Maximilian Neidhardt*, Nils Gessert, Tobias Gosau, Julia Kemmling, Susanne Feldhaus, Udo Schumacher and Alexander Schlaefer

Force estimation from 4D OCT data in a human tumor xenograft mouse model

Abstract: Minimally invasive robotic surgery offer benefits such as reduced physical trauma, faster recovery and lesser pain for the patient. For these procedures, visual and haptic feedback to the surgeon is crucial when operating surgical tools without line-of-sight with a robot. External force sensors are biased by friction at the tool shaft and thereby cannot estimate forces between tool tip and tissue. As an alternative, vision-based force estimation was proposed. Here, interaction forces are directly learned from deformation observed by an external imaging system. Recently, an approach based on optical coherence tomography and deep learning has shown promising results. However, most experiments are performed on ex-vivo tissue. In this work, we demonstrate that models trained on dead tissue do not perform well in *in vivo* data. We performed multiple experiments on a human tumor xenograft mouse model, both on *in vivo*, perfused tissue and dead tissue. We compared two deep learning models in different training scenarios. Training on perfused, *in vivo* data improved model performance by 24% for *in vivo* force estimation.

Keywords: 4D deep learning; force estimation; optical coherence tomography; xenograft mouse model.

Problem

Robot-assisted surgery for minimally invasive interventions has become popular since physical trauma can be reduced through motion compensation and scaling [1]. Over the last decade, visual feedback for these systems has considerably improved through image fusion of preoperative data and head-mounted displays. However, many systems still lack force feedback which can be beneficial during surgical tasks to avoid malfunction or damage to organs as well as in distinguishing tissues with respect to type and condition [2].

Force feedback can be enabled through electromechanical force sensors that are attached to the tool base outside of the surgical field. However, biased force measurement due to friction forces at the tool shaft is undesirable. Gessert et al. [3] proposed a miniature force sensor integrated in the tool tip. These force sensors can be problematic due to sterilization, biocompatibility and integration in microsurgical instruments with a working channel. Therefore, vision-based force estimation was proposed as contact free alternative.

Previous approaches in vision-based force estimation included deformable template matching methods [4] or mechanical deformation models [5]. These methods are mainly based on single shots of the sample. A different more recent approach is to include temporal information in force estimation models. This approach provides a more realistic scenario since *in vivo* tissue is always in motion due to pulsation, breathing and force interaction between surgical tools and tissue. This can be modeled efficiently with convolutional neural networks (CNNs) or recurrent neural networks (RNN) and was demonstrated with RGB(D)-images [6].

Recently, optical coherence tomography (OCT) was proposed as an imaging modality which provides a high spatial and temporal resolution for vision-based force estimation. Feasibility in mapping the OCT surface deformation to forces was demonstrated [7]. Also, learning force estimates from full OCT volumes with CNNs has been studied [8, 9] where promising results were achieved on ex-vivo data.

Predicting forces acting on ex-vivo tissue surrogates is always limited to a feasibility approach in the laboratory. These measurements do not reflect the complex physiological and biomechanical properties of *in vivo* tumor tissue which leads to a different elastic response than ex-vivo

---

Maximilian Neidhardt and Nils Gessert were contributed equally to this article.

---

*Corresponding author: Maximilian Neidhardt, Institute of Medical Technology and Intelligent Systems, Hamburg University of Technology, Hamburg, Germany,
E-mail: Maximilian.Neidhardt@tuhh.de
Nils Gessert and Alexander Schlaefer, Institute of Medical Technology and Intelligent Systems, Hamburg University of Technology, Hamburg, Germany
Tobias Gosau, Julia Kemmling, Susanne Feldhaus and Udo Schumacher, Department of Anatomy and Experimental Morphology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

tissue [10]. A static laboratory setup neglects properties such as the surrounding soft tissue of the tumor, pulsation, breathing, muscle twitches or speckle characteristics.

In this paper, we investigate vision-based force estimation in a human tumor xenograft mouse model. We employed a high-speed OCT imaging device to acquire OCT volumes at a high temporal rate. We employ two different 4D CNNs that process the high-dimensional 4D spatio-temporal OCT data for predicting forces acting on the tissue. We investigate how the deep learning models performed in an *in vivo* setting when being trained on either perfused or dead tissue data. Force estimation has been studied with different tissue types, however, there are no studies with tumor tissue so far.

Material and methods

Experimental setup

For data acquisition, the following experimental setup shown in Figure 1 was designed. A robot (H-820.D1, Physik Instrumente) for positioning the OCT field of view (FOV) relative to the tumor was employed. Note, the position of a volume does not change relative to a world reference system. Rather, the mouse which is fixed with tape to a heated bed (37°C) was driven. The heated bed can be easily mounted to the robot with a 3D printed adapter and prevents cooling of the narcotized mouse. A high-speed OCT imaging system with an A-scan rate of approximately 1.5 MHz was used. An A-scan was defined as a one-dimensional resolved depth signal. By moving an A-Scan along both lateral axes an OCT volume can be acquired. Each volume includes 32×32 scanlines in both lateral directions and 430 px along the depth dimension. The physical size of each volume is approximately $2.5 \times 2.5 \times 3.5$ mm in air and the temporal resolution was set to 100 Hz. As a surgical tool, a needle with a diameter of

2 mm attached to a force sensor (Nano 43, ATI) for ground-truth annotation was employed. The needle can be forwarded with a stepper motor along the needle axis which represents a typical surgical pushing task [6].

Xenograft mouse model

Experiments were conducted on pathogen-free balb/c severe combined immunodeficient (SCID) mice (Charles River, Wilmington, MA, USA). They were housed in individually ventilated cages and provided with sterile water and food ad libitum. For injection, 1×10^6 viable human HT29 colon cancer cells in 200 µm cell culture medium were injected subcutaneously into the right flank. Experiments were performed on mice if the primary tumors exceeded 1.2 cm or ulcerated the mouse skin. All experiments were approved by the local licensing authority (Freie und Hansestadt Hamburg, Behörde für Gesundheit und Verbraucherschutz, Amt für Verbraucherschutz, project N037/2019) and supervised by the institutional animal welfare officer.

Data acquisition and datasets

Each mouse was anesthetized and the skin above the subcutaneous tumor was carefully removed with a scalpel. Next, the mouse was fixed to the heated bed which can be easily mounted to the robot. For each experiment, the robot positions the FOV on the tumor (Figure 1, right). Next, the surface was detected by forwarding the needle along the needle shaft direction until a force of 0.02 N was registered. The tumor was palpated by moving and retracting the needle with a distance of 2 mm while continuously OCT volumes were acquired. For data variation, we performed palpation at five different velocities ranging from 0.3 mm/s to 0.7 mm/s. Tissue deformation on perfused tissue was compared to experiments performed on dead tissue. We refer to acquired data as Ante-Mortem (AM) datasets and Post-Mortem (PM) datasets, respectively. In total 10 AM and 10 PM datasets were acquired from five mice.
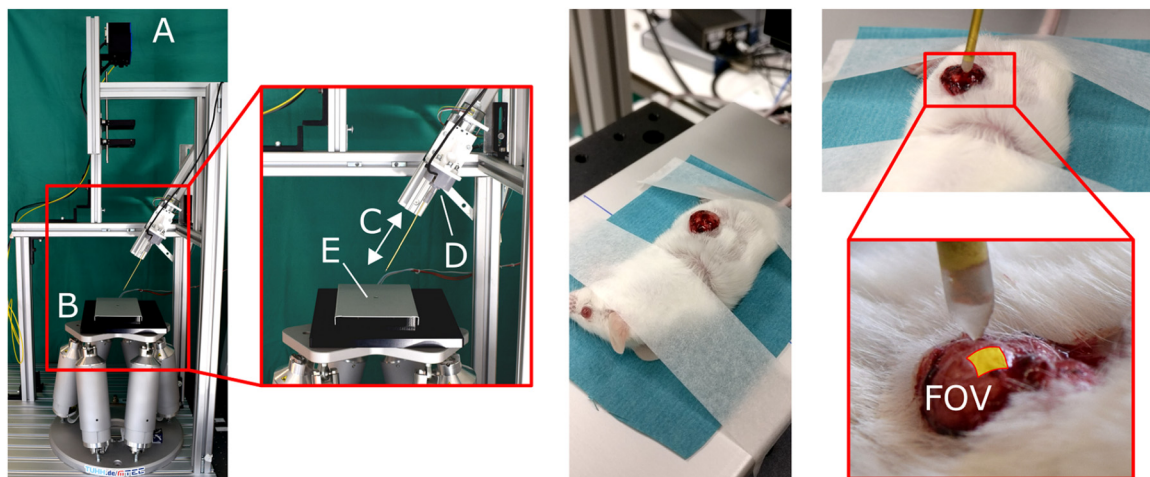

Figure 1: Experimental setup for data acquisition. *Left*: A OCT scan head (A) for volume acquisition was employed. A robot (B) drives a heated bed (E) on which the mouse is fixed. Tumor tissue is deformed with a needle (C) which is driven by a stepper motor (C) along the needle axis. The force sensor (D) is mounted between the stepper motor and needle. *Right*: The skin was carefully dissected and the anesthetized mouse was fixed with tape to the heated bed (left image). The right image indicates the approximate position of the field of view (FOV).

For deep learning model training and evaluation, a 10-fold cross-validation (CV) scheme was employed. Each fold contains approximately 6,000 vol and represents one experiment with the five different velocities and a different location on the tumor. Iteratively, we leave out one fold for validation, one fold for testing and train the deep learning model on all other folds. The final performance is expressed as the mean across all test folds.

Deep learning architectures

A sequence of 3D OCT volumes represents 4D data that needed to be processed. For this purpose, we employed 4D spatio-temporal CNNs that performed simultaneous spatial and temporal processing. Due to their high-dimensional nature, 4D CNNs are very parameter-intensive which might lead to a risk of overfitting. Therefore, a more efficient variant that uses factorized convolutions was employed [11]. Here, spatial and temporal processing were decomposed by using separate kernels. Thus, a fully 4D convolutional kernel of size $k \in \mathbb{R}^{t \times h \times d \times w \times c}$ was split into a spatial kernel $k_s \in \mathbb{R}^{1 \times h \times d \times w \times c}$ and a temporal kernel $k_t \in \mathbb{R}^{t \times 1 \times 1 \times 1 \times c}$ where t is the temporal kernel size, h, w, and d are spatial kernel sizes and c is the feature channel dimension. This decomposition led to a reduced number of trainable parameters, however, at the same time, representational power was reduced as the decomposed kernel can only represent separable kernels. To ensure proper gradient propagation throughout the network, both CNNs were built on the ResNet principle [12] where skip connections enable a reliable gradient flow throughout network training. The two architectures ResNet4D and fResNet4D are shown in Figure 2. In each CV iteration, we trained the models for 50 epochs using Adam optimizer and a starting learning rate of $l_r = 10^{-4}$ and batch size of $b = 64$. The learning rate was halved every 20 epochs. Hyperparameters such as the number of layers, learning rate and batch size were chosen based on validation performance.

Results

Table 1 shows results for tool tip force predictions from the two 4D CNN architectures. For all experiments, we report the mean absolute error (MAE) in mN, the mean absolute

Figure 2: The two 4D CNN architectures employed. The red boxes represent ResNet blocks with 4D convolutions.

Table 1: Results for two deep learning architectures. The training and evaluation sets include either AM or PM data sets referring to data acquisition in perfused and dead tissue respectively.

Architecture	Train.	Eval.	MAE (mN)	rMAE	ACC
ResNet4D	PM	PM	5.77 ± 1.83	0.62 ± 0.17	0.69
	AM	AM	4.88 ± 0.85	0.66 ± 0.17	0.67
	PM	AM	6.06 ± 1.84	0.85 ± 0.41	0.36
fResNet4D	PM	PM	5.97 ± 1.87	0.65 ± 0.19	0.64
	AM	AM	4.90 ± 0.88	0.68 ± 0.27	0.69
	PM	AM	6.66 ± 2.27	0.93 ± 0.49	0.34

error relative to the target's standard deviation (rMAE), and average correlation coefficient (ACC). Only including PM-data or AM-data for training and evaluation show a similar ACC of 0.69 and 0.67, respectively. If PM-data is used for training and evaluation is performed on AM-data the ACC is only 0.36. The results are similar for both architectures. Example predictions can be seen in Figure 3. Clearly, predictions are best if training is performed on AM datasets.

Discussion and conclusion

Our results show that 4D CNNs can predict forces in perfused tumor tissue with an error of 4.8 mN. Further, our results indicate that CNN's trained on dead tissue perform poorly when applied to perfused tissue. Note, all experiments were performed on tumors that were embedded in soft tissue. Hence, the change of physiological properties such as perfusion and breathing motion between the vital and dead tissue state strongly influences acquired data sets.

Summarized, we find that vision-based force estimation in *in vivo* data with deep learning models is heavily influenced by training data. In-vivo force estimation performs substantially better when models are trained on perfused, *in vivo* data.
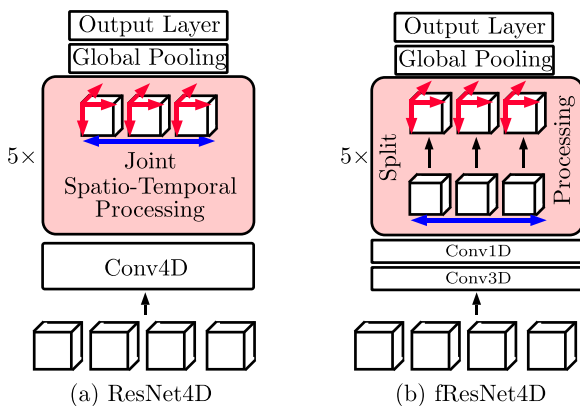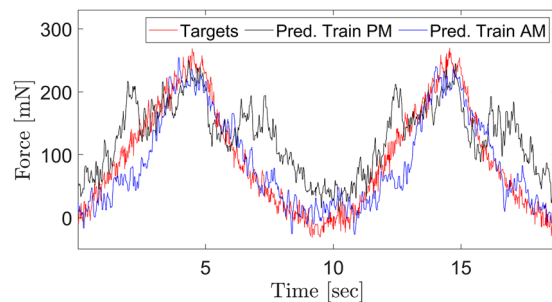

Figure 3: Example predictions of tool tip forces acting on perfused tumor tissue. The model was trained with Post-Mortem data (blue) or Ante-Mortem data (black).

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: Authors state no conflict of interest.

Informed consent: Informed consent has been obtained from all individuals included in this study.

Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the ethics committee of the Medical Center Hamburg-Eppendorf.

References

1. Song SE. Robotic interventions. In: Handbook of medical image computing and computer assisted intervention: Elsevier; 2020: 841–60 pp.
2. Diana M, Marescaux J. Robotic surgery. Br J Surg 2015;102:e15–28.
3. Gessert N, Priegnitz T, Saathoff T, Antoni ST, Meyer D, Hamann M, et al. Needle tip force estimation using an oct fiber and a fused convgru-cnn architecture. In: MICCAI: Springer; 2018: 222–9 pp.
4. Greminger MA, Nelson BJ. Vision-based force measurement. IEEE TPAMI 2004;26:290–8.
5. Mozaffari A, Behzadipour S, Kohani M. Identifying the tool-tissue force in robotic laparoscopic surgery using neuro-evolutionary fuzzy systems and a synchronous self-learning hyper level supervisor. Appl Soft Comput 2014;14:12–30.
6. Marban A, Srinivasan V, Samek W, Fernández J, Casals A. A recurrent convolutional neural network approach for sensorless force estimation in robotic surgery. Biomed Signal Process Contr 2019;50:134–50.
7. Otte C, Beringhoff J, Latus S, Antoni ST, Rajput O, Schlaefer A, et al. Towards force sensing based on instrument-tissue interaction. In: MFI 2016: IEEE; 2016:180–5 pp.
8. Gessert N, Beringhoff J, Otte C, Schlaefer A. Force estimation from oct volumes using 3d cnns. IJCARS 2018;13:1073–82.
9. Gessert N, Bengs M, Schlüter N, Schlaefer A. Deep learning with 4d spatio-temporal data representations for oct-based force estimation. Med Image Anal 2020;64:101730.
10. Carter FJ, Frank TG, Davies PJ, McLean D, Cuschieri A. Measurements and modelling of the compliance of human and porcine organs. Med Image Anal 2001;5:231–6.
11. Sun L, Jia K, Yeung DY, Shi BE. Human action recognition using factorized spatio-temporal convolutional networks. In: Proceedings of the IEEE ICCV; 2015:4597–605 pp.
12. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE CVPR; 2016: 770–8 pp.