



# A deep learning based fusion of RGB camera information and magnetic localization information for endoscopic capsule robots

Mehmet Turan<sup>1</sup> · Jahanzaib Shabbir<sup>2</sup> · Helder Araujo<sup>2</sup> · Ender Konukoglu<sup>3</sup> · Metin Sitti<sup>1</sup>

Received: 10 August 2017 / Accepted: 14 November 2017 / Published online: 23 November 2017  
© The Author(s) 2017. This article is an open access publication

## Abstract

A reliable, real time localization functionality is crucial for actively controlled capsule endoscopy robots, which are an emerging, minimally invasive diagnostic and therapeutic technology for the gastrointestinal (GI) tract. In this study, we extend the success of deep learning approaches from various research fields to the problem of sensor fusion for endoscopic capsule robots. We propose a multi-sensor fusion based localization approach which combines endoscopic camera information and magnetic sensor based localization information. The results performed on real pig stomach dataset show that our method achieves sub-millimeter precision for both translational and rotational movements.

**Keywords** Deep Learning based Sensor Fusion · Endoscopic Capsule Robots · RNN-CNN (RNN:Recurrent Neural Network, CNN: Convolutional Neural Network)

## 1 Introduction

Robot localization denotes the robot's ability to establish its position and orientation within the frame of reference. Different sensors used in medical milliscale robot localization have their own particular strengths and weaknesses, which makes sensor data fusion an attractive solution. Monocular visual-magnetic odometry approaches, for example, have received considerable attention in mobile robotic sensor fusion literature. In general, localization techniques for endoscopic capsule robots can be categorized into three main groups: electromagnetic wave-based techniques; magnetic field strength-based techniques and hybrid techniques Umay et al. (2017).

In recent years, numerous electromagnetic wave-based approaches like time of flight and difference of arrival (ToF and TDoA)-, received signal strength (RSS)-, RF

identification (RFID)- and angle of arrival (AoA) based methods have been proposed Wang et al. (2011); Fischer et al. (2004); Wang et al. (2009); Ye (2013); Hou et al. (2009).

In magnetic localization systems, the magnetic source and magnetic sensor system are the essential components. The magnetic source can be designed in different ways: a permanent magnet, an embedded secondary coil, or a tri-axial magnetoresistive sensor. Magnetic sensors located outside the human body detect the magnetic flux density in order to estimate the location of the capsule (e.g., Popek et al. (2013), Natali et al. (2016), Yim and Sitti (2013)). One of the major advantages of utilizing magnetic field strength-based localization techniques is their successful coupling with magnetic locomotion systems. This could be achieved using magnetic steering, magnetic levitation, and remote magnetic manipulation. Other advantages include their robustness against attenuation by the human body. However, the disadvantage is that they experience interference from the environment. This could be handled by implementing additional hardware for handling the localization problem.

Another group of endoscopic capsule robot localization techniques is the hybrid techniques. These implement an integration of different sources at once such as RF sensors, magnetic sensors, and RGB sensors. The core idea is to integrate data from different sources which strengthen each other and can produce more accurate localization data. As a common

✉ Mehmet Turan  
turam@is.mpg.de

Metin Sitti  
sitti@is.mpg.de

<sup>1</sup> Physical Intelligence Department, Max Planck Institute for Intelligent Systems, Stuttgart, Germany

<sup>2</sup> Institute for Systems and Robotics, Universidade de Coimbra, Coimbra, Portugal

<sup>3</sup> Computer Vision Laboratory, ETH Zurich, Zurich, Switzerland

approach Kalman filter and its derivatives are proposed to fuse RF electromagnetic signal data, magnetic sensor data, and video data. The first group of hybrid methods fuses RF and video signal Geng and Pahlavan (2016); Bao et al. (2015), whereas the second group focus on fusion of magnetic and RF signal data Umay and Fidan (2016); Geng and Pahlavan (2016); Umay and Fidan (2017) and the last group on fusion of magnetic and video data Gumprecht et al. (2013).

Some other methods of localization utilize X-rays, MRI, computed tomography (CT), or ultrasound sensing Arshak and Adepoju (2006) and  $\gamma$  rays Than et al. (2014). However, they all have their respective drawbacks of radiation hazards. MRI hardware is costly and presents (Fig. 1) additional design restrictions, and ultrasound sensing acquire planar pictures that might not intersect with the capsule robot. Inspired by the recent success of deep-learning models for processing raw, high-dimensional data, we propose in this paper a sequence-to-sequence deep sensor fusion approach for endoscopic capsule robot localization.

## 2 System architecture details

Regardless of the algorithm, traditional monocular visual odometry solutions are subject to scale drift and ambiguity. It is proven that sophisticated loop closure methods can

be helpful to reduce scale drift. However, scale ambiguity requires fusion of external information to be solved which not only resolves the scale ambiguity but also increases the pose estimation accuracy. With that motivation, we developed a deep neural network approach which fuses hand-eye calibrated and synchronized RGB camera information with magnetic localization information. Figure 3 shows the system architecture diagram of our sequence-to-sequence learning approach consisting of:

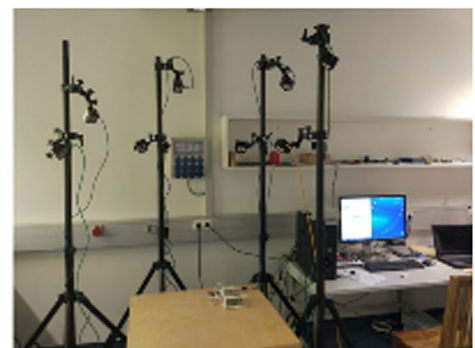
1. Optical Flow estimation.
2. CNN based feature vector extraction.
3. LSTMs based sensor fusion.

Before optical flow estimation, input frames are preprocessed by a vessel enhancement module which aims to emphasize unique features on the organ tissue (Fig. 2). As a next step, keyframe detection module examines each endoscopic camera frame and identify keyframes. Magnetic localization data coming from 2D Hall sensor array is a 6 dimensional vector containing the x, y, z position components and x-, y- and w orientation parameters in quaternion format (rotational z-degree is missing). The output of the network is a 7 dimensional vector consisting of x-, y-,

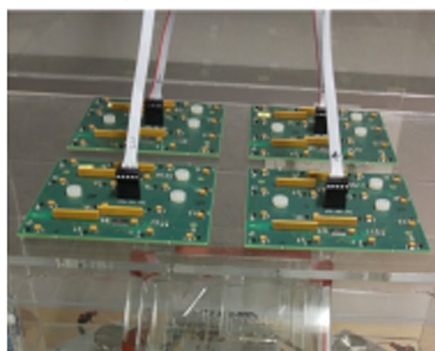
Fig. 1 Experimental setup



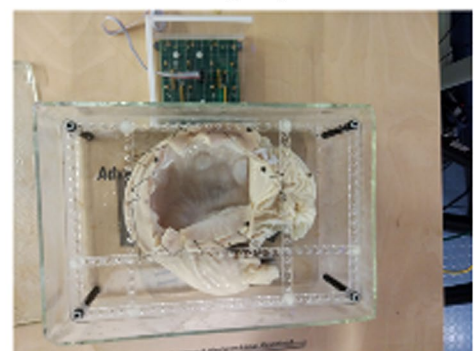
Magnetically actuated soft capsule endoscopes



Optitrack Prime 13 Tracking System

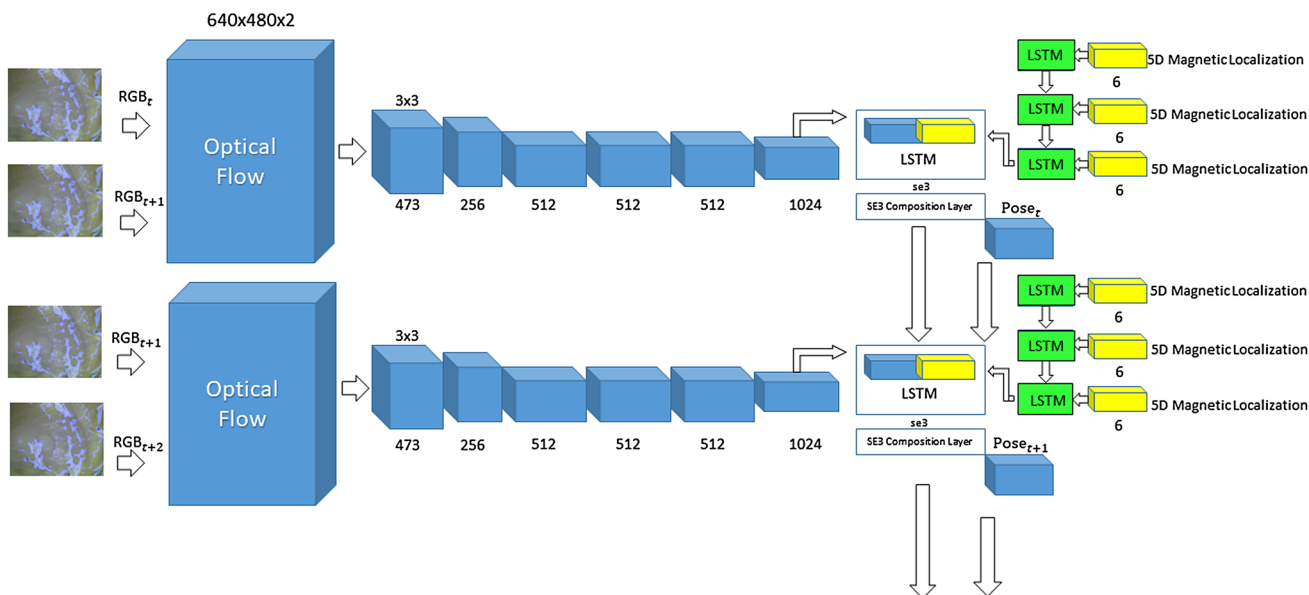
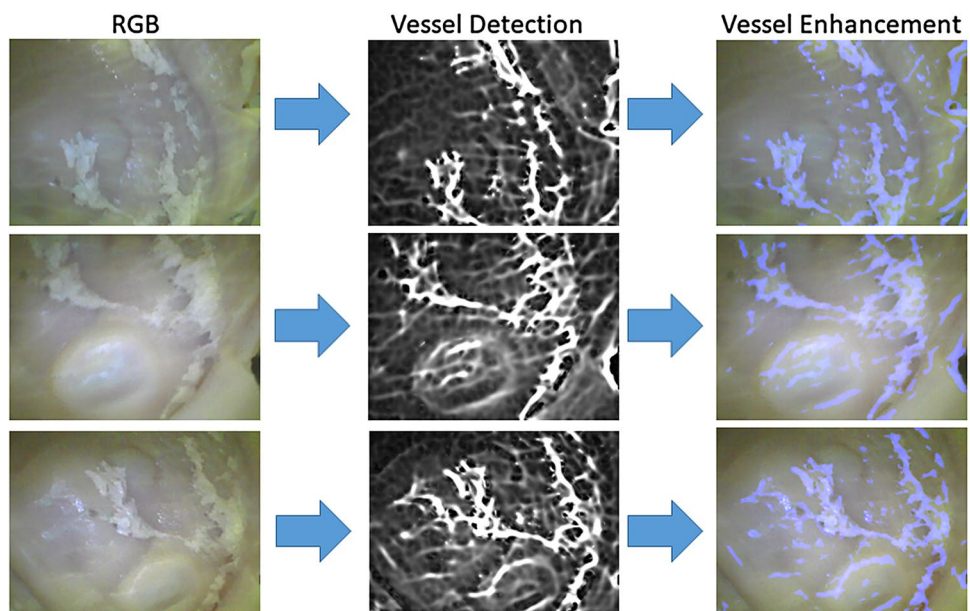


Magnetic localization sensors



A preserved pig stomach

**Fig. 2** Vessel detection and enhancement



**Fig. 3** System architecture diagram

z- translation and 4 orientation parameters in quaternion format.

**2.1 Preprocessing**

Even though the beauty of deep learning is told to be its success and easiness to process raw input data without inquiring any pre-and post processing, we do preprocessing since it increases the accuracy of our method upon our observations we made during evaluations. This section explains the preprocessing operations we applied on the raw RGB

image data before passing it into the deep neural network. The operations include vessel detection, enhancement and keyframe selection.

**2.1.1 Multi-scale vessel enhancement**

Since endoscopic images have mostly homogeneous and poorly textured areas, our framework starts with a vessel enhancement operation inspired from Frangi et al. (1998). Proposed approach enhances blood vessels by analyzing the

multiscale second order local structure of an image. First, we extract the Hessian matrix:

$$H = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix} \quad (1)$$

where  $I$  is the input image, and  $I_{xx}$ ,  $I_{xy}$ ,  $I_{yx}$ ,  $I_{yy}$  the second order derivatives, respectively. Secondly, eigenvalues  $|\lambda_1| \leq |\lambda_2|$  and principal directions  $u_1$ ,  $u_2$  of the Hessian matrix are extracted. The eigenvalues and principal directions are then ordered and analyzed to decide whether the region belongs to a vessel. To identify vessels in different scales and sizes, multiple scales are created by convolving the input image and the final output is taken as the maximum of the vessel filtered image across all scales. For further details of the mathematical equations, the reader is referred to the original paper of [?]. Figure 2 shows input RGB images, vessel detection and vessel enhancement results for different frames.

### 2.1.2 Keyframe selection

Due to the incremental and slow motion of the capsule robot inside the inner organ, endoscopic videos generally contain numerous frames with similar and redundant content. Thus, an algorithm has to be developed to skip frames with similar overlapping features. This procedure is called keyframe selection. We developed a method based on optical flow interpretation between consecutive frame pairs. The output of optical flow algorithm is the vector values of each pixel which we sum the magnitudes of and normalize by dividing it with the total number of pixels. If the normalized value exceeds a pre-defined threshold value of 20 pixels, the overlap between the corresponding frames is less than 75% and it is identified as a keyframe. Conversely, if it does not exceed the threshold then there is a high overlap between the frame pair. This procedure is summarized below:

1. Choose a candidate keyframe and extract Farneback optical flow between this and the reference keyframe.
2. Compute the magnitude of the extracted optical flow vector for each pixel.
3. Calculate the cumulative value by summing up all the magnitude values.
4. Normalize the cumulative value by the total number of pixels.
5. If the normalized cumulative value is less than  $\tau$  then go to the next frame. Otherwise, identify the candidate key frame as a key frame and repeat the process.

## 2.2 Optical flow extraction

In deep learning based applications, there is in general a tendency to serve raw input images into the neural network without any preprocessing so that neural network can decide by itself how to organize the raw information. Contrary to that, we do not use raw images, instead we extract optical flow from consecutive keyframes. This way, we want to force CNN to focus more directly on the motion dynamics between frames rather than redundant unnecessary information. To achieve real time performance, we make use of GPU for optical flow estimation. Several optical flow algorithms such as Lucas-Kanade method (Cornelius and Kanade 1984), Buxton-Buxton method (Beauchemin and Barron 1995) and the Black Jepson method were tested. The Farneback optical flow estimation based on polynomial expansion (Farneback 2003) out-performed all other optical flow methods by having the lowest reprojection error (Fig. 4). The input information of the CNN needs to be discrete which arises the need to quantify the optical flow vector. The resolution of the input image is  $640 \times 480$ , so the maximum value of the quantification is devised to be the diagonal length of the image resolution, and minimum quantification value is set to zero. The quantification range is divided into 1024 intervals and the resulting quantized x and y values were concatenated for each pixel (Fig. 3).

## 2.3 Magnetic localization system

As described in Fig. 4, our magnetic localization system consists of a magnetic Hall sensor array for localization and electromagnets for actuation of the endoscopic capsule robot. Our magnetic localization technique is able to measure 5-DoF absolute pose of the untethered meso-scale magnetic robot in real-time. A Hall-effect sensor array measures magnetic field at several locations from the capsule robot,

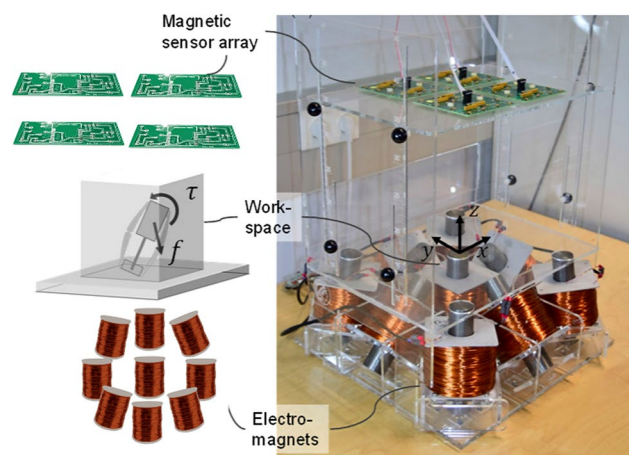


Fig. 4 Magnetic localization system



whereas a computer-controlled electromagnetic coil array provides actuator's magnetic field. The core idea of our localization technique is separation of capsule's magnetic field from actuator's known magnetic field, which is realized by subtracting actuator's magnetic field component from the acquired magnetic data. Finally, noise effects are reduced by second-order directional differentiation. For curious readers, further details of our magnetic localization technique can be found in Son et al. (2016).

## 2.4 Deep CNN-RNN architecture for sensor fusion

We propose a sensor fusion architecture consisting of CNN layers for feature extraction from quantized optical flow vector, multi-rate long short-term memories (LSTMs) for frequency adjustment, and a core LSTM unit for fusion of flattened feature vector and magnetic localization information. The details of the system architecture can be seen in Fig. 3. The core part of the architecture is inspired and modified from Clark et al. (2017). For the implementation, Keras library with Theano back-end was used which provides a modifiable framework, enables multi-GPU training which accelerates the computational procedure. The learning rate was initialized to 0.001 reducing as the epochs of the training continues. Adaptive moment estimation (Adam) method was used to optimize the goal function. We trained our algorithm on an Amazon EC2 p2.xlarge GPU compute instance. The list of the parameters is as follow:

- learning rate: 0.001
- momentum1: 0.9
- momentum2: 0.999
- epsilon:  $10^{-8}$
- solver type: Adam
- batch size: 64
- GPU: NVIDIA K80

With using LSTM, we pursue to learn the complex motion dynamics of endoscopic capsule robot and try to let the neural network describe sequential dependencies across frames which require extensive engineering in case of manual modeling. Contrary to the traditional LSTM, we connect the output pose of the current core LSTM as input to the core LSTM of the next timestep so that odometry can benefit from the information of past frames thanks to its hidden memory lasting over time. With the help of this hidden memory, LSTM can encode the previously gained knowledge up to time step  $t$  and use it for posterior estimations. An exponential map in the  $SE(3)$  composition layer is used to convert  $se(3)$  data to the special euclidean group  $SE(3)$  Clark et al. (2017). In our architecture, each LSTM layer has

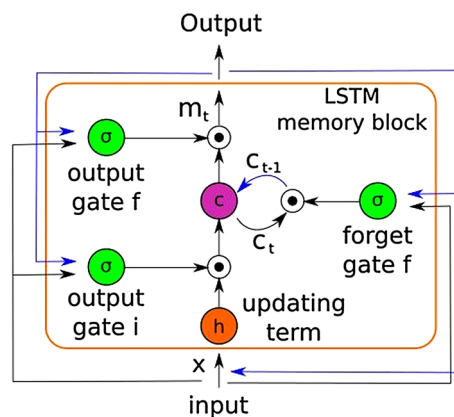


Fig. 5 Information flow through the hidden units of the LSTM

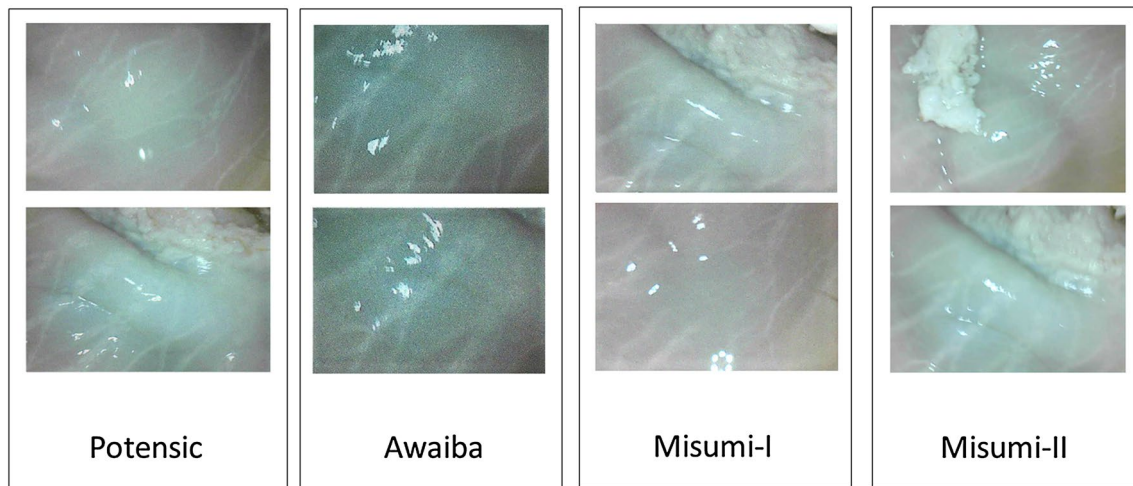
200 hidden states. To regress the 6-DoF pose, we trained the architecture on the following objective loss function:

$$\text{loss}(I) = \|\hat{\mathbf{x}} - \mathbf{x}\|_2 + \beta \|\hat{\mathbf{q}} - \mathbf{q}\|_2 \quad (2)$$

where  $\mathbf{x}$  is the translation vector and  $\mathbf{q}$  is the quaternion vector for a rotation. A balance  $\beta$  must be kept between the orientation and translation loss values which are highly coupled as they are learned from the same model weights. Experimental results show that the optimal  $\beta$  is given by the ratio between expected error of position and orientation at the end of training session (Fig. 5). The back-propagation algorithm is used to determine the gradients of the network weights which are passed into the Adam optimization. The moments of the gradient are calculated using exponential moving average in addition to exponentially decaying average of past gradients, which also corrects the bias.

## 3 Dataset

This section introduces the experimental setup and explains how the training and testing datasets were created. The dataset was recorded on five different real pig stomachs (see Fig. 1). In order to ensure that our algorithm is not tuned to a specific camera model, four different commercial endoscopic cameras were employed. For each pig stomach and camera combination, 3000 frames were acquired, which makes 60,000 frames for four cameras and five pig stomachs in total. 40,000 frames were used for training, whereas the remaining 20,000 frames were used for evaluation. Sample real pig stomach frames are shown in Fig. 6 for visual reference. During video recording, an Optitrack motion tracking system consisting of eight Prime-13 cameras was utilized to obtain 6-DoF localization ground-truth-data with sub-millimeter accuracy (see Fig. 1).



**Fig. 6** Sample images from dataset

## 4 Evaluation

We evaluate the performance of our system both quantitatively and qualitatively in terms of trajectory estimation. We also report the computational time requirements of the method.

### 4.1 Trajectory estimation

The absolute trajectory (ATE) root-mean-square error metric (RMSE) is used for quantitative comparisons, which measures the root-mean-square of Euclidean distances between all estimated endoscopic capsule robot poses and the ground truth poses. We created six different trajectories with various complexity levels. Overfitting, which would make the resulting pose estimator inapplicable in other scenarios, was prevented using dropout and early stopping techniques. The dropout regularization technique, which samples a part of the whole network and updates its parameters based on the input data, is an extremely effective and simple method to avoid overfitting.

Early stopping is another widely used technique to prevent overfitting of a complex neural network architecture optimized by a gradient-based method. We strictly avoided the use of any image frames from the training session for the testing session. We compared the performance of our deep fusion approach with the odometry approach proposed in Turan et al. (2017) which we call endoscopic

visual odometry (EVO), magnetic localization approach proposed in Son et al. (2017) and with the same CNN-RNN system in Fig. 3 except we disabled the magnetic localization fusion path. We call this last configuration deep visual odometry (DVO). The average translational and rotational RMSEs for deep sensor fusion, EVO, magnetic localization and DVO against different path lengths are shown in Figs. 8 and 9, respectively. The results indicate that deep sensor fusion clearly outperforms all other configurations, while magnetic localization outperforms EVO and DVO. Some qualitative tracking results and corresponding ground truth trajectories for deep fusion approach, DVO and magnetic localization are demonstrated in Fig. 7 for visual reference. As seen in sample trajectories, deep fusion is able to stay close to the ground-truth pose values for even complex, fast rotational and translational motions, where both EVO and magnetic localization by themselves clearly deviate from the ground-truth trajectory. Thus, we can conclude that deep fusion makes effective use of both sensor data streams. Based on our evaluations, we presume that the hybrid use of the CNN-LSTM architecture enabled learning from both magnetic and visual information effectively which led to optimal localization results. We run our training and testing sessions both on an Amazon EC2 p2.xlarge machine instance. The duration of the training session was 8 h and 37 min. The 6-DoF pose estimation per image pair and magnetic data took 35 ms Figs. 8 and 9.

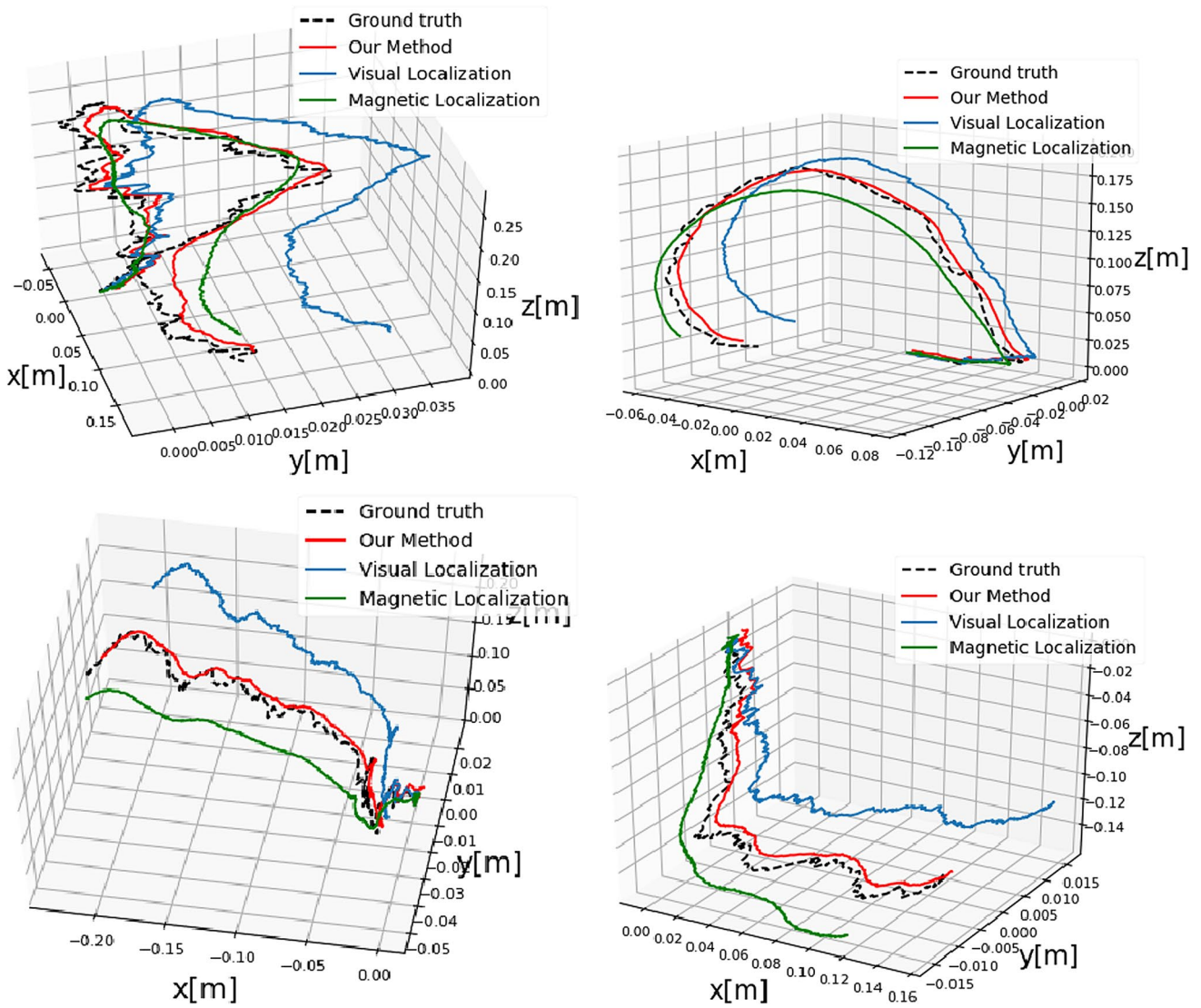
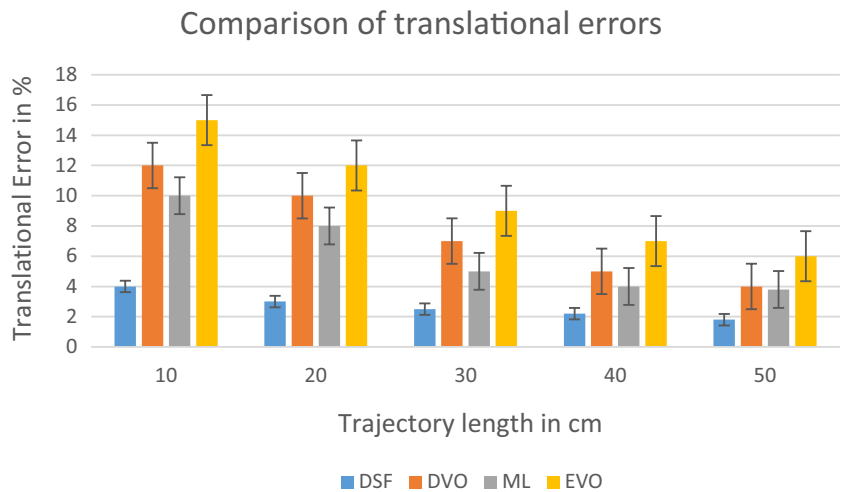
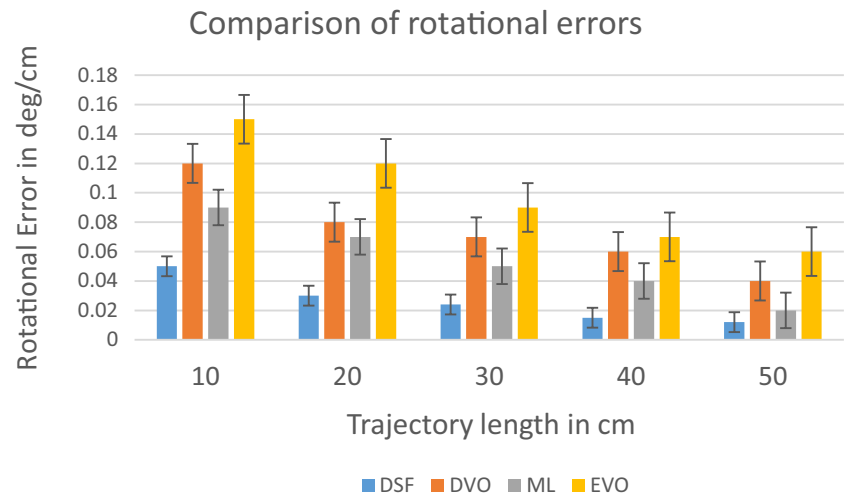


Fig. 7 Plotted trajectories

Fig. 8 Trajectory length versus translation error



**Fig. 9** Trajectory length versus rotation error



## 5 Conclusion

In this study, we presented, to the best of our knowledge, the first sensor fusion method based on deep learning for endoscopic capsule robots. The proposed CNN-RNN architecture based fusion approach is able to achieve simultaneous learning and sequential modeling of motion dynamics across frames and magnetic data streams. Since it is trained in an end-to-end manner, there is no need to carefully hand-tune the parameters of the system except the hyperparameters. In the future, we will incorporate controlled actuation into the scenario to investigate a more complete system, and additionally we will seek ways to make the system more robust against representational singularities in the rotation data.

**Acknowledgements** Open access funding provided by Max Planck Society.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Arshak, K., Adepoju, F.: Capsule tracking in the gi tract: a novel micro-controller based solution. In: *Sensors Applications Symposium, 2006. Proceedings of the 2006 IEEE*, pp. 186–191. IEEE (2006)
- Bao, G., Pahlavan, K., Mi, L.: Hybrid localization of microrobotic endoscopic capsule inside small intestine by data fusion of vision and RF sensors. *IEEE Sens. J.* **15**(5), 2669–2678 (2015)
- Beauchemin, S.S., Barron J.L.: The computation of optical flow. *ACM Comput Surv* **27**(3), 433–466 (1995)
- Cornelius, N., Kanade, T.: Adapting optical-flow to measure object motion in reflectance and X-ray image sequences. In: *Proceedings of ACM SIGGRAPH Computer Graphics*, vol. 18, pp. 24–25. ACM, New York, NY (1984). <https://doi.org/10.1145/988525.988537>

- Clark, R., Wang, S., Wen, H., Markham, A., Trigoni, N.: Vinet: visual-inertial odometry as a sequence-to-sequence learning problem. In: *AAAI*, pp. 3995–4001 (2017)
- Di Natali, C., Beccani, M., Simaan, N., Valdastri, P.: Jacobian-based iterative method for magnetic localization in robotic capsule endoscopy. *IEEE Trans. Robot.* **32**(2), 327–338 (2016)
- Fischer, D., Schreiber, R., Levi, D., Eliakim, R.: Capsule endoscopy: the localization system. *Gastroint. Endosc. Clin.* **14**(1), 25–31 (2004)
- Farneback, G.: Two-frame motion estimation based on polynomial expansion. In: Bigun, J., Gustavsson, T. (eds.) *Image Analysis. Lecture Notes in Computer Science*, vol. 2749, pp. 363–370. Springer, Heidelberg (2003)
- Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells W.M., Colchester A., Delp S. (eds.) *Medical Image Computing and Computer-Assisted Intervention—MICCAI’98. Lecture Notes in Computer Science*, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
- Geng, Y., Pahlavan, K.: Design, implementation, and fundamental limits of image and RF based wireless capsule endoscopy hybrid localization. *IEEE Trans. Mob. Comput.* **15**(8), 1951–1964 (2016)
- Gumprecht, J.D., Lueth, T.C., Khamesee, M.B.: Navigation of a robotic capsule endoscope with a novel ultrasound tracking system. *Microsyst. Technol.* **19**(9–10), 1415–1423 (2013)
- Hou, J., Zhu, Y., Zhang, L., Fu, Y., Zhao, F., Yang, L., Rong, G.: Design and implementation of a high resolution localization system for in-vivo capsule endoscopy. In: *Dependable, Autonomic and Secure Computing, 2009. DASC’09. Eighth IEEE International Conference on*, pp. 209–214. IEEE (2009)
- Popek, K.M., Mahoney, A.W., Abbott, J.J.: Localization method for a magnetic capsule endoscope propelled by a rotating magnetic dipole field. In: *2013 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5348–5353. IEEE (2013)
- Son, D., Dogan, M.D., Sitti, M.: Magnetically actuated soft capsule endoscope for fine-needle aspiration biopsy. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1132–1139. IEEE (2017)
- Son, D., Yim, S., Sitti, M.: A 5-d localization method for a magnetically manipulated untethered robot using a 2-d array of hall-effect sensors. *IEEE/ASME Trans. Mechatron.* **21**(2), 708–716 (2016)
- Than, T.D., Alici, G., Harvey, S., O’Keefe, G., Zhou, H., Li, W., Cook, T., Alam-Fotias, S.: An effective localization method for robotic endoscopic capsules using multiple positron emission markers. *IEEE Trans. Robot.* **30**(5), 1174–1186 (2014)
- Turan, M., Almalioglu, Y., Araujo, H., Konukoglu, E., Sitti, M.: A non-rigid map fusion-based rgb-depth slam method for endoscopic capsule robots. *arXiv preprint arXiv:1705.05444* (2017)



- Umay, I., Fidan, B.: Adaptive magnetic sensing based wireless capsule localization. In: 2016 10th International Symposium on Medical Information and Communication Technology (ISMICT), pp. 1–5. IEEE (2016)
- Umay, I., Fidan, B.: Adaptive wireless biomedical capsule tracking based on magnetic sensing. *Int. J. Wirel. Inf. Netw.* **24**(2), 189–199 (2017)
- Umay, I., Fidan, B., Barshan, B.: Localization and tracking of implantable biomedical sensors. *Sensors* **17**(3), 583 (2017)
- Wang, L., Hu, C., Tian, L., Li, M., Meng, M.Q.H.: A novel radio propagation radiation model for location of the capsule in gi tract. In: 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 2332–2337. IEEE (2009)
- Wang, Y., Fu, R., Ye, Y., Khan, U., Pahlavan, K.: Performance bounds for RF positioning of endoscopy camera capsules. In: 2011 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems (BioWireSS), pp. 71–74. IEEE (2011)
- Ye, Y.: Bounds on RF cooperative localization for video capsule endoscopy. Ph.D. thesis, Worcester Polytechnic Institute (2013)
- Yim, S., Sitti, M.: 3-d localization method for a magnetically actuated soft capsule endoscope and its applications. *IEEE Trans. Robot.* **29**(5), 1139–1151 (2013)



**Mehmet Turan** received his Diploma Degree from the Information technology and Electronics engineering department of RWTH Aachen, Germany in 2012. He was a research scientist at UCLA (University of California Los Angeles) between 2013–2014 and a research scientist at the Max Planck Institute for Intelligent Systems between 2014–present. He is currently enrolled as a PhD Student at the ETH Zurich, Switzerland. He is also affiliated with Max Planck-ETH Center for Learning Sys-

tems, the first joint research center of ETH Zurich and the Max Planck Society. His research interests include SLAM (simultaneous localization and mapping) techniques for milli-scale medical robots and deep learning techniques for medical robot localization and mapping. He received DAAD fellowship between years 2005–2011 and Max Planck Fellowship between 2014–present. He has also received MPI-ETH Center fellowship between 2016–present.

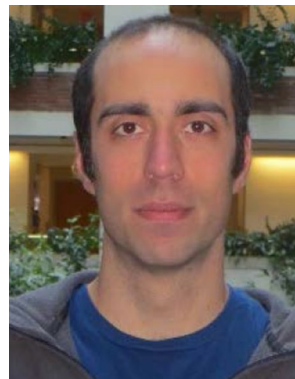


**Jahanzaib Shabbir** finished his BSc (Hons.) degree in electrical and electronics engineering from National University of Computer and Emerging Sciences, Lahore, Pakistan. Since then he conducted research under UNSW department of robotics, Sydney, Australia. Being in affiliation with the UNSW robotics department, he is currently a research intern at Max Planck Institute for Intelligent Systems in Germany and Department of Electrical and Computer Engineering of the University of Coimbra in Portugal.

His research interests include small to medium-scale robotics, image processing, machine learning, neural networking and mobile robot localization.



**Helder Araujo** is a Professor at the Department of Electrical and Computer Engineering of the University of Coimbra. His research interests include Computer Vision applied to Robotics, robot navigation and visual servoing. In the last few years he has been working on non-central camera models, including aspects related to pose estimation, and their applications. He has also developed work in Active Vision, and on control of Active Vision systems. Recently he has started work on the development of vision systems applied to medical endoscopy.



**Ender Konukoglu**, PhD, finished his PhD at INRIA Sophia Antipolis in 2009. From 2009 till 2012 he was a post-doctoral researcher at Microsoft Research Cambridge. From 2012 till 2016 he was a junior faculty at the Athinoula A. Martinos Center affiliated to Massachusetts General Hospital and Harvard Medical School. Since 2016 he is an Assistant Professor of Biomedical Image Computing at ETH Zurich. His is interested in developing computational tools and mathematical methods for analyzing

medical images with the aim to build decision support systems. He develops algorithms that can automatically extract quantitative image-based measurements, statistical methods that can perform population comparisons and biophysical models that can describe physiology and pathology.



**Dr. Metin Sitti** received the BSc and MSc degrees in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 1992 and 1994, respectively, and the PhD degree in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1999. He was a research scientist at UC Berkeley during 1999–2002. He has been a professor in the Department of Mechanical Engineering and Robotics Institute at Carnegie Mellon University, Pittsburgh, USA since 2002. He is currently

a director at the Max Planck Institute for Intelligent Systems in Stuttgart. His research interests include small-scale physical intelligence, mobile microrobotics, bio-inspired materials and miniature robots, soft robotics, and micro-/nanomanipulation. He is an IEEE Fellow. He received the SPIE Nanoengineering Pioneer Award in 2011 and NSF CAREER Award in 2005. He received many best paper, video and poster awards in major robotics and adhesion conferences. He is the editor-in-chief of the *Journal of Micro-Bio Robotics*.