Fine-Grained Cerebrovascular Parsing in DSA via Structurally-Grounded Semantic Disentanglement

Kai Zhu[1,2], Le Cao[3], Li Chen[1], Jun Cheng[4], Lei Mou[2,✉], and Yitian Zhao[2,✉]

[1] Wuhan University of Science and Technology, Wuhan, China
[2] Laboratory of Advanced Theranostic Materials and Technology, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China
[3] West China Hospital of Sichuan University, Chengdu, China
[4] Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore
{moulei, yitian.zhao}@nimte.ac.cn

Abstract. Accurate quantitative cerebrovascular assessment and clinical intervention rely fundamentally on vessel parsing in Digital Subtraction Angiography (DSA) to segment and distinctly label the vascular tree. However, the intricate vascular topology, pronounced vessel overlap, and inconsistent contrast enhancement inherent in DSA present formidable obstacles; consequently, most existing works have been limited to binary vessel segmentation, overlooking the fine-grained anatomical categorization required for detailed clinical analysis. To address this, we propose AngioParse, a multi-encoder collaborative framework explicitly designed for fine-grained vascular parsing. AngioParse ensures parsing precision through three key mechanisms: (1) a dual-stream encoder synergy module that improves localization accuracy by fusing a high-recall global binary prior with fine-grained anatomical features from heterogeneous encoders; (2) a self-consistent structural grounding mechanism that regularizes fine-grained parsing with a high-recall binary vascular prior by constraining the union of anatomical-wise predictions to lie within the vessel support, improving connectivity and topological validity; and (3) a decoupled orthogonal prediction module that enhances class discrimination by separating multi-target learning into parallel branches to minimize feature confusion. Experiments on a public DSA dataset demonstrate that AngioParse not only achieves state-of-the-art segmentation performance but also exhibits superior capability in accurately parsing and differentiating six distinct vascular categories. The code is available at https://github.com/kylechuuuuu/AngioParse.git.

Keywords: DSA · Vascular Parsing · Multi-Encoder Fusion

1 Introduction

Digital Subtraction Angiography (DSA) stands as the clinical gold standard for visualizing cerebrovascular architecture, offering the high-fidelity spatial resolution essential for the diagnosis and management of neurovascular disorders [6].

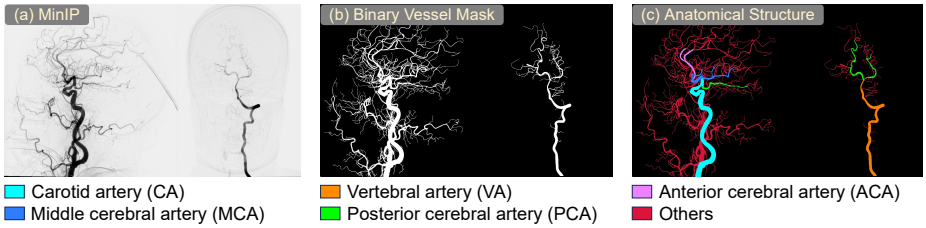

Fig. 1: Visualization of cerebrovascular complexities and vasculature parsing: (a) Minimum intensity projection (MinIP) of a DSA sequence; (b) the corresponding binary vessel mask; and (c) vasculature parsing into distinct anatomical structures.

Beyond mere visualization, as shown in Fig. 1(a-b), the clinical utility of DSA relies on the anatomical categorization of the vascular tree into several distinct functional segments, including the carotid artery (CA), vertebral artery (VA), anterior cerebral artery (ACA), middle cerebral artery (MCA), posterior cerebral artery (PCA), and others, as depicted in Fig. 1(c). This fine-grained parcellation is clinically pivotal, as specific vascular segments dictate disparate therapeutic protocols and prognostic outcomes.

Despite the clinical necessity of semantic parsing, current automated analysis of DSA images has remained predominantly focused on binary vascular segmentation. These approaches aim to extract the geometric structure of the vessel tree from complex background, with state-of-the-art models [9,1,5,11,20] achieving high fidelity in pixel-wise extraction. However, these methods typically stop at the structural level, restricted to coarse categorizations such as distinguishing major arteries from peripheral branches [19,16] or binary artery-vein separation [14]. Consequently, there remains a significant semantic gap between these binary vascular maps and the comprehensive anatomical parsing required for advanced clinical workflows.

The tendency of existing research to prioritize segmentation over parsing is largely driven by the inherent complexity of the cerebrovascular hierarchy and the formidable challenges of resolving it in 2D space. The projective superposition of 3D vascular hierarchies onto 2D planes creates a profound disconnect between local visual morphology and global anatomical identity. Tortuous branching patterns and pronounced overlaps obscure clear delineation, creating ambiguities where segments share similar local textures but possess distinct anatomical identities. Current models, lacking the global topological constraints required to resolve these conflicts, fail to distinguish these overlapping structures. As a result, the labor-intensive and subjective manual identification of these segments remains the bottleneck, underscoring the urgent demand for automated parsing solutions capable of overcoming these projective ambiguities.

To address the limitations of existing binary segmentation methods in meeting clinical diagnostic needs, we propose AngioParse, a novel framework that redefines cerebrovascular parsing as a task of structure-aware semantic disentanglement. By reconciling the discord between 2D visual morphology and 3D anatomical identity, AngioParse directly addresses the bottlenecks of projective

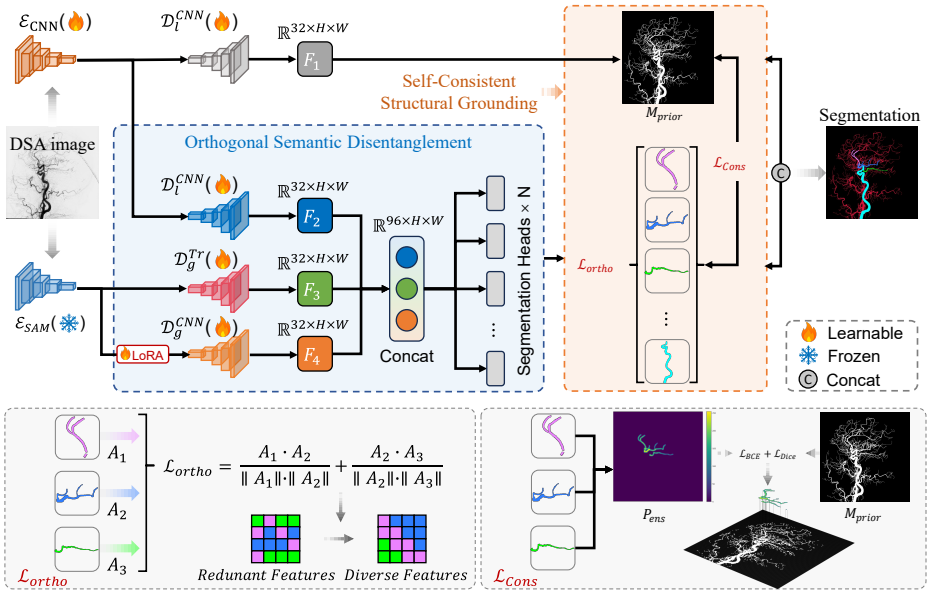

Fig. 2: Architecture of the proposed AngioParse.

overlap and vessel fragmentation that hinder automated clinical workflows. The main contributions of this work are summarized as follows:

- We propose AngioParse, the first framework dedicated to closing the semantic gap between binary vessel extraction and fine-grained anatomical parsing. By synergizing global morphological cues with local anatomical signatures, it enables the precise categorization of the cerebrovascular tree into six clinically distinct functional segments.
- We introduce a self-consistent structural grounding strategy to maintain the topological continuity of individual arterial segments. By enforcing prior-guided consistency between the global vascular tree and local anatomical predictions, it mitigates vessel fragmentation, ensuring that parsed segments remain topologically valid for downstream clinical quantification.
- We propose an orthogonal semantic disentanglement module to resolve the projective ambiguity inherent in 2D DSA. By enforcing feature orthogonality in the latent space, it minimizes semantic confusion at vessel crossovers, effectively disentangling overlapping arteries to ensure high-fidelity anatomical labeling.

2 Method

The proposed AngioParse is a structure-aware framework that reconciles morphological anchors with anatomical signatures through a dual-pathway architecture. It primarily consists of two core components: self-consistent structural grounding (SCSG) and orthogonal semantic disentanglement (OSD). The overall pipeline is illustrated in Fig. 2.

2.1 Self-Consistent Structural Grounding

The SCSG structure enforces self-consistency between the decoupled anatomical predictions and a global binary vascular prior. This coupling improves anatomical parsing in two aspects: it strengthens the spatial localization of each label and preserves topological agreement with the overall vessel tree. In DSA parsing, a common failure mode is spatial confusion, where pixels of one anatomical segment can be spuriously activated at distant locations due to projective overlap and morphological similarity. We specifically design the binary stream to favor high recall, ensuring it captures the complete vascular topology to serve as a reliable anchor for the subsequent parsing branches.

To handle the diverse shapes and scales of vascular structures, we employ a hybrid multi-encoder stream that integrates complementary architectural priors. Given an input DSA image X, a CNN-based encoder ($\mathcal{E}_{CNN}$) leverages local inductive biases to extract fine-grained feature representations F_{local}. Simultaneously, we utilize a SAM2 [12] encoder to capture long-range structural dependencies. By injecting LoRA [7] into the frozen SAM2 encoder, we generate a domain-aligned global representation F_{global}. The feature extraction process is formulated as:

$$F_{local} = \mathcal{E}_{CNN}(X), \quad F_{global} = \mathcal{E}_{SAM}(X; \theta_{lora}), \tag{1}$$

where θ_{lora} denotes the trainable adapter parameters. F_{local} preserves high-resolution edge details essential for boundary delineation, while F_{global} encapsulates the high-level topology of the cerebrovascular tree.

These heterogeneous representations are subsequently processed through distinct decoding pathways. The local features F_{local} are decoded by a binary decoder $\mathcal{D}_{CNN}$ into a holistic vessel prior M_{prior}, which serves as a structural anchor delineating the valid vascular support:

$$M_{prior} = \sigma(\mathcal{D}_{CNN}(F_{local})), \tag{2}$$

where σ denotes the sigmoid activation. Conversely, the global features F_{global} guide the OSD module to generate decoupled anatomical-wise probability maps $\{A_i\}_{i=1}^{N}$ for the N anatomical structures.

Finally, to enforce coherence between the binary prior and anatomical-wise predictions, we introduce a self-consistent grounding constraint requiring the union of anatomical-wise probabilities to agree with the vessel support indicated by M_{prior}. Since OSD outputs N parallel sigmoid maps, we aggregate them into an ensemble probability P_{ens}. Under the mild assumption that different anatomy heads are approximately decorrelated (encouraged by Orthogonality Loss in Sec. 2.2), we use a probabilistic noisy-OR to approximate the union:

$$P_{ens} = 1 - \prod_{i=1}^{N}(1 - A_i). \tag{3}$$

We constrain P_{ens} to match the high-recall vessel prior via a consistency loss:

$$\mathcal{L}_{Cons} = \mathcal{L}_{BCE}(P_{ens}, M_{prior}) + \mathcal{L}_{Dice}(P_{ens}, M_{prior}). \tag{4}$$

This alignment suppresses anatomy predictions outside the vascular support and stabilizes fine-grained parsing to remain structurally consistent.

2.2 Orthogonal Semantic Disentanglement

The OSD module is designed to alleviate semantic ambiguity induced by 2D projective overlaps by disentangling anatomical structures in a decoupled latent space. Given the local encoder feature F_{local} and the global grounding feature F_{global}, OSD first performs multi-branch decoding to obtain complementary, spatially aligned representations. Specifically, we employ three parallel decoders: a local CNN decoder ($\mathcal{D}_l^{CNN}$) and two global decoders, one CNN-based ($\mathcal{D}_g^{CNN}$) and one Transformer-based ($\mathcal{D}_g^{Tr}$), to capture both fine boundary details and long-range topology cues. Each decoder projects features to a common dimension $\mathbb{R}^{32\times H\times W}$. The three decoded features are then concatenated to form a diverse fused representation $P \in \mathbb{R}^{96\times H\times W}$:

$$P = \mathrm{Concat}\big(\mathcal{D}_l^{CNN}(F_{local}),\ \mathcal{D}_g^{CNN}(F_{global}),\ \mathcal{D}_g^{Tr}(F_{global})\big). \tag{5}$$

This design preserves high-frequency vessel boundaries from the local stream while injecting global, topology-aware context from the grounding stream, which is crucial for resolving overlaps and mitigating fragmentation in low-contrast DSA images.

Standard multi-class segmentation models typically rely on a single competitive layer, which often leads to feature suppression, where the dominant signal of a major vessel interferes with the identification of overlapping or adjacent peripheral structures. To resolve this, OSD decomposes the multi-target parsing task into N parallel binary heads, treating each anatomical structure as an independent task within its own feature subspace. Given the fused feature P, the i-th head produces an anatomy-specific probability map A_i:

$$A_i = \sigma\big(\Phi_i(P)\big), \quad i = 1, \ldots, N, \tag{6}$$

where Φ_i denotes the i-th prediction branch consisting of a 3×3 convolution, ReLU activation, and a 1×1 convolution, and σ is the sigmoid function. This parallel design avoids mutual suppression inherent in traditional output layers.

To further promote feature diversity and minimize cross-branch redundancy, we introduce an Orthogonality Loss $\mathcal{L}_{ortho}$:

$$\mathcal{L}_{ortho} = \sum_{1\le j<k\le N} \left(\frac{A_j \cdot A_k}{\|A_j\|\|A_k\|} \right)^2, \tag{7}$$

where A_j and A_k are the flattened probability vectors. This constraint forces each branch to capture unique anatomical characteristics, effectively disentangling the intertwined vascular hierarchies.

2.3 Loss Function

We optimize AngioParse via a multi-task objective supervising both the global vessel prior and N anatomical predictions. Given the ground-truth binary mask Y_{bin} and structure-wise masks $Y_i \in \{0,1\}^{H\times W}$, the total loss is:

$$\mathcal{L}_{total} = \mathcal{L}_{seg}(M_{prior}, Y_{bin}) + \sum_{i=1}^{N} \mathcal{L}_{seg}(A_i, Y_i) + \alpha\,\mathcal{L}_{ortho} + \beta\,\mathcal{L}_{Cons}, \quad (8)$$

where $\mathcal{L}_{seg}(P,Y) = \mathcal{L}_{BCE}(P,Y) + \mathcal{L}_{Dice}(P,Y)$ combines pixel-wise binary cross-entropy and soft Dice loss. $\mathcal{L}_{Cons}$ and $\mathcal{L}_{ortho}$ (Sec. 2.1-2.2) enforce structural grounding and semantic disentanglement, respectively, with weighting hyperparameters α and β.

3 Experiments

3.1 Dataset and Implementation Details

To validate the parsing capabilities of AngioParse, we established a new semantic annotation standard based on the DSCA benchmark [19]. While the original benchmark focused on binary segmentation, this extended dataset features high-fidelity, pixel-level annotations performed by senior interventional neuroradiologists. The annotations classify the cerebrovascular architecture into six distinct anatomical segments: CA, VA, ACA, MCA, PCA, and others. The dataset consists of high-resolution DSA MinIP images, stratified into a training set of 175 images and a test set of 45 images. Fig. 1 illustrates the label distribution and color coding used throughout this study.

Implemented in PyTorch, the framework was trained on an NVIDIA RTX 5090 for 350 epochs using a batch size of 1. We employed the AdamW optimizer (initial $LR = 1 \times 10^{-4}$) with a cosine annealing scheduler and standard data augmentations. For fair comparison, all baselines were evaluated under identical settings. For SAM-based methods (SAM-Adapter), we adopt the ViT-L pretrained weight. For SAM2-based approaches (SAM2-Unet and our SAM2 encoder backbone), we utilize the SAM 2.1 Hiera Large pretrained weights. Performance was measured via Dice and IoU for overlap accuracy, and clDice for topological connectivity.

3.2 Comparison with State-of-the-Art Methods

We compare AngioParse against representative state-of-the-art methods, including CNN-based (U-Net [13], CS^{2}Net [10], nnU-Net [8], BCU-Net [18]), Transformer-based (Swin-unet [2], TransUNet [3], Segformer [15]), and foundation model-based architectures (SAM-Adapter [4], SAM2-Unet [17]).

As shown in Table 1, CNN-based architectures such as nnU-Net and BCU-Net perform robustly on this structured task. In contrast, pure Transformer

Table 1: Performance comparison on the DSCA dataset (%). "Cervic. V." refers to cervical vessels (CA and VA segments).

Methods	Mean			Cervic. V.			ACA			MCA			PCA			Others		
	Dice	IoU	clD.	Dice	IoU	clD.	Dice	IoU	clD.	Dice	IoU	clD.	Dice	IoU	clD.	Dice	IoU	clD.
U-Net [13]	44.79	32.45	23.89	65.81	49.04	24.81	9.84	5.18	8.70	43.48	27.78	17.86	21.63	12.13	10.38	80.23	66.99	63.04
CS^2Net [10]	50.66	37.77	30.78	72.22	56.51	34.82	5.73	2.95	6.20	40.39	25.31	19.36	36.02	21.97	20.81	78.88	65.12	62.13
nnU-Net [8]	75.48	62.33	51.51	89.88	81.62	59.62	66.23	49.51	**44.57**	64.25	47.33	39.26	58.84	41.68	36.76	83.39	71.51	65.44
BCU-Net [18]	72.88	59.09	48.02	86.32	75.93	53.54	58.81	41.65	34.26	60.91	43.79	35.05	59.13	41.98	38.72	84.82	73.64	67.70
Swin-unet [2]	19.97	12.39	4.42	31.16	18.46	4.88	4.38	2.24	1.23	0.83	0.051	0.05	0.34	0.03	0.03	51.97	35.10	15.45
TransUNet [3]	28.84	19.43	7.97	52.90	35.96	14.66	3.34	2.12	1.12	7.50	3.90	0.67	0.70	0.04	0.004	55.70	38.60	16.72
Segformer [15]	47.10	32.62	22.97	66.80	50.15	29.05	23.06	13.03	12.12	36.53	22.35	16.75	29.38	17.22	14.26	62.17	45.11	32.48
SAM-Adapter [4]	64.79	50.16	32.25	84.34	72.92	42.48	48.71	32.20	25.11	50.09	33.42	18.99	48.63	32.12	23.78	71.49	55.63	37.54
SAM2-Unet [17]	60.94	47.13	19.92	87.64	78.01	37.49	47.54	31.18	11.51	50.56	33.84	12.43	42.25	26.78	10.64	47.99	31.57	7.87
AngioParse	**78.51**	**67.79**	**53.70**	**94.03**	**88.74**	**63.14**	**68.75**	**52.38**	44.11	**67.74**	**51.22**	**42.77**	**65.71**	**48.93**	**41.22**	**86.82**	**76.70**	**67.83**

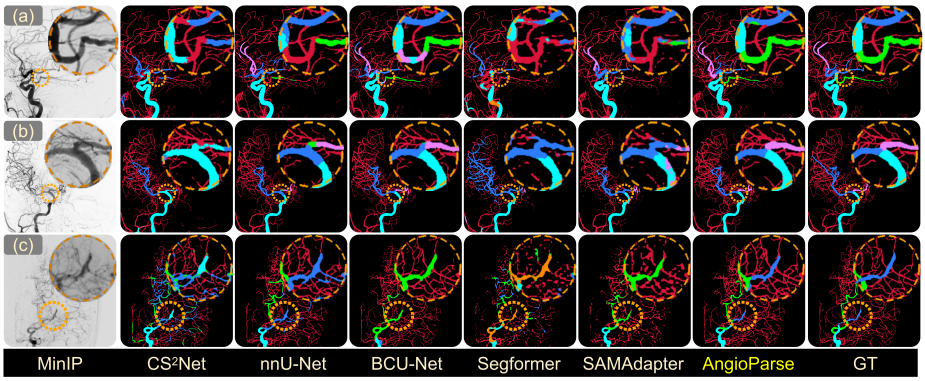

Fig. 3: Qualitative comparison of anatomical parsing results on DSCA dataset.

and foundation model-based approaches, including Swin-Unet and SAM2-Unet, struggle significantly with fine-grained multi-target labels. AngioParse achieves state-of-the-art results across nearly all metrics. Notably, AngioParse excels in segmenting delicate structures, reaching 86.82% Dice for Others and 94.03% for cervical vessels. Furthermore, its high clDice scores across most categories demonstrate a superior ability to preserve vascular connectivity, outperforming the runner-up by a significant margin in anatomical classification.

The qualitative results across three representative scenarios are visualized in Fig. 3. In case (a), the magnified regions demonstrate that AngioParse successfully identifies distal vascular branches (highlighted in green), which are frequently omitted or misclassified by baseline models. In case (b), our approach exhibits superior precision at complex bifurcation points, maintaining accurate category assignment and eliminating the inter-class leakage prevalent in competing methods. For case (c), AngioParse yields the most robust anatomical parsing by effectively mitigating both false positives (leakage) and false negatives (missing segments). These visual results confirm that the synergy of complementary multi-encoder features and structural priors enables AngioParse to produce more reliable, topologically consistent vascular maps compared to existing state-of-the-art frameworks.

Table 2: Ablation of components.

SCSG	OSD	Dice	mIoU	clD.
✗	✗	65.81	54.58	45.86
✓	✗	78.04	65.40	51.72
✗	✓	75.10	63.90	49.60
✓	✓	**78.51**	**67.79**	**53.70**

Table 3: Ablation of decoders.

$\mathcal{D}_l^{CNN}$	$\mathcal{D}_g^{Tr}$	$\mathcal{D}_g^{CNN}$	Dice	mIoU	clD.
✓	✗	✗	77.35	64.15	51.62
✓	✓	✗	77.62	64.48	51.65
✓	✓	✓	**78.51**	**67.79**	**53.70**

Fig. 4: Analysis of α and β.

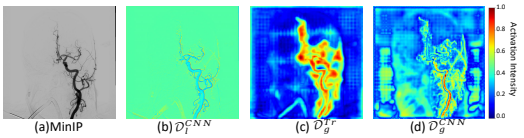

Fig. 5: Visualization of decoder activations.

3.3 Ablation Study

We validated the effectiveness of each component through a stepwise ablation study (Table 2). Starting with the baseline, the integration of the SCSG module yielded a substantial gain (+12.23% Dice), underscoring the critical role of global binary priors in anchoring fine-grained parsing. The OSD module also provided significant improvement (+9.29% Dice) by effectively mitigating inter-class feature redundancy. The synergic integration of both modules yielded the best overall performance, specifically boosting topological continuity as evidenced by the high clDice and mIoU scores. Hyperparameter sensitivity analysis (Fig. 4) identified an optimal performance peak (78.51% mean Dice) at $\alpha = 0.6$ and $\beta = 0.1$, striking an effective balance between structural consistency and semantic orthogonality.

A granular analysis of the multi-source decoders (Table 3) reveals the additive benefit of our decoding architecture. The sequential incorporation of the local CNN decoder ($\mathcal{D}_l^{CNN}$), global Transformer decoder ($\mathcal{D}_g^{Tr}$), and global CNN decoder ($\mathcal{D}_g^{CNN}$) results in progressive performance gains. This confirms that combining local high-frequency details with global topological context is essential for resolving complex vascular hierarchies. Qualitative visualizations in Fig. 5 further corroborate these findings, distinguishing the specific contributions of each decoder branch in capturing sharp gradients versus semantic continuity.

4 Conclusions

In this study, we presented AngioParse, a novel framework specifically designed for fine-grained anatomical parsing of the cerebrovascular tree in DSA. By synergizing local CNN encoders with global foundation model priors, our approach successfully bridges the semantic gap between binary vessel extraction and detailed anatomical categorization. The proposed mechanisms, including SCSG and OSD, effectively address the challenges of topological connectivity and projective ambiguity inherent in 2D imaging. Experimental results on the DSCA

dataset demonstrate that AngioParse achieves state-of-the-art performance, accurately differentiating six vascular categories while maintaining structural integrity. Despite these advancements, the computational complexity inherent in the multi-encoder architecture and the requirement for high-quality pixel-level annotations represent practical constraints. Future work will focus on optimizing architectural efficiency and exploring semi-supervised learning strategies to reduce the reliance on manual annotation.

Acknowledgments. This work is supported by the National Science Foundation Program of China (62422122, 62272444, 62271359), the Key Research and Development Program of Zhejiang Province (2024C03281), the Ningbo 'KeChuang Yongjiang 2035' Key R&D Program (2024Z230), and the Agency for Science, Technology and Research (A*STAR) under its MTC Programmatic Funds (Grant No. M23L7b0021).

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

References

1. Ahmed, R.M., Iltaf, A., Elmanna, M., Zhao, G., Li, H., Du, Y., Li, B., Zhou, S.: Msa-unet3+: Multi-scale attention unet3+ with new supervised prototypical contrastive loss for coronary dsa image segmentation. Biomedical Signal Processing and Control **123**, 110539 (2026)
2. 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. pp. 205–218. Springer (2022)
3. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis p. 103280 (2024)
4. Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., Li, Z., Sun, L., Zang, Y., Mao, P.: Sam-adapter: Adapting segment anything in underperformed scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3367–3375 (2023)
5. Deng, H., Liu, X., et al.: Dfa-net: Dual multi-scale feature aggregation network for vessel segmentation in x-ray digital subtraction angiography. Journal of Big Data **11**(1), 57 (2024)
6. Goyal, M., Ospel, J.M., Ganesh, A., Dowlatshahi, D., Volders, D., Möhlenbruch, M.A., Jumaa, M.A., Nimjee, S.M., Booth, T.C., Buck, B.H., et al.: Endovascular treatment of stroke due to medium-vessel occlusion. New England Journal of Medicine **392**(14), 1385–1395 (2025)
7. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (2022)
8. 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), 203–211 (2021)

9. Luo, Y., Sun, L.: Digital subtraction angiography image segmentation based on multiscale hessian matrix applied to medical diagnosis and clinical nursing of coronary stenting patients. Journal of Radiation Research and Applied Sciences **16**(2), 100603 (2023)
10. Mou, L., Zhao, Y., Fu, H., Liu, Y., Cheng, J., Zheng, Y., Su, P., Yang, J., Chen, L., Frangi, A.F., et al.: Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. Medical image analysis **67**, 101874 (2021)
11. Qin, J., Luo, H., et al.: Dsa-former: A network of hybrid variable structures for liver and liver tumour segmentation. The International Journal of Medical Robotics and Computer Assisted Surgery (2024), published online Oct 2024
12. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. In: International Conference on Learning Representations. vol. 2025, pp. 28085–28128 (2025)
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
14. Su, R., van der Sluijs, P.M., et al.: Cave: Cerebral artery-vein segmentation in digital subtraction angiography. Computerized Medical Imaging and Graphics **115**, 102392 (2024)
15. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. In: Neural Information Processing Systems (NeurIPS) (2021)
16. Xie, Q., Zhang, D., et al.: 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 (MICCAI) (2023)
17. Xiong, X., Wu, Z., Tan, S., Li, W., Tang, F., Chen, Y., Li, S., Ma, J., Li, G.: Sam2-unet: Segment anything 2 makes strong encoder for natural and medical image segmentation. Visual Intelligence **4**(1), 2 (2026)
18. Zhang, H., Zhong, X., Li, G., Liu, W., Liu, J., Ji, D., Li, X., Wu, J.: Bcu-net: Bridging convnext and u-net for medical image segmentation. Computers in Biology and Medicine **159**, 106960 (2023)
19. Zhang, J., Xie, Q., et al.: Dsca: A digital subtraction angiography sequence dataset and spatio-temporal model for cerebral artery segmentation. IEEE Transactions on Medical Imaging **44**(6), 2515–2527 (2025)
20. Zhang, L., Jiang, X., et al.: 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 (MICCAI) (2024)