









# Development and Analysis of a Pipeline for Cardiac Ultrasound Simulation for Deep Learning Segmentation Methods

M. Bauer<sup>1</sup> , C. Manini<sup>2,3</sup> , S. Klemmer<sup>4</sup> , T. Meyer<sup>4</sup> , M. Ivantsits<sup>2,3</sup> , L. Walczak<sup>2,3,5</sup> ,  
A. Hennemuth<sup>2,3,5</sup> , H. Tzschätzsch<sup>1</sup> 

<sup>1</sup>Institute of Medical Informatics, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, 10117, Germany

<sup>2</sup>Institute of Computer-Assisted Cardiovascular Medicine, Deutsches Herzzentrum der Charité (DHZC), Berlin, Germany

<sup>3</sup>Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, 10117, Germany

<sup>4</sup>Department of Radiology, Charité - Universitätsmedizin Berlin, 10117, Germany

<sup>5</sup>Fraunhofer MEVIS Berlin, 13353, Germany

## Abstract

Accurate and efficient segmentation of anatomical structures in medical images, e.g. ultrasound images, is crucial for diagnosis. Deep Learning methods can provide automatic reproducible segmentation, and simulation of medical images with their intrinsic ground truth could help to develop and tune these methods. We introduce a simulation pipeline for the example of mitral valve segmentation in Transesophageal Echocardiography (TEE) images including different valve opening states. As anatomical ground truth, we used a CT based patient phantom with simulated mitral valve closure. For each region within the phantom, scatter intensities and reflections between tissue boundaries were set, and ultrasound images were simulated with incorporation of attenuation and noise. To further improve realism of the simulated images a speckle reduction filter was used. The adjustments applied to improve realism were assessed by testing the segmentation performance (including Dice score) of a deep learning method trained on real TEE data. The initial Dice score for the simulation was 31 %. This value increased with image post-processing (37 %), exclusion of surrounding cardiac structures (45 %) and the combination of both (46 %). In comparison, the initial Dices score for real TEE was 72 %. On both simulated and real TEE images, the deep learning method performed better on fully closed valve states (42 % and 77 %) than on fully open valves (27 % and 66 %). This work introduced a novel pipeline for the realistic simulation of TEE images with different valve opening states. Our analysis demonstrated feasibility of the proposed pipeline and highlighted the importance of accurate and dynamic valve phantoms, comprehensive simulations and specific post-processing for the simulation of realistic TEE images. In the future, with further improvements of the simulation, we will evaluate the pipeline for the training of Deep Learning methods on simulated data for the application on real data.

## CCS Concepts

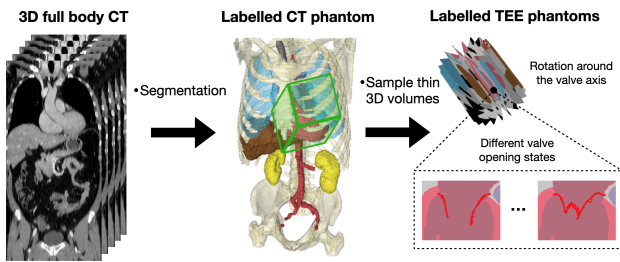
• **Computing methodologies** → Modeling and simulation; • **Human-centered computing** → Visualization;

## 1. Introduction

The segmentation of medical images to identify and classify anatomical structures enables a quantitative assessment of these structures. Challenges include anatomical variability and fluctuations in image quality [MHL\*24]. To overcome these challenges, Deep Learning (DL) methods have been used in the past to improve the efficiency and, in certain cases, the accuracy of medical image analysis [WLW\*21]. A wide range of DL methods are available for segmenting medical images, tuned for different medical imaging modalities, procedures and the preferred anatomical structure. These segmentation models are created by training with image data and the corresponding ground truth labels of the anatomical structure. The simulation of training data for medical image data is challenging [RIN\*24], because:

1. Available collections of medical image data are biased towards specific diseases and demographics. However, representative data is required in order for DL models to learn accurately segmenting all patient datasets.
2. No real ground truth data is available since the real underlying anatomical structures depicted in the image data are not known, which makes it difficult to do accurate benchmarking of the models.

To address the first problem, data augmentation have led to higher accuracy when applying the models to test data sets [GSLM23]. One idea to solve both problems is the simulation of medical images. Simulations offer the possibility to mimic different clinical scenarios. In addition, they can be created on the basis of constructed phantoms representing a real ground truth and allow accurate benchmarking as successfully demonstrated by Behboodi



**Figure 1:** Generation of labelled Transesophageal Echocardiography (TEE) phantoms with different valve opening states. The 3D CT was segmented to obtain a labelled phantom. A bounding box (green cube) was positioned covering the TEE acquisition region. Thin 3D volumes were extracted with the rotation around the valve axis. Different mitral valve opening states were incorporated into the TEE phantoms.

et al. [BR19]. Employing causality-inspired data augmentation can enhance model robustness against unseen domains [OCL\*22]. The application example that we examined in this work is the segmentation of the mitral valve in Transesophageal Echocardiography (TEE). Due to the small size of the valve, the complex subvalvular apparatus, its central location among numerous tissues as well as its movement, precise segmentation of the valve can be quite difficult and requires a significant amount of expertise and effort. To analyze the potential of simulations to train DL methods we aimed to

1. Introduce a simulation pipeline for TEE ultrasound images from CT-based 3D phantoms with different valve opening states.
2. Utilize a pre-trained DL method on the simulations to identify the relative importance of the used pipeline parts for realistic simulation.

## 2. Related Work

### 2.1. CT based 3D patient phantoms with different valve opening states

Phantoms of the human body could serve as a helpful basis for simulating medical images and providing real ground truth data [WGK23]. Manini et al. [MNA\*23] presented a computational phantom (Figure 1) created with anatomical structures extracted from patient 3D full body CT in combination with a mitral valve closure simulation [WGT\*21]. However, to our knowledge, no simulation pipeline for cardiac ultrasound images based on 3D CT phantoms is available. To establish this pipeline, we used this labelled phantom model as the anatomical structures ground truth. We extracted the 3D volume through which the TEE passes (190 mm × 143 mm × 7 mm) and sampled thin 3D volumes oriented with different rotation angles of the TEE. For the simulation, 3 phantoms extracted from different patients were used in which we considered 17 different rotation angles at 10-degree intervals. In each angle, we had slices for 30 valve opening states from fully open to fully closed. This resulted in 1530 simulated TEE images with known ground truth.

### 2.2. Ultrasound simulation software

For the simulation of 2D ultrasound images from phantoms, Jensen developed the software solution Field II (current Version 3.30) [Jen]. 3D positions and intensities of the point scatterers are used as input and the program utilizes the concept of spatial impulse response to simulate the RF data, which can be then further processed to a B-mode image.

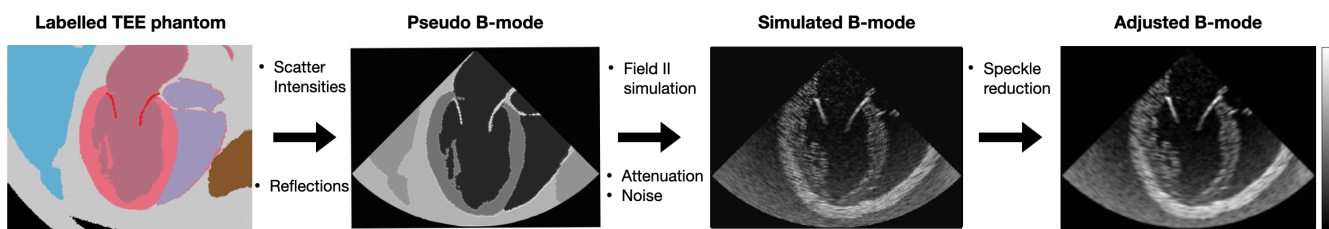
### 2.3. DL segmentation method

The U-Net architecture proposed by Ronneberger et al. proved to achieve high accuracy segmentations with limited datasets [RFB15]. Numerous U-Net models now exist for image analysis across various applications [SPED21]. In this work, the nnU-Net by Isensee et al. was used. The nnU-Net is a versatile method for various medical images and provides automatically configuring pre-processing, network architecture, training, and post-processing for a given data set [IJK\*21].

## 3. Method

### 3.1. Ultrasound simulation pipeline

The pipeline for the simulation of TEE images based on 3D phantoms consists of pre-processing, simulation and adjustments, (Figure 2). For pre-processing, the individual phantom slices from each rotation angle and valve opening state were used as a map to assign the intensities to the labelled tissues. The intensity values used were chosen to visually cover the range of real TEE images taken with different devices and different image quality. In addition, to mimic the ultrasound reflections at the tissue boundaries, we calculated the axial gradient of the scatter intensity map and added this to the scatter intensity map (Pseudo B-mode). A second axial gradient of the tissue outside the myocardium was empirically dilated by 8 mm and multiplied by 400 to match the reference TEE images reflections. For the simulation part, point scatters ( $263 / \text{cm}^3$ ) were randomly positioned and intensities were assigned according to the pseudo B-mode images. To match the visual appearance of real valves 5,000 additional scatters were positioned in the mitral valve. The parameters were set to a common TEE ultrasound probe (phased array transducer with center frequency 7 MHz, sampling frequency 40 MHz, focus 50 mm, 64 elements with height 5 mm and width 0.11 mm, sector with 143 mm depth and 190 mm width) and the corresponding RF data was simulated with Field II. Afterwards, the ultrasound attenuation was mimicked by an empirically derived exponential axial decay (attenuation coefficient of  $8.05 / \text{m}$ ). Normal distributed noise with an empirically derived standard deviation of  $1 / 25,000$  of the average B-mode intensity near the transducer was added to the RF data. The B-mode was then logarithmically compressed and compensated for mimicked attenuation. In total, 1530 2D ultrasound images were simulated on a 2.7 GHz Dual-Core Intel Core i5 CPU with a computation time of around 3 minutes per image. To analyze the impact of image post-processing, we applied further adjustments on the simulated B-mode. To mimic image post-processing (speckle reduction) as used in common TEE, the speckle reduction filter from the Image Despeckle Filtering Toolbox (Version 1.0) in MATLAB R2023a



**Figure 2:** Pipeline for the simulation of a Transesophageal Echocardiography (TEE) ultrasound B-mode. Based on the labelled TEE phantom, the scatter intensities were set for each anatomical region and reflections at tissue boundaries were incorporated, resulting in a pseudo B-mode. The simulated B-mode image was generated using Field II, incorporating ultrasound attenuation and noise. For an adjusted B-mode, image post-processing (speckle reduction filter) was applied.

with a  $3 \times 3$  neighborhood and the iteration parameter 2 was added to the simulated B-mode image [LTPK14].

### 3.2. Training and running inferences with the nnU-Net

The real TEE images were collected from 21 patients (11 type I mitral insufficiency with dilated annulus, 9 type II with prolapse, 1 volunteer) by a GE Vingmed Ultrasound Vivid E9 system (frame rate 15-83 Hz, in-plane resolution 0.47-1.43 mm) using the same 17 rotation angles as in the phantoms. Each angle had a time series of 3-6 frames capturing the valve closing process leading to a total of 1904 ultrasound images with ground truth segmented manually by cardiologists. The nnU-Net first extracted a dataset fingerprint including dataset-specific properties such as image sizes (250 mm  $\times$  250 mm), pixel spacing (0.4 mm  $\times$  0.4 mm) and intensity information. After selecting the 2D configuration, the training was started with the pre-processed data and own splits. To prevent bias during the training process, the images were not split individually, but split by volume [TEHH22] with a random 15 - 3 - 3 split of the valves for training, validation and testing respectively. During the training phase, the nnU-Net was used with different data augmentations including rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring [IJK\*21]. The trained model was then applied to the desired test data set and inference was run on the simulated B-mode images to analyze the performance. To analyze the impact of image post-processing and to focus on only the valve region, we trained the nnU-Net again on the cropped real TEE data and applied it on cropped B-mode images. Cropping was realized by placing a bounding box (93 mm  $\times$  81 mm) randomly around the mitral valve matching the next larger patch size of the nnU-Net. Training and inference were run on a NVIDIA A40 GPU.

### 3.3. Statistics

Dice score, average symmetric surface distance (ASD) and Hausdorff distance (HD) [YV] were used to analyze the performance of segmentation predictions made by the nnU-Net including calculating average values and standard deviation. To compare the Dice score distribution for fully open and closed valves a Wilcoxon signed-rank test was carried out with a P-value  $< 0.05$  considered as significant. Statistic evaluation was performed in python 3.10 with

the nnU-Net evaluation module as well as the seg\_metrics (1.2.6) and scipy library (1.13.1).

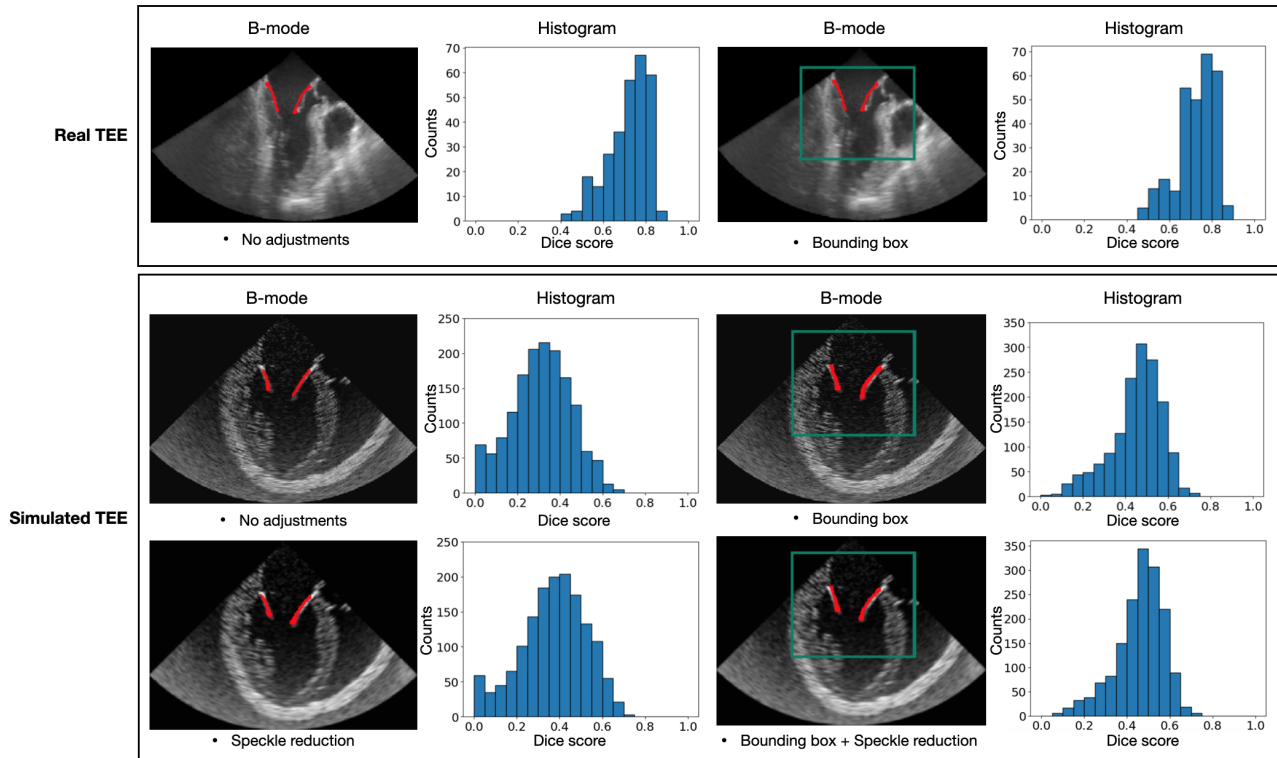
## 4. Results

### 4.1. Qualitative results

A qualitative comparison of the real TEE images and the simulated images in the general approach revealed that the simulated images had a coarser-grained speckle pattern, a darker blood pool and the tissues outside of the heart appear with a higher contrast. The simulated images had no image artefacts such as shadows, that could be seen in some real images. After adding the speckle reduction filter in the simulated images, a less grainy speckle pattern could be recognised.

### 4.2. Quantitative results

For each metric and adjustment, the average values and standard deviation are shown in Table 1 and histograms for the respective Dice score distributions are shown in Figure 3. These include the 289 real TEE images and the 1530 simulated TEE images before and after adjustments (bounding box for real TEE as well as speckle reduction and bounding box for simulated TEE). Before adjustments, the trained DL model achieved on average about half the values for Dice score, seven times higher values for ASD and six times higher values for HD on the simulated TEE images compared to the real TEE images. Moreover, the model was unable to predict any segmentation at all for 20 of the simulated images (15 after speckle reduction) and accordingly neither the ASD nor HD could be determined. For the simulated TEE images the average relative improvement of all metrics was 22 % for only speckle reduction, 56 % for only bounding box and 57 % for both combined. For the real TEE images the average relative improvements of all metrics was 4 % for using the bounding box. Furthermore, after using the bounding box, the model was able to predict a segmentation in all images. After all adjustments, the average relative improvement of all metrics for fully closed valves compared to the fully open valves was 40 % for simulated TEE images and 27 % for real TEE images (Table 1 and Figure 4) with significant P-values  $< 0.05$  for each metric.



**Figure 3:** Comparison of the mitral valve segmentation performance between real and simulated transesophageal echocardiography (TEE) images before and after different adjustments (bounding box for real TEE as well as speckle reduction and bounding box for simulated TEE). Valve segmentation in red and bounding box in green.

**Table 1:** Comparison of the valve segmentation performance between real and simulated transesophageal echocardiography (TEE) images. Comparison metrics (mean values  $\pm$  standard deviation) are differentiated before and after adjustments (bounding box for real TEE as well as speckle reduction and bounding box for simulated TEE) as well as for the fully open and closed mitral valve state. Average symmetric surface distance (ASD); Hausdorff-distance (HD).

Metric unit	Real TEE			Simulated TEE			
	Dice score in %	ASD in mm	HD in mm	Dice score in %	ASD in mm	HD in mm	
No adjustments	72 $\pm$ 10	0.19 $\pm$ 0.07	1.0 $\pm$ 0.6	31 $\pm$ 14	1.53 $\pm$ 2.20*	6.4 $\pm$ 4.8*	
Speckle reduction				37 $\pm$ 15	1.14 $\pm$ 1.53*	5.1 $\pm$ 4.1*	
Bounding box	73 $\pm$ 9	0.19 $\pm$ 0.08	1.1 $\pm$ 0.7	45 $\pm$ 12	0.52 $\pm$ 0.24	2.8 $\pm$ 1.5	
Speckle reduction + Bounding box				46 $\pm$ 11	0.51 $\pm$ 0.22	2.8 $\pm$ 1.5	
After all adjustments	Fully open valve	66 $\pm$ 9**	0.23 $\pm$ 0.16**	1.5 $\pm$ 0.8**	27 $\pm$ 14**	0.70 $\pm$ 0.40**	3.4 $\pm$ 2.0**
	Fully closed valve	77 $\pm$ 6**	0.16 $\pm$ 0.04**	1.0 $\pm$ 0.3**	42 $\pm$ 12**	0.46 $\pm$ 0.10**	2.4 $\pm$ 0.8**

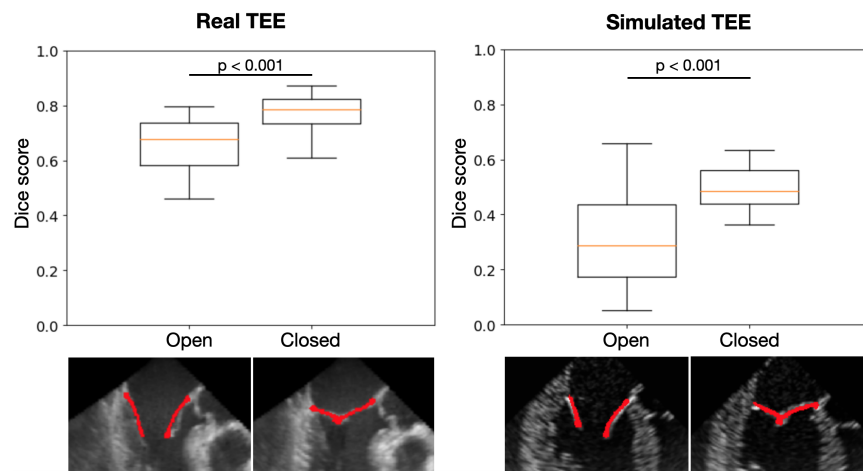
\*For certain cases no segmentation could be predicted by the model therefore the values had to be omitted.

\*\* A Wilcoxon signed-rank test was carried out for fully open and fully closed valves leading to significant P-values  $<$  0.05 .

## 5. Discussion

In this work, a first approach to a novel pipeline for the simulation of 2D TEE images from 3D phantoms with different valve opening states has been proposed. The nnU-Net, trained on real TEE images, provided surrogate information on the realism of the simulated images as well as on the influencing factors. The results have shown that the simulated images were able to achieve a comparable appearance as real images, but still had a lower overall per-

formance. This discrepancy may be attributed to the limited quantity of training data and the fact that the simulations were adjusted to represent a wide range of TEE images apart from the training data. Still, post-processing and the exclusion of tissue outside the mitral valve were identified as major factors to improve the performance of the DL method on simulated images. This suggests that even structures far from the valve have a significant impact on segmentation performance which could be handled by using an



**Figure 4:** Comparison of the valve segmentation performance between the real Transesophageal Echocardiography (TEE) and adjusted simulated B-mode images (bounding box for real TEE as well as speckle reduction and bounding box for simulated TEE) for the fully open and closed mitral valve states. Valve segmentation in red.

object identifier prior to the DL model. Additionally, it could be observed that the valve opening state influences the performance of the DL segmentation suggesting, that this is a problem of the DL model rather than due to differences in image quality due to valve state. Further improvements for realism could be enhancements to the phantoms such as incorporating myocardium movement, fine-tuning simulation parameters, implementing additional ultrasound properties like image artefacts, and employing more sophisticated image post-processing techniques. Also, more datasets are needed for a more representative analysis. With these improvements and the availability of numerous and more diverse sets of phantoms, the ultimate goal, to train the DL method directly on the simulations with known ground truth, could be achieved.

## 6. Acknowledgements

Support of the German Research Foundation (SFB1340 Matrix in Vision and Research Training Group BIOQIC) is gratefully acknowledged.

## References

- [BR19] BEHBOODI B., RIVAZ H.: Ultrasound segmentation using u-net: learning from simulated data and testing on real data. *EMBC 2019* (2019). 2
- [GSLM23] GARCEA F., SERRA A., LAMBERTI F., MORRA L.: Data augmentation for medical imaging: A systematic literature review. *CBM 152* (2023), 106391. 1
- [IJK\*21] ISENSEE F., JAEGER P. F., KOHL S. A. A., PETERSEN J., MAIER-HEIN K. H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods 18* (2021), 203–211. 2, 3
- [Jen] JENSEN J.: Field ii software for matlab. URL: [https://field-ii.dk/?downloading\\_2021.html](https://field-ii.dk/?downloading_2021.html). Last access: 2024-06-05. 2
- [LTPK14] LOIZOU C., THEOFANOUS C., PANTZIARIS M., KASPARIS T.: Despeckle filtering software toolbox for ultrasound imaging of the common carotid artery. *CMPB 114* (2014), 109–124. 3
- [MHL\*24] MA J., HE Y., LI F., HAN L., YOU C., WANG B.: Segment anything in medical images. *Nat Commun 15* (2024), 654. 1
- [MNA\*23] MANINI C., NEMCHYNA O., AKANSEL S., WALCZAK L., TAUTZ L., KOLBITSCH C., FALK V., SÜNDERMANN S., KÜHNE T., SCHULZ-MENGER J., HENNEMUTH A.: A simulation-based phantom model for generating synthetic mitral valve image data—application to mri acquisition planning. *IJCARS 19* (2023), 553–569. 2
- [OCL\*22] OUYANG C., CHEN C., LI S., LI Z., QIN C., BAI W., RUECKERT D.: Causality-inspired single-source domain generalization for medical image segmentation. *IEEE Transactions on Medical Imaging 42* (2022), 11095 – 1106. 2
- [RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. *MICCAI 2015* (2015), 234–241. 2
- [RIN\*24] RAYED M. E., ISLAM S. M. S., NIHA S. I., JIM J. R., KABIR M. M., MRIDHA M. F.: Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Informatics in Medicine Unlocked 47* (2024), 101504. 1
- [SPED21] SIDDIQUE N., PAHEDING S., ELKIN C., DEVABHAKTUNI V.: U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access 9* (2021), 82031–82057. 2
- [TEHH22] TAMPU I., EKLUND A., HAJ-HOSSEINI N.: Inflation of test accuracy due to data leakage in deep learning-based classification of oct images. *Sci Data 9* (2022), 580. 3
- [WKG23] WEGNER M., GARGIONI E., KRAUSE D.: Classification of phantoms for medical imaging. *Procedia CIRP 119* (2023), 1140–1145. The 33rd CIRP Design Conference. 2
- [WGT\*21] WALCZAK L., GEORGH J., TAUTZ L., NEUGEBAUER M., WAMALA I., SÜNDERMANN S., FALK V., HENNEMUTH A.: Using position-based dynamics for simulating mitral valve closure and repair procedures. *Computer Graphics Forum 41* (2021), 270–287. 2
- [WLW\*21] WANG S., LI C., WANG R., ELIU Z., WANG M., TAN H., WU Y., LIU X., SUN H., YANG R., LIU X., CHEN J., ZHOU H., BEN AYED I., ZHENG H.: Annotation-efficient deep learning for automatic medical image segmentation. *Nat Commun 12* (2021), 5915. 1
- [YV] YEGHIAZARYAN V., VOICULESCU I.: An overview of current evaluation methods used in medical image segmentation. URL: <https://www.cs.ox.ac.uk/files/7732/CS-RR-15-08.pdf>. Last access: 2024-06-05. 3