Adjusting U-Net for the aortic abdominal aneurysm CT segmentation case

R.U. Epifanov¹, N.A. Nikitin², A.A. Rabtsun³, L.N. Kurdyukov³, A.A. Karpenko², R.I. Mullyadzhanov^{1,4}

¹Novosibirsk State University, 63090, Russia, Novosibirsk, Pirogov str. 2;

² Meshalkin National Medical Research Center, 630055, Russia, Novosibirsk, Rechkunovskaya str. 15;

³ Novosibirsk State Medical University, 630091, Russia, Novosibirsk, Krasny ave. 52;

⁴ Institute of Thermophysics SB RAS, 630090, Russia, Novosibirsk, Lavrentyev ave. 1

Abstract

In this paper, we address the issue of developing of a convolutional neural network for the problem of aneurysm segmentation into three classes and of exploring ways for improving the quality of final segmentation masks. As a result of our study, macro dice score for classes of interest reaches $83.12\% \pm 4.27\%$. We explored different augmentation styles and showed the importance of applying intensity augmentation style to improve segmentation algorithm robustness in conditions of clinical data diversity. Augmentation with spatial and insensitive styles increase macro dice score up to 3%. The comparison of various inference mode indicate that combination of overlapping inference and segmentation window enlargement ameliorate macro dice up to 1.4%. Overall improvement of the quality of segmentation masks by macro dice score amounted up to 6% using combination of data-based augmentation style and advanced inference technique.

Keywords: abdominal aortic aneurysm, neural network, semantic segmentation, calcifications.

<u>Citation</u>: Epifanov RU, Nikitin NA, Rabtsun AA, Kurdyukov LN, Karpenko AA, Mullyadzhanov RI. Adjusting U-Net for the aortic abdominal aneurysm CT segmentation case. Computer Optics 2024; 48(3): 418-424. DOI: 10.18287/2412-6179-CO-1338.

Introduction

Abdominal aortic aneurysm is a disease manifested as an abnormal expansion of the abdominal aorta diameter [1], which potentially leads to aortic wall rupture causing life-threatening conditions. In this regard, timely surgical treatment has become a standard to prevent further disease progression upon reaching the critical size of the aneurysm [2]. Current main surgical tactics in the form of aneurysmectomy or stent implantation [3] require careful preoperative assessment of aneurysm anatomy in order to choose the best and safest surgical treatment. On the one hand, when identifying the best spots for clamping in order to reduce intraoperative embolism during aneurysmectomy, surgeons need to know where thrombotic masses and calcifications are located [4]. On the other hand, when determining stent graft implantation opportunities and assessing possible postoperative risks, surgeons need to quantitatively describe the aortic lumen morphology [4]. In current clinical practice, qualitative and quantitative descriptions of abdominal aortic aneurysm geometry are based on ultrasound and computer tomographic imaging. However, in order to facilitate surgery planning, it is advisable to perform three-dimensional aneurysm modelling based on segmentation of CT-images and model quantitative description.

In the biomedical field, U-Net has become the most popular choice to obtain segmentation [5]. Also, U-Net is used to segment abdominal aortic aneurysms in CT-images [6–13]. However, the papers [6–8] consider segmentation of the aneurysm as a whole, without separating the aortic lumen and adjacent tissues, which

does not allow one to assess the geometry of the aortic lumen, which is of paramount importance when planning stent graft implantations. The papers [9-11] discuss algorithms for a more detailed aneurysm segmentation, where the aortic lumen and adjacent tissue are separated, but those algorithms are not able to differentiate tissue into thrombotic masses and calcifications, which is critical when planning an aneurysmectomy. The paper [12] discusses an algorithm for extracting from CT images of aortic lumen, an aortic wall that includes thrombotic mass and calcifications that may be incorporated into the aortic wall, but the author does not provide any quality metric for calcifications. Moreover, a significant limitation of the above research is that the algorithms described in it are not in the public domain, nor are the datasets on which these algorithms were trained and tested. In addition, in the above papers, the authors do not describe the applicability boundaries of their algorithms in the context of [14] domain shift, which can greatly reduce the performance of the segmentation algorithm [15]. All of the above explains the relevance of developing an algorithm for segmentation of abdominal aortic aneurysms on CT images capable of distinguishing the aortic lumen, the aortic wall including one with thrombotic masses, and calcifications incorporated into the wall with documented application conditions.

In this paper, we developed an algorithm for segmentation of CT images into three classes: the aortic lumen, the aortic wall including one with thrombotic masses, and calcifications incorporated into the aortic wall. For the developed algorithm, we documented application conditions by describing train and test data distributions and obtained metrics for test data. Also, we analysed how different augmentation styles and inference modes affect the final segmentation quality.

1. Dataset

The dataset used in this article consists of 30 CT volumes provided by the Meshalkin National Medical Research Center (Novosibirsk, Russia). Each CT volume contains an abdominal aortic aneurysm recorded with contrast enhancement. Scanning CT volume was performed on Toshiba Aquilion One tomograph with convolution kernels FC03 or FC08 used for reconstruction algorithm (Fig. 1*b*, 2*b*). CTDIvol was ranged from 8.1 to 38.3 (Fig. 1*c*, 2*c*). Exposure time was 500 ms or 600 ms (Fig. 1*d*, 2*d*). Data discretization was performed along the axial direction (*z*-direction) with a slice thickness of 1 mm and an

increment of 0.8 mm. Each slice (x-y plane) of CT volume is 512×512 pixels, with the number of slices varying from 128 to 272. In x-y plane, pixel spacing varies from 0.518 to 0.976 mm per pixel (Fig. 1*a*, 2*a*).

Employing ITK-SNAP software [16], three medical experts independently prepared segmentation masks for three classes of interest, i.e. the aortic lumen (Class 1), aortic wall including one with thrombotic masses (Class 2), and calcifications in the aortic wall (Class 3). Medical experts used 350HU level with the window width of 700HU to label aortic lumen and calcifications incorporated into the aortic wall, and level set to 70HU with the window width of 400HU for the aortic wall including thrombotic masses. The final segmentation masks correspond to the aggregation of expert masks by a majority vote.



Fig. 1. CT image variation depending on scanning parameters. Four parameters are considered: pixel spacing, convolution kernel, computed tomography dose index (CTDIvol) and exposure time. For each pair, only one parameter is changed. a) Increasing of pixel spacing enlarges organ size. b) Changing convolution kernel from FC08 to FC03 adds sharped small grain on CT image. c) Decreasing CTDIvol adds smooth large grain on CT image. d) Enlargement exposure time adds sharped large grain on CT image

2. Neural network architecture

To develop the segmentation model, we relied on Unet architecture [5] (Fig. 3). We enhanced the vanilla Unet encoder with residual connections [17] and the decoder with squeeze and excitation blocks [18]. We decided to select an encoder similar to resnext101 32x8d [19] because the vanilla Unet encoder suffers from the vanishing gradient problem [20]. Adding residual connection to the encoder helps to ensure a more stable gradient flow. Our encoder implementation is based on torchvision one [21]. Also, we decided to exclude upconvolution layers, replacing them with linear interpolation layers, because up-convolutions are prone to the check board problem [22]. Strengthening the model by squeeze and excitation blocks allows one to precisely detect small details, such as calcifications [23]. To structure Unet layers and connections order, we lean on SMP library [24].

To train the neural network, we used AdamW optimizer [25]. Training process amounted to 80 epochs, with early stopping after 25 epochs without improvements. Each epoch consisted of 1000 iterations. For each iteration, 16 randomly cropped patches of the size of 128×128×16 from CT volumes with corresponding ground-truth patches were employed for the training. Trainings run on an Nvidia RTX 2070 SUPER GPU with 8 GB memory. As medical data is highly imbalanced, we select a combination of focal [26] and dice [27] losses for training:

$$L = -\sum_{i \in I} \sum_{k \in K} t_{ik} (1 - p_{ik}) \log p_{ik} + \frac{2}{|K|} \sum_{k \in K} \frac{\sum_{i \in I} t_{ik} p_{ik}}{\sum_{i \in I} t_{ik} + \sum_{i \in I} p_{ik}}, \quad (1)$$

where t and p stand for the ground truth and predicted segmentation probabilities for patches with shape I.

Indexes i and k iterate over I shape and K classes, respectively.



Fig. 3. Neural network architecture used for segmentation

3. Experiment descriptions

To conduct experiments, we trained Unet in three styles, namely a style without augmentation (WA), with spatial augmentations (SA), and with a combination of spatial and intensity augmentation (SA+IA). Spatial augmentations include vertical and horizontal flips, affine transformation, elastic transform and, grid distortion. Spatial augmentations were selected to compensated variation of pixel spacing. Intensity augmentations include gauss noise and blur. Intensity augmentations were selected to cover all type of noisiness and its sharpness on CT images. All augmentations [28] are applied in the x-y plane. To ensure statistical correctness, we applied the 5-fold cross validation, which are 60% training, 20% validating and 20% testing.

For each augmentation style, we inferred data with segmentation window size and step as $128 \times 128 \times 16$ and $128 \times 128 \times 16$ (I1), $128 \times 128 \times 16$ and $64 \times 64 \times 8$ (I2), $256 \times 256 \times 16$ and $256 \times 256 \times 8$ (I3). We selected I1

configuration to score because we apply I1 for early stopping during training. We selected I2 configuration to compare with I1 because half-overlapping is typically used [29] for improvement of segmentation mask quality. As we used Unet with convolutional layers without transformers, we applied I3 as a technical trick in order to decrease the segmentation border problem.

Moreover, there are usually prerequisites for segmented masks, as no holes in them or their singleness. In our experiments, we employed post-processing techniques such as largest connected component filtration (LCF) and filling holes filtration (FHF). LCF extracts the largest connected component from binarized segmented mask and multiply the extracted component with the segmented mask (Fig. 4b). Application of LCF allows removing small error areas from segmented masks. FHF fills the inner holes of the segmented mask (Fig. 4c). To generate class candidates for hole voxels, FHF uses k-nearest neighbour algorithm [30] with k equal to 3 and Euclidean metric for distance estimation.



Fig. 4. Explanation of post-processing steps. In the figure, a synthetic mask is used. Aortic lumen is marked red. The yellow and blue colours denote the aortic wall, including one with thrombotic masses, and calcifications, respectively. a) The mask is generated for explanation. b) The mask after applying LCF. c) The mask after applying FHF

To evaluate performance, we used volumetric dice score [31]:

$$VDS = \frac{2TP}{2TP + FP + FN},$$
 (2)

where TP (true positive) is the number of the true positive, FP (false positive) is that of the false positive and FN (false negative) is that of the false negative. We measure and provide metrics for each class of interest with organ windowing 256×256 in the x-y plane. Also, we use macro averaging to aggregate metric between classes.

5. Results

Tab. 1 provides dice metrics for different augmentation styles and inference modes. As seen on the table, inference with overlapping (I2) outperformed with no overlapping (I1) in all augmentation styles. The score improvement in overlapping inference can be explained by self-regulation based on averaging pixel classes probabilities obtained from different segmentation windows. The above parameter in overlapping inference varies from 1 % to 4 %, depending on classes. Less clear conclusions can be drawn about enlarging a segmentation window (I2 vs I3). In SA+IA augmentation style, enlarging segmentation window improves macro dice from 0.817 ± 0.061 to 0.831 ± 0.042 . While for WA and SA augmentation styles, dice score changes only slightly from 0.799 ± 0.085 to 0.801 ± 0.091 for WA, and from 0.777 ± 0.101 0.777 ± 0.108 for SA. Score to improvement due to enlarging a segmentation window can be explained by decreasing the influence of edge effects on final segmentation by reducing overall number of border pixels in a segmentation window.

Tab. 1. Test performance for different augmentation styles and inference modes. Dice score is provided in the mean value ± standarddeviation value format. AS and IM stand for augmentation style and inference mode, respectively. PP mean additional applying LCFand FHF

AS	IM	Class 1	Class 2	Class 3
WA	I1	0.914 ± 0.091	0.731 ± 0.072	0.667 ± 0.137
	I2	0.929 ± 0.075	0.761 ± 0.066	0.708 ± 0.125
	13	0.928 ± 0.078	0.768 ± 0.069	0.707 ± 0.133
	I3 + PP	0.926 ± 0.084	0.769 ± 0.071	0.704 ± 0.137
SA	I1	0.879 ± 0.126	0.715 ± 0.089	0.659 ± 0.115
	I2	0.885 ± 0.118	0.762 ± 0.079	0.683 ± 0.123
	13	0.879 ± 0.129	0.773 ± 0.079	0.680 ± 0.129
	I3 + PP	0.875 ± 0.131	0.780 ± 0.091	0.680 ± 0.135
SA+IA	I1	0.923 ± 0.068	0.737 ± 0.092	0.665 ± 0.118
	I2	0.946 ± 0.047	0.791 ± 0.069	0.714 ± 0.094
	13	0.956 ± 0.029	0.821 ± 0.056	0.716 ± 0.088
	I3 + PP	0.956 ± 0.029	0.826 ± 0.057	0.733 ± 0.068

Further, we analyse the dependence of dice score on the style of augmentation used during training. The SA+IA style increases macro dice score from 0.801 ± 0.091 to 0.831 ± 0.043 compared to WA, whereas the SA style decreases macro that from 0.801 ± 0.091 to 0.777 ± 0.011 compared to WA. A dramatic score reduction with SA compared to WA happens in cases: when aneurysm is adjacent to the spine and when the model mistakes vertebra parts with high HU in CT for a contrasted aortic lumen

(Fig. 5). In our opinion, these types of missegmentation can be caused by combinations of variations of inserted contrast in aorta and increased variability in aneurysm anatomy due to elastic transform and grid distortion. These factors lead models to assign lumen class probabilities based on local large intensity. Adding IA to SA (SA+IA) helps to avoid mistakes by using CT images granularity variations during training, reducing the importance of local large intensity to assign lumen class probabilities.



Fig. 5. Segmentation of aneurysm adjacent to spine. Ground truth and neural network masks on left and right, respectively. Aortic lumen (Class 1) are marked red. The yellow and blue colours denote aortic wall (Class 2), including one with thrombotic masses, and calcifications (Class 3), respectively. a) Spine area with high HU is missegmented as aortic lumen. b) Spine area without high HU is well segmented

As aneurysm is a single-connected object, it is natural to apply the largest connected component filtration in order to remove the FP areas. As aneurysm is a solid object, after the largest connected component filtration, it is logical to apply filling holes filtration in order to remove the FN areas. Applying post-processing steps is more efficient when in a segmentation mask there are small separate areas or small holes inside aneurysm, then post-processing steps help to prepare a mask for further automatization analysis. For SA+IA augmentation style and I3 inference mode, post-processing steps increase dice score from 0.831 ± 0.042 to 0.837 ± 0.039 . For poorly segmented CT images, mask post-processing could even decrease dice score, but such masks still require expert intervention to correct the mask for further processing.

Table 2 provides dice metrics of algorithms described in literature for the aneurysm segmentation task. Since the problem of aneurysm segmentation is formulated in different ways, we split the table into three parts based on problem formulation. Also, as for the algorithms described in the literature there are not provided its realization or test data, we compare the algorithm based on dice score provided in the papers indicating the number of CT images used for the test phase. In formulating the segmentation problem into one class, our algorithm outperforms by mean value dice metric the algorithm of [6-8]. Brutti et al. demonstrate smaller standard deviation value compared to others, but the number of CT test images are only 8. In formulating the segmentation problem into two classes, all algorithms listed in table show a similar standard deviation value, but our algorithm poses the largest mean dice value compared to the others [9-11]. In formulating the segmentation problem into three classes, we compare our algorithm with that of Lareyre et al [12] one. Lareyre et al. do not provide metrics for calcifications because we only compare dice score of aortic lumen and thrombotic masses. For aortic lumen and thrombotic masses, the compared algorithms demonstrate similar standard deviation values. Our algorithm is better at segmenting aortic lumen, but it falls behind Lareyre et al. algorithm on thrombotic masses.

Tab. 2. Dice scores of aneurysm segmentation algorithms described in literature

			-		
	Alg	orithms segmented a	neurysm on three cl	asses	
	class 1	class 2	class 3	Averaged	CT count
Lareyre et al [12]	0.93 ± 0.04	0.88 ± 0.12	-	-	40
Our	0.96 ± 0.06	0.82 ± 0.11	0.72 ± 0.15	0.83 ± 0.08	30
Algorithms segment	ed aneurysm on two	classes			
	class 1	class 2 + class 3		Averaged	CT count
Caradu et al [9]	0.93 ± 0.05	0.81 ± 0.10		-	100
López et al [10]	-	-		0.84 ± 0.068	12
Lalys et al [11]	-	-		0.86 ± 0.06	92
Our	0.95 ± 0.06	0.83±0.11		0.89 ± 0.07	30
Algorithms segment	ed aneurysm on one	class			
	cl	ass 1 + class 2 + clas	Averaged	CT count	
Lu et al [6]	0.873 ± 0.129			0.873 ± 0.129	321
Habijan et al [7]	0.91±0.156			0.910 ± 0.156	19
Brutti et al [8]	0.89 ± 0.04			0.89 ± 0.04	8
Our	0.92 ± 0.08			0.92 ± 0.08	30

Conclusion

In this paper, we developed an algorithm for aortic aneurysm segmentation for three classes problem formulation. We showed the importance of using not only spatial augmentations, but also intensity augmentations for training. Using the combination of spatial and intensity augmentations increases dice score up to 3 %. Moreover, we demonstrated that a combination of overlapping inference and enlarged segmentation window increases the quality of final segmentation. Overlapping inference and segmentation window enlargement improve dice score by 1.4% for combination of spatial and intensity augmentation styles. We reached the accuracies by dice score for aortic lumen 0.96 ± 0.06 , for aortic wall including the wall with thrombotic masses 0.82 ± 0.11 and for calcifications incorporated into a rtic wall 0.72 ± 0.15 . For the algorithm developed and metrics obtained, we provided CT-data acquisition parameters, which is essential during algorithm application stage as it helps to avoid a reduction in segmentation quality caused by domain shift.

Acknowledgements

The work is supported by the Russian Science Foundation grant No. 21-15-00091.

References

- Ernst CB. Abdominal aortic aneurysm. N Engl J Med 1993; 328(16): 1167-1172.
- [2] Pecoraro F, et al. Mortality rates and risk factors for emergent open repair of abdominal aortic aneurysms in the endovascular era. Updates Surg 2018; 70(1): 129-136.
- [3] Sharafuddin MJ, Man JH. Management of aortic aneurysms. In Book: Shammas NW, ed. Peripheral arterial interventions. Springer; 2022: 309-318.
- [4] Wanhainen A, et al. European Society for Vascular Surgery (ESVS) 2019 clinical practice guidelines on the management of abdominal aorto-iliac artery aneurysms. Acta Angiol 2022; 28(3): 69-146.
- [5] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Int Conf on Medical Image Computing and Computer-Assisted Intervention 2015: 234-241.
- [6] Lu J-T, et al. DeepAAA: clinically applicable and generalizable detection of abdominal aortic aneurysm using deep learning. Int Conf on Medical Image Computing and Computer-Assisted Intervention 2019: 723-731.
- [7] Habijan M, et al. Abdominal aortic aneurysm segmentation from ct images using modified 3D U-Net with deep supervision. 2020 Int Symp ELMAR 2020: 123-128.
- [8] Brutti F, et al. Deep learning to automatically segment and analyze abdominal aortic aneurysm from computed tomography angiography. Cardiovasc Eng Technol 2022; 13(4): 535-547.
- [9] Caradu C, et al. Fully automatic volume segmentation of infrarenal abdominal aortic aneurysm computed tomography images with deep learning approaches versus physician controlled manual segmentation. J Vasc Surg 2021; 74(1): 246-256.
- [10] López-Linares K, et al. 3D convolutional neural network for abdominal aortic aneurysm segmentation. arXiv

Preprint. 2019. Source:

<https://arxiv.org/abs/1903.00879>.

- [11] Lalys F, et al. Generic thrombus segmentation from preand post-operative CTA. Int J Comput Assist Radiol Surg 2017; 12(9): 1501-1510.
- [12] Lareyre F, et al. A fully automated pipeline for mining abdominal aortic aneurysm using image segmentation. Sci Rep 2019; 9(1): 13750.
- [13] Fedotova Y, et al. Automatically hemodynamic analysis of AAA from CT images based on deep learning and CFD approaches. J Phys Conf Ser 2021; 2119: 012069.
- [14] Kloenne M, et al. Domain-specific cues improve robustness of deep learning-based segmentation of CT volumes. Sci Rep 2020; 10(1): 10712.
- [15] Kurmukov A, et al. Challenges in building of deep learning models for glioblastoma segmentation: Evidence from clinical data. Stud Health Technol Inform 2021; 27(281): 298-302.
- [16] Yushkevich PA, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage 2006; 31(3): 1116-1128.
- [17] Qamar S, Ahmad P, Shen L. Dense encoder-decoderbased architecture for skin lesion segmentation. Cogn Comput 2021; 13(2): 583-594.
- [18] Rundo L, et al. USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. Neurocomputing 2019; 365: 31-43.
- [19] Ibtehaz N, Rahman MS. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw 2020; 121: 74-87.
- [20] He K, et al. Deep residual learning for image recognition. Proc IEEE Conf on Computer Vision and Pattern Recognition 2016: 770-778.
- [21] Paszke A, et al. Automatic differentiation in PyTorch. 31st Conf on Neural Information Processing Systems (NIPS 2017) 2017: 1-4.
- [22] Odena A, Dumoulin V, Olah C. Deconvolution and checkerboard artifacts. Distill 2016; 1(10): e3.
- [23] Roy AG, Navab N, Wachinger C. Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. IEEE Trans Med Imaging 2018; 38(2): 540-549.
- [24] Iakubovskii P. Segmentation_models.pytorch. 2019. Source: https://github.com/qubvel/segmentation models.pytorch>.
- [25] Loshchilov I, Hutter F. Fixing weight decay regularization in Adam. ICLR 2018 Conf Blind Submission. 2018. Source: https://openreview.net/forum?id=rk6qdGgCZ>.
- [26] Lin T-Y, et al. Focal loss for dense object detection. Proc IEEE Int Conf on Computer Vision 2017: 2980-2988.
- [27] Cardoso MJ, et al, eds. Deep learning in medical image analysis and multimodal learning for clinical decision support. Cham: Springer International Publishing AG; 2017.
- [28] Buslaev A, et al. Albumentations: fast and flexible image augmentations. Information 2020; 11(2): 125.
- [29] Cardoso MJ, et al. MONAI: An open-source framework for deep learning in healthcare. arXiv Preprint. 2022. Source: https://arxiv.org/abs/2211.02701>.
- [30] Mucherino A, et al. K-nearest neighbor classification. In Book: Mucherino A, Papajorgji PJ, Pardalos PM, eds. Data mining in agriculture. Dordrecht: Springer Science+Business Media LLC; 2009: 83-106.
- [31] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 2015; 15(1): 29.

Authors' information

Rostislav UIrevich Epifanov, (b. 1996) graduated from Novosibirsk State University. Scientific interests include medical image processing. E-mail: <u>*rostepifanov@gmail.com*</u>

Nikita Aleksandrovich Nikitin, (b. 1990), graduated from Siberian Medical State University. Research interests: radiology and cardiology. E-mail: <u>n nikitin@outlook.com</u>

Artem Aleksandrovich Rabtsun, (b. 1983), graduated from Siberian Medical State University. Research interests: vascular surgery, cardiology. E-mail: <u>rabtsun@xaker.ru</u>

Leonid Nicolaevich Kurdyukov, (b. 1996), graduated from Novosibirsk State Medical University. Research interests: cardiology. E-mail: *leonidkurdyukov@gmail.com*

Andrey Anatolievich Karpenko, (b. 1962), graduated from Altai State Medical Institute. Research interests: vascular surgery, cardio surgery. E-mail: <u>andreikarpenko@rambler.ru</u>

Rustam Ilhamovich Mullyadzhanov, (b. 1987), graduated from Novosibirsk State University. Research interests: hemodynamics, numerical simulations. E-mail: *rustammul@gmail.com*

Received December 20, 2022. The final version – June 16, 2023.