Multicenter Aortic Vessel Tree Extraction using Deep Learning

Bernhard Scharinger[a,b,c], Antonio Pepe[a,b], Yuan Jin[a,b,d], Christina Gsaxner[a,b], Jianning Li[a,b,c], and Jan Egger[a,b,c,*]

[a]Graz University of Technology, Institute for Computer Graphics and Vision, Inffeldgasse 16c/II, A-8010 Graz, Austria
[b]Computer Algorithms for Medicine Laboratory, Graz, Austria
[c]Institute for AI in Medicine (IKIM), University Medicine Essen (A oR), Girardetstrasse 2, 45131, Essen, Germany
[d]Research Center for Connected Healthcare Big Data, Zhejiang Lab, Hangzhou, 311121 Zhejiang, China

ABSTRACT

The aorta is the largest vessel of the human body and its pathological degenerations, such as dissections and aneurysms, can be life threatening. An automatic and fast segmentation of the aorta can therefore be a helpful tool to quickly identify an abnormal anatomy. The segmentation of the aortic vessel tree (AVT) typically requires extensive manual labor, but, in recent years, progress in deep learning techniques made the automation of this process viable. For this purpose, we tested different deep learning networks to segment the aortic vessel tree from computed tomography angiography (CTA) scans with a deep neural network consisting of an encoder-decoder architecture with skip connections and an optional self-attention block. The networks were trained on a dataset of 56 CTA scans from three different sources and resulted in Dice-score similarities between $0.043 - 0.897$. Generally, the classical U-Nets performed better than the ones containing a self-attention block, indicating that they might diminish performance for AVT segmentation. The quality of the resulting segmentations was highly dependent on the CTA image quality, especially on the contrast between the aorta and the surrounding tissues. However, the trained deep neural network can segment CTA scans well with limited computational resources and training data.

Keywords: Convolutional Neural Network, CNN, Patch-Based, Aorta, Vessel Tree, CTA, Deep Learning

1. DESCRIPTION OF PURPOSE

Over the last years, new medical imaging modalities were developed to help detect, diagnose and monitor different illnesses.[1] Historically, the interpretation and analysis of the images obtained with these methods was conducted by trained radiologists or physicians, but, in recent years, with the development of more and more sophisticated computational image analysis methods and machine learning techniques, computer-aided medical image analysis finds its way into the clinical practice.[2–4] To highlight anatomical structures or pathological changes in images and therefore make the diagnostic process easier, more accurate and more efficient, segmentation is the method to go with.[5] In the last 10 years, the number of papers describing automatic heart and vessel segmentation in medical images has greatly increased, showing that with the improvement over classical methods and the significant increase in computational capabilities, the popularity of automated cardiovascular image analysis has also increased.[6–9]

Pathologies of the cardiovascular system, like dissections and aneurysms, can be life-threatening and require prompt attention.[10,11] Therefore automatic segmentation can be a helpful tool to promptly identify an abnormal anatomy. This process usually requires a significant amount of manual labour with traditional segmentation methods.[2] To simplify this process, we developed a 3D deep neural network that consists of an encoder-decoder

*Corresponding author: Jan Egger, E-mail: egger@tugraz.at

network together with a self-attention block and evaluated the role of the attention block. A collection of 56 computational tomography angiography (CTA) scans,[12] preprocessed with windowing, re-sampling, cropping and normalization was used to train, validate and test the networks.

2. METHODS

2.1 Implementation

The overall dataset consists of 56 CTA-scans from three different public sources with ground truth segmentations and also includes aortas with abdominal aortic aneurysms and aortic dissections.[12] Since the sub-datasets have differing voxel resolution and slice thicknesses, we first re-sampled the images to a voxel size of $1 \times 1 \times 1$ mm using the module `Resample Scalar Volume` of 3D Slicer.[13] For the Images B-spline-interpolation was used and for the ground truth nearest neighbor interpolation. Next, the images were cropped to a size of $256 \times 256 \times z$ around the center of the image and subsequently windowing was applied to remove unnecessary information from the image. Scans from the Dongyang dataset were windowed between 100 and 700 Hounsfield Units (HU), images from the Rider dataset were windowed between 1100 and 1350 HU and images from the KiTS dataset were windowed between 1100 and 2000 HU. After that, the images were normalized between the interval $[0, 1]$ and patches of the size $256 \times 256 \times 128$ were extracted.

The suggested network models are based on the U-Net architecture[14] and incorporate the Channel and Spatial Attention Module (CSAM) proposed by Mou et al.[15,16] that is inspired by the Dial Attention Network (DANet).[17] They consist of an encoder and a decoder part, consisting of 3 convolution blocks each. A 3D convolution with a kernel size of $3 \times 3 \times 3$ is performed on the input image, followed by batch normalization[18] and ReLu-activation. This sequence is performed twice, before the image is down-sampled via max-pooling with size 2 and stride 2. This is repeated for the second and third encoder block. As bottom block, either another convolution block was used, or a spatial and a channel attention module (CSAM). The decoder path begins with a 3D-upsampling of size 2, followed by a concatenation with the output of the encoder-convolution-block of the same depth. Next, a convolution block, identical to the encoder convolution block, is applied. This is repeated twice more and finally the output segmentation map is obtained via a 3D-convolution with kernel size 1 and 1 filter followed by a sigmoid activation layer.

The first networks tested were the U-Net16 with and without the CSAM, where all convolutions in the encoder and decoder path have 16 filters. Secondly, we tested the U-Net176, named for its 176 filters, with and without CSAM. Here, the number of filters increases with the network depth, more precisely 8, 16, 32 and 64 filters. Finally we tested a U-Net with 16, 32, 64 and 128 filters, called U-Net352, for 352 filters. This U-Net model was the biggest classical U-Net tested, and it was not possible to combine it with a CSAM due to video memory limitations. As an example, the U-Net16 with the Channel and Spatial Attention Module (CSAM) is visualized in figure 1.

Six images of the dataset were set aside as test data, and the remaining data, 250 patches, was split into 30% validation data (75 patches) and 70% training data (175 patches). The network was trained on a PC with a Nvidia GeForce RTX 2070 with 8GB of video memory, using the Adamax-optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 0.01$)[19] and a modified Dice-loss,[20] where the loss is multiplied by a large negative number to improve the learning process (see equation 1). The initial learning rate was set to 0.001 and was reduced by a factor of 0.1 after 10 epochs without validation loss improvement. The training was stopped early after 20 epochs without validation loss improvement.

$$\mathcal{L}_{Dice} = 2 \cdot \frac{|X \cap Y|}{|X| + |Y|} \cdot -100 \cdot 10^3 \tag{1}$$

For testing, the images were preprocessed identically to the training images. The patches were fed into the model for inference and thresholded afterwards. The test images performed differently well and therefore required different threshold values to obtain an optimal dice score. The thresholding values were found empirically and can be found in Table 1. Next, the patches were assembled again and morphological closing was performed to eliminate any artifacts at the borders of the patches and connect the patches together. Finally the connected components were extracted using the python package `connected-components-3d` in Version 3.10.1. Since the

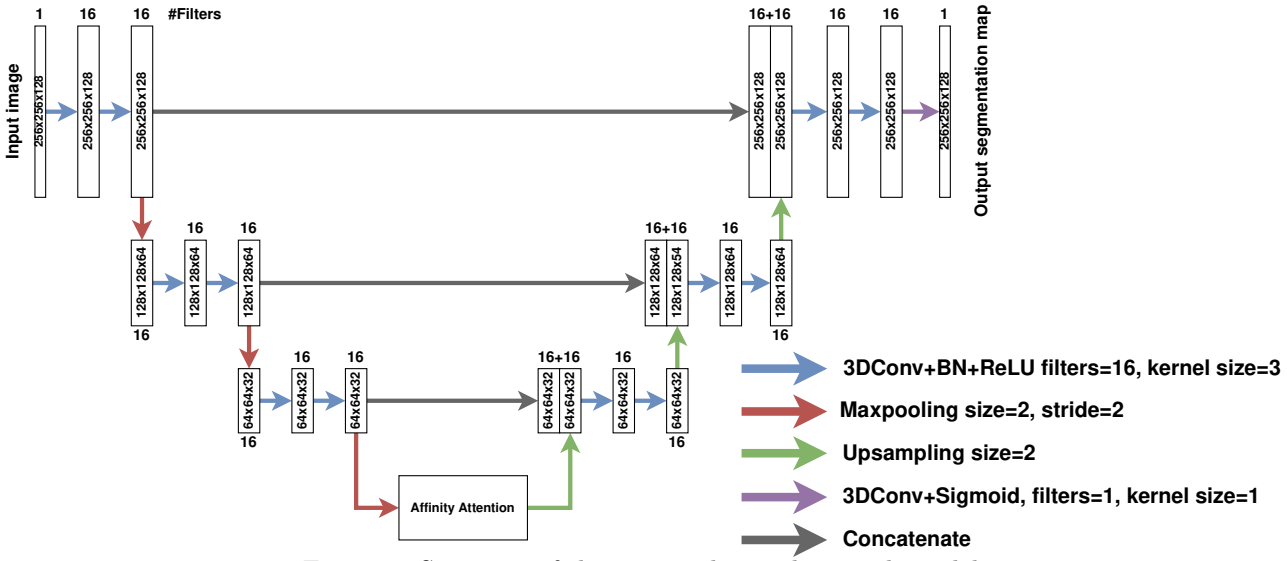


Figure 1: Structure of the proposed neural network model.

Background is the largest connected component, the second largest connected component was extracted and saved in an `.nrrd`-file. The test segmentations were evaluated using 3D slicer[13] with the *Segment Comparison* function of the Radiotherapy module.[21]

3. RESULTS AND DISCUSSION

The training of the networks was done in 38 to 66 epochs in a time of 3-8 hours. Only the weights of the model at minimal validation loss were saved and used for predictions. The proposed methods yielded the results described in table **??** for the test dataset.

The images D1, R3 and K1 achieved relatively high dice scores throughout all segmentation methods, while R1 performed well only in the standard U-Nets with 176 and 352 filters. Image K3 was segmented well by the U-Net176 with and without CSAM as well as U-Net352. Image D3 could not be segmented well by any method, the highest achieved dice-score was 0.541 by the U-Net16+CSAM. The average Hausdorff-distance ranges from $1.62 - 68.99$ mm.
Figure 2 shows the ground truth, the segmentation and the difference between ground truth and segmentation for the images R3 and D3 as exemplary segmentation results. Segmentation R3 includes most of the aortic arch, the thoracic and the abdominal aorta. Additionally the iliac arteries are included at least partially, as well as the left carotid, the brachiocepahlic and the left subclavian artery. Additionally, the major branches of the thoracic and abdominal aorta can be seen partially, although some significant offshoots are not segmented. The segmentation of image D3 contains the thoracic and abdominal aorta as well as parts of the aortic arch. The iliac arteries and the aortic bifurcation are not segmented. D3 shows a high amount of false positives around the heart and the abdominal aorta is not segmented completely.
In the Rider dataset, multiple samples were included that contained abdominal aortic aneurysms and aortic dissections. One image with an aortic dissection was used for testing, but the networks were not able to segment it correctly. The true and false lumen were connected, and only on the outside of the aorta a small rift could be seen. This could stem from a lack of training samples with aortic defects, thus a bigger dataset containing more samples with aortic dissections and aortic aneurysms could improve the performance of the model regarding aortic pathologies.

In general, the networks without the attention module performed better than the ones where it was included. In their paper,[15] Mou et al. used their mechanism for retinal and brain blood vessels, where their CS^2-Net performed better than a standard U-Net. This might suggest, that this self-attention mechanism performs better

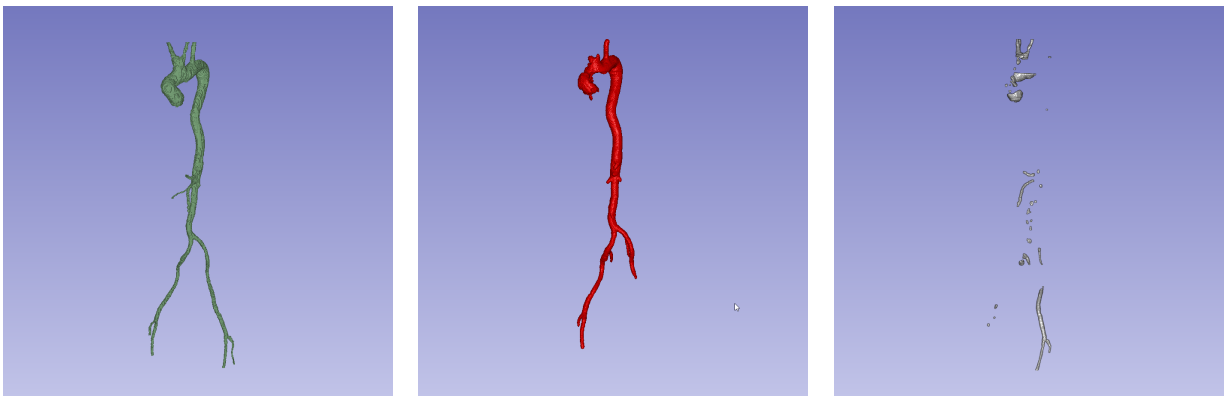
		Dongyang		Rider		KiTS	
U-Net-16	**Image**	D1	D3	R1	R3	K1	K3
	Threshold	0.05	0.05	0.05	0.05	0.05	0.01
	Dice-Score	**0.889**	0.189	0.607	0.794	0.840	0.589
	True negatives [%]	99.37	98.83	95.74	99.17	99.07	98.84
	True positives [%]	0.50	0.12	1.86	0.55	0.68	0.67
	False negatives [%]	0.06	0.003	2.40	0.24	0.18	0.31
	False positives [%]	0.07	1.04	0.00	0.04	0.08	0.26
	Avg. Hausdorff-Distance [mm]	10.06	55.25	43.04	6.59	4.47	20.05
U-Net-16 + CSAM	**Image**	D1	D3	R1	R3	K1	K3
	Threshold	0.02	0.005	0.2	0.1	0.05	0.02
	Dice-Score	0.856	**0.541**	0.613	0.830	0.828	0.485
	True negatives [%]	99.30	97.71	99.05	99.31	99.12	99.11
	True positives [%]	0.52	0.85	0.42	0.49	0.62	0.28
	False negatives [%]	0.12	1.13	0.07	0.11	0.12	0.04
	False positives [%]	0.05	0.31	0.45	0.10	0.14	0.56
	Avg. Hausdorff-Distance [mm]	5.68	17.3	9.59	5.72	5.06	47.34
U-Net-176	**Image**	D1	D3	R1	R3	K1	K3
	Threshold	0.05	0.005	0.05	0.25	0.05	0.05
	Dice-Score	0.883	0.369	**0.897**	**0.883**	0.847	**0.836**
	True negatives [%]	99.37	98.41	98.55	99.31	99.05	99.12
	True positives [%]	0.50	0.36	0.76	0.55	0.70	0.63
	False negatives [%]	0.07	0.80	0.57	0.04	0.06	0.21
	False positives [%]	0.06	0.42	0.11	0.10	0.19	0.04
	Avg. Hausdorff-Distance [mm]	9.95	41.94	11.94	1.62	2.93	17.16
U-Net-176 + CSAM	**Image**	D1	D3	R1	R3	K1	K3
	Threshold	0.05	0.005	0.05	0.4	0.05	0.05
	Dice-Score	0.877	0.318	0.585	0.779	0.803	0.749
	True negatives [%]	99.38	98.30	99.07	99.36	99.03	99.14
	True positives [%]	0.49	0.32	0.38	0.41	0.65	0.51
	False negatives [%]	0.05	0.84	0.05	0.05	0.21	0.33
	False positives [%]	0.08	0.54	0.49	0.18	0.11	0.01
	Avg. Hausdorff-Distance [mm]	9.48	33.35	12.48	6.97	5.12	19.04
U-Net-352	**Image**	D1	D3	R1	R3	K1	K3
	Threshold	0.05	0.05	0.05	0.05	0.05	0.05
	Dice-Score	0.843	0.043	0.715	0.867	**0.882**	0.821
	True negatives [%]	99.40	98.84	99.01	99.25	99.12	99.14
	True positives [%]	0.43	0.03	0.55	0.57	0.70	0.25
	False negatives [%]	0.14	9e-6	0.11	0.16	0.13	0.01
	False positives [%]	0.03	1.14	0.33	0.02	0.06	0.56
	Avg. Hausdorff-Distance [mm]	10.37	68.99	13.73	3.72	2.63	17.39

Table 1: Results of the segmentation with different models on 6 unseen images from the three datasets.

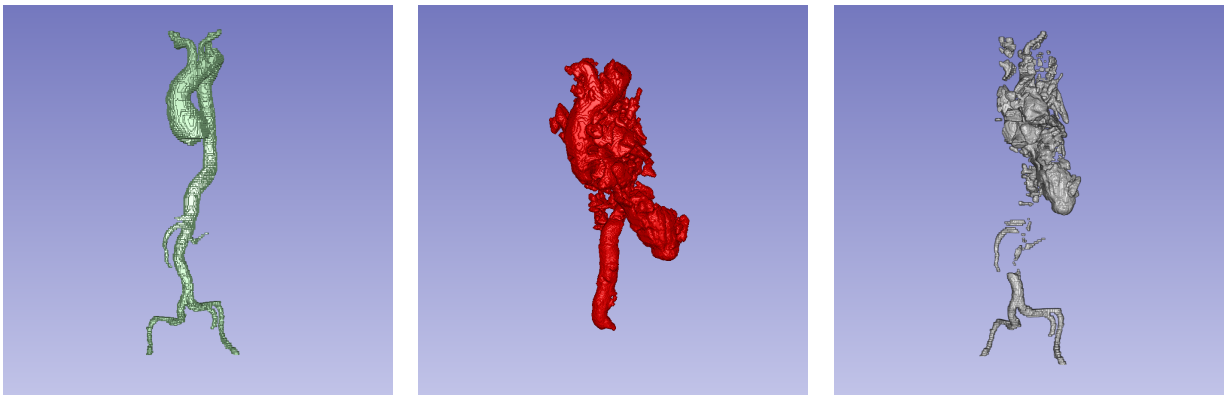
on datasets with small structures, rather than for example an aorta, which has significantly bigger structures, than retinal or brain vessels.

The results for image D3 might be worse because of intensity differences in the samples, such as differences in the intensity value of the aorta due to differing amounts of contrast agent or varying image acquisition protocols.

A major drawback of segmentation with deep learning are the substantial amounts of computational resources required, especially video memory and CUDA-cores. Therefore, an attempt was made to keep the computational requirements low enough for the networks to run on a home PC with a consumer graphics card. With more resources on hand, it would be possible to deepen the networks, increase the number of filters or feed them with bigger patches or increased batch size. This might improve the performance, but make it impossible to run on



(a) R3 ground truth (b) R3 segmentation (c) R3 difference



(d) D3 ground truth (e) D3 segmentation (f) D3 difference

Figure 2: The images (a)-(c) show the ground truth, the segmentation and the difference between segmentation and ground truth of image R3, images (d)-(f) those of D3 as an example of segmentation results of U-Net16+CSAM.

end-user machines and would require high performance computational resources.
Another problem with aortic vessel tree segmentation is the lack of annotated datasets. Some of this can be alleviated by data augmentation, but a massive dataset of similarly acquired CTA scans that can also be used to compare different segmentation approaches would be better. Especially parameters like spatial resolution, contrast agent administration and image size should be kept as similar as possible, to simplify the amount of pre-processing required and enhance network performance.

4. CONCLUSIONS

In this contribution, we developed and tested a deep learning network inspired by U-Net and the CS^2-Net for aortic vessel tree segmentation. The model was tested against six unknown CTA-scans and compared against other U-Net models. Generally, the classical U-Nets preformed better without the Spatial and Channel Attention Module, suggesting that this attention mechanism might not work well for AVT segmentation. All the networks delivered good segmentations in at least three of six cases, where all of the aortic arch, the thoracic and abdominal aorta, as well as the bifurcation, parts of the iliac arteries and in part most of the major branches were segmented. Some segmentations delivered bad results possibly because of differing image parameters, like image intensity at the aorta. These models were developed with low computational requirements, therefore the model could be tested with increased patch size, network width and depth and higher number of filters in the future. The small amount of data proved a challenge, therefore a big dataset of similarly acquired CTA scans might improve performance, as well as provide a baseline dataset to compare different AVT segmentation approaches.

REFERENCES

[1] Owens, B., "Scans: Enhanced medical vision," *Nature* **502**, 82–83 (Oct 2013).

[2] Shen, D. et al., "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.* **19**(1), 221–248 (2017).

[3] Egger, J., Gsaxner, C., Pepe, A., Pomykala, K. L., Jonske, F., Kurz, M., Li, J., and Kleesiek, J., "Medical deep learning–a systematic meta-review," *Computer Methods and Programs in Biomedicine* , 106874 (2022).

[4] Gao, J., Jiang, Q., Zhou, B., and Chen, D., "Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview," *Mathematical Biosciences and Engineering* **16**(6), 6536–6561 (2019).

[5] Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., and Nandi, A. K., "Medical image segmentation using deep learning: A survey," *IET Image Processing* **16**, 1243–1267 (jan 2022).

[6] Chen, C. et al., "Deep learning for cardiac image segmentation: A review," *Front. Cardiovasc. Med.* **7** (2020).

[7] Hesamian, M. H. et al., "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Dig. Imaging* **32**(4), 582–596 (2019).

[8] Pepe, A., Li, J., Rolf-Pissarczyk, M., Gsaxner, C., Chen, X., Holzapfel, G. A., and Egger, J., "Detection, segmentation, simulation and visualization of aortic dissections: A review," *Medical image analysis* **65**, 101773 (2020).

[9] Jin, Y., Pepe, A., Li, J., Gsaxner, C., Zhao, F.-h., Kleesiek, J., Frangi, A. F., and Egger, J., "Ai-based aortic vessel tree segmentation for cardiovascular diseases treatment: status quo," *arXiv preprint arXiv:2108.02998* (2021).

[10] Yu, T., Zhu, X., Tang, L., Wang, D., and Saad, N., "Review of ct angiography of aorta," *Radiologic Clinics of North America* **45**(3), 461–483 (2007). Emergency Cross-Sectional Imaging.

[11] Nienaber, C. A., Clough, R. E., Sakalihasan, N., Suzuki, T., Gibbs, R., Mussa, F., Jenkins, M. P., Thompson, M. M., Evangelista, A., Yeh, J. S., and et al., "Aortic dissection," *Nature Reviews Disease Primers* **2**(1) (2016).

[12] Radl, L. et al., "Avt: Multicenter aortic vessel tree cta dataset collection with ground truth segmentation masks," *Data in Brief* **40**, 107801 (2022).

[13] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., and Kikinis, R., "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic Resonance Imaging* **30**(9), 1323–1341 (2012). Quantitative Imaging in Cancer, Online available at: https://slicer.org.

[14] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*MICCAI*], 234–241, Springer International Publishing, Cham (2015).

[15] Mou, L. et al., "Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging," *Medical Image Analysis* **67**, 101874 (2021).

[16] Mou, L., Zhao, Y., Chen, L., Cheng, J., Gu, Z., Hao, H., Qi, H., Zheng, Y., Frangi, A., Liu, J., and et al., "Cs-net: Channel and spatial attention network for curvilinear structure segmentation," *Lecture Notes in Computer Science* , 721–730 (2019).

[17] Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., and Lu, H., "Dual attention network for scene segmentation," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], (June 2019).

[18] Ioffe, S. and Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR* **abs/1502.03167** (2015).

[19] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," (2014).

[20] Sampat, M. P., Wang, Z., Markey, M. K., Whitman, G. J., Stephens, T. W., and Bovik, A. C., "Measuring intra-and inter-observer agreement in identifying and localizing structures in medical images," in [*2006 International Conference on Image Processing*], 81–84, IEEE (2006).

[21] Pinter, C. et al., "SlicerRT – Radiation therapy research toolkit for 3D Slicer," *Med. Phys.* **39**(10), 6332–6338 (2012).