Weak-supervised Attention Fusion Network for Carotid Artery Vessel Wall Segmentation

Haijun Lei[1], Guanjiie Tong[1], Huaqiang Su[1], and Baiying Lei*[2]

[1] Key Laboratory of Service Computing and Applications, Guangdong Province Key Laboratory of Popular High Performance Computers, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China.
[2] Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Shenzhen University Medical school, Shenzhen 518060, China. (*Email: leiby@szu.edu.cn)

Abstract. The automatic and accurate segmentation of the carotid artery vessel wall can assist doctors in clinical diagnosis. Medical images often have complex and blurry features, which makes manual data annotation very difficult and time-consuming. 3D CNN can utilize three-dimensional spatial information to more accurately identify diseased tissues and organ structures, but its segmentation performance is limited due to the lack of global contextual information correlation. This paper proposes a network based on CNN and Transformer to segment the carotid artery vessel wall. By combining the effectiveness of CNN in dealing with 3D image segmentation problems and the global attention mechanism of Transformer, it is possible to better capture and process the features of this information. By designing Joint Attention Structure Block (JAS), semantic information in skip connections can be enhanced. The feature fusion block (FF) is used to associate input information with each layer of feature maps, enhancing the detailed information of the feature maps. The effectiveness of this method has been verified through a large number of comparative experiments.

Keywords: Weak-supervision · Carotid vessel wall segmentation · Transformer.

1 Introduction

Atherosclerosis is a progressive disease. In the early stage, it is manifested as thickening of the vascular wall and narrowing of lumen, and in the late stage, it is manifested as plaque lesions. There is a risk of rupture of atherosclerotic plaque in the internal carotid artery (ICA), which embolizes cerebral vessels and leads to stroke [1]. Since carotid atherosclerosis is the main source of ischemic stroke and the main indicator of systemic atherosclerosis, the carotid artery has always been an important target of image-based diagnostic procedures [2]. In recent decades, the incidence rate and mortality of atherosclerosis in most parts

of the world have continued to increase. In recent years, researchers have explored a variety of atherosclerotic treatment methods. However, the incidence rate of cardiovascular and cerebrovascular events is still high [3]. Several medical imaging methods are commonly used for plaque imaging, such as magnetic resonance imaging (MRI), computed tomography (CT), X-ray, and ultrasound (US). Although carotid CTA provides rapid and detailed imaging of extracranial and intracranial carotid systems, contrast agent administration helps to detect atherosclerosis plaque with high accuracy. However, due to the improved spatial resolution, reduced saturation effects and voxel dephasing, and better evaluation of vascular lumen and plaque features, specific MRA research protocols provide more detailed imaging that can provide structural and functional information, including luminal stenosis [4], blood vessel wall measurements, plaque composition, blood flow velocity, and flow velocity [5]. Meanwhile, MRA imaging is non-invasive and more patient-friendly compared to other diagnostic methods. MRI is more preferable than CT [6]. The degree of carotid artery stenosis is an important factor affecting clinical surgical treatment decisions, Despite the powerful capabilities of MRI, automated, accurate, objective, and reproducible quantitative analysis of MRI data has been proven to be a challenging goal in computer science over the years. In the field of vascular wall imaging, almost all image processing algorithms are applied to accurately and objectively depict the boundaries of vascular walls for the task of segmenting blood vessels. Therefore, there is an urgent need for automatic and accurate carotid bifurcation segmentation methods to better quantify carotid artery stenosis and provide effective decision-making opinions for doctors' diagnoses [7] [8]. At present, a large number of researchers have proposed methods based on traditional image processing. For example, Christos P. Loizou et al. proposed and evaluated an integrated system for atherosclerotic plaque segmentation in carotid ultrasound imaging, which is based on normalization, speckle reduction filtering and four different snake segmentation methods [9]. Vukadinovic D et al. proposed a new method for segmenting carotid artery vessel walls in computed tomography (CTA) data [10]. Rocha R et al. proposed a new method to segment the proximal and distal intima-media regions of the common carotid artery in ultrasound images [11]. However, these methods often require complex and time-consuming feature processing, making it difficult to achieve real-time and accurate results.

With the rapid development of deep learning, a large number of deep learning models are currently being applied to segmentation tasks in the medical field. As Yanchao Yuan proposed a flexible method CSM-Net for joint segmentation of IMC and lumen in carotid artery ultrasound images [12]. Lin Y et al. proposed a two-dimensional V-Net model that can automatically segment the intima-media in carotid artery ultrasound images [13]. Lain é N et al. developed a region-based segmentation method that learns the appearance of IMC from data annotated by human experts [14]. Lain é N et al. proposed a region-based segmentation method called careSegDeep and a supervised deep learning method based on the extended U-net architecture [15]. Although these methods have achieved certain results, they are all based on precisely annotated datasets. However, it

is difficult to effectively address rough labels in complex scenes. Therefore, the contributions of this paper are as follows:

This article proposes a network model based on CNN and Transformer for precise segmentation of carotid artery vessel walls. The Joint Attention Structure Block (JAS) was designed to enhance the semantic information in skip connections and used a Feature Fusion Block (FF) to associate input information with each layer of feature maps, enhancing the detail information of feature maps. A large number of experiments have demonstrated the effectiveness of the design method in this study, which can segment relatively complete carotid artery vessel walls from rough labels

2 Method

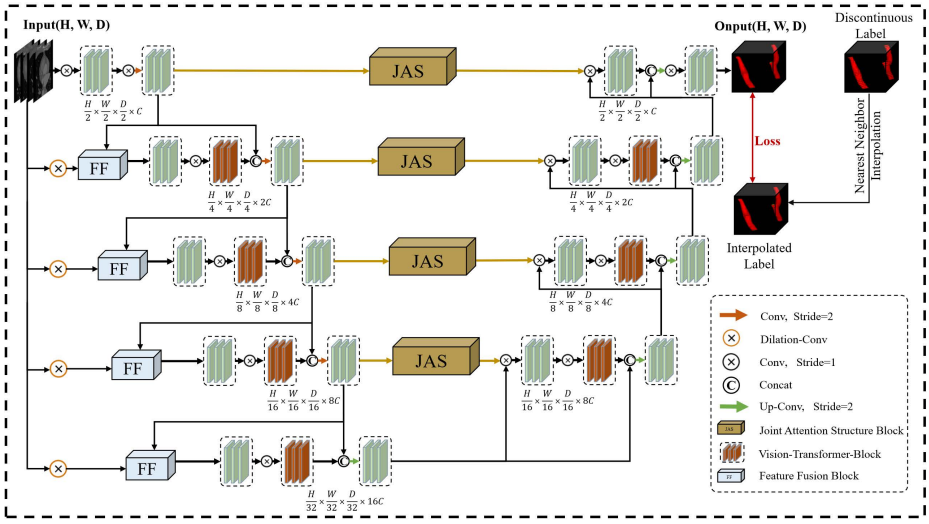

Fig. 1. Network architecture diagram for carotid vessel wall segmentation based on CNN and Transformer attention fusion network.

As is known to us all, 3D CNN can use three-dimensional spatial information to more accurately identify diseased tissues and organ structures, but its segmentation performance is limited due to the lack of global contextual information correlation. This paper proposed a network based on CNN and Transformer [16] to segment the carotid artery vessel wall. As shown in Fig.1, the model constructed in this study is based on an improvement of VNet [17], which replaces the convolutional layer between the encoder and decoder with a visual converter to enhance global information correlation. Firstly, uniformly scale the input to a fixed size, then normalize the maximum and minimum values to adjust pixel

values between 0 and 1. During the encoding process, the downsampled feature maps at each layer are connected to the features that have already passed through the Transformer structure, and then connected to the decoder's feature maps through a skip-connected joint attention structure module. Each decoder layer uses a feature fusion module to associate the feature maps convolved from different scale holes with the feature maps from the previous layer. It is worth mentioning that the labels of the carotid artery vessel walls obtained from our dataset are discontinuous. In the data processing stage, this study used linear interpolation to change the size of the data and corresponding labels from 432x432x432 to 512x512x512, because linear interpolation produces jagged vessel boundaries. Therefore, in this study, Three-dimensional middle finger filtering was used to smooth the interpolated carotid artery boundaries, and finally fill the voids in the vessel walls to obtain the final training labels through image morphology opening and closing operations.

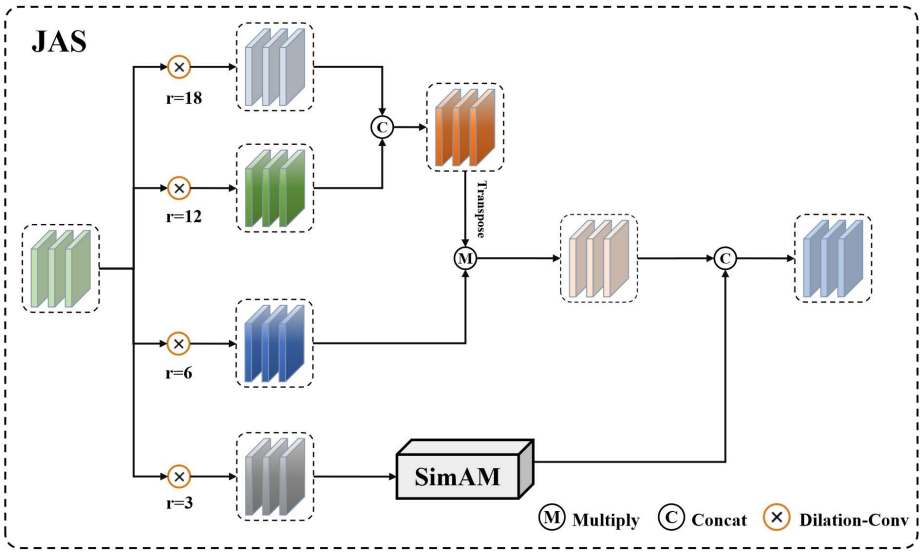

Fig. 2. The detail of the joint attention structure feature fusion module, Where r represents the step size of dilated convolution

2.1 Joint Attention Structure Block

This paper proposed the joint attention structure feature fusion module adopts a multi-level dilated convolution combined with attention structure, as shown in Fig.2, which is similar to ASPP [18] in structure. ASPP module is a commonly used module in deep learning, mainly used for image segmentation tasks, aiming

to solve the problem of insufficient contextual information in semantic segmentation. This module uses multiple parallel dilated convolution layers with different sampling rates to further process the extracted features for each sampling rate in separate branches and fuse them to generate the final result. The formulation of the entire procedure can be expressed as:

$$F_{jas} = Cat(Cat(DC_{l=18}, DC_{l=12})^T \times DC_{l=6}, SimAM(DC_{l=3})) \quad (1)$$

Where DC represents dilated convolution, l represents stride, Cat represents concatenate.

Having multi-scale feature extraction and fusion capabilities can effectively improve the generalization performance of the network. Unlike ASPP, this study combines the feature maps obtained from multi-scale dilated convolution with self-attention fusion, and then concatenates them with the feature maps enhanced by SimAM [19]. The SimAM module is an attention mechanism-based module used to calculate the weight of each feature map. This module consists of two parts: global average pooling and fully connected layer. Global average pooling is used to average the feature values of each channel in the feature map to obtain a global feature vector. The fully connected layer maps the global feature vector to an attention weight vector, which is used to represent the importance of each feature map.For each feature map, calculate its global feature vector through global average pooling operation. Then the fully connected layer maps the global feature vector to the attention weight vector, indicating the importance of the feature map in the final result. Finally, each value in the attention weight vector is used to weight the corresponding channel of the input feature map, resulting in the final output feature map.

2.2 Feature Fusion Block

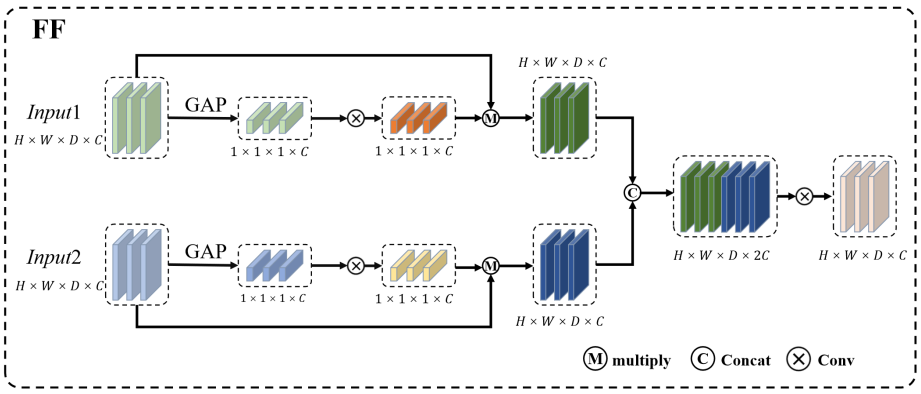

Fig. 3. The detail of the feature fusion block.

In the encoder stage, we use a feature fusion module to fuse the output of each layer's dilated convolution with the previous layer's feature map. To avoid semantic loss of input information and high model complexity, we adjust the channels of the two input branches separately. After average pooling and sigmoid activation function, these features were multiplied with the original feature map, Finally, the feature maps obtained from the two branches are concatenated to obtain the final output. Dual branch feature fusion can fuse global and local features, allowing the model to capture both environmental and detailed information, thereby improving the performance of the model. Through feature fusion, multiple features can be transformed into a more easily interpretable form, thereby improving the interpretability and visualization of the model. The detailed diagram is shown in the Fig.3, The process can be formulated as:

$$F_{ff} = Conv(Cat(F_1 \times Conv(GAP(F_1)), F_2 \times Conv(GAP(F_2)))) \quad (2)$$

Where F_1 and F_2 represents input features, GAP represents global average pooling.

3 Experiments and Results

3.1 Implementation details

The dataset of this study comes from carotid vascular wall segmentation and atherosclerosis diagnosis challenge (COSMOS 2022), and the experiment of this study was conducted on a workstation using the Pytorch deep learning framework equipped with RTX 8000 48GB graphics memory. In addition, all models were trained using the AdamW optimizer with an initial learning rate of 0.001 and choosing the Dice loss be the cost function. The entire training process takes 5 hours, with a total of 200 Epochs. During the training process, the annealing cosine method is used to dynamically adjust the learning rate. For the pre-processing of the dataset, this study first used the utils of the SimpleITK to convert the medical digital imaging and communication (DICOM) format to NIFTI format and adjusted it to a fixed size. Truncate the HU value to 0-1000 as the normalized input. In the post-processing stage, this study uses median filtering to smoothly resample the predicted image back to its original size and construct the largest connected region to eliminate some erroneous segmentation results in the model. In addition, this study also used image morphology to fill and segment the holes in the carotid artery vessel wall to obtain the final output result.

3.2 Result

The evaluation indicators of this study include Dice similarity, mIoU, recall, and accuracy Pre. Compared with traditional segmentation networks, this study achieved good performance improvements in all indicators. Please refer to Table 1 for details.

Table 1. Performance comparison between our method and other methods.

Method	DSC(%)	mIou(%)	Recall(%)	Pre(%)
SegNet	63.11	57.88	60.13	64.07
U-Net [20]	73.78	65.59	75.53	69.04
Res-UNet	74.21	65.96	81.69	73.03
Dense-UNet	74.50	66.21	81.06	70.23
VNet [17]	75.49	67.09	80.15	72.08
STUNet [21]	76.47	67.98	79.78	73.82
KIUNet [22]	75.32	66.12	79.47	71.44
Ours	**76.52**	**69.03**	**82.24**	**74.26**

Table 2. Performance comparison between our method and other methods.

Method	DSC(%)	mIou(%)	Recall(%)	Pre(%)
Baseline	75.49	67.09	80.15	72.08
Baseline+ViT	75.64	67.42	80.87	73.17
Baseline+ViT+JAS	76.09	68.41	82.10	73.92
Baseline+ViT+JAS+FF	76.52	69.03	82.24	74.26

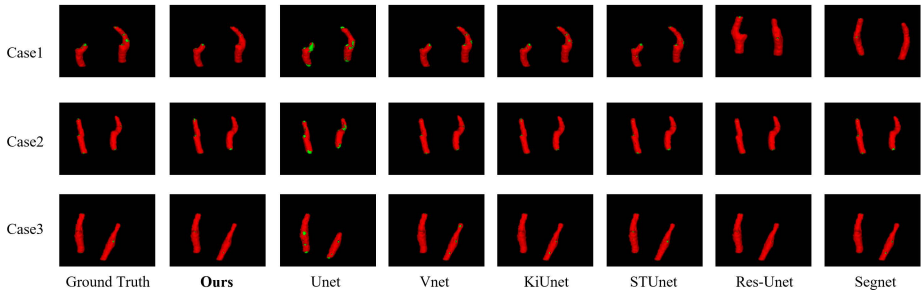

Fig. 4. Visualization results of segmentation for three typical examples (1). Our method has segmented a more complete carotid artery vessel wall.

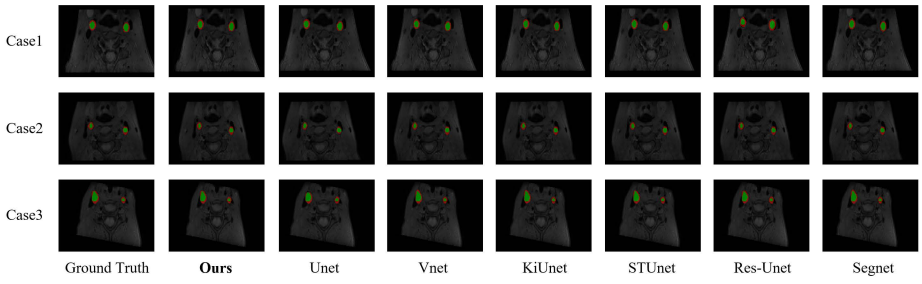

Fig. 5. Visualization results of segmentation for three typical examples (2). Our method provides a more complete segmentation of blood vessel walls from cross-sectional views.

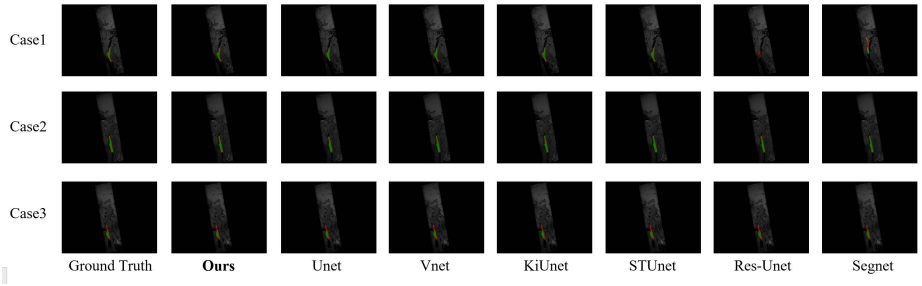

Fig. 6. Visualization results of segmentation for three typical examples (3). Our method achieves better continuity in segmentation from a cross-sectional perspective.

Through the table 1, it can be seen that the model constructed in this study has achieved the best results in all indicators, with a DSC value of 76.52%, mIou of 69.03%, Recall of 82.24%, and Pre of 74.26%.

This study demonstrated the effectiveness of the design module by setting up corresponding smile experiments. The ablation experiment results are detailed in Table 2:

The ablation experiment in the above table can verify the effectiveness of the proposed module in this study, and it can be observed that the fusion module constructed in this study can have certain performance improvements in various indicators. Because the model constructed in this study is based on an improvement of VNet, the baseline of this study is VNet, and corresponding modules are added layer by layer to form the ablation experiment of this study.

This study drew segmentation visualization diagrams for three typical examples, as shown in Fig.4 - Fig.6. It can be seen that the method proposed in this study has better segmentation performance compared to other methods in comparative experiments.It can be seen from the segmentation visualization that the proposed method has better segmentation performance compared to other methods in comparative experiments, and the segmented carotid artery vessel wall is more complete.

4 Conclusion

This study proposes a segmentation network based on convolutional neural networks and Transformers for precise segmentation of arterial blood vessels. This study combines the Vision Transformer module with the VNet model. Compared to traditional U-Net, VNet can directly process 3D data in processing 3D image segmentation problems, while U-Net is commonly used for processing 2D images and improves its structure by introducing residual connections. Residual connections allow networks to learn complex mapping relationships more easily, thereby improving their expressive power. Due to VNet working independently on each pixel, it is unable to effectively utilize long-range dependency information. In contrast, Transformer, due to its global attention mechanism, can better

capture and process this information. This study replaced the intermediate convolution module in the Vnet module with the Vision Transformer module, and introduced the Joint Attention Structure Block (JAS) to enhance the semantic information in skip connections. The feature fusion module (FF, Feature fusion block) is used to associate input information with each layer of feature maps, enhancing the detailed information of the feature maps.

Acknowledgement. This work was supported partly by the National Natural Science Foundation of China (Grant Nos. 62171312, U22A2024, 62276172 and 62271328), Shenzhen Science and Technology Program (Grant No. JCYJ2022081 8095809021 JCYJ20230808105602005), Shenzhen Medical Research Funds(No.C 2301005) National Natural Science Foundation of Guangdong Province (Nos. 2024A1515011950, 2023A1515011378).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Huang, X., Wang, J., Li, Z.: 3d carotid artery segmentation using shape-constrained active contours. Computers in Biology and Medicine **153**, 106530 (2023)
2. Jiang, M., Chiu, B.: A dual-stream centerline-guided network for segmentation of the common and internal carotid arteries from 3d ultrasound images. IEEE Transactions on Medical Imaging (2023)
3. Lin, Y., Huang, J., Xu, W., Cui, C., Xu, W., Li, Z.: Method for carotid artery 3-d ultrasound image segmentation based on cswin transformer. Ultrasound in Medicine & Biology **49**(2), 645–656 (2023)
4. Nederkoorn, P.J., van der Graaf, Y., Hunink, M.M.: Duplex ultrasound and magnetic resonance angiography compared with digital subtraction angiography in carotid artery stenosis: a systematic review. Stroke **34**(5), 1324–1331 (2003)
5. Markl, M., Schnell, S., Wu, C., Bollache, E., Jarvis, K., Barker, A., Robinson, J., Rigsby, C.: Advanced flow mri: emerging techniques and applications. Clinical radiology **71**(8), 779–795 (2016)
6. Dakis, K., Nana, P., Athanasios, C., Spanos, K., Konstantinos, B., Giannoukas, A., Kouvelos, G.: Carotid plaque vulnerability diagnosis by cta versus mra: A systematic review. Diagnostics **13**(4), 646 (2023)
7. Samber, D.D., Ramachandran, S., Sahota, A., Naidu, S., Pruzan, A., Fayad, Z.A., Mani, V.: Segmentation of carotid arterial walls using neural networks. World Journal of Radiology **12**(1), 1 (2020)
8. Ziegler, M., Alfraeus, J., Bustamante, M., Good, E., Engvall, J., de Muinck, E., Dyverfeldt, P.: Automated segmentation of the individual branches of the carotid arteries in contrast-enhanced mr angiography using deepmedic. BMC medical imaging **21**, 1–10 (2021)
9. Loizou, C.P., Pattichis, C.S., Pantziaris, M., Nicolaides, A.: An integrated system for the segmentation of atherosclerotic carotid plaque. IEEE Transactions on Information technology in Biomedicine **11**(6), 661–667 (2007)
10. Vukadinovic, D., van Walsum, T., Manniesing, R., Rozie, S., Hameeteman, R., de Weert, T.T., van der Lugt, A., Niessen, W.J.: Segmentation of the outer vessel wall of the common carotid artery in cta. IEEE transactions on medical imaging **29**(1), 65–76 (2009)

11. Rocha, R., Campilho, A., Silva, J., Azevedo, E., Santos, R.: Segmentation of the carotid intima-media region in b-mode ultrasound images. Image and vision computing **28**(4), 614–625 (2010)
12. Yuan, Y., Li, C., Xu, L., Zhu, S., Hua, Y., Zhang, J.: Csm-net: Automatic joint segmentation of intima-media complex and lumen in carotid artery ultrasound images. Computers in Biology and Medicine **150**, 106119 (2022)
13. Lin, Y., Huang, J., Chen, Y., Chen, Q., Li, Z., Cao, Q.: Intelligent segmentation of intima–media and plaque recognition in carotid artery ultrasound images. Ultrasound in Medicine & Biology **48**(3), 469–479 (2022)
14. Lainé, N., Liebgott, H., Zahnd, G., Orkisz, M.: Carotid artery wall segmentation in ultrasound image sequences using a deep convolutional neural network. In: International Conference on Computer Vision and Graphics. pp. 73–84. Springer (2022)
15. Lainé, N., Zahnd, G., Liebgott, H., Orkisz, M.: Segmenting the carotid-artery wall in ultrasound image sequences with a dual-resolution u-net. In: 2022 IEEE International Ultrasonics Symposium (IUS). pp. 1–4. IEEE (2022)
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
17. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
18. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
19. Yang, L., Zhang, R.Y., Li, L., Xie, X.: Simam: A simple, parameter-free attention module for convolutional neural networks. In: International conference on machine learning. pp. 11863–11874. PMLR (2021)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
21. Huang, Z., Wang, H., Deng, Z., Ye, J., Su, Y., Sun, H., He, J., Gu, Y., Gu, L., Zhang, S., et al.: Stu-net: Scalable and transferable medical image segmentation models empowered by large-scale supervised pre-training. arXiv preprint arXiv:2304.06716 (2023)
22. Valanarasu, J.M.J., Sindagi, V.A., Hacihaliloglu, I., Patel, V.M.: Kiu-net: Towards accurate segmentation of biomedical images using over-complete representations. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23. pp. 363–373. Springer (2020)