STHP-Net:A Spatio-Temporal Collaborative High-Frequency Purification Network for Intracranial Artery Segmentation in DSA Sequences

Qianwen Zhao[1], Peng Zhang[1], Zeyu Lv[1], Fengbo Xie[1*], Hongxin Dong[1], Pinle Qin[1], Xiaoxia Zhao[2*]

[1]School of Computer Science and Technology, North University of China, Taiyuan, China

[2]Shanxi Key Lab. of Neurobehavior and Brain Disease Control, Shanxi Provincial People's Hospital, Taiyuan, China

Email: Fengbo Xie (20230119@nuc.edu.cn), Xiaoxia Zhao (wsplbb@163.com)

***Abstract*—Digital subtraction angiography (DSA) is regarded as the gold standard for the diagnosis of cerebrovascular diseases owing to its high spatial resolution and dynamic blood flow imaging capability. Accurate segmentation of intracranial arteries is critical for the diagnosis, treatment, and prognosis of ischemic stroke. Nevertheless, existing methods still suffer from several challenges: traditional convolutional attention mechanisms are prone to missing small vessels due to inherent limitations in modeling local features, and current noise suppression strategies can hardly maintain a balance between preserving fine vascular details and eliminating noise at the feature level.To address these issues, we propose a Spatio-Temporal Collaborative High-Frequency Purification Network (STHP-Net). The network first exploits cross-frame complementary information in DSA sequences to maintain the integrity of vascular structures. We then design a Spatial-Channel Combined Attention Module (SCAM) to enhance the weak feature representation of thin vessels. Based on this, a High-Frequency Adaptive Purification Module (HAPM) is introduced to achieve accurate decoupling of vascular details and high-frequency noise in the feature space. Finally, a full-resolution decoding network is employed to generate the segmentation result.Experiments on the public DIAS dataset show that the Dice and clDice coefficient of STHP-Net reach 0.7909 and 0.7386, respectively, significantly outperforming state-of-the-art methods in terms of small vessel detection rate and vascular topological connectivity.**

***Index Terms*—digital subtraction angiography, intracranial artery segmentation, attention mechanism, high-frequency feature purification**

I. Introduction

Ischemic stroke is one of the leading causes of death and disability globally, accounting for approximately eighty-five percent of all stroke cases. Its core etiology is cerebral blood supply deficiency caused by intracranial artery stenosis or occlusion [1]. Precise segmentation of intracranial artery structures and quantification of the extent of vascular lesions are of significant clinical importance for the early diagnosis, treatment decision-making, and prognostic evaluation of ischemic stroke. Digital Subtraction Angiography (DSA), with its high spatial resolution and dynamic blood flow imaging capabilities, can clearly present the vascular structure during the arterial phase and is widely recognized as the gold standard for cerebrovascular imaging [2]. The DSA sequence can reflect the temporal flow process of contrast medium within the blood vessels, with each frame capturing only a small portion of the contrast medium, as shown in Fig. 1. This characteristic helps reduce interference from veins and background tissues, providing a more favorable information basis for precise segmentation of intracranial arteries. However, traditional DSA segmentation relies on visual analysis and integration of vascular information by neuroradiologists frame by frame, which is not only labor-intensive and time-consuming but also susceptible to inter-observer variability, thereby compromising diagnostic consistency. Therefore, the development and application of automated segmentation are particularly crucial [3].

Despite the fruitful achievements in deep learning-driven cerebrovascular segmentation research, there has been a long-term scarcity of specialized research on DSA, with the core crux lying in the lack of high-quality public datasets to support it. Early related research had limitations: Meng et al. [4] proposed a multi-scale dense CNN network, which achieved single-frame DSA intracranial artery segmentation but did not utilize sequence temporal information; Zhang et al. [5] designed a boundary enhancement and feature denoising module to improve the accuracy of vascular boundary segmentation, but their dataset focused on coronary DSA; in the direction of sequence segmentation, Hao et al. [6] developed SVS-Net, which relied on fixed-length input frames and could not adapt to clinically variable-length DSA sequences; Su et al. [7] developed ST U-Net, which achieved DSA cerebral arteriovenous sequence segmentation, but the model had poor scalability. Liu et al. [8] constructed the first dedicated DSA intracranial artery segmentation dataset (DIAS), filling the research gap and providing a standardized benchmark, which facilitated targeted research such as TSI-Net [9], DSNet [10], DSANet [11], and the temporal pre-training method TemSAM [12]. However, existing work still faces core bottlenecks: it is difficult to accurately capture weak signals from small blood vessels, and the feature layer decoupling between noise and vascular details is insufficient, limiting the further improvement of segmentation accuracy.

*Corresponding author.

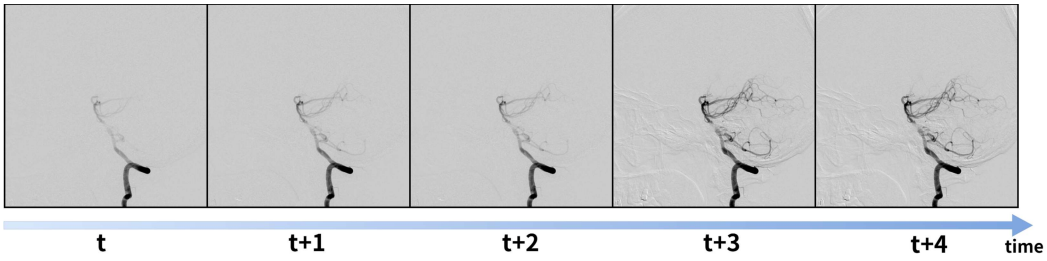


Fig. 1. A DSA sequence, where frames from t to t+4 represent consecutive frames during the arterial phase.

To address the aforementioned challenges, we propose the Spatio-Temporal Collaborative High-Frequency Purification Network (STHP-Net). By deeply mining the temporal correlation information of DSA sequences, it enhances the distinguishability of weak signal vascular features, while accurately distinguishing between noise interference and vascular details, ultimately improving the completeness and accuracy of intracranial artery segmentation. Our contributions can be summarized as follows:

- To overcome the bottleneck of local feature extraction in traditional attention modules, we propose a spatial-channel joint attention mechanism based on coordinate attention, providing an efficient solution for enhancing the weak signals of fine blood vessels and modeling the long-range topological dependencies of blood vessels.
- To address the core contradiction in DSA image segmentation—denoising tends to weaken the signals of small blood vessels while detail preservation easily leaves residual artifacts—we design a high-frequency feature purification strategy that fuses Laplacian pyramid decomposition with KANConv. This strategy effectively alleviates the trade-off between noise suppression and detail preservation.
- We construct a functionally decoupled and flexibly extensible end-to-end framework, and verify that STHP-Net achieves performance improvement compared with existing methods on the public DIAS dataset.

II. Methods

To address the core challenges of weak signals from small vessels and interference from noise artifacts in intracranial artery segmentation of DSA sequences, we propose the Spatio-Temporal Collaborative High-Frequency Purification Network (STHP-Net) and construct an end-to-end segmentation framework consisting of temporal feature modeling, multi-dimensional attention collaborative enhancement, high-frequency feature adaptive purification, and full-resolution decoding. The overall model architecture is shown in Fig. 2.

The core of this framework comprises four key components with complementary functions: the Sequence Feature Extraction Module (SFEM) for capturing the flow patterns of contrast agents, the Spatial-Channel Attention Module (SCAM) for enhancing spatial and channel sensitivity, the High-Frequency Adaptive Purification Module (HAPM) for detail restoration and noise suppression, and the Full-Resolution U-Net (FR-UNet) for ensuring topological continuity. The design principles and implementation methods of each module will be introduced in the following sections.

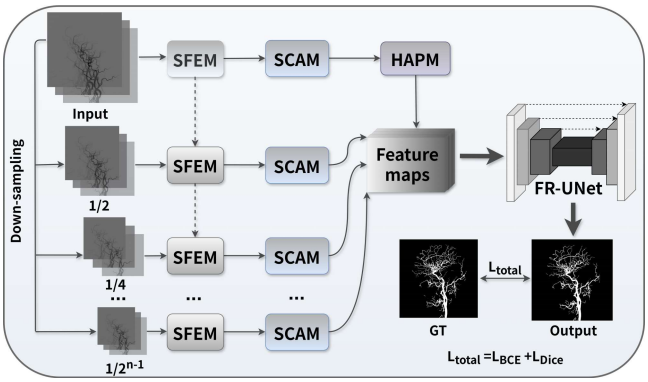


Fig. 2. Overall architecture of STHP-Net.

A. Sequence Feature Extraction Module

In DSA sequences, the contrast agent fills the blood vessels frame by frame. A single frame cannot fully present the vascular tree structure. The lack of modeling dependent on temporal sequence can lead to issues such as missing branches and discontinuous distal details in segmentation results. Based on this imaging characteristic, this study adopts the Sequence Feature Extraction Module (SFEM) proposed by VSS-Net [8] . It captures the spatiotemporal dependencies of contrast agent flow through bidirectional ConvGRU and aggregates complementary information from multiple frames, laying the foundation for the integrity of vascular structures. Although SFEM plays a crucial role in constructing temporal sequence features, its output still has certain limitations. Due to significant differences in development intensity between sequence frames, the temporal fusion process tends to emphasize main vessels with high brightness and coarse structures, while small vessels with weak signals are easily weakened in deep features. Furthermore, global max pooling in the encoding stage further amplifies the responses of salient regions, suppressing fine branch features in the representation, which affects the integrity of subsequent segmentation results.

Based on the aforementioned analysis, our study further designs a feature enhancement mechanism on the basis of SFEM output features to enhance the model's representation ability for weakly responsive vascular regions.

B. Spatial-Channel Attention Module

To address the issue of weak signals from small blood vessels being obscured, SCAM utilizes a synergistic mechanism combining coordinate attention spatial localization with multi-branch channel attention feature selection to enhance weak signal responses and improve feature discriminability. Its structure is illustrated in Fig. 3.

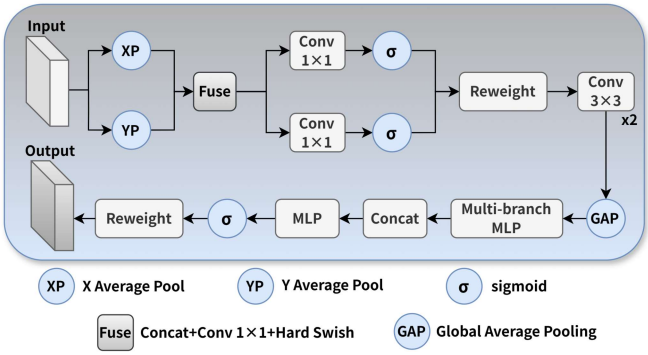


Fig. 3. Structure of Spatial-Channel Attention Module.

Coordinate Attention. Traditional spatial attention mechanisms typically rely on convolution operations to model spatial information, but this approach performs poorly when handling elongated vascular structures. Coordinate Attention [13] decomposes the pooling operation along both horizontal and vertical directions, giving the attention mechanism both directional sensitivity and spatial localization capabilities. This makes it particularly suitable for enhancing structural features such as blood vessels, which are elongated and highly directional. Therefore, this study introduces Coordinate Attention to enhance the model's ability to focus on small vascular regions. To our knowledge, this represents the first use of Coordinate Attention to cerebral vascular segmentation tasks, providing a new and effective approach for enhancing small blood vessels. The specific implementation steps are as follows:

Firstly, the input feature map $X \in \mathbb{R}^{C\times H\times W}$ undergoes one-dimensional global pooling along both horizontal and vertical directions to extract direction-aware features. Its mathematical expression is as follows:

$$Z_c^{\mathrm{h}}(h) = \frac{1}{W}\sum_{i=0}^{W-1} x_c(h,i), \quad Z_c^{\mathrm{w}}(w) = \frac{1}{H}\sum_{j=0}^{H-1} x_c(j,w) \quad (1)$$

where $z_c^h$and z_c^w represent the pooling features in the horizontal and vertical directions, respectively. H and W correspond to the height and width of the feature map, and x_c denotes the c-th channel of the input feature map.

Subsequently, to compress redundant information and retain key directional features, the feature maps from two directions are concatenated along the channel dimension. 1×1 convolution is applied to achieve feature dimensionality reduction and fusion. Meanwhile, an activation function is introduced to enhance nonlinear expression capability. The corresponding transformation process is as follows:

$$f = \delta\left(W_1\left([z_h, z_w]\right)\right) \quad (2)$$

where W_1 denotes a 1×1 convolution operation, δ represents the Hard-Swish activation function.

Next, the concatenated feature map f is split into a horizontal component f_h and a vertical component f_w along the spatial dimension. Attention maps for the horizontal and vertical directions are generated respectively, as follows:

$$\mathbf{g}_h = \sigma\left(W_h(f_h)\right), \quad \mathbf{g}_w = \sigma\left(W_w(f_w)\right) \quad (3)$$

where W_h and W_w are both 1×1 convolutions, σ denotes the Sigmoid activation function, and $\mathbf{g}_h$ and $\mathbf{g}_w$ are the generated bidirectional attention maps.

Ultimately, the bidirectional attention map is applied to the input feature map by multiplying element by element to obtain the output feature map.Coordinate attention primarily enhances the spatial localization capability of features, providing reliable spatial priors for subsequent feature enhancement. Following this, two 3×3 convolution operations are performed. On one hand, this can further refine the direction-aware features output by coordinate attention. On the other hand, with the help of batch normalization and ReLU activation, the nonlinear expression capability of features is enhanced, laying a more robust foundation for subsequent feature selection in the channel dimension.

Channel Attention. Based on spatial localization, the importance of different channels in vascular semantic representation still varies. Thus, weight adaptive modeling in the channel dimension is incorporated to optimize channel responses. The core idea is to enhance the diversity and robustness of channel weight learning through a multi-branch structure, avoiding feature bias and weight distribution deviation caused by a single path. The specific implementation process is as follows:

Firstly, perform global average pooling on the input feature map $X \in \mathbb{R}^{C\times H\times W}$ (C represents the number of channels, H and W represent the dimensions of the feature map) to obtain the channel feature vector y_c:

$$y_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{c,i,j} \quad (4)$$

where $X_{c,i,j}$ corresponds to the feature value of the c-th channel at the spatial position (i,j), and y_c is the global average response of this channel.

To balance expressive power and computational efficiency, four parallel branches are designed with a dimensionality reduction coefficient r. Each branch achieves nonlinear transformation and dimensionality compression through a fully connected layer and ReLU activation, uniformly expressed as:

$$y_k = \mathrm{ReLU}\left(W_k \cdot y\right) \quad (k = 1,2,3,4) \quad (5)$$

where W_k is the weight matrix of the k-th branch, and y is the vector composed of y_c.

Afterwards, the compressed feature vectors from multiple branches are concatenated along the channel dimension. Finally, the channel weights are obtained through a fully connected layer and multiplied element by element with the original input feature map, achieving adaptive weighting in the channel dimension and obtaining an enhanced feature map:

$$\begin{cases} y_{\text{concat}} = \text{Concat}\left(y_1, y_2, y_3, y_4\right) \\ w = \text{Sigmoid}\left(W_{\text{out}} \cdot y_{\text{concat}}\right) \\ X_{ca,c,i,j} = X_{i,j} \odot w_c \end{cases} \tag{6}$$

where W_{out} is the weight matrix of the output layer, w is the channel weight vector, w_c is the weight of the c-th channel, and $\odot$ denotes the pixel-wise multiplication operation.

By adaptively modeling channel weights, this mechanism can highlight channel responses related to vascular structures and suppress redundant features. Combined with the spatial localization ability of coordinate attention, SCAM synergistically focuses on key features in both spatial and channel dimensions, forming a more discriminative representation of vascular features.

C. High-Frequency Adaptive Purification Module

After temporal feature modeling and spatial-channel joint attention enhancement, the localization accuracy and signal significance of the vascular region have been improved. However, the noise components have not been explicitly processed. If directly inputted into the segmentation network, it will not only interfere with the identification of weak signals but also easily cause problems such as boundary blurring and distal branch breakage.

To this end, this study designs a High-Frequency Adaptive Purification Module (HAPM). Based on the precise spatial priors provided by SCAM, this module realizes the enhancement of structural high-frequency components and the selective suppression of noise-induced high-frequency components through an integrated pipeline consisting of multi-scale decomposition, feature adaptive purification, and cross-scale fusion, thus improving the robustness of feature representation while preserving the integrity of fine vascular structures. Its architecture is illustrated in Fig. 4. The feature map is decomposed into high-frequency and low-frequency layers using the Laplacian pyramid, and processed hierarchically using the KANConv module [14]. Finally, cross-scale fusion is employed to achieve complementary enhancement of details and structures. Considering the issue of noise amplification in low-resolution layers, this module is only deployed on the maximum resolution feature map. The specific implementation is as follows:

The input feature map $F_{\text{in}} \in \mathbb{R}^{C\times H\times W}$ is processed via Laplacian pyramid decomposition:

$$\begin{cases} L_0 = G_0(F_{\text{in}}) \\ L_k = G_{k-1}(F_{\text{in}}) - \text{Up}\left(G_k(F_{\text{in}})\right) \quad (k = 1, 2, 3) \end{cases} \tag{7}$$

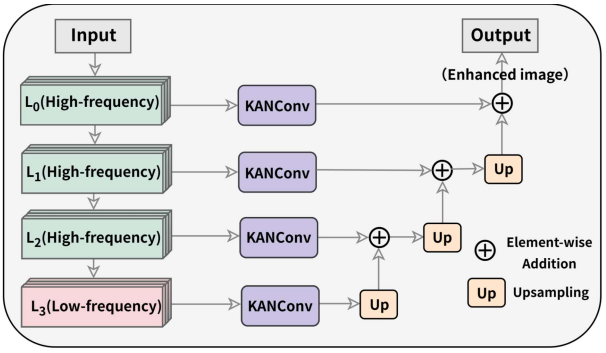


Fig. 4. Structure of High-Frequency Adaptive Purification Module.

where $G_k(\cdot)$ denotes the Gaussian blurring and downsampling operation, and $\text{Up}(\cdot)$ represents the bilinear interpolation upsampling operation. Through the difference operation between Gaussian pyramids of adjacent scales, accurate separation of high and low-frequency components is achieved.

For the decomposed layer features, the KANConv module is introduced for hierarchical adaptive processing. The core advantage of KANConv lies in its ability to adaptively learn complex feature selection rules without the need for manually designed feature templates. Compared to traditional convolution, it is more suitable for distinguishing highly similar high-frequency components such as vascular details and noise.

To avoid feature fragmentation caused by layered processing, the purified low-frequency structural layer L_3' is integrated across scales with the high-frequency detail layers L_0', L_1', and L_2':

$$\begin{cases} F_{2,3} = F_1\left(\text{Concat}\left(\text{Up}(L_3'), L_2'\right)\right) \\ F_{1,2,3} = F_1\left(\text{Concat}\left(\text{Up}(F_{2,3}), L_1'\right)\right) \\ F_{\text{out}} = L_0' \oplus \text{Up}(F_{1,2,3}) \end{cases} \tag{8}$$

where $F_1(\cdot)$ denotes a 1×1 convolution operation, used to integrate cross-scale features and maintain consistent channel dimensions, the symbol $\oplus$ denotes element-wise addition.

D. Full-Resolution Network

After feature purification by HAPM, a vascular feature map with high signal-to-noise ratio and complete details has been obtained. Subsequently, precise pixel-level segmentation needs to be achieved through an efficient decoding network. Referring to the Full-Resolution U-Net (FR-UNet) proposed by Liu et al. [15], its core advantage lies in maintaining high-resolution feature representation throughout the process, effectively alleviating the problem of vascular spatial information loss caused by downsampling in traditional encoder-decoder architectures. It can not only accurately preserve the global topological structure of the main blood vessels but also avoid missed detection or fracture of small blood vessels due to weak signals and compressed spatial information. This is highly aligned with the core goal of this study, which is to achieve complete and precise segmentation of intracranial

arterial structures at all scales. Therefore, it is adopted as the decoding path of the model.

III. Experiments

A. Implementation details and evaluation indicators

This study conducted experiments based on the DIAS dataset, with the data divided into 30 training samples, 10 validation samples, and 20 test samples at the patient level. To adapt to the model input format, all DSA sequences were resampled to a length of 8 frames, and Z-score normalization was performed on each frame image to standardize the data.

The experiment was implemented based on the PyTorch framework, with the model deployed on a single NVIDIA A800 GPU and a total of 20,000 training iterations. The AdamW optimizer was adopted, and the loss function combined cross-entropy loss with Dice loss to balance pixel-level classification accuracy and the overlap of segmented regions, thereby addressing the imbalance between background and vascular samples in the intracranial artery segmentation task.

To comprehensively evaluate the performance of STHP-Net for intracranial artery segmentation, we select the following evaluation metrics: Dice Similarity Coefficient (Dice), Accuracy (Acc), Recall, Specificity (Spe), and Intersection over Union (IOU). Additionally, the connectivity-aware Dice (clDice) [16] is introduced to quantify the coherence of vascular topological structures and specifically evaluate the continuous segmentation performance of slender branches.

B. Comparison experiment

To comprehensively verify the segmentation performance of STHP-Net, we compared it with multiple mainstream advanced methods in the field of medical image segmentation, including UKAN [17], Transunet [18], ConDSeg [19], CS²Net [20], and ST-UNet [7]. The comparison results are shown in Table I.

TABLE I
COMPARISON OF STHP-NET WITH OTHER METHODS

Methods	Dice↑	Acc↑	Recall↑	Spe↑	IOU↑	clDice↑
UKAN[17]	0.6990	0.9546	0.6899	0.9765	0.5373	0.6409
TransUNet[18]	0.7138	0.9603	0.7024	0.9848	0.5629	0.6597
Condseg[19]	0.7293	0.9600	0.7041	0.9812	0.5739	0.6634
CS^2Net[20]	0.7589	0.9659	0.7250	**0.9865**	0.6213	0.6831
ST-UNet[7]	0.7621	0.9657	0.7241	0.9860	0.6238	0.6956
STHP-Net	**0.7909**	**0.9689**	**0.7879**	0.9855	**0.6541**	**0.7386**

Quantitative comparisons presented in Table I reveal that the proposed method outperforms other methods. Specifically, both the Dice coefficient and IOU of STHP-Net are higher than those of ST-UNet and CS²-Net. Notably, the clDice coefficient reaches 0.7386, significantly surpassing all comparative methods, fully demonstrating the model's advantage in ensuring vascular topological connectivity.

Fig. 5 illustrates the qualitative segmentation results of different models on the DIAS dataset, visually demonstrating the performance advantages of STHP-Net over other methods. From the overall segmentation results, the output of the proposed STHP-Net exhibits superior conformity to the vascular morphology of the gold standard (GT). The enlarged areas corresponding to the red and green boxes in the figure more clearly reveal the differences in fine structure restoration among various methods. STHP-Net can more comprehensively restore the details of vascular branches, effectively mitigating issues such as vascular discontinuity and detail loss commonly observed in other models, fully substantiating its superior performance in intracranial artery segmentation tasks.

C. Ablation experiment

To verify the effectiveness of the modules proposed in this study in the task of intracranial artery segmentation, we conducted ablation experiments to assess the contribution of each component to the model's performance. The experiments used VSS-Net as the baseline model, gradually adding the Spatial-Channel Attention Module (SCAM) and the High-Frequency Adaptive Purification Module (HAPM) to compare the impact of each module. All experiments were conducted on the same dataset and with the same hyperparameter configuration to ensure experimental fairness. Table II presents the experimental results under different network configurations.

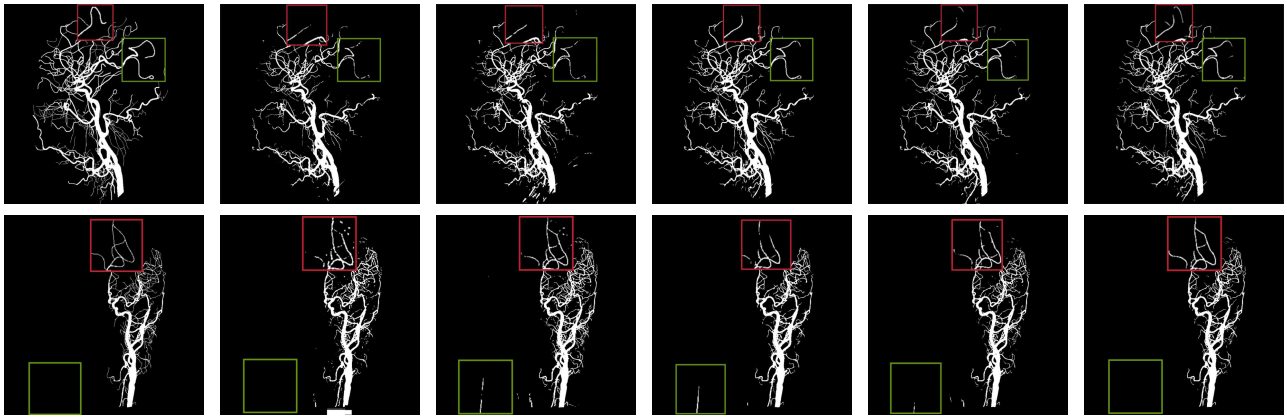
TABLE II
ABLATION STUDY OF OUR MODEL

Methods	Metrics					
	Dice↑	Acc↑	Recall↑	Spe↑	IOU↑	clDice↑
baseline	0.7633	0.9658	0.7323	**0.9893**	0.6242	0.6879
w/o HAPM	0.7790	0.9675	0.7481	0.9856	0.6380	0.7095
w/o SCAM	0.7826	0.7678	0.7591	0.9849	0.6429	0.7140
ours	**0.7909**	**0.9689**	**0.7879**	0.9855	**0.6541**	**0.7386**

The experimental results show that introducing SCAM or HAPM alone can both improve performance, with the Dice coefficient increasing to 0.7790 and 0.7826, respectively. When the two are combined, all indicators reach their optimal values, with the Dice coefficient and clDice coefficient rising to 0.7909 and 0.7386, respectively. The precise spatial localization provided by SCAM effectively reduces the burden of invalid computations in HAPM, while the high-frequency detail enhancement provided by HAPM further amplifies the localization effect of SCAM. The two complement each other, forming a synergistic effect of "spatial focusing + detail enhancement", ultimately achieving comprehensive improvement in all indicators.

IV. Conclusion

Focusing on the clinical requirements for intracranial artery segmentation from DSA sequences and addressing core challenges including the poor detectability of small vessels and severe interference from noise artifacts, we propose an innovative four-stage technical approach consisting of temporal aggregation, weak signal enhancement, high-frequency decoupling, and full-resolution decoding, and construct a Spatiotemporal Collaborative High-Frequency Purification Net-



(a) GT (b) TransUNet (c) Condseg (d) CS^{2}Net (e) ST-UNet (f) STHP-Net

Fig. 5. Qualitative comparison results of intracranial artery segmentation on the DIAS dataset. (a) represents the gold standard (GT), and (b)-(f) show the segmentation results of different models. The red and green boxed areas in the figure are enlarged locally to clearly demonstrate the segmentation details differences between various methods.

work (STHP-Net). Experimental results show that the proposed network achieves a synergistic improvement in segmentation accuracy and detail integrity on the DIAS dataset. Specifically, the spatial-channel joint enhancement mechanism of SCAM and the high-frequency feature decoupling strategy of HAPM yield complementary advantages, providing an optimization paradigm of spatial focusing followed by detail enhancement for similar vascular segmentation tasks. In the future, we will focus on improving the robustness of the model and incorporate semi-supervised/weakly supervised learning methods to further enhance the clinical adaptability and practicality of the algorithm.

REFERENCES

[1] X. Li, Y. He, D. Wang, and M. J. Rezaei, "Stroke rehabilitation: from diagnosis to therapy," *Frontiers in neurology*, vol. 15, p. 1402729, 2024.

[2] S. Shaban, B. Huasen, A. Haridas, M. Killingsworth, J. Worthington, P. Jabbour, and S. M. M. Bhaskar, "Digital subtraction angiography in cerebrovascular disease: current practice and perspectives on diagnosis, acute treatment and prognosis," *Acta Neurologica Belgica*, vol. 122, no. 3, pp. 763–780, 2022.

[3] H. Li, K. Song, P. Zhang, H. Yu, and H. Bian, "Vmc-unet: A u-shaped structure combining mamba and cnn for medical image segmentation," in *2025 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2025.

[4] C. Meng, K. Sun, S. Guan, Q. Wang, R. Zong, and L. Liu, "Multiscale dense convolutional neural network for dsa cerebrovascular segmentation," *Neurocomputing*, vol. 373, pp. 123–134, 2020.

[5] M. Zhang, F. Yu, J. Zhao, L. Zhang, and Q. Li, "Befd: Boundary enhancement and feature denoising for vessel segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 775–785, Springer, 2020.

[6] D. Hao, S. Ding, L. Qiu, Y. Lv, B. Fei, Y. Zhu, and B. Qin, "Sequential vessel segmentation via deep channel attention network," *Neural Networks*, vol. 128, pp. 172–187, 2020.

[7] R. Su, M. van der Sluijs, S. Cornelissen, P. J. van Doormaal, R. Broek, W. van Zwam, J. Hofmeijer, D. Ruijters, W. Niessen, A. van der Lugt, *et al.*, "Spatio-temporal unet for cerebral artery and vein segmentation in digital subtraction angiography," *arXiv preprint arXiv:2208.02355*, 2022.

[8] W. Liu, T. Tian, L. Wang, W. Xu, L. Li, H. Li, W. Zhao, S. Tian, X. Pan, Y. Deng, *et al.*, "Dias: A dataset and benchmark for intracranial artery segmentation in dsa sequences," *Medical Image Analysis*, vol. 97, p. 103247, 2024.

[9] L. Wang, W. Liu, W. Xu, H. Li, H. Yang, and F. Gao, "Tsi-net: A timing sequence image segmentation network for intracranial artery segmentation in digital subtraction angiography," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, IEEE, 2024.

[10] Q. Xie, D. Zhang, L. Mou, S. Wang, Y. Zhao, M. Guo, and J. Zhang, "Dsnet: A spatio-temporal consistency network for cerebrovascular segmentation in digital subtraction angiography sequences," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 199–208, Springer, 2024.

[11] J. Zhang, Q. Xie, L. Mou, D. Zhang, D. Chen, C. Shan, Y. Zhao, R. Su, and M. Guo, "Dsca: A digital subtraction angiography sequence dataset and spatio-temporal model for cerebral artery segmentation," *IEEE Transactions on Medical Imaging*, 2025.

[12] L. Zhang, X. Jiang, X. Ding, Z. Huang, T. Zhao, and X. Yang, "Temsam: Temporal-aware segment anything model for cerebrovascular segmentation in digital subtraction angiography sequences," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 609–618, Springer, 2025.

[13] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13713–13722, 2021.

[14] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "Kan: Kolmogorov-arnold networks," *arXiv preprint arXiv:2404.19756*, 2024.

[15] W. Liu, H. Yang, T. Tian, Z. Cao, X. Pan, W. Xu, Y. Jin, and F. Gao, "Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation," *IEEE journal of biomedical and health informatics*, vol. 26, no. 9, pp. 4623–4634, 2022.

[16] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. Pluim, U. Bauer, and B. H. Menze, "cldice-a novel topology-preserving loss function for tubular structure segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16560–16569, 2021.

[17] A. Moradzadeh, L. Wawrzyniak, M. Macklin, and S. G. Paliwal, "Ukan: Unbound kolmogorov-arnold network accompanied with accelerated library," *arXiv preprint arXiv:2408.11200*, 2024.

[18] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[19] M. Lei, H. Wu, X. Lv, and X. Wang, "Condseg: A general medical image segmentation framework via contrast-driven feature enhancement," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 4571–4579, 2025.

[20] L. Mou, Y. Zhao, H. Fu, Y. Liu, J. Cheng, Y. Zheng, P. Su, J. Yang, L. Chen, A. F. Frangi, *et al.*, "Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging," *Medical image analysis*, vol. 67, p. 101874, 2021.