Gaussian kernel.

Results

The results of the Gradient Boosting classifier are shown in Table 1. For the training of all the multiple labels as independent studies an accuracy of 88% in the training and an accuracy of 79% in the testing dataset were obtained, which already outperforms the PCLs classification rate of medical experts which is between 40 and 65% [2].

Comparing both training and testing accuracies, it can be seen that the higher two belong to the dataset with no merging. Moreover, the majority voting gets a higher accuracy in training but a lower in testing, which maybe shows a worse performance of a merging algorithm that turns out in a worse training dataset. It can also be observed that all the specificities tend to be a bit lower than the sensitivities which means that the classification tends to over-classify potentially malignant cysts and under-classify benign cysts. Even if this is not the best performance, it is better to over-classify potentially malignant cyst due to the impact that ignoring a potentially malignant cyst may cause.

Conclusion

Preliminary results indicate that the classification tends to over-classify potentially malignant cysts and under-classify benign cysts. This could be beneficial in terms of minimizing false negatives, but further work is needed to improve the classification scheme.

The presence of multiple labels per image presented a significant challenge. In order to study more thoroughly its influence on the PCLs classification, a dataset with more than two labelers per CT is needed.

In order to address these issues and improve the accuracy of the classification, future work will be focused on expanding and improving the dataset and evaluating a wider range of merging algorithms.

References

[1] Mirikharaji Z, Barata C, Abhishek K, Bissoto A, Avila S, Valle E, Celebi E, Hamarneh G (2022). A Survey on Deep Learning for Skin Lesion Segmentation. arXiv preprint arXiv:2206.00356.

[2] Song, S. J., Lee, J. M., Kim, Y. J., Kim, S. H., Lee, J. Y., Han, J. K., Choi, B. I. (2007). Differentiation of intraductal papillary mucinous neoplasms from other pancreatic cystic masses: Comparison of multirow-detector CT and MR imaging using ROC analysis. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 26(1), 86–93

Automatic and Accurate Epileptogenic Zone Localization with SEEG-Informed Structural Connectivity for Focal Epilepsy

L. Xiao[1], Q. Zheng[2], C. Li[2], L. Yu[3], W. Si[2]

[1]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Institute of Advanced Computing and Digital Engineering, Shenzhen, China [2]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Institute of Advanced Integration Technology, Shenzhen, China [3]The University of Hong Kong, Department of Statistics & Actuarial Science, Faculty of Science, Hong Kong, China

Keywords Epileptogenic zone localization, SEEG, Structural connectivity, Focal epilepsy.

Purpose

Accurate epileptogenic zone (EZ) localization is key to successful epilepsy surgery for patients with medically refractory focal epilepsy. Currently, Electrophysiological recording and medical imaging are two fundamental techniques in preoperative evaluation of epilepsy surgery. Stereoelectroencephalogram (SEEG) is a gold standard for EZ localization by monitoring the abnormal electrophysiological neural activity from a holistic 3D perspective. However, SEEG can only be utilized to detect the seizure electrodes independently, and its limitation of local brain coverage fails to capture the global brain connection relationship. While medical imaging can demonstrate the whole brain structural abnormalities, it is incapable of detecting functional electrophysiological changes. Due to the heterogeneous nature of the two modals, it is challenging to localize EZ precisely by comprehensively analyzing the combination of SEEG signal and structural imaging.

Methods

This paper proposed an automatic and accurate EZ localization method for focal epilepsy, which subtly integrated the SEEG and Diffusion Tensor Imaging (DTI) to simultaneously detect the functional and structural abnormalities for EZ localization. The proposed method consisted of two steps: the determination of the region of interest (ROI) through SEEG and the accurate EZ localization through structural connectivity based on the SEEG-informed ROI. In order to determine ROI with abnormal functional neural activity, we first presented a spectral-temporal-aware fusion network for the SEEG signal identification to determine which region the implanted

Table 1 The average accuracy and precision (%) of different methods to identify SEEG from EZ, PZ and NIZ with tenfold cross-validation

Method	Accuracy	Precision
EMD + SVM	85.69 ± 0.73	86.65 ± 0.87
CWT + SVM	88.03 ± 0.68	89.40 ± 0.41
CWT + CNN	90.89 ± 0.35	91.90 ± 0.70
Ours	95.94 ± 1.97	98.25 ± 0.83

electrode contact belonged to. These regions were the EZ_{SEEG}, propagation zone (PZ_{SEEG}), and brain regions not involved with epileptiform activity (NIZ_{SEEG}), which were defined as whether the electrodes demonstrate initial, subsequent or no seizure activity [1]. The SEEG-informed ROI is defined by EZ_{SEEG} and PZ_{SEEG}. Specifically, the proposed spectral-temporal-aware fusion network integrated the spectral and temporal features of SEEG subband signals and utilized the hierarchical multi-domain representations with squeeze-and-excitation (SE) block to extract features for fusion. Subsequently, the feature representation was fed into the convolutional neural network (CNN) for SEEG seizure detection, then the SEEG electrode of these three regions can be recognized according to the Epileptogenicity Index (EI) of SEEG seizure signals. Though the SEEG recording can help to identify the local seizure points of EZ, it still failed to determine the EZ boundary. Inspired by [1], which demonstrated that the structural connectivity between EZ and non-seizure regions was significantly decreased, while the connectivity of EZ itself was not significantly changed, we aimed to accurately localize the EZ by the brain structural connectivity. Therefore, we proposed a regional structural connectivity network based on the SEEG-informed ROI to further accurately localize the EZ. The electrode contact localization of SEEG-defined ROI were co-registrated through a post-implantation Computer Tomography (CT) scan with DTI. Due to the significant structural connectivity differences between nodes of the epileptogenic network, we performed the analysis of whole-brain structural connectivity and the regional structural connectivity within and between the SEEG-informed ROI, and then calculated the connectivity strengths from different regions.

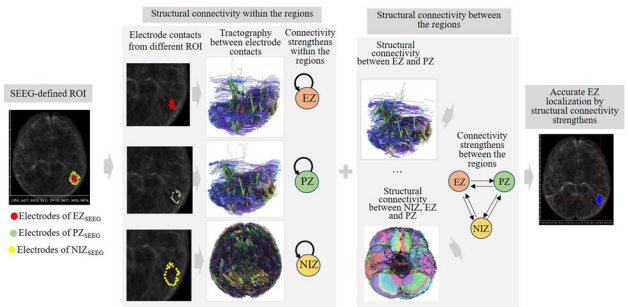

Fig. 1 The EZ localization process by structural connectivity differences from SEEG defined different regions. The EZ localization is determined through structural connectivity strengthens within and between EZ_{SEEG}, PZ_{SEEG} and NIZ_{SEEG}. The connectivity of EZ itself is not significantly changed, whereas the strengthens of the structural connectivity within seizure zone and non-seizure zones is significantly decreased. The region with the red circles denotes the EZ_{SEEG}, the green circles denotes the PZ_{SEEG}, the yellow circles denotes the PZ_{SEEG}, the circles denotes the electrode contacts, and the blue region denotes the final EZ localization

Finally, the EZ boundary can be further accurately determined according to these connectivity strengths from different regions.

Results

We conducted our experiments using the dataset of the SEEG signals and DTI imaging of 10 patients. The experimental results illustrated that the SEEG identification results of our method can achieve the average accuracy of 95.94% and the average precision of 98.25%, shown in Table 1, which outperformed the state-of-the-art methods for SEEG identification. The experimental results of EZ localization determined by structural connectivity illustrated that the significant structural connectivity differences within and between three SEEG-informed brain regions can assist to localize the EZ boundary, shown in Fig. 1. Meanwhile, we conducted the ablation study of SEEG determination, DTI determination, and the combination determination of SEEG and DTI for EZ localization, and the results demonstrated that our combination determination method showed better performance for EZ localization, and the EZ localization determined by our method was almost consistent with the result of doctor's decision.

Conclusion

Our method can assist neurosurgeons in more automatic and accurate EZ localization for focal epilepsy via SEEG informed structural connectivity by the combination of functional electrophysiological information with global structural connectivity information, which can provide an important auxiliary way to precisely locate the EZ for epileptic preoperative assessment. While for the preoperative assessment, more modal data can be considered for the localization of EZ from functional and metabolic aspects, including resting-state functional magnetic resonance imaging (rsfMRI), positron emission tomography-Computer Tomography (PET-CT), etc., so in the future, our immediate plan is to integrate more modal information for the preoperative assessment of epilepsy.

Acknowledgement

This work was supported in part by Shenzhen Fundamental Research Program (JCYJ20200109110420626, JCYJ20200109110208764), in part by National Natural Science Foundation of China (U22A2034) and in part by Guangdong Basic and Applied Basic Research. Foundation (2022A1515110996, 2023A1515030268, 2021A1515012604).

References

[1] Besson P, Bandt S K, Proix T, Lagarde S, Jirsa V K, Ranjeva J P, Bartolomei F, Guye M. (2017) Anatomic consistencies across epilepsies: a stereotactic-EEG informed high-resolution structural connectivity study. Brain, 140(10), 2639–2652.

Classification Trial of Anaphylaxis Caused by Nonionic Iodine Contrast Agents Using Machine Learning

H. Hattori[1], Y. Oshita[2], A. Teramoto[3], R. Takagi [4], T. Hanamatsu[5], H. azuma[4], S. Ota[5], A. Watanabe[5], E. Sakaguchi [1], Y. Ohno[5], H. Naruse[1], S. Kobayashi[6], H. Toyama[5]

[1]Fujita Health University School of Health Sciences, Department of Medical Laboratory Science, Toyoake, Japan [2]Fujita Health University, Graduate School of Health Sciences, Toyoake, Japan [3]Fujita Health University School of Health Sciences, Department of Intelligent Information Engineering, Toyoake, Japan [4]Fujita Health University Hospital, Department of Radiology, Toyoake, Japan [5]Fujita Health University School of Medicine, Department of Radiology, Toyoake, Japan [6]Fujita Health University School of Health Sciences, Department of Medical Imaging Technology, Toyoake, Japan

Keywords Anaphylaxis, non-ionic iodine contrast med, side effect, safety.

Purpose

Severe anaphylactic reactions to non-ionic iodine contrast media administered intravenously are rare, but it is known that there are cases in which the initial symptoms are mild but rapidly become severe, and the initial response is often difficult. The World Allergy Organization has summarized the criteria in 2020[1] into two. Acute onset of an illness (minutes to several hours) with simultaneous involvement of the skin, mucosal tissue, or both and at least one of the following: Respiratory compromise, Circulatory compromise, Severe gastrointestinal symptoms. Acute onset of hypotension or bronchospasm or laryngeal involvement after exposure to a known or highly probable allergen for that patient (minutes to several hours), even in the absence of typical skin involvement. Severe anaphylactic reactions have been considered that most cases do not belong to the type-I hypersensitivity reactions [2]. However, there is an older theory of IgE-mediated reactions to contrast media, other pathophysiological mechanisms (including direct secretory actions on mast cells and basophils, and activation of the complement system with or without association with the plasma contact system) are also much debated. Anaphylaxis and anaphylactoid reactions are ultimately clinically indistinguishable. It is currently no established method knows to avoid anaphylaxis due to contrast media. Although training and prevention of these side effects are important, but because they occur so rarely, it is difficult to diagnose and cure them in a clinical situation. Thus, if not handled correctly, it can lead to death.

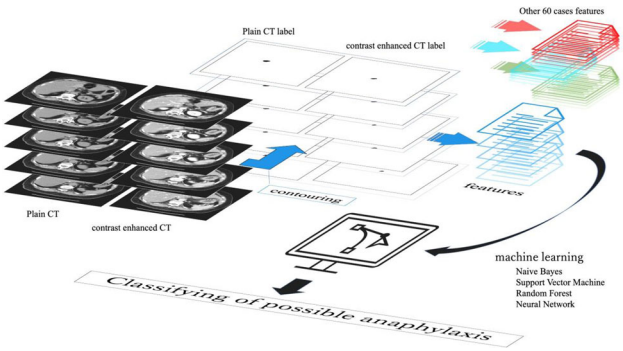

Fig. 1 Outline of the proposed method

Although little is known about imaging findings at the onset of anaphylaxis, it has been reported that the inferior vena cava (IVC) is diminished. Therefore, in this study, we developed a method to automatically detect the diminished of IVCs due to anaphylaxis by machine learning.

Methods

In this study, 61 cases with contrast-enhanced computed tomography (CT) scans using non-ionic iodine contrast media between April 2011 and April 2022 were included in the study. The 61 cases included 20 patients (male-to-female patient ratio, 10:10; mean age: 61.2 years ± 14.3 [standard deviation]) who had no contrast-induced symptoms at the time of CT examination (Normal group), 20 cases (male-to-female patient ratio, 13:7; mean age: 60.3 years ± 21.3) who had mild to moderate allergic symptoms (Allergic group), and 21 cases (male-to-female patient ratio, 10:11; mean age: 65.1 years ± 17.2) who required initial treatment for anaphylactic reactions (Anaphylactic group).

In each case, 5 contiguous slices of 5 mm-thick images were extracted from both plain and contrast-enhanced CT on the same examination date. Extracted areas were 10 mm to 35 mm headword from the celiac artery branch. The board-certified radiologist

Table 1 Anaphylaxis predictive results

	Random forest						SVM	Neural network	Naive Bayes
	1	2	3	4	5	Average			
TP	16	17	18	17	17	85	6	14	11
FP	3	2	2	1	2	10	0	5	19
FN	5	4	3	4	4	20	15	7	10
TN	37	38	38	39	38	190	40	35	21
Binary accuracy	86.9%	90.2%	91.8%	91.8%	90.2%	90.2% ± 2.0	75.4%	80.3%	52.5%
Balanced accuracy	84.3%	88.0%	90.4%	89.2%	88.0%	88.0% ± 2.3	64.3%	77.1%	52.4%
Classification error rate	13.1%	9.8%	8.2%	8.2%	9.8%	9.8% ± 2.0	24.6%	19.7%	47.5%
Precision	84.2%	89.5%	90.0%	94.4%	89.5%	89.5% ± 3.6	100.0%	73.7%	36.7%
F-score	80.0%	85.0%	87.8%	87.2%	85.0%	85.0% ± 3.1	44.4%	70.0%	43.1%
Sensitivity	76.2%	81.0%	85.7%	81.0%	81.0%	81.0% ± 3.4	28.6%	66.7%	52.4%
Specificity	92.5%	95.0%	95.0%	97.5%	95.0%	95.0% ± 1.8	100.0%	87.5%	52.5%

TP true positive, *FP* false positive, *FN* false negative, *TN* true negative

manually contoured the IVC to the extracted images under the blind for the symptoms, etc., and created label images. From the label images, features such as area, major diameter, minor diameter, and angle were extracted. These were machine-learned (Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Neural Network (NN)) by Orange (Ver. 3.32. 0) using the five-part validation method, and the cases causing anaphylaxis were classified based on the results. In RF, five iterations were performed, and the average value was utilized.

Results

The prediction accuracy of anaphylaxis by five-part validation was analyzed in 61 patients who underwent contrast-enhanced CT. The Binary Accuracy and Balanced Accuracy of prediction by four machine learning methods were 52.5% and 52.4% (NB), 75.4% and 64.3% (SVM), 90.2% ± 2 and 88.8% ± 2.3 (RF), and 80.3% and 77.1% (NN) (Table 1). In this study, RF performance was the best, although results varied.

Conclusion

In this study, we developed a method to detect anaphylaxis from contrast-enhanced CT images obtained by machine learning the label images of IVC before and after imaging, and demonstrated its usefulness in predicting the onset (Fig. 1).

References

[1] Cardona V, Ansotegui IJ, Ebisawa M, El-Gamal Y, Fernandez Rivas M, Fineman S, Geller M, et al. World allergy organization anaphylaxis guidance 2020. World Allergy Organ J 2020; 13:100,472.

[2] Eloy R., Corot C., Belleville J. Contrast media for angiography: physicochemical properties, pharmacokinetics and biocompatibility. Clin. Mat. (1991) 7 89–197.

Diagnostic support system for coronary artery disease using dynamic heart shapes and metadata

S. Takashima[1], S. Miyauchi[2], K. Morooka[3], R. Kurazume[2]

[1]Kyushu University, Department of Electrical Engineering and Computer Science, Fukuoka, Japan [2]Kyushu University, Faculty of Information Science and Electrical Engineering, Fukuoka, Japan [3]Okayama University, Faculty of Natural Science and Technology, Okayama, Japan

Keywords Coronary artery disease, Dynamic heart shape, Lifestyle, Single nucleotide polymorphism.

Purpose

When diagnosing coronary artery disease (CAD) which is the main cause of death in the world, medical doctors check two factors: one is the shape and motion of the patient's heart (referred to as "dynamic cardiac shape") and the other is the metadata of patient including patient's lifestyle and family history of CAD. One noninvasive method for capturing the dynamic heart shape is to use cardiac MR images (CMR images). The CMR image consists of a sequence of 3D images during a cardiac cycle. The dynamic 3D heart shape changes are observed by the CMR images. Using the CMR images, Biffi et al. [1] proposed a diagnostic support system for hypertrophic cardiomyopathy composed of ladder variational autoencoder (LVAE) and multilayer perceptron (MLP). However, the system employed only the dynamic heart shape as input to the system.

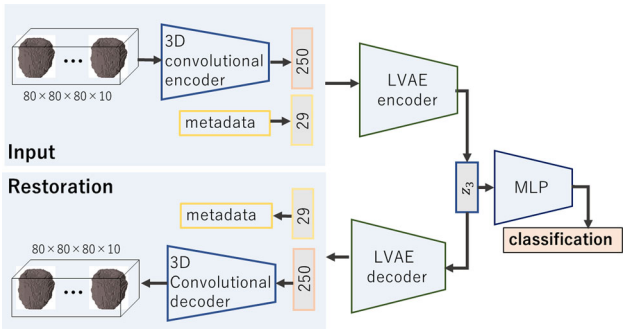

Fig. 1 Overview of our diagnose system for CAD

Therefore, we propose a new diagnostic support system for CAD by using dynamic heart shapes and multiple metadata including lifestyle and genetic information.

Methods

Figure 1 shows the overview of our system which classifies CAD patients and healthy subjects using the dynamic heart shape and metadata. The dynamic heart shape is described with a set of voxel models of the left ventricle. Each voxel model with 80 × 80 × 80 [voxel] is generated from the CMR image in a frame during a cardiac cycle by using joint learning model [2]. Here, in the previous research [1], only 2-frame voxel models are used as the dynamic heart shape (2-frame dynamic heart shape). On the other hand, our system uses 10-frame voxel models as the dynamic heart shape (10-frame dynamic heart shape) to consider more detailed shape changes during the cardiac cycle.

The metadata used in our system are three types: 6 continuous data (systolic/diastolic blood pressures, body fat percentage, BMI, forced vital capacity, and pulse rate), 3 discrete data (alcohol frequency and past/current tobacco frequencies), and 7 single nucleotide polymorphisms (SNPs) which are highly relevant to CAD. Among the metadata, the continuous data are converted to a 6-dimentional vector after applying z-score normalization to each datum, and the discrete data are converted to one-hot vectors. Also, 7 SNPs are described as a 7-dimentional vector where each element represents the number of minor alleles (0, 1, or 2) in its corresponding SNP. By collecting the data, the input metadata is a 29-dimentional vector.

In our system, when the dynamic heart shape is given, a 3D convolutional encoder extracts the feature vector from the dynamic heart shape. The extracted feature vector and the metadata vector are inputted to a 3-layer LVAE encoder to obtain a latent variable $\boldsymbol{z_3}$ corresponding to the input vectors. Using $\boldsymbol{z_3}$, MLP conducts a binary classification between CAD patients and healthy subjects.

To train our system, we define our loss function L as follows:

$$L = \gamma(\Sigma_{i=1}^{3} k_i \boldsymbol{K}_i + c\boldsymbol{C}) + \boldsymbol{D} + \mathrm{m}\boldsymbol{M},$$

where $\boldsymbol{K_i}$ is a Kullback–Leibler divergence to bring a distribution of the latent variable $\boldsymbol{z_i}$ of the ith-layer of LVAE closer to a Gaussian distribution. $\boldsymbol{C}$ is a cross entropy loss to evaluate error of binary classification. In the training of our system, to construct the LVAE encoder, the input dynamic heart shape and metadata vector are restored by using 3-layer LVAE decoder and 3D convolutional decoder. Therefore, $\boldsymbol{D}$ is the dice coefficient between the input and restored dynamic heart shapes. $\boldsymbol{M}$ is squared errors between the input and restored metadata vectors. Through a preliminary experiment, scale factors m, k_i, c, γ of each term are set to m = 3.0, k_1 = 0.02, k_2 = 0.001, k_3 = 0.0001, c = 3.0, and γ = 1. Also, latent variables z_1, z_2, and z_3 of each layer are set to 48, 32, and 2, respectively.

Table 1 Experimental results with different inputs to the system

	Accuracy	Precision	Recall
Our system (10-frame dynamic heart shape + metadata)	0.86 ± 0.02	0.87 ± 0.07	0.85 ± 0.08
Only metadata	0.76 ± 0.05	0.76 ± 0.06	0.75 ± 0.05
Only 10-frame dynamic heart shape	0.76 ± 0.04	0.80 ± 0.09	0.72 ± 0.10
Previous system [1] (only 2-frame dynamic heart shape)	0.70 ± 0.04	0.72 ± 0.05	0.65 ± 0.07

Results

To validate the availability of our system, we conducted four experiments by using four systems with different inputs: (1) 10-frame dynamic heart shapes and metadata (our system), (2) only metadata, (3) only 10-frame dynamic heart shapes, and (4) only 2-frame dynamic heart shapes which is the previous system [1]. In the experiments, we used the dataset composed of 100 healthy subjects and 100 patients with CAD published in UK Biobank. By dividing the dataset into four sub-datasets so that each sub-dataset includes 50 data, we performed fourfold cross-validation by using one sub-dataset as test data and others as training data.

Table 1 shows the average accuracy, precision, and recall of each experiment. From rows 4 and 5 of Table 1, increasing the number of frames in the dynamic heart shape from 2 (previous system [1]) to 10 improved the classification accuracy. Moreover, as shown in rows from 2 to 4 of Table 1, our system inputting both the dynamic heart shape and metadata greatly improved the classification accuracy compared to the cases inputting them separately.

Conclusion

We proposed a diagnostic support system for CAD using the dynamic heart shape and patient's metadata. Through the experiment, we confirmed that the classification accuracy of our system outperforms that of the baseline.

References

[1] Biffi C, Cerrolaza JJ, Tarroni G, Bai W, de Marvao A, Oktay O, Ledig C, Le Folgoc L, Kamnitsas K, Doumou G, Duan J, Prasad SK, Cook SA, O'Regan DP, Rueckert D (2020) Explainable anatomical shape analysis through deep hierarchical generative models. IEEE transactions on medical imaging, 39(6), pp. 2088–2099, 2020.

[2] Qin C, Bai W, Schlemper J, Petersen SE, Piechnik SK, Neubauer S, Rueckert D (2018) Joint learning of motion estimation and segmentation for cardiac MR image sequences, MICCAI, pp. 472–480

Development of practical education system for intraoral radiography using Extended Reality

K. I. Ejima[1], T. Sasaki[1], K. Honda[1], Y. Arai[1]

[1]Nihon University School of Dentistry, Oral and Maxillofacial Radiology, Tokyo, Japan

Keywords Extended Reality, VR headset, educational system, intraoral radiography.

Purpose

The principle of intraoral radiography is difficult for undergraduate students to understand. We have therefore introduced a system in which, after lectures, undergraduate students take images of each other during practical sessions of intraoral radiography. The expansion of Covid-19 has made it difficult for undergraduate students to practice radiography with each other in intraoral radiography due to the risk of infection. Under these circumstances, we have tried to develop an environment that facilitates students' understanding of the principles. In recent years, educational methods using Extended Reality (XR) have been introduced in various fields [1]. We have developed an educational system for intraoral radiography practice using XR, Fig. 1.

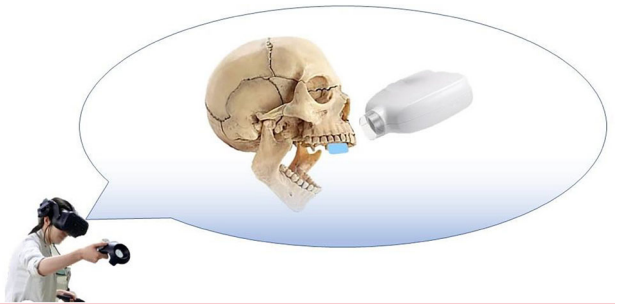

Fig. 1 Diagram of the education system

Methods

This educational system involves undergraduates wearing a Virtual Reality (VR) headset and practicing intraoral radiography within the VR. In the VR, there is a skull, a film, and an xray tube, and undergraduate students can learn the appropriate bisecting technique and orthoradial projection for the film and teeth on which the dentition is placed. For the data of this educational system, first, CT scanned skull image data was output as a polygon file using a workstation, and then the data was converted to VR headset data. Those datasets were loaded into a VR headset. In a similar way, the film and xray tube datasets were loaded into the VR headset. The VR headset used was the Meta Quest 2(Meta,USA), and the conversion of polygon files to VR headset data was done using the website(Holoeyes Inc.,Tokyo,Japan).

One of our staff members wore the VR headset of this system to check the operability and operational status of the system. Other staff members checked the images from the VR headset output externally to a PC.

Results

This educational system operated without major problems and was generally easy to operate. Instructors commented that they had a better understanding of intraoral radiography and that they could learn efficiently in the VR space, regardless of their location. On the other hand, some said that they could not reproduce the feeling of holding

the xray tube, that they needed to get used to the operation, and that they could not get instructions from the instructor in the VR.

Conclusion

We have developed an educational system for the practice of intraoral radiography using XR, and confirmed the system in terms of its usability. In our opinion, the system was largely satisfactory. In the future, we plan to customize the system and solve the problems. We believe that this educational system will make a significant contribution to students" future use of intraoral radiography in real patients.

References

[1] Sugahara K, Koyachi M, Koyama Y, Sugimoto M, Matsunaga S, Odaka K, Abe S, Katakura A (2021) Mixed reality and three dimensional printed models for resection of maxillary tumor: a case report. Quant Imaging Med Surg 11(5):2187–2194

Comparison of tumor detection models in laryngeal endoscopic still images using deep learning

S. H. Lee[1], Y. J. Kim[2], J. W. Seo[1], S. B. Mun[1], K. G. Kim[2]

[1]Gachon University, Health Sciences and Technology, Incheon, South Korea [2]Gachon University, Biomedical Engineering, Incheon, South Korea

Keywords deep learning, object detection, Computer Aided Diagnosis, laryngeal endoscopic

Purpose

Laryngeal cancer is the most common malignant tumor of the head and neck except for thyroid cancer. In the case of laryngeal cancer, the presence or absence of lesions is determined through endoscopy, and malignancy is confirmed through biopsy. Results read from endoscopic images may vary depending on the proficiency of the clinician, and even skilled people may not find microtumors. Accordingly, there is a need to develop a diagnostic assistance technology to reduce the deviation of image reading and maintain constant accuracy.

Currently, research on artificial intelligence technology that analyzes and classifies MRI and CT images to assist diagnosis is being actively conducted, but few studies have incorporated artificial intelligence technology in the field of endoscopy [1].

Therefore, in this study aims to improve clinicians' reading accuracy and reduce misdiagnosis by comparing the Yolov5 and Yolov6 models and developing a lesion detection model using them.

Table 1 Model performance results according to threshold

Model	Prob	TP	FN	FP	Recall	precision	FPPI
YOLOv5	0.5	137	63	8	0.68	0.94	0.04
	0.8	131	69	7	0.65	0.94	0.03
YOLOv6	0.5	148	52	10	0.74	0.94	0.04
	0.8	123	77	4	0.61	0.96	0.02

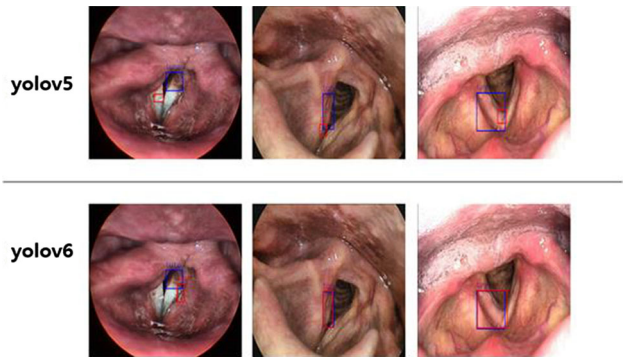

Fig. 1 The location of the lesion predicted by the model in data where lesion detection is difficult; the red box is what the model predicted, and the blue box is the Ground-truth

Methods

In this study, laryngeal endoscopy was performed on adult patients who underwent laryngeal endoscopy. The collected 1500 pieces of data consist of normal data, benign data, and malignant data, and the learning was conducted by dividing them into 900 training data and 600 model performance evaluation data data. To detect the presence or absence of lesions, the Yolov5 and Yolov6 models, which are the first-stage object detection frameworks, were used. The deep learning environment was learned by setting the number of epochs to 200, the batch size to 16, and the input image size to 640 × 640.

Results

The performance evaluation of the model was expressed in terms of precision, recall, and average precision (AP). According to the threshold, when checking the precision and recall values of the data predicted by the Yolov5 and Yolov6 models to have lesions, if the threshold is set to 0.5, the accuracy of the Yolov5 model is 94%, the recall rate is 68%, the accuracy of the Yolov6 model is 93%, and the recall rate is 74%, Table 1. When the critical value was set to 0.8, it showed 94% accuracy and 65% recovery rate for the Yolov5 model, 96% accuracy and 61% recovery rate for the Yolov6 model. In addition, the difference between the two models was confirmed in the data that is difficult to locate the lesion, and the Yolv6 model was found to detect the lesion closer to the actual lesion site than the Yolov5 model.

Conclusion

Both Yolov5 and Yolov6 showed that the area of the lesion predicted by the model was detected close to the area of the actual lesion. Since benign tumors are more prominent than malignant tumors, the trained model captures the benign data almost the same as the ground truth, while for malignant data, we detect wider than the ground truth, and this difference is greater in the Yolov5 model than in the Yolov6 model, Fig. 1. In addition, all data with lesions were not found among the model performance evaluation data, and it is judged that it is necessary to supplement them through future research.

In future studies, it is expected that generalized model performance will be obtained through cross-validation, and the detection ability for various types of tumors will be improved through model advancement by increasing the number of cases for tumors that have not been found.

References

[1] Sunyoung Park (2021) Laryngeal cancer diagnosis prediction through deep learning. Industrial Engineering Magazine.

Breast mass detection using paired mammogram views based on artificial intelligence algorithm

J. W. Seo[1], Y. J. Kim[2], S. H. Lee[1], M. Sae Byeol[1], K. G. Kim[2,1]

[1]Gachon University, Health Sciences and Technology, GAIHST, Incheon, South Korea [2]Gachon University, Biomedical Engineering, Gil Medical Center, Incheon, South Korea

Keywords breast mass detection, deep learning, mammogram, computer aided diagnosis

Purpose

In clinical practice, using two standard mammogram views (cranial–caudal (CC) and mediolateral-oblique (MLO)) helps radiologists make more reliable clinical decisions by providing information of correspondence. Some studies proposed deep learning algorithms based on matching mass patch images extracted from each view for FP reduction. However, these methods could ignore each breast mass's surrounding information and location characteristics. Therefore, this study aimed to investigate the effectiveness of a proposed AI algorithm for improving the performance of breast mass detection by considering the relational information in the paired whole mammogram [1].

Methods

We collected 1050 mammograms from Hologic Lorad Selenia 2D Digital Mammography machine (Hologic) and 586 mammograms from General Electric healthcare Senographe Essential (GE). We used 1476 data (944 Hologic and 532 GE data) for training models. The 160 data (106 Hologic data, 54 GE data) were not used for training the models but for evaluating the model performance. The convolutional neural network (CNN) extracts a feature map based on the information for each channel by performing convolution operations according to the channel. Therefore, we generated synthesized pseudo-color images from the gray-scale original mammogram by adjusting the histogram distribution and fusing each image along the channel axis. The collected mammograms have 2560 × 3328 (pixel) or 3328 × 4096 (pixel) resolution, which is a considerable size comparing other data in deep learning algorithms of computer vision (256 × 256 (pixel) or 512 × 512 (pixel)). Hence, when the original images are used as they are, excessive computation and memory problems occur due to their large size. To prevent this situation, we resampled the image size to 512 × 512 (pixel). The height of the image was reduced to 512 pixels, and the width was resized to maintain the original ratio based on the height. Then we applied zero-padding to fit and scaled the width and height to 512 pixels. Since the number of training data is proportional to the model's performance, a data augmentation technique was applied to compensate for the small number of data. To simultaneously utilize the correlated information from the whole paired views, we propose a CNN algorithm using cosine similarity and the Squeeze and Excitation method, which applies weights using each channel of the feature map extracted from the convolution layer. The proposed network consists of a U-shape structure which has shown excellent performance in medical image segmentation. The proposed structure comprises encoders and decoders, and we applied VGGnet16 as an encoder. To demonstrate the effectiveness of the proposed method, we compared it with the performance of the single view model, which has one U-shape structure and extracts feature maps using a single image to not share the pair-view-based information.

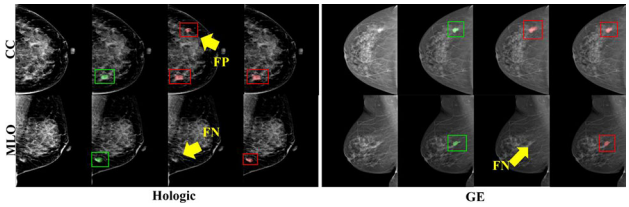

Fig. 1 Example images of the prediction results from models. The green rectangles and regions represent to ground truth and the red rectangles and regions represent to prediction from models. **a** Original images. **b** Ground truth. **c** Results from single-view model. **d** Results from paired-view model. False positive, FP; False negative, FN

Results

We compared the performance results with the single view models and the paired view model in 160 test sets. We extracted the segmentation results from model output and the detection results of rectangle boxes from mass area. In the segmentation results, the proposed paired-view model achieved 0.709 dice similarity score (DSC), and the single-view model (the contrast model) reached 0.579 DSC. From the segmented map from models and ground truth, we get box region including mass and calculated the detection performance for verifying the localization ability of models. The proposed model achieved 0.950 sensitivity and 0.900 precision at 0.156 false positive per images (FPPI). The single-view model achieved 0.813 sensitivity and 0.747 precision at 0.188 FPPI in single view model. The Table 1 show the results of detection performance from each model. Our study shows that the proposed algorithm effectively reduced FPPI and increased sensitivity by sharing the feature map between the two views, increasing the weight of channels that could be missed from one side and lowering the weights of insignificant channels.

Conclusion

In this paper, we proposed a deep learning-based artificial intelligence algorithm that enhances breast mass detection performance by extracting significant features by using two standard mammogram images simultaneously. The validity of the proposed

Table 1 The performance results from models

View	Sensitivity (95% CI)	Precision (95% CI)	FPPI (95% CI)
Pair	0.95 (0.912–0.988)	0.9 (0.859–0.941)	0.156 (0.092–0.22)
Single	0.813 (0.749–0.876)	0.747 (0.684–0.809)	0.188 (0.118–0.257)

algorithm based on the dual images is verified by comparing the performance of a single view based algorithm. As shown in Fig. 1, the result demonstrates the effectiveness of the proposed algorithm for reducing the false positives and preventing missing true positives comparing based on single view algorithm. We expect that this proposed method can provide practical diagnostic solutions for improving patients' treatment as a computer-aided diagnosis system.

References

[1] R.M. Warren, S. Duffy, and S. Bashir, The value of the second view in screening mammography. Br. J. Radiol. Vol.69, pp. 105–108, 1996.