DSNet: A Spatio-Temporal Consistency Network for Cerebrovascular Segmentation in Digital Subtraction Angiography Sequences

Qihang Xie[1,2], Dan Zhang[3], Lei Mou[1], Shanshan Wang[4], Yitian Zhao[1], Mengguo Guo[5], and Jiong Zhang[1,*]

[1] Laboratory of Advanced Theranostic Materials and Technology, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China
jiong.zhang@ieee.org
[2] Cixi Biomedical Research Institute, Wenzhou Medical University, Ningbo, China
[3] School of Cyber Science and Engineering, Ningbo University of Technology, Ningbo, China
[4] Shenzhen Institute of Advanced Technology, Chinese Academy of Science, Shenzhen, China
[5] First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Abstract. Digital Subtraction Angiography (DSA) sequences serve as the foremost diagnostic standard for cerebrovascular diseases (CVDs). Accurate cerebrovascular segmentation in DSA sequences assists clinicians in analyzing pathological changes and pinpointing lesions. However, existing methods commonly utilize a single frame extracted from DSA sequences for cerebrovascular segmentation, disregarding the inherent temporal information within these sequences. This rich temporal information has the potential to achieve better segmentation coherence while reducing the interference caused by artifacts. Therefore, in this paper, we propose a spatio-temporal consistency network for cerebrovascular segmentation in DSA sequences, named DSNet, which fully exploits the information of DSA sequences. Specifically, our DSNet comprises a dual-branch encoder and a dual-branch decoder. The encoder consists of a temporal encoding branch (TEB) and a spatial encoding branch (SEB). The TEB is designed to capture dynamic vessel flow information and the SEB is utilized to extract static vessel structure information. To effectively capture the correlations among sequential frames, a dynamic frame reweighting module is designed to adjust the weights of the frames. In bottleneck, we exploit a spatio-temporal feature alignment (STFA) module to fuse the features from the encoder to achieve a more comprehensive vascular representation. Moreover, DSNet employs unsupervised loss for consistency regularization between the dual output from the decoder during training. Experimental results demonstrate that DSNet outperforms existing methods, achieving a Dice score of 89.34% for cerebrovascular segmentation.

Keywords: Cerebrovascular Segmentation · DSA Sequence · Spatio-Temporal

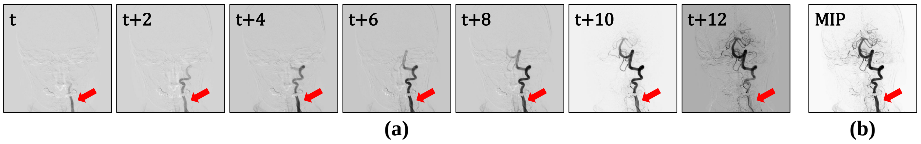

Fig. 1. Schematic diagram of the DSA sequence. (a) illustrates a DSA sequence depicting the angiography process. (b) presents the MIP image derived from the DSA sequence. Red arrows indicate the contrast agent from appearance to dissipation.

1 Introduction

Cerebrovascular diseases (CVDs) are major causes of global mortality and disability, causing profound physical and psychological anguish upon patients and imposing substantial financial strains on families [1]. These conditions mainly result from abnormalities in cerebrovascular structures. For instance, occlusion or stenosis of the cerebral arteries precipitates stroke, while abnormal dilation of cerebral arteries may result in aneurysms. Digital Subtraction Angiography (DSA) comprises a series of sequential 2D frames captured over time, as demonstrated in Fig. 1-(a). DSA has emerged as an indispensable technique for dynamic imaging of cerebrovascular structure and precise revealing of lesion details, owing to its inherent high spatial and temporal resolution capabilities [2]. Nevertheless, the interpretation of DSA images predominantly relies on visual assessment by radiologists, a process characterized by its time-consuming and labor-intensive nature. Furthermore, junior clinicians with limited expertise are susceptible to the risks of misdiagnosis and underdiagnosis. Therefore, pursuing the fully automated, high-precision segmentation of cerebrovascular structure from DSA sequences is important. This advancement can assist clinicians in obtaining complete and valuable cerebral vascular structures and quantifying pathological changes to support clinical diagnosis and treatment of CVDs.

The rapid developments in deep learning have motivated several studies on DSA cerebrovascular segmentation. Zhang *et al.* [3] firstly proposed a straightforward U-shaped network for cerebrovascular segmentation in single-frame DSA images. To segment cerebrovascular structures with different diameters, Meng *et al.* [4] introduced a segmentation framework called MDCNN, utilizing convolutional neural network (CNN) with multi-scale dense connections, along with an entropy-sampled patch method for segmenting single-frame DSA images. Xu *et al.* [5] proposed an edge regularization network (ERNet) for segmenting cerebrovascular structures in DSA images. ERNet uses erosion and dilation processes on the initial binary annotations to create pseudo-ground-truths for false negative and false positive cases. In addition, to alleviate the challenge with limited labeled images, Vepa *et al.* [6] proposed a weakly supervised approach using an active contour model to generate pseudo-labels for cerebrovascular structure segmentation in DSA images. However, whether employing a fully supervised or a weakly supervised method, existing approaches typically rely on training with a single frame selected from DSA sequences.

The DSA sequences (Fig. 1-(a)) present dynamic changes in vessels, while a single frame only captures a portion of the contrast agent. Notably, the angiogram at the bottom of the last frame nearly disappears, while its corresponding minimum intensity projection (MIP) image (Fig. 1-(b)) exhibits good visibility. Hence, conventional 2D DSA segmentation methods, relying solely on single frames from DSA sequences, often fail to fully exploit temporal dynamic information and may inadequately capture cerebrovascular nuances. Additionally, solely considering the DSA sequences as video data for segmentation could lead to information redundancy and resource consumption. To address these limitations, we aim to leverage spatio-temporal information to improve segmentation performance.

In this work, we propose a spatio-temporal consistency network to segment cerebrovascular structures in DSA sequences, namely DSNet. It takes a DSA sequence and its MIP image as input to produce a 2D mask. The proposed method is composed of four main components: a Dual-Branch Encoder (DEB), a Dynamic Frame reWeighting (DFW) module, a Spatio-Temporal Feature Alignment (STFA) module, and a dual-branch decoder. The DEB is designed to extract spatial and temporal features, and DFW is proposed to capture the correlations among sequential frames. To effectively integrate the spatio-temporal information, we propose STFA to fuse the features from the DEB. The main contributions are summarized as follows:

(a) We propose a spatio-temporal consistency network for accurate cerebrovascular segmentation in DSA sequences (DSNet), which simultaneously utilizes dynamic flow and static vessel structure information in DSA sequences.
(b) We design a Dynamic Frame reWeighting module, which dynamically adjusts the weights of features generated by the sequences. This module serves to effectively capture the correlations among sequential frames and mitigate the presence of redundant information.
(c) To comprehensively incorporate spatio-temporal information, we propose a Spatio-Temporal Feature Alignment module, which integrates features extracted by the encoder. Additionally, we employ an unsupervised optimization strategy to enforce spatio-temporal consistency regularization between the dual outputs from the decoder during training.

2 Proposed Method

2.1 Dual-branch encoder

The DEB contains two encoding branches, namely the temporal encoding branch (TEB) and spatial encoding branch (SEB), as demonstrated in Fig. 2. The TEB is designed to capture dynamic vessel flow information and the SEB is utilized to extract static vessel structure information. Each encoding branch is comprised of five residual blocks, with the last four blocks incorporating downsampling operations. Although these branches are structurally identical, they operate independently without weight sharing. The primary distinction between the two

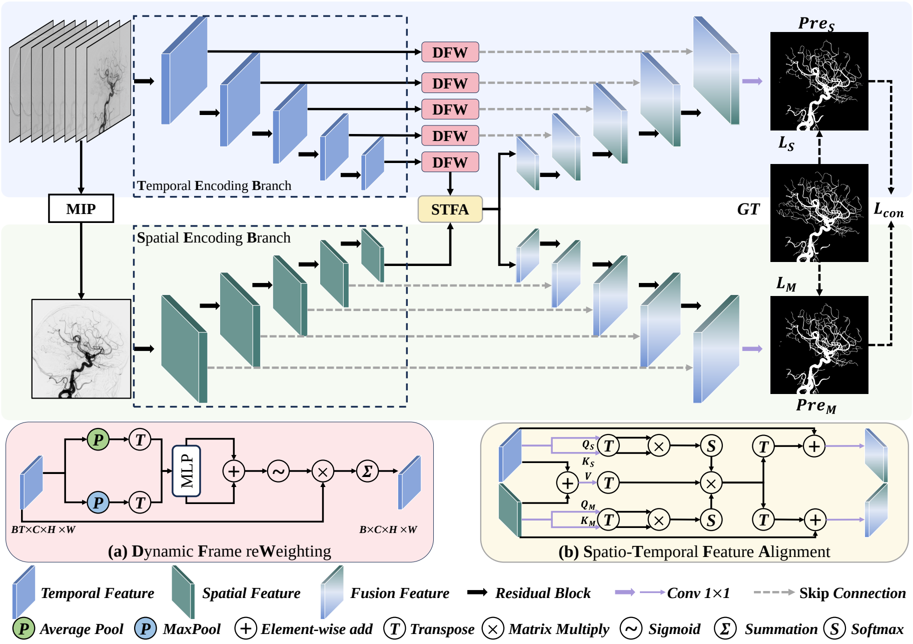

Fig. 2. Schematic diagram of the proposed DSNet, which contains a dual-branch encoder, five dynamic frame reweighting (DFW) modules, a spatio-temporal feature alignment (STFA) module, and a dual-branch decoder.

encoding branches lies in their inputs, with the sequence and its MIP image being fed into the TEB and SEB, respectively.

Temporal encoding branch: Given an input DSA sequence $S \in \mathbb{R}^{B \times C \times T \times H \times W}$, where B, C, T, H, W represent the batch sizes, the number of channels, the number of frames, the height, and the width of S, respectively. The process of temporal encoding has been optimized for training convenience by merging B and T into N, where $N = B \times T$. After being processed by the TEB, the high-dimensional temporal feature can be obtained and denoted as $F_S^l \in \mathbb{R}^{N \times C \times \frac{H}{2^l} \times \frac{W}{2^l}}$, where $l \in \{0,\ 1,\ 2,\ 3,\ 4\}$.

Spatial encoding branch: Similarly, given an input MIP image $M \in \mathbb{R}^{B \times C \times H \times W}$, which is produced by the corresponding DSA sequence S. We can obtain the high-dimensional spatial feature $F_M^l \in \mathbb{R}^{B \times C \times \frac{H}{2^l} \times \frac{W}{2^l}}$.

2.2 Dynamic frame reweighting

It is important to utilize the information on dynamic contrast flow in sequential frames to assist the segmentation of cerebrovascular structures. Furthermore, the feature F_S^l obtained from the TEB cannot be directly used for skip connections, necessitating compression of dimensional T. Therefore, we designed the DFW module, which is located in the skip connection of TEB, to capture the correlations among sequential frames and avoid redundant information. The DFW module can dynamically adjust the weight of feature F_S^l in dimensional T. The structure of DFW is shown in Fig. 2-(a). First, F_S^l is subjected to both max pooling and average pooling operations across the dimensional H and W to get $f_{S_m}^l$ and $f_{S_a}^l \in \mathbb{R}^{N\times C}$. Subsequently, the dimensional N, C is converted to B, C, T, followed by individual passes through a Multi-Layer Perceptron (MLP) block, and then is added to obtain f_S^l. By applying a sigmoid function to f_S^l, dynamic weights are derived, which are subsequently multiplied with F_S^l and aggregated through weighted summation along the dimensional T. Finally we obtain the enhanced feature $\tilde{F}_S^l \in \mathbb{R}^{B\times C\times \frac{H}{2^l}\times \frac{W}{2^l}}$. The complete computational process can be illustrated as follows:

$$f_S^l = \mathcal{W}(T(MP(F_S^l)) + \mathcal{W}(T(AP(F_S^l)), \tag{1}$$

$$\tilde{F}_S^l = \sum(\mathcal{S}(f_S^l) \times F_S^l), \tag{2}$$

where $\mathcal{W}, \mathcal{S}, T, MP$, and AP represent MLP block, sigmoid function, transpose operation, max pooling, and average pooling, respectively.

2.3 Spatio-temporal feature alignment

The high-dimensional temporal feature $\tilde{F}_S^4$ from the last layer of TEB and the high-dimensional spatial feature F_M^4 from the last layer of SEB exhibit certain disparities. Simple addition may potentially result in adverse effects. Therefore, to effectively integrate these two features into complete features, we introduce the STFA module. The STFA module aligns the features from TEB and SEB in the spatial dimension to generate fusion features with both spatial structure and temporal flow information. The STFA module is based on the self-attention mechanism [7], which is a linear combination of self-values to refine the input features. As shown in Fig. 2-(b), $\tilde{F}_S^4$ and F_M^4 are individually processed through 1×1 convolutions to derive Q_s, K_s, Q_m and K_m, respectively. Simultaneously, we combine them and employ another 1×1 convolution to obtain V in the STFA module. Subsequently, the final fusion features can be obtained by computing Q_s, K_s, Q_m, K_m and V using Eq. (3) and Eq. (4).

$$\tilde{f}_S = Softmax(\frac{Q_s \times {K_s}^{\mathtt{T}}}{\sqrt{c}})V, \tilde{f}_M = Softmax(\frac{Q_m \times {K_m}^{\mathtt{T}}}{\sqrt{c}})V, \tag{3}$$

$$\hat{F}_S^4 = Conv_{1\times 1}(\tilde{f}_S + \tilde{F}_S^4), \hat{F}_M^4 = Conv_{1\times 1}(\tilde{f}_M + F_M^4), \tag{4}$$

where $\tilde{f}_S$, $\tilde{f}_M \in \mathbb{R}^{B\times C\times \frac{H}{16}\times \frac{W}{16}}$ represent the aligned features. $\hat{F}_S^4$ and $\hat{F}_M^4 \in \mathbb{R}^{B\times C\times \frac{H}{16}\times \frac{W}{16}}$ represent the fusion features.

2.4 Spatio-temporal consistency regularization

The dual-branch decoder uses fusion features to produce segmentation results. The fusion features $\hat{F}_S^4$ and $\hat{F}_M^4$ from the STFA module are directed into their respective decoding branches to progressively decode and restore the original resolution. As previously explained the sequence contains rich temporal flow information, while the MIP image contains complete spatial structural information. Therefore, each decoding branch pays distinct attention to spatial and temporal information, thereby creating a difference in the fine details of cerebrovascular structure between the predictions of the dual-branch decoder. To utilize the difference, we propose a spatio-temporal consistency regularization (STCR) strategy to optimize the model. Inspired by XNet [8], we employ an unsupervised approach to minimize the dual-output consistency loss. Given $\mathcal{L}_{con}$ as the consistency loss, which is achieved by cross pseudo supervision loss [9]. We use one branch prediction as a pseudo-label to supervise the other branch, and vice versa. $\mathcal{L}_{con}$ can be defined as:

$$\mathcal{L}_{con} = \mathcal{L}_{con}^S(Pre_S, \hat{Pre}_M) + \mathcal{L}_{con}^M(Pre_M, \hat{Pre}_S), \tag{5}$$

where $\hat{Pre}_S$ and $\hat{Pre}_M$ represent the pseudo-label of sequence and MIP image generated by Pre_S and Pre_M, respectively.

In this study, both the supervised loss $\mathcal{L}_S$ and $\mathcal{L}_M$, and the consistency loss $\mathcal{L}_{con}^S$ and $\mathcal{L}_{con}^M$, all use dice loss and cross-entropy loss. The overall loss is formulated as follows:

$$\mathcal{L}_{total} = \mathcal{L}_S + \mathcal{L}_M + \mathcal{L}_{con}^S + \mathcal{L}_{con}^M \tag{6}$$

3 Experimental Results

3.1 Dataset and implementation details

The proposed DSNet was evaluated on a dataset consisting of 70 DSA sequences. We extracted the frames that contain vessels in the arterial phase [10] from the original sequence and registered them together. The registered sequential frames were subjected to a minimum-intensity projection operation to obtain the MIP image. Additionally, to facilitate training, the registered sequential frames were resampled to 8 frames. The MIP images underwent pixel-by-pixel annotation by an experienced radiologist and then reviewed and corrected by another senior radiologist. All DSA sequences described in this paper are from studies approved by institutional ethics committees, and written informed consent was acquired from each participant by the Declaration of Helsinki.

Our method was implemented based on the PyTorch framework with NVIDIA GeForce RTX 3090. Stochastic Gradient Descent (SGD) was employed for model optimization during training, with a momentum of 0.99 and a weight decay of 3e-5 for 300 epochs. The initial learning rate was set to 0.01 with a polynomial decay strategy. The batch sizes were set as 2 and patch sizes were set as 512×512 pixels. During training, common data augmentation methods, including random

Table 1. Performance comparisons for cerebrovascular segmentation.

Methods	Dice(%)	clDice(%)	IoU(%)	SEN(%)	AUC(%)
UNet [11]	84.78±0.99	79.00±1.46	74.04±1.39	79.72±1.87	89.60±0.91
UNet++ [12]	83.28±1.53	78.39±1.89	72.33±1.98	76.46±2.29	88.05±1.13
AttentionUNet [13]	82.56±2.26	76.71±2.45	71.19±2.93	75.71±3.29	87.65±1.63
CSNet [14]	84.96±2.17	80.11±2.58	74.48±3.06	79.90±4.08	89.72±2.00
CENet [15]	81.17±1.58	74.89±2.00	69.08±1.95	73.70±2.40	86.63±1.18
Segformer [16]	79.24±1.24	70.11±1.21	65.95±1.63	73.86±2.31	86.51±1.11
SwinUNet [17]	81.56±0.80	71.24±1.65	69.40±1.13	76.47±1.89	87.90±0.92
TransUNet [18]	86.48±0.74	83.17±0.85	76.51±1.07	87.80±0.80	93.40±0.35
nnUNet [19]	87.45±0.22	83.80±0.04	77.93±0.37	85.98±0.94	92.65±0.45
Ours-DSNet	**89.34**±0.24	**86.26**±0.55	**80.96**±0.40	**88.85**±0.80	**94.10**±0.38

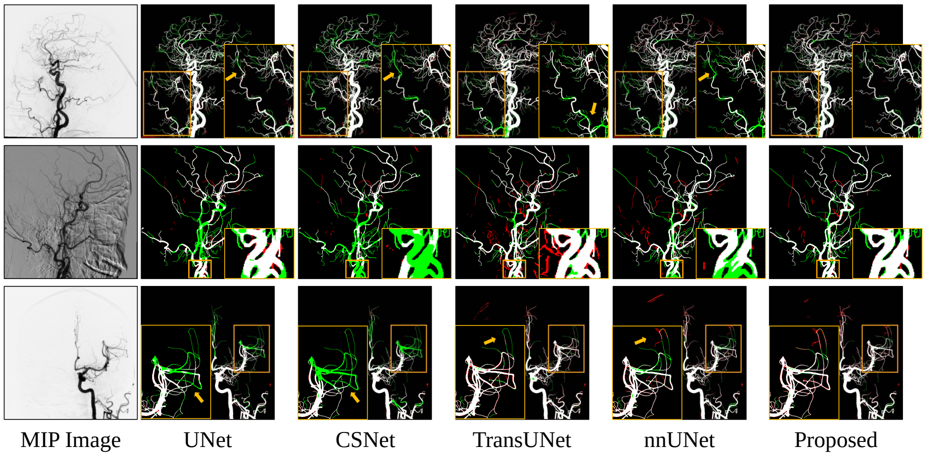

Fig. 3. Qualitative results of cerebrovascular segmentation on MIP images. The withe, green, and red denote the true positive, false negative, and false positive, respectively. Enlarged viewing for better clarity.

rotation, elastic deformation, random scaling, random cropping, gamma augmentation, and mirroring, were applied. We utilized a 5-fold cross-validation approach to evaluate the performance.

3.2 Comparison with the state-of-the-art methods

To demonstrate the superiority of the proposed DSNet, we compared it with several state-of-the-art methods in the medical image segmentation field, including UNet [11], UNet++ [12], AttentionUNet [13], CSNet [14], CENet [15], Segformer [16], SwinUNet [17], TransUNet [18], and nnUNet [19]. We use the Dice coefficient (Dice), clDice [20], intersection over union (IoU), sensitivity (SEN), and area under the ROC curve (AUC) to evaluate the segmentation performance.

Table 2. Ablation results of our DSNet for cerebrovascular segmentation.

Methods	Dice(%)	clDice(%)	IoU(%)	SEN(%)	AUC(%)
Backbone	88.20±0.34	84.81±0.40	79.16±0.54	86.69±0.96	93.03±0.46
Backbone + M1	88.87±0.32	85.74±0.36	80.23±0.51	87.86±0.91	93.62±0.44
Backbone + M2	88.99±0.31	86.12±0.27	80.42±0.50	87.83±0.74	93.61±0.36
Backbone + M3	**89.34**±0.24	**86.26**±0.55	**80.96**±0.40	**88.85**±0.80	**94.10**±0.38

The comparison results are shown in Table 1. Based on quantitative comparison results, we can observe that the proposed method outperforms the other methods. Specifically, our method demonstrates superior performance compared to nnUNet [19], achieving improvements of 1.89% and 3.03%, in Dice and IoU, respectively. DSNet also surpasses TransUnet [18], by 1.05% on SEN and 0.7% on AUC. We further conducted a comparison using the clDice [20] metric, where our method significantly leads with a score of 86.26%. This indicates that our approach performs exceptionally well in maintaining vascular connectivity.

Fig. 3 visualizes the qualitative segmentation results of DSNet with other competitors [19,18,14,11]. Thanks to the spatio-temporal consistency framework, our DSNet can not only preserve vascular connectivity, which is failed in other methods (as indicated by the yellow arrows in the 1^{st} row), but it also can capture small vessels in areas of low contrast (highlighted by yellow arrows in the 3^{rd} row). Moreover, our DSNet demonstrates robust performance by balancing false positives and false negatives in challenging scenarios (2^{nd} row).

3.3 Ablation study

To investigate the effectiveness of all components of the proposed DSNet, we conducted the following ablation studies. We employed a network stripped of all components as the backbone, systematically reintegrating each component in sequence to conduct comprehensive ablation studies. We successively incorporate M1: Dynamic Frame reWeighting, M2: M1 and Spatio-Temporal Feature Alignment, and M3: M2 and Spatio-Temporal Consistency Regularization, into the Backbone. The results for ablation are summarized in Table 2. Compared to the backbone, the network with DFW achieves better performance, with an increase of approximately 0.67%, 0.93%, 1.07%, 1.17%, and 0.59% in Dice, clDice, IoU, SEN, and AUC, respectively. Moreover, integrating STFA and STCR into the network also results in a certain degree of improvement. Overall, by combining the proposed components, our model can achieve superior performance.

4 Conclusion

In this paper, we have proposed a spatio-temporal consistency network for cerebrovascular segmentation in DSA sequences. In contrast to other segmentation methods that only focus on individual frames, our DSNet exploits the temporal dynamical information of sequences and incorporates the complete spatial structural information derived from MIP images to effectively segment cerebrovascular structure. The DSNet can learn the correlations among DSA sequential frames via the dynamic frame reweighting module. Meanwhile, to facilitate the learning and fusion of spatial and temporal features, the DSNet integrates a spatio-temporal feature alignment module with a consistency regularization strategy. Extensive experiments confirmed the effectiveness of our method, demonstrating the potential for practical use in clinical applications.

Acknowledgments. This work was supported in part by the National Science Foundation Program of China (62371442, 62103398, 62272444, U22A2040), in part by the Zhejiang Provincial Natural Science Foundation (LQ23F010002, LZ23F010002, LR24F010002), in part by the Ningbo Natural Science Foundation (2022J143).

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

References

1. Roth, G.A., Mensah, G.A., Johnson, C.O., Addolorato, G., Ammirati, E., Baddour, L.M., Barengo, N.C., Beaton, A.Z., Benjamin, E.J., Benziger, C.P., et al.: Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the gbd 2019 study. Journal of the American College of Cardiology **76**(25) (2020) 2982–3021
2. Hess, C.P.: Imaging in cerebrovascular disease. In: Molecular, Genetic, and Cellular Advances in Cerebrovascular Diseases. World Scientific (2018) 1–23
3. Zhang, M., Zhang, C., Wu, X., Cao, X., Young, G.S., Chen, H., Xu, X.: A neural network approach to segment brain blood vessels in digital subtraction angiography. Computer methods and programs in biomedicine **185** (2020) 105159
4. Meng, C., Sun, K., Guan, S., Wang, Q., Zong, R., Liu, L.: Multiscale dense convolutional neural network for dsa cerebrovascular segmentation. Neurocomputing **373** (2020) 123–134
5. Xu, W., Yang, H., Shi, Y., Tan, T., Liu, W., Pan, X., Deng, Y., Gao, F., Su, R.: Ernet: Edge regularization network for cerebral vessel segmentation in digital subtraction angiography images. IEEE Journal of Biomedical and Health Informatics (2023)
6. Vepa, A., Choi, A., Nakhaei, N., Lee, W., Stier, N., Vu, A., Jenkins, G., Yang, X., Shergill, M., Desphy, M., et al.: Weakly-supervised convolutional neural networks for vessel segmentation in cerebral angiography. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. (2022) 585–594
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
8. Zhou, Y., Huang, J., Wang, C., Song, L., Yang, G.: Xnet: Wavelet-based low and high frequency fusion networks for fully-and semi-supervised semantic seg-

mentation of biomedical images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2023) 21085–21096

9. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2021) 2613–2622
10. Su, R., Cornelissen, S.A., Van Der Sluijs, M., Van Es, A.C., Van Zwam, W.H., Dippel, D.W., Lycklama, G., Van Doormaal, P.J., Niessen, W.J., Van Der Lugt, A., et al.: autotici: Automatic brain tissue reperfusion scoring on 2d dsa images of acute ischemic stroke patients. IEEE transactions on medical imaging **40**(9) (2021) 2380–2391
11. 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, Springer (2015) 234–241
12. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, Springer (2018) 3–11
13. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. Medical image analysis **53** (2019) 197–207
14. Mou, L., Zhao, Y., Chen, L., Cheng, J., Gu, Z., Hao, H., Qi, H., Zheng, Y., Frangi, A., Liu, J.: Cs-net: Channel and spatial attention network for curvilinear structure segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22, Springer (2019) 721–730
15. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: Ce-net: Context encoder network for 2d medical image segmentation. IEEE transactions on medical imaging **38**(10) (2019) 2281–2292
16. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems **34** (2021) 12077–12090
17. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision, Springer (2022) 205–218
18. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
19. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2) (2021) 203–211
20. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2021) 16560–16569