Temporal Phase-Difference Guided Spatiotemporal Learning for DSA Vessel Segmentation

Kun Liu[1,⋆], Ziyang He[1,⋆], Bin Zheng[1], Wenyi Zhao[1], Mengke Zhu[2], Weijin Xu[3], Wentao Liu[4], Zijun He[5], Hanlin Chen[6], Bofeng Lu[1], Zhiyuan Liang[1], and Huihua Yang[1](✉)

[1] School of Intelligent Engineering and Automation, Beijing University of Posts and Telecommunications, Beijing, China
[2] State Key Laboratory of Space Information System and Integrated Application, Beijing Institute of Satellite Information Engineering, Beijing, China
[3] Department of Intelligent Police Technology, Beijing Police College, Beijing, China
[4] Dynamic Products Department, Wandong Medical Technology Co., Ltd., Beijing, China
[5] Department of Interventional Neuroradiology, Beijing Tiantan Hospital, Beijing, China
[6] Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China
yhh@bupt.edu.cn

Abstract. Vessel segmentation in digital subtraction angiography (DSA) is crucial for intraoperative visualization and quantitative assessment during endovascular interventions. However, accurate segmentation remains challenging due to complex thin vascular branches and strong background interference from residual anatomical structures after subtraction. Although recent methods incorporate temporal modeling using CNNs, recurrent networks, or transformers, they often treat temporal information implicitly and fail to explicitly exploit the contrast flow dynamics in DSA sequences. In this work, we present a spatiotemporal DSA vessel segmentation framework that combines a forward phase-difference prior with Mamba-based temporal sequence modeling. The Mamba-based temporal encoder models long-range temporal dependencies and contrast flow dynamics with linear computational complexity. To capture contrast-driven intensity changes, we introduce a forward phase-difference projection that enhances contrast-responsive vascular structures while suppressing temporally stable background tissues. A minimum intensity projection (MinIP) is further used as a complementary static spatial prior to provide coarse vessel structural cues. These priors are integrated with temporal features through a lightweight domain-aware fusion module before decoding. Extensive experiments on the DIAS and DSCA datasets demonstrate that the proposed method consistently outperforms state-of-the-art approaches, particularly in preserving vascular connectivity and segmenting thin distal vessels. Code is available at https://github.com/kkk123lll/FPD-SegDSA.

⋆ K. Liu and Z. He contributed equally to this work.

Keywords: Digital Subtraction Angiography · Vessel Segmentation · Spatiotemporal Modeling · Forward Phase-Difference Prior

1 Introduction

Digital subtraction angiography (DSA) is the clinical gold standard for cerebral vascular assessment [8,9], providing dynamic visualization of contrast-enhanced vessels during endovascular procedures. Accurate vessel segmentation from DSA sequences is essential for quantitative analysis and computer-assisted intervention. However, automated DSA vessel segmentation remains challenging due to the presence of numerous thin vascular branches and strong background interference from residual soft tissue and bony structures after subtraction. These challenges highlight the need to effectively leverage both spatial appearance and temporal dynamics inherent in DSA sequences.

Early DSA vessel segmentation methods primarily focus on single-frame images and rely on CNN-based U-shaped architectures to model spatial features [4,7,12], generally overlooking the inherent temporal characteristics of DSA sequences. To incorporate temporal information, DIAS [5] introduced a benchmark dataset for intracranial artery segmentation in DSA sequences, together with a baseline method based on recurrent neural networks. Subsequent studies have further explored sequence modeling architectures, including Transformers and LSTMs, to learn long-range temporal relationships in DSA data [11]. Despite these advances, most existing methods fail to fully exploit the dynamic flow characteristics of contrast agents that fundamentally characterizes DSA imaging. Moreover, directly treating DSA sequences as volumetric (3D) data may be suboptimal, as the temporal dimension in DSA reflects asymmetric propagation of contrast agents rather than spatial continuity [5]. Several recent approaches therefore adopt explicit spatio-temporal fusion strategies [14,17], extracting temporal features from raw sequences while incorporating complementary spatial priors from minimum intensity projections (MinIP). More recently, SAM-based approaches have been explored for joint spatial–temporal modeling in DSA vessel segmentation [18].

However, from a clinical imaging perspective, vessel structures in DSA are primarily revealed through contrast agent propagation over time. Vessel regions exhibit pronounced intensity changes between early and late phases, whereas most background tissues and bony structures remain largely unchanged after subtraction, a property routinely exploited during clinical interpretation. As illustrated in Fig. 1, modeling phase-wise temporal intensity differences enables effective characterization of contrast agent dynamics across DSA sequences that are barely visible in MinIP. Despite its interpretability, such contrast-induced temporal differences are rarely modeled explicitly in existing DSA vessel segmentation methods.

Motivated by this observation, we propose a forward phase-difference prior that encodes contrast-induced temporal intensity variations into a spatial structural cue, enhancing contrast-responsive vascular regions while suppressing per-

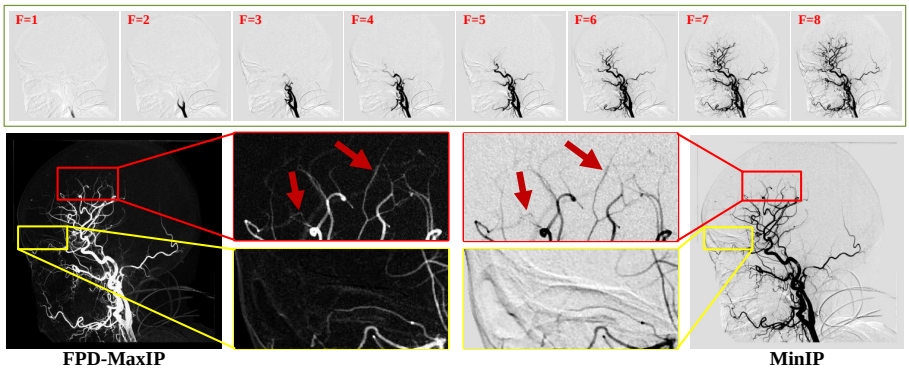

Fig. 1. Visual comparison between FPD-MaxIP and MinIP. The top row shows representative DSA frames across temporal phases (F=1–8). FPD-MaxIP suppresses background structures (yellow box) while enhancing distal vascular details (red box). Compared with MinIP, FPD-MaxIP reduces non-vascular background signals, including soft tissue and bone, and improves the visibility of low-contrast distal microvessels.

sistent tissue and bony artifacts via maximum intensity projection(FPD-MaxIP). Building upon this prior, we develop a unified spatiotemporal segmentation framework that integrates Mamba-based temporal modeling [13, 15, 20] with a lightweight domain-aware fusion module to combine dynamic temporal features and static spatial priors. By decoupling temporal modeling from spatial prior guidance, the resulting framework enables robust and interpretable integration of blood flow dynamics and vessel morphology, achieving state-of-the-art performance on two public DSA datasets with a favorable accuracy-efficiency trade-off.

The main contributions of this work are as follows:

1. We introduce a phase-difference-based spatial prior for DSA vessel segmentation that explicitly encodes contrast-induced temporal intensity variations and enhances contrast-responsive vessel structures.
2. We develop a spatiotemporal segmentation approach that integrates the proposed spatial prior with temporal sequence modeling, enabling effective combination of blood flow dynamics and vessel morphology.
3. We validate the proposed method on DIAS and DSCA datasets and demonstrate consistent improvements in segmentation accuracy, vascular connectivity, and overall performance across both DSA benchmarks compared with existing methods.

2 Method

2.1 Overall Architecture

As illustrated in Fig. 2, we propose a U-shaped spatiotemporal vessel segmentation framework that decouples temporal dynamics modeling from spatial prior

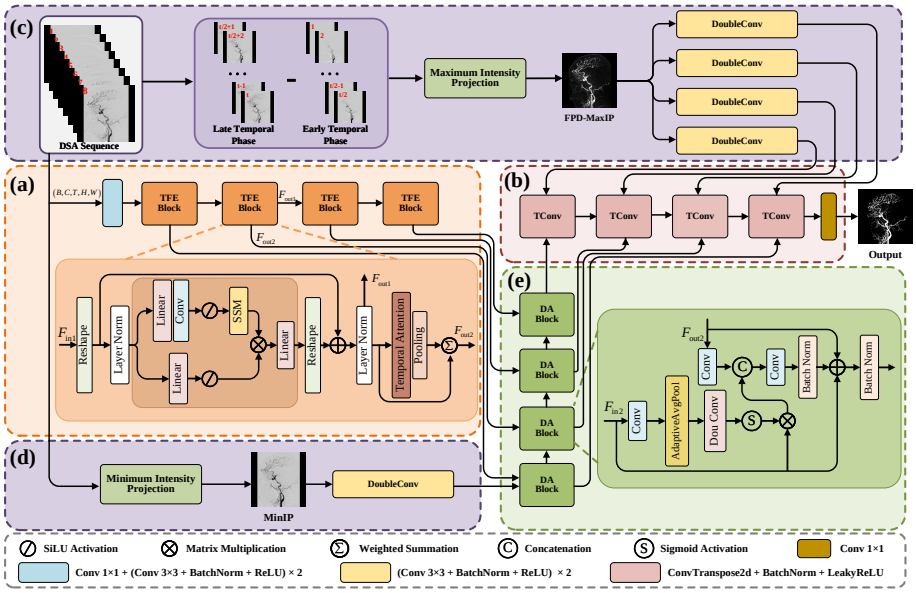

Fig. 2. Overview of the proposed framework, consisting of (a) a Mamba-based temporal encoder, (b) a lightweight decoder, (c) a forward phase-difference MaxIP (FPD-MaxIP) prior, (d) a static vessel MinIP prior, and (e) a domain-aware fusion module.

guidance. More specifically, each DSA clip is first represented as a quasi-3D tensor $x \in \mathbb{R}^{B\times C\times T\times H\times W}$, which is reshaped into a flattened spatiotemporal sequence $\mathbf{X} \in \mathbb{R}^{B\times N\times C}$ with $N = T\cdot H\cdot W$ for temporal modeling. Before being fed into the temporal encoder, the input features are processed by a shallow convolutional stem for initial feature projection. The resulting feature sequence is then processed by a temporal encoder composed of stacked Temporal Feature Encoding (TFE) blocks. At the l-th stage, each TFE block outputs a spatiotemporal feature map $\mathbf{F}_l \in \mathbb{R}^{B\times C_l\times T\times H_l\times W_l}$ along with a scale-aligned 2D representation $\mathbf{S}_l \in \mathbb{R}^{B\times C_l\times H_l\times W_l}$. The multi-level 2D representations $\{\mathbf{S}_l\}$ are subsequently fused with static MinIP priors via a domain-aware fusion module, thereby integrating temporal semantics and spatial vessel cues. The fused features are then forwarded to the decoder as scale-aligned inputs $\{\tilde{\mathbf{S}}_l \in \mathbb{R}^{B\times C_l\times H_l\times W_l}\}$, while FPD-MaxIP features are injected only at the corresponding decoder scales to guide structural refinement. Finally, the decoder progressively upsamples the fused representations to produce the vessel prediction $\hat{y} \in \mathbb{R}^{B\times 1\times H\times W}$.

2.2 Mamba-based Temporal Encoder

DSA sequences exhibit long-range temporal dependencies induced by contrast agent propagation, motivating efficient temporal modeling over high-dimensional spatiotemporal inputs. We therefore adopt a Mamba-based temporal encoder

with stacked TFE blocks that leverages selective state space modeling to capture long-range temporal dependencies with linear complexity, as illustrated in Fig. 2(a). The first TFE block preserves the input feature resolution to maintain fine-grained spatial details during early temporal modeling, while subsequent TFE blocks progressively construct multi-scale representations by doubling the channel dimension and halving the spatial resolution. Within each TFE block, spatiotemporal features are modeled using a Mamba-based selective state space model (SSM), whose state update is given by

$$s_n = A_n s_{n-1} + B_n z_n, \quad \hat{z}_n = C_n s_n. \tag{1}$$

where s_n denotes the hidden state and z_n, $\hat{z}_n$ are the input and output tokens, respectively. A residual connection is applied to refine temporal representations, yielding a spatiotemporal feature map $\mathbf{F}_l$, which is forwarded to the next TFE block for hierarchical temporal modeling. To obtain a compact 2D representation $\mathbf{S}_l$ for subsequent fusion and decoding, we apply temporal attention pooling to aggregate features along the temporal dimension at l-th stage:

$$a_t = \mathrm{Conv}_2\big(\sigma(\mathrm{Conv}_1(\mathrm{AvgPool}_{H,W}(\mathbf{F}_l(t))))\big) \tag{2}$$

$$\mathbf{S}_l = \sum_{t=1}^{T} \mathrm{Softmax}(a_t) \cdot \mathbf{F}_l(t), \tag{3}$$

where a_t denotes the attention weight of the t-th frame estimated from its global spatial context.

2.3 Spatial Prior Guidance Module

While temporal modeling captures dynamic flow patterns in DSA sequences, accurate vessel segmentation also requires explicit spatial structural cues, particularly for thin vessels and low-contrast regions. We therefore introduce a spatial prior guidance module to provide explicit spatial cues during decoding, as illustrated in Fig. 2(c). This module is centered on a forward phase-difference prior to highlight contrast-induced vessel structures, while a simple MinIP-based vessel prior $\mathbf{P}_l^{\min}$ is incorporated as auxiliary static guidance (Fig. 2(d)).

Forward Phase-Difference Prior. Although the MinIP prior provides stable structural cues, it is less sensitive to transient contrast changes in DSA sequences. We therefore introduce a forward phase-difference projection module as a dynamic spatial prior to highlight contrast-induced intensity variations between early and late phases. The resulting phase-difference prior enhances weak vessel responses without introducing additional temporal modeling overhead. Given an input DSA sequence x, the forward phase-difference between early and late frames is computed as

$$\Delta x_t = x_{t+T/2} - x_t, \quad t = 1, 2, \ldots, T/2. \tag{4}$$

A maximum intensity projection is then applied along the temporal dimension to obtain the spatial prior

$$\mathbf{P}_l^{\max}(h_l, w_l) = \max_{t=1,\ldots,T/2} \Delta x_t(h_l, w_l). \tag{5}$$

To provide multi-scale guidance consistent with the decoder hierarchy, the phase-difference prior is processed by *DoubleConv* blocks with different strides to generate scale-aligned features $\mathbf{P}_l^{\max}$, which are injected into the decoder at corresponding l stage to refine vessel reconstruction.

2.4 Domain-aware Fusion Module

As shown in Fig. 2(e), we further introduce a lightweight domain-aware fusion module to combine temporal features $\mathbf{S}_l$ with static spatial priors $\mathbf{P}_l^{\min}$. As the two feature streams are derived from different sources, naive fusion may introduce redundant or irrelevant responses. To mitigate this issue, we apply a channel-wise fusion scheme. Both inputs are first projected to the same channel dimension using 1×1 convolutions. The spatial prior feature $\mathbf{P}_l^{\min}$ is then modulated by a lightweight channel attention mechanism to emphasize vessel-related responses. The modulated prior is concatenated with the temporal feature and fused through a convolutional layer, followed by a residual connection. The fusion operation is formulated as:

$$\tilde{\mathbf{S}}_l = BN\big(\mathcal{F}_{\text{fuse}}\big[\mathcal{A}(\mathbf{S}_l) \,\|\, \mathcal{A}(\mathbf{P}_l^{\min}) \odot \mathcal{W}(\mathcal{A}(\mathbf{P}_l^{\min}))\big] + \mathbf{S}_l + \mathbf{P}_l^{\min}\big)\,, \tag{6}$$

where $\mathcal{A}(\cdot)$ denotes channel projection, $\mathcal{W}(\cdot)$ denotes channel attention, $\mathcal{F}_{\text{fuse}}(\cdot)$ denotes a fusion convolution, and $BN(\cdot)$ denotes batch normalization. The fused features are forwarded to the decoder to guide vessel reconstruction. Thus, at each decoder stage, the transposed convolution ($TConv$) takes the fused feature $\tilde{\mathbf{S}}_l$ as input, together with the corresponding phase-difference prior $\mathbf{P}_l^{\max}$, enabling scale-consistent vessel reconstruction guided by both static structure and dynamic contrast.

3 Experiments

Datasets. We evaluated our method on two DSA sequence datasets for cerebral artery segmentation. The DIAS [5] dataset consists of 60 expert-validated sequences with pixel-level annotations, each with a spatio-temporal resolution of $8 \times 800 \times 800$. The DSCA [17] dataset contains 224 annotated DSA sequences, all uniformly resampled to the same resolution. Following the standard data split, 180 sequences were used for training and the remaining 44 for testing.
Training Details. For fair comparison, all models were trained and evaluated using identical data splits, preprocessing, and training settings. During training, input sequences were randomly cropped into 64×64 patches with a batch size of 64. Model parameters were optimized using AdamW, with momentum

Table 1. Performance comparison of different methods on the DIAS and DSCA datasets.

Models	DSC		Acc		IoU		AUC		Pre		clDice		Params	FLOPs
	DIAS	DSCA	DIAS	DSCA	DIAS	DSCA	DIAS	DSCA	DIAS	DSCA	DIAS	DSCA		
UNet [10]	0.6502	0.8867	0.9266	0.9885	0.5153	0.8009	0.9300	0.9960	0.8475	0.9263	0.5772	0.8595	31.04M	3.43G
UNet++ [19]	0.7368	0.8861	0.9620	0.9884	0.5875	0.8001	0.9751	0.9957	0.8068	0.9204	0.6434	0.8607	26.17M	37.89G
SVS-Net [1]	0.6727	0.8649	0.9373	0.9865	0.5260	0.7670	0.9561	0.9945	0.6906	0.9102	0.6291	0.8311	8.48M	0.43G
PSC [3]	0.7360	0.8822	0.9608	0.9878	0.5869	0.7934	0.9722	0.9956	0.8070	0.9207	0.6520	0.8573	6.43M	4.34G
FR-UNet [6]	0.7502	0.8846	0.9643	0.9881	0.6041	0.7972	0.9777	0.9962	0.8255	0.9080	0.6560	0.8598	5.15M	20.11G
H2Former [2]	0.6766	0.8807	0.9589	0.9879	0.5332	0.7913	0.9656	0.9946	0.8393	0.9191	0.5742	0.8569	33.67M	2.09G
VSS-Net [5]	0.7687	0.8931	0.9666	0.9891	0.6269	0.8110	0.9811	**0.9970**	**0.8515**	0.9164	0.6816	0.8742	18.52M	21.50G
CAVE [11]	0.7646	0.8971	0.9661	0.9894	0.6229	0.8179	0.9787	0.9968	0.8343	0.9282	0.6671	0.8771	20.93M	10.05G
CCViM [21]	0.7127	0.8602	0.9576	0.9860	0.5637	0.7595	0.9640	0.9946	0.8056	0.9159	0.6133	0.8268	32.17M	0.29G
VSNet [16]	0.7677	0.8841	0.9657	0.9886	0.6257	0.7967	0.9791	0.9963	0.8160	**0.9284**	0.6829	0.8583	2.40M	20.64G
Ours	**0.7845**	**0.8990**	**0.9678**	**0.9895**	**0.6472**	**0.8200**	**0.9819**	**0.9970**	0.8212	0.9162	**0.7054**	**0.8819**	11.69M	6.47G

Table 2. Ablation study on projection-based spatial priors on the DIAS dataset.

Backbone	MinIP	MaxIP*	DSC	Acc	Sen	IoU	AUC	Pre	clDice	Params	FLOPs
Mamba	-	-	0.7718	0.9661	0.7405	0.6305	0.9816	0.8149	0.6896	8.95M	5.91G
	✓	-	0.7747	0.9665	0.7443	0.6346	0.9804	0.8208	0.6944	9.65M	6.01G
	-	✓	0.7735	0.9673	0.7240	0.6329	0.9802	**0.8402**	0.6811	11.00M	6.34G
	✓	✓	**0.7845**	**0.9678**	**0.7569**	**0.6472**	**0.9819**	0.8212	**0.7054**	11.69M	6.47G
Transformer	-	-	0.7626	0.9618	0.7381	0.6237	0.9750	0.8057	0.6925	9.86M	8.03G
	✓	-	0.7793	0.9673	0.7450	0.6407	0.9822	0.8295	0.6950	10.56M	8.16G
	-	✓	0.7746	0.9670	0.7345	0.6344	0.9809	**0.8349**	0.6878	11.91M	8.49G
	✓	✓	**0.7828**	**0.9674**	**0.7512**	**0.6449**	**0.9816**	0.8259	**0.6977**	12.60M	8.61G

* MaxIP denotes the proposed forward phase-difference projection prior.

parameters set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. A cosine learning rate scheduler with linear warm-up was adopted, where the learning rate was linearly increased from 5×10^{-7} to 1×10^{-4} over the first 10 epochs. All experiments were implemented in PyTorch with fixed hyperparameters and conducted on a single NVIDIA GeForce RTX 3090 GPU.

Comparison Results.

The proposed method is compared with CNN-based approaches [2, 3, 6, 10, 16, 19], and methods that incorporate temporal encoding [1, 5, 11, 21], with performance evaluated using DSC, clDice, IoU, Acc, AUC, and Pre. As shown in Table 1, the proposed method demonstrates consistently strong performance on the DIAS dataset and remains competitive on the DSCA dataset. Compared with existing approaches, it exhibits improved segmentation accuracy while maintaining a favorable trade-off between accuracy and efficiency. Qualitative results in Fig. 3 further illustrate clearer vessel delineation and reduced false predictions.

Ablation Study.

We conduct ablation experiments to examine the effect of projection-based spatial priors under different temporal modeling backbones. As shown in Table 2, introducing projection-based priors leads to consistent performance improvements when applied to both Mamba-based and Transformer-based frameworks, indicating that the proposed priors generalize well across different temporal modeling choices. MinIP and FPD-MaxIP exhibit complementary effects on vessel representation, and using both priors together results in the best performance

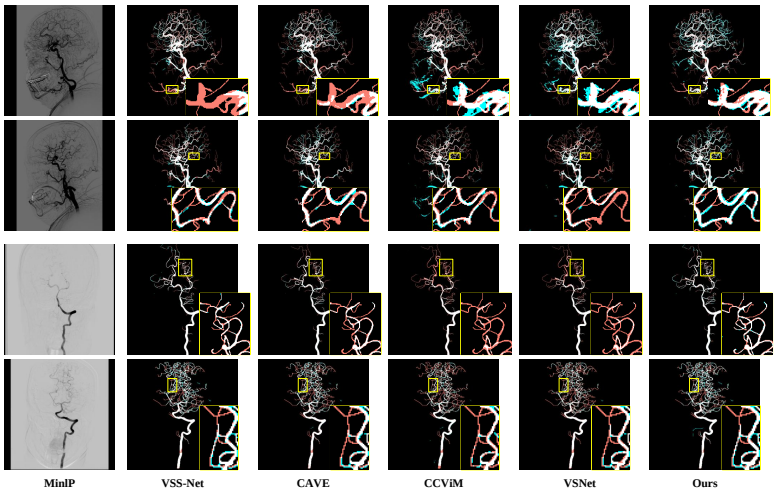

Fig. 3. Qualitative comparison results of cerebral vascular segmentation in the DIAS (the first two rows) and DSCA dataset(the last two rows). White, blue, and red represent true positives, false positives, and false negatives, respectively. Localized magnification provides clearer visualization.

Table 3. Ablation study of the number of FPD-MaxIP layers in the Mamba-based framework on the DIAS dataset.

Layers	DSC	Acc	Sen	IoU	AUC	Pre	clDice	Params	FLOPs
1	0.7652	0.9644	0.7395	0.6231	0.9804	0.7586	0.6818	10.16M	6.13G
2	0.7727	0.9669	0.7265	0.6320	0.9687	**0.8377**	0.6843	10.67M	6.25G
3	0.7784	0.9665	**0.7620**	0.6391	0.9804	0.8103	0.7017	11.18M	6.36G
4	**0.7845**	**0.9678**	0.7569	**0.6472**	**0.9819**	0.8212	**0.7054**	11.69M	6.47G

under both backbones. It is worth nothing that the proposed FPD-MaxIP improves precision under both backbones, suggesting its effectiveness in suppressing false-positive responses caused by persistent tissue and bony structures. Table 3 further explores the influence of the number of FPD-MaxIP layers in the Mamba-based framework. Performance improves as additional FPD-MaxIP layers are introduced, reflecting the benefit of aggregating contrast information from multiple temporal phases. With deeper stacking, the model exhibits only a marginal increase in parameters and FLOPs, while achieving a substantial performance improvement, indicating the effectiveness of the FPD-MaxIP branch.

Discussion. Our framework combines projection-based spatial priors and temporal modeling for DSA vessel segmentation. However, FPD-MaxIP may be less effective under mismatched contrast dynamics, particularly in cases of delayed filling or severe motion. In the future, we will explore adaptive phase selection strategies tailored to patient-specific contrast dynamics.

4 Conclusion

In this paper, we propose a unified spatiotemporal framework for cerebral vessel segmentation from DSA sequences. By integrating projection-based spatial priors with temporal modeling, the proposed method improves DSA vessel segmentation by enhancing distal vessel visibility and suppressing persistent background interference. Experiments on two public datasets demonstrate superior performance and a favorable accuracy–efficiency trade-off.These results indicate that combining temporal modeling with projection-based spatial priors is a promising direction for vessel segmentation in dynamic angiographic imaging.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China (No. 62376038), the National Key R&D Program of China (No. 2018AAA0102600), the Beijing–Tianjin–Hebei Natural Science Foundation Cooperation Project (No. H2024102009), and the BUPT Excellent Ph.D. Students Foundation (No. CX20251022).

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

References

1. Hao, D., Ding, S., Qiu, L., Lv, Y., Fei, B., Zhu, Y., Qin, B.: Sequential vessel segmentation via deep channel attention network. Neural Networks **128**, 172–187 (2020)
2. He, A., Wang, K., Li, T., Du, C., Xia, S., Fu, H.: H2former: An efficient hierarchical hybrid transformer for medical image segmentation. IEEE Transactions on Medical Imaging **42**(9), 2763–2775 (2023)
3. Lachinov, D., Seeböck, P., Mai, J., Goldbach, F., Schmidt-Erfurth, U., Bogunovic, H.: Projective skip-connections for segmentation along a subset of dimensions in retinal OCT. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 431–441. Springer International Publishing, Cham (2021)
4. Li, R.Q., Bian, G.B., Zhou, X.H., Xie, X., Ni, Z.L., Hou, Z.: Cau-net: A novel convolutional neural network for coronary artery segmentation in digital substraction angiography. In: International Conference on Neural Information Processing. pp. 185–196. Springer (2020)
5. Liu, W., Tian, T., Wang, L., Xu, W., Li, L., Li, H., Zhao, W., Tian, S., Pan, X., Deng, Y., et al.: Dias: A dataset and benchmark for intracranial artery segmentation in dsa sequences. Medical Image Analysis **97**, 103247 (2024)
6. Liu, W., Yang, H., Tian, T., Cao, Z., Pan, X., Xu, W., Jin, Y., Gao, F.: Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation. IEEE journal of biomedical and health informatics **26**(9), 4623–4634 (2022)

7. Meng, C., Sun, K., Guan, S., Wang, Q., Zong, R., Liu, L.: Multiscale dense convolutional neural network for dsa cerebrovascular segmentation. Neurocomputing **373**, 123–134 (2020)
8. Mensah, G.A., Fuster, V., Murray, C.J., Roth, G.A., Abate, Y.H., Abbasian, M., Abd-Allah, F., Abdollahi, A., Abdollahi, M., Abdulah, D.M., et al.: Global burden of cardiovascular diseases and risks, 1990-2022. Journal of the American College of Cardiology **82**(25), 2350–2473 (2023)
9. Phipps, M.S., Cronin, C.A.: Management of acute ischemic stroke. Bmj **368** (2020)
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
11. Su, R., van der Sluijs, P.M., Chen, Y., Cornelissen, S., van den Broek, R., van Zwam, W.H., van der Lugt, A., Niessen, W.J., Ruijters, D., van Walsum, T.: Cave: Cerebral artery–vein segmentation in digital subtraction angiography. Computerized Medical Imaging and Graphics **115**, 102392 (2024)
12. 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. pp. 585–594 (2022)
13. Wang, Z., Zheng, J.Q., Zhang, Y., Cui, G., Li, L.: Mamba-unet: Unet-like pure visual mamba for medical image segmentation. arXiv preprint arXiv:2402.05079 (2024)
14. Xie, Q., Zhang, D., Mou, L., Wang, S., Zhao, Y., Guo, M., Zhang, J.: 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)
15. Xing, Z., Ye, T., Yang, Y., Cai, D., Gai, B., Wu, X.J., Gao, F., Zhu, L.: Segmamba-v2: Long-range sequential modeling mamba for general 3d medical image segmentation. IEEE Transactions on Medical Imaging (2025)
16. Xu, J., Dong, A., Yang, Y., Jin, S., Zeng, J., Xu, Z., Jiang, W., Zhang, L., Dong, J., Wang, B.: VSNet: Vessel Structure-aware Network for hepatic and portal vein segmentation. Medical Image Analysis **101**, 103458 (2025)
17. Zhang, J., Xie, Q., Mou, L., Zhang, D., Chen, D., Shan, C., Zhao, Y., Su, R., Guo, M.: Dsca: A digital subtraction angiography sequence dataset and spatio-temporal model for cerebral artery segmentation. IEEE Transactions on Medical Imaging (2025)
18. Zhang, L., Jiang, X., Ding, X., Huang, Z., Zhao, T., Yang, X.: 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)
19. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: International workshop on deep learning in medical image analysis. pp. 3–11. Springer (2018)
20. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417 (2024)
21. Zhu, Y., Zhang, D., Lin, Y., Feng, Y., Tang, J.: Merging context clustering with visual state space models for medical image segmentation. IEEE Transactions on Medical Imaging (2025)