BiM-GeoAttn-Net: Linear-Time Depth Modeling with Geometry-Aware Attention for 3D Aortic Dissection CTA Segmentation

Yuan Zhang[1,†], Lei Liu[2,3,†], Jialin Zhang[4], Ya-Nan Zhang[1], Ling Wang[1,⋆], and Nan Mu[1,⋆]

[1] College of Computer Science, Sichuan Normal University, Chengdu, Sichuan, China.
[2] Ant Group, Hangzhou, Zhejiang, China.
[3] Zhejiang University, Hangzhou, Zhejiang, China.
[4] State Key Laboratory of Public Big Data, Guizhou University, Guiyang, China.
[†] Equal contribution.

Abstract. Accurate segmentation of aortic dissection (AD) lumens in CT angiography (CTA) is essential for quantitative morphological assessment and clinical decision-making. However, reliable 3D delineation remains challenging due to limited long-range context modeling, which compromises inter-slice coherence, and insufficient structural discrimination under low-contrast conditions. To address these limitations, we propose **BiM-GeoAttn-Net**, a lightweight framework that integrates linear-time depth-wise state-space modeling with geometry-aware vessel refinement. Our approach features a **Bidirectional Depth Mamba (BiM)** module to efficiently capture cross-slice dependencies and a **Geometry-Aware Vessel Attention (GeoAttn)** module that combines orientation-sensitive anisotropic filtering with lightweight attention for vessel refinement. Experiments on the current multi-source AD CTA dataset show that BiM-GeoAttn-Net achieves a Dice score of 93.35% and an HD95 of 12.36 mm, outperforming representative CNN-, Transformer-, and SSM-based baselines in overlap metrics while maintaining competitive HD95. These results suggest that coupling linear-time depth modeling with geometry-aware refinement provides an effective and computationally efficient strategy for 3D AD segmentation.

Keywords: Aortic dissection · 3D CTA segmentation · State-space modeling · Depth-wise sequence modeling · Geometry-aware attention

1 Introduction

Accurate segmentation of aortic dissection (AD) lumens is critical for morphological assessment and longitudinal monitoring of Type-B AD [7,3], enabling cohort-level analysis via standardized volumetric descriptors [4]. However, robust segmentation in CT angiography (CTA) is hindered by the aorta's complex, tortuous geometry and substantial appearance variations along the centerline.

⋆ Corresponding authors: lingwang@sicnu.edu.cn (L. Wang) and nanmu@sicnu.edu.cn (N. Mu).

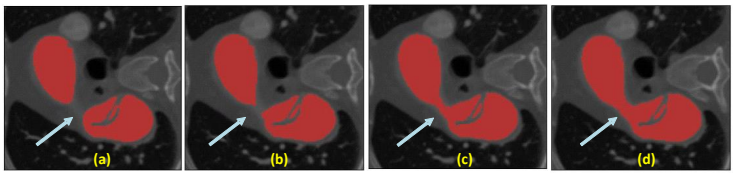


Fig. 1. Challenges in AD CTA segmentation. Four consecutive axial slices exhibit strong cross-slice continuity, subtle boundary contrast, and morphological variation. Ground-truth masks are in red.

Conventional 3D CNNs effectively capture local texture and boundary cues [11], yet their limited receptive fields hinder long-range dependency modeling. Transformer-based architectures improve global context aggregation via self-attention [5], but full 3D attention is computationally expensive, and practical approximations may compromise boundary precision. Recently, state-space models (SSMs) such as Mamba offer linear-time long-sequence modeling [8]. However, existing SSM-based medical segmentation methods [14] typically adopt generic sequence construction without explicitly aligning modeling direction with anatomical continuity, and purely global modeling remains insufficient for low-contrast boundary refinement without structural priors [16].

To address these limitations, we propose BiM-GeoAttn-Net, a bottleneck enhancement framework that integrates efficient depth-axis state-space modeling with geometry-aware vessel refinement. Specifically, a Bidirectional Depth Mamba (BiM) module performs linear-time modeling along the depth axis to strengthen cross-slice consistency. Built upon BiM, a Geometry-Aware Vessel Attention (GeoAttn) module incorporates plane-aligned anisotropic filtering with dual attention mechanisms to enhance vessel-structure representations under low contrast. The cascaded BiM-GeoAttn design directly addresses two dominant failure modes in AD CTA segmentation: depth-wise discontinuity and boundary degradation under low contrast. The implementation of BiM-GeoAttn-Net is publicly available at https://github.com/bbbwang/BiM_GeoAttn_Net.git.

The main contributions are summarized as follows:

(1) We propose **BiM-GeoAttn-Net**, a lightweight 3D segmentation framework tailored for aortic dissection lumens, addressing challenges in long-range context modeling and low-contrast boundary delineation.

(2) We introduce an anatomy-aware bottleneck design that couples **Bidirectional Depth Mamba** with **Geometry-Aware Vessel Attention** for depth-axis modeling and plane-aligned vessel refinement.

(3) Experiments demonstrate improved overlap accuracy over representative CNN-, Transformer-, and SSM-based baselines, while maintaining competitive boundary accuracy and computational efficiency.

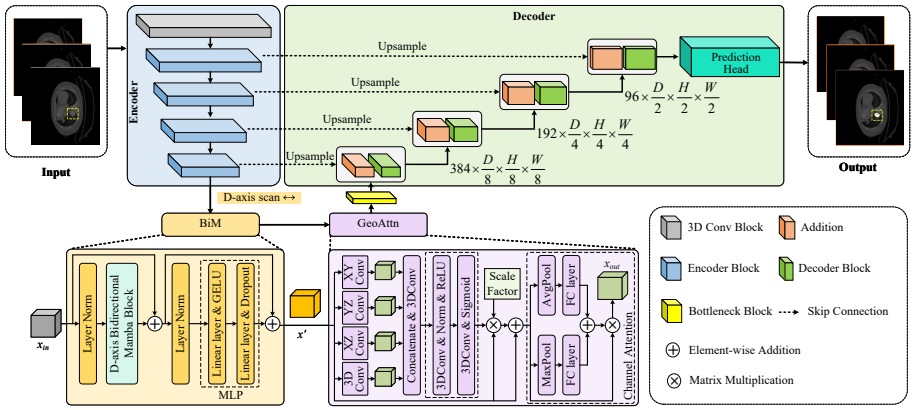


Fig. 2. Overall architecture of the proposed BiM-GeoAttn-Net.

2 Methodology

2.1 Motivation and Overview

As illustrated in Fig. 2, we build upon the 3D nnU-Net [6], which adopts a U-shaped encoder–decoder architecture with multi-scale feature extraction and skip connections. Although nnU-Net provides strong volumetric representation capability, AD segmentation remains challenged by two key issues: (i) inter-slice inconsistency in anisotropic CTA volumes, leading to fragmented lumen predictions, and (ii) boundary ambiguity under low contrast, which degrades thin-structure delineation.

To explicitly address these limitations, we introduce BiM-GeoAttn-Net by augmenting the bottleneck of nnU-Net with two complementary modules. A Bidirectional Depth Mamba (BiM) module performs depth-axis state-space modeling to enhance long-range cross-slice coherence with near-linear complexity. Built upon this global modeling stage, a Geometry-Aware Vessel Attention (GeoAttn) module incorporates vessel-structure priors via orientation-sensitive filtering and dual attention mechanisms, refining tubular boundaries and suppressing background interference.

The cascaded BiM–GeoAttn design unifies depth-wise continuity modeling and local structure-aware refinement within a lightweight framework. The enhanced features are residually fused and forwarded to the decoder for final lumen segmentation, with training supervised by a hybrid objective to mitigate foreground–background imbalance.

2.2 Bidirectional Depth Mamba

The BiM module models long-range dependencies along the depth axis without incurring the quadratic cost of full 3D self-attention. This is particularly important for maintaining coherent lumen predictions across adjacent slices.

Let $\boldsymbol{x}_{in} \in \mathbb{R}^{B\times C\times D\times H\times W}$ denote the input feature map. We reshape spatial locations into depth-wise sequences:

$$\widetilde{\boldsymbol{x}} = \mathrm{Reshape}(\boldsymbol{x}_{in}) \in \mathbb{R}^{(BHW)\times D\times C},$$

where each spatial position forms a 1D sequence along D. A Mamba state-space model $\mathcal{M}(\cdot)$ is applied bidirectionally, where $\mathrm{Rev}(\cdot)$ reverses the sequence along the depth dimension:

$$\boldsymbol{z}^{\rightarrow} = \mathcal{M}(\widetilde{\boldsymbol{x}}), \quad \boldsymbol{z}^{\leftarrow} = \mathrm{Rev}(\mathcal{M}(\mathrm{Rev}(\widetilde{\boldsymbol{x}}))), \tag{1}$$

$$\boldsymbol{z} = \boldsymbol{z}^{\rightarrow} + \boldsymbol{z}^{\leftarrow}. \tag{2}$$

The fused representation is reshaped back and added residually:

$$\boldsymbol{x}_1 = \boldsymbol{x}_{in} + \mathrm{Reshape}^{-1}(\boldsymbol{z}). \tag{3}$$

Then a lightweight channel-mixing MLP with pre-normalization further refines the features:

$$\boldsymbol{x}' = \boldsymbol{x}_1 + \mathrm{MLP}(\mathrm{Norm}(\boldsymbol{x}_1)), \tag{4}$$

where $\mathrm{MLP}(\boldsymbol{u}) = W_2\,\delta(W_1\boldsymbol{u})$, δ denotes GELU, and r is the expansion ratio.

Based on SSM-based scanning, BiM achieves near-linear complexity with respect to D while aggregating context from both directions, thereby stabilizing inter-slice continuity and boundary prediction.

2.3 Geometry-Aware Vessel Attention

The GeoAttn module refines vessel representations by combining direction-aware filtering with spatial and channel attention. Given $\boldsymbol{X} \in \mathbb{R}^{B\times C\times D\times H\times W}$, GeoAttn outputs $\boldsymbol{Y}$ of the same shape.

To capture orientation-sensitive tubular structures, we employ three plane-aligned anisotropic convolutions together with a standard 3D convolution:

$$\boldsymbol{F}_{xy} = \mathrm{Conv}_{1\times3\times3}(\boldsymbol{X}), \quad \boldsymbol{F}_{yz} = \mathrm{Conv}_{3\times3\times1}(\boldsymbol{X}), \quad \boldsymbol{F}_{xz} = \mathrm{Conv}_{3\times1\times3}(\boldsymbol{X}), \tag{5}$$

$$\boldsymbol{F}_{3d} = \mathrm{Conv}_{3\times3\times3}(\boldsymbol{X}). \tag{6}$$

where each branch produces C_b channels. The multi-branch features are concatenated and fused via a $1\times1\times1$ convolution:

$$\boldsymbol{F} = \mathrm{Conv}_{1\times1\times1}\left([\boldsymbol{F}_{xy}, \boldsymbol{F}_{yz}, \boldsymbol{F}_{xz}, \boldsymbol{F}_{3d}]\right). \tag{7}$$

Then, a lightweight bottleneck generates a spatial attention map:

$$\boldsymbol{A} = \sigma\left(\mathrm{Conv}_{1\times1\times1}\left(\delta\left(\mathrm{IN}\left(\mathrm{Conv}_{1\times1\times1}(\boldsymbol{F})\right)\right)\right)\right), \tag{8}$$

where IN denotes instance normalization, δ denotes GELU activation, and σ denotes the sigmoid function. The attention map modulates the input through a residual gate:

$$\boldsymbol{X}_s = \boldsymbol{X} \odot (1 + \gamma\boldsymbol{A}), \tag{9}$$

where $\odot$ denotes element-wise multiplication and γ is a learnable scaling coefficient. Finally, channel attention is applied using global average pooling (GAP) and global max pooling (GMP):

$$\boldsymbol{W} = \sigma\left(\mathrm{MLP}(\mathrm{GAP}(\boldsymbol{X}_s)) + \mathrm{MLP}(\mathrm{GMP}(\boldsymbol{X}_s))\right), \tag{10}$$

$$\boldsymbol{Y} = \boldsymbol{X}_s \odot \boldsymbol{W}. \tag{11}$$

By combining plane-aligned anisotropic filtering with dual attention, GeoAttn enhances geometry-consistent vessel representations and local boundary awareness. Positioned after BiM, it complements depth-wise coherence modeling with orientation-aware boundary refinement.

2.4 Loss Function

We adopt a hybrid Dice–Cross Entropy objective to address foreground–background imbalance. Let $p_{i,c}$ denote the predicted probability of voxel i belonging to class c, and $g_{i,c} \in \{0, 1\}$ the corresponding ground truth indicator. The total loss is

$$\mathcal{L}_{\text{total}} = \lambda_{\text{dice}}\mathcal{L}_{\text{Dice}} + \lambda_{\text{ce}}\mathcal{L}_{\text{CE}}. \tag{12}$$

The soft Dice loss averaged over K classes is:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{1}{K}\sum_{c=1}^{K}\frac{2\sum_i p_{i,c}g_{i,c} + \epsilon}{\sum_i p_{i,c}^2 + \sum_i g_{i,c}^2 + \epsilon}, \tag{13}$$

where ϵ ensures numerical stability. The cross-entropy term is computed voxel-wise:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_i\sum_c g_{i,c}\log(p_{i,c}). \tag{14}$$

We set $\lambda_{\text{dice}} = \lambda_{\text{ce}} = 1$ to balance global shape alignment and voxel-level accuracy.

3 Experiments

This section describes the experimental setup, quantitative evaluation, and ablation studies used to assess the effectiveness of our BiM-GeoAttn-Net.

3.1 Dataset

We train and evaluate our model on a curated dataset for Stanford Type-B Aortic Dissection (TBAD) vessel segmentation, termed *Dataset500_ VesselTBAD*. The dataset combines cases from two publicly available sources: (1) **ImageTBAD** [1]: This dataset provides expert-annotated 3D CTA scans with labels for the true lumen (TL), false lumen (FL), and false lumen thrombosis (FLT). We selected 32 cases based on annotation quality and pathological diversity. (2)

TBD-CTA [9]: A large-scale TBAD CTA dataset covering varying degrees of thrombosis and dissection extent. We included 39 cases with complete TL/FL annotations.

In total, 71 cases are split at the patient level into training, validation, and test sets with $N_{\text{train}} = 50$, $N_{\text{val}} = 7$, and $N_{\text{test}} = 14$. Each subset contains samples from both sources: ImageTBAD (22/3/7) and TBD-CTA (28/4/7) for train/val/test, respectively. The patient-level split ensures that scans from the same subject do not appear across different subsets.

We formulate the task as binary segmentation, extracting the vessel foreground (TL+FL) from the background. FLT annotations are not treated as an independent class in this study.

3.2 Implementation Details and Evaluation Metrics

Our method was implemented in PyTorch within the nnU-Net framework and trained on NVIDIA RTX 3090 GPUs with CUDA support. Automatic mixed-precision (AMP) training [10] was adopted to improve computational efficiency and memory usage. We followed the nnU-Net 3d_fullres configuration and employed a patch-based training scheme with randomly cropped $128 \times 128 \times 128$ patches and a batch size of 2. The network was trained for 400 epochs using SGD with a cosine annealing learning rate schedule. Weighted deep supervision was applied at multiple resolution levels, and the final model was selected based on the best validation performance.

Segmentation performance was evaluated using Dice Similarity Coefficient (Dice), Intersection over Union (IoU), Recall, Precision, and 95% Hausdorff Distance (HD95). Dice, IoU, Recall, and Precision measure volumetric overlap accuracy (↑), while HD95 assesses boundary agreement (↓). All metrics were computed on the test set and averaged across cases. Statistical significance was assessed using a paired Wilcoxon signed-rank test on case-wise Dice scores between BiM-GeoAttn-Net and nnU-Net, with $p < 0.05$ considered significant.

3.3 Comparison with Representative Methods

We compare BiM-GeoAttn-Net with representative 3D segmentation models spanning CNN-, Transformer-, and SSM-based paradigms, including Attention U-Net [12], nnU-Net v2 [6], Swin-UNet [2], SegFormer3D [13], and Mamba-UNet [15]. All methods are trained and evaluated under identical settings for fair comparison.

Quantitative Results. As shown in Table 1, our method achieves the best overall performance across overlap metrics. It attains the highest Dice (93.35%), IoU (87.53%), Recall (94.85%), and Precision (92.04%). Compared with nnU-Net, Dice and IoU improve by 2.51% and 3.89%, respectively, with a statistically significant case-wise Dice improvement ($p = 0.021$). Regarding boundary accuracy, Attention U-Net achieves the lowest HD95 (10.77 mm), while

Table 1. Quantitative comparison with representative methods on the AD dataset. Best results are shown in bold. (↑ higher is better; ↓ lower is better.)

Models	Segmentation Performance					Training Cost	
	Prec. ↑	Rec. ↑	Dice ↑	IoU ↑	HD95 (mm) ↓	Time (min/epoch) ↓	Memory (GB) ↓
Attention U-Net	89.71	90.76	90.22	82.26	**10.77**	3.2	8.3
nnU-Net	90.46	91.49	90.84	83.64	18.51	2.8	7.6
Swin-UNet	87.78	91.73	89.62	81.62	20.01	4.1	10.7
SegFormer3D	87.53	88.12	88.43	79.30	24.15	4.6	11.2
Mamba-UNet	88.83	91.52	89.43	83.12	12.20	3.5	8.9
BiM-GeoAttn-Net	**92.04**	**94.85**	**93.35**	**87.53**	12.36	2.9	8.0

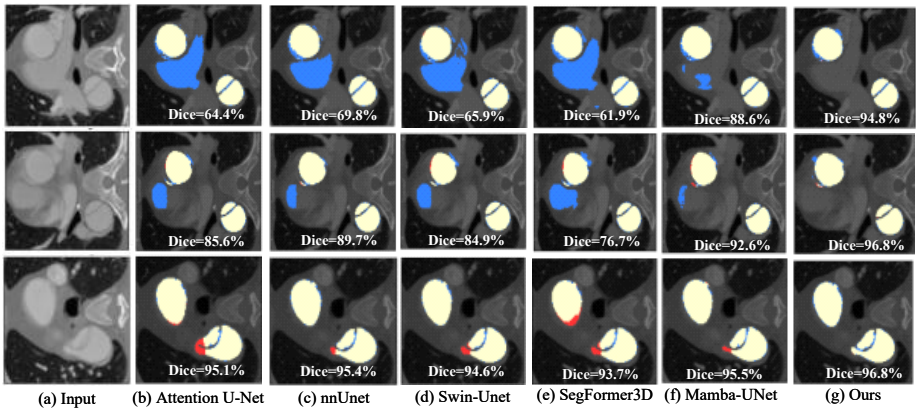


Fig. 3. Qualitative comparison on representative AD test cases. Yellow indicates correct overlap, blue denotes false positives, and red denotes false negatives. Compared with CNN-, Transformer-, and SSM-based baselines, our BiM-GeoAttn-Net shows improved lumen overlap and depth-wise continuity.

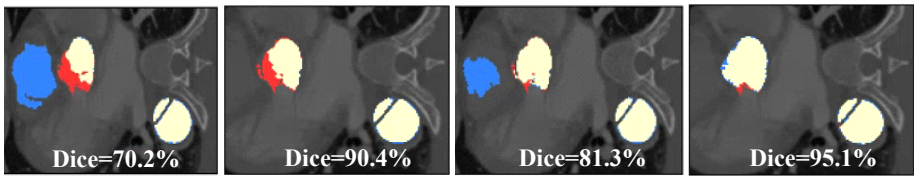
our model (12.36 mm) remains competitive and substantially outperforms nnU-Net (18.51 mm) and SegFormer3D (24.15 mm). These results indicate that our method primarily improves volumetric overlap and inter-slice consistency, while achieving competitive rather than the best HD95 performance.

Computational Efficiency. As reported in Table 1, our model requires 2.9 min/epoch and 8.0 GB memory, which is comparable to nnU-Net (2.8 min, 7.6 GB). In contrast, Transformer-based models such as SegFormer3D and Swin-UNet incur higher computational cost. This demonstrates that the proposed method achieves improved accuracy without introducing substantial overhead.

Qualitative Results. Fig. 3 presents representative visual comparisons. Transformer-based methods tend to produce over-segmentation around ambiguous boundaries, while nnU-Net and Mamba-UNet occasionally miss fine vascular structures in low-contrast regions. In the shown examples, our model yields

Table 2. Ablation results on the AD dataset. Best results are shown in bold. (↑ higher is better; ↓ lower is better.)

Models	Dice (%) ↑	IoU (%) ↑	Recall (%) ↑	Precision (%) ↑	HD95 (mm) ↓
B	89.77	81.60	92.11	87.73	29.08
B+BiM	91.12	83.69	94.41	89.65	20.35
B+GeoAttn	90.67	83.25	93.17	88.46	25.48
B+BiM+GeoAttn	**93.35**	**87.53**	**94.85**	**92.04**	**12.36**



(a) B (b) B+BiM (c) B+GeoAttn (d) B+BiM+GeoAttnB

Fig. 4. Qualitative ablation comparison. Yellow indicates correct overlap, blue false positives, and red false negatives. The combined model yields the most consistent continuity and boundary delineation.

fewer false positives and false negatives, resulting in more consistent lumen continuity. These observations are consistent with the quantitative improvements in overlap metrics and the competitive HD95 performance.

3.4 Ablation Study

To assess the contribution of each component, we evaluate four configurations: (1) baseline (B), (2) B+BiM, (3) B+GeoAttn, and (4) B+BiM+GeoAttn (*i.e.*, BiM-GeoAttn-Net). Quantitative results are summarized in Table 2.

Quantitative Analysis. The baseline achieves a Dice of 89.77% with relatively high HD95 (29.08 mm), indicating limited boundary precision. Incorporating BiM improves Dice to 91.12% and reduces HD95 to 20.35 mm, reflecting enhanced cross-slice continuity. Adding GeoAttn alone yields moderate gains (Dice 90.67%) and improved Recall, suggesting better local structure sensitivity. Combining both modules delivers the best performance, with Dice reaching 93.35% and HD95 decreasing substantially to 12.36 mm. These results demonstrate the complementarity of the two components: BiM strengthens depth-wise global coherence, while GeoAttn refines directional vessel boundaries.

Qualitative Analysis. Fig. 4 further illustrates the effect of each module. The baseline exhibits leakage and fragmented predictions. BiM reduces false positives and improves cross-slice consistency, whereas GeoAttn enhances boundary adherence but may not fully suppress large-scale interference. The combined model

achieves the most accurate delineation, with fewer missed regions and improved structural continuity, corroborating the quantitative findings.

4 Conclusions

We presented BiM-GeoAttn-Net, a lightweight framework for AD CTA segmentation that couples depth-wise long-range modeling with geometry-aware vessel refinement. By integrating BiM and GeoAttn at the bottleneck, the method improves depth-axis context modeling and plane-aligned structure refinement for slice-wise discontinuity and low-contrast vessel ambiguity. Experiments and ablations on the current multi-source AD CTA dataset show improved volumetric overlap and depth-wise coherence over representative 3D baselines, with competitive boundary accuracy. The main limitations are the relatively small single-split evaluation, the use of established anisotropic filtering and attention operations in GeoAttn, and the non-best HD95, which motivate future multi-center validation and boundary-outlier refinement.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (62006165) and the Natural Science Foundation of Sichuan Province (2025ZNSFSC1477).

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

References

1. Abaid, A., Ilancheran, S., Iqbal, T., Hynes, N., Ullah, I.: Exploratory analysis of type b aortic dissection (tbad) segmentation in 2d cta images using various kernels. Computerized Medical Imaging and Graphics **118**, 102460 (2024)
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. Cao, L., Shi, R., Ge, Y., Xing, L., Zuo, P., Jia, Y., Liu, J., He, Y., Wang, X., Luan, S., et al.: Fully automatic segmentation of type b aortic dissection from cta images enabled by deep learning. European journal of radiology **121**, 108713 (2019)
4. Chen, D., Zhang, X., Mei, Y., Liao, F., Xu, H., Li, Z., Xiao, Q., Guo, W., Zhang, H., Yan, T., et al.: Multi-stage learning for segmentation of aortic dissections using a prior aortic anatomy simplification. Medical image analysis **69**, 101931 (2021)
5. Hu, Y., Mu, N., Liu, L., Zhang, L., Jiang, J., Li, X.: Slimmable transformer with hybrid axial-attention for medical image segmentation. Computers in biology and medicine **173**, 108370 (2024)
6. 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)
7. Li, Z., Feng, J., Feng, Z., An, Y., Gao, Y., Lu, B., Zhou, J.: Lumen segmentation of aortic dissection with cascaded convolutional network. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 122–130. Springer (2018)

8. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
9. Mayer, C., Pepe, A., Hossain, S., Karner, B., Arnreiter, M., Kleesiek, J., Schmid, J., Janisch, M., Hannes, D., Fuchsjäger, M., et al.: type b aortic dissection cta collection with true and false lumen expert annotations for the development of ai-based algorithms. Scientific data **11**(1), 596 (2024)
10. Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., et al.: Mixed precision training. In: International Conference on Learning Representations (2018)
11. Mu, N., Lyu, Z., Rezaeitaleshmahalleh, M., Tang, J., Jiang, J.: An attention residual u-net with differential preprocessing and geometric postprocessing: Learning how to segment vasculature including intracranial aneurysms. Medical image analysis **84**, 102697 (2023)
12. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
13. Perera, S., Navard, P., Yilmaz, A.: Segformer3d: an efficient transformer for 3d medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4981–4988 (2024)
14. Song, R., Liu, L., Zhang, Y.N., Wang, C., Li, X., Mu, N.: Vfgs-net: Frequency-guided state-space learning for topology-preserving retinal vessel segmentation. In: Proceedings of the 2026 International Conference on Multimedia Retrieval. pp. 2210–2218 (2026)
15. 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)
16. Xie, Z., Huang, X., Chen, S., Shi, Y.: Vesselmamba: 3d vessel segmentation in cta images using mamba with enhanced spatial-channel attention. Biomedical Signal Processing and Control **110**, 107982 (2025)