

HybridMamba: A Dual-domain Mamba for 3D Medical Image Segmentation

Weitong Wu¹, Zhaohu Xing¹, Jing Gong^{2,3}, Qin Peng^{2,3}, and Lei Zhu^{1,4} ✉

¹ The Hong Kong University of Science and Technology (Guangzhou),
Guangzhou, China

² Department of Radiology, Fudan University Shanghai Cancer Center,
Shanghai, China

³ Department of Oncology, Shanghai Medical College, Fudan University,
Shanghai, China

⁴ The Hong Kong University of Science and Technology, Hong Kong, China
`leizhu@ust.hk`

Abstract. In the domain of 3D biomedical image segmentation, Mamba exhibits the superior performance for it addresses the limitations in modeling long-range dependencies inherent to CNNs and mitigates the abundant computational overhead associated with Transformer-based frameworks when processing high-resolution medical volumes. However, attaching undue importance to global context modeling may inadvertently compromise critical local structural information, thus leading to boundary ambiguity and regional distortion in segmentation outputs. Therefore, we propose the HybridMamba, an architecture employing dual complementary mechanisms: 1) a feature scanning strategy that progressively integrates representations both axial-traversal and local-adaptive pathways to harmonize the relationship between local and global representations, and 2) a gated module combining spatial-frequency analysis for comprehensive contextual modeling. Besides, we collect a multi-center CT dataset related to lung cancer. Experiments on MRI and CT datasets demonstrate that HybridMamba significantly outperforms the state-of-the-art methods in 3D medical image segmentation.

Keywords: State space model · Mamba · Frequency and spatial feature modeling · 3D medical image segmentation.

1 Introduction

In clinical diagnostic workflows, achieving voxel-level precision in pathological region segmentation from 3D medical imaging modalities (e.g., CT, MRI) constitutes a critical prerequisite for quantitative disease characterization [25]. While deep learning architectures have demonstrated remarkable potential in medical image segmentation, thereby reducing inter-observer variability and clinician workload, there still exist fundamental limitations. Traditional Convolutional Neural Networks (CNNs) suffer from structural constraints in modeling long-range dependencies within medical image segmentation due to their intrinsic

locality [3]. Deep learning architectures have shown great promise in medical image segmentation by minimizing inter-observer variability and clinician workload. However, traditional Convolutional Neural Networks (CNNs) are fundamentally limited by structural constraints that hinder their ability to model long-range dependencies due to their inherent locality [3]. Recent attempts like 3D UX-Net [12] endeavor to mitigate this issue by expanding effective receptive fields through large-kernel convolutions, but still have constraints in modeling global relationships.

The emergence of Transformer architecture [19] has notably enhanced the modeling of global context through self-attention mechanisms, effectively addressing the limitations of CNNs in capturing long-range dependencies. For instance, UNETR [8] merges the Vision Transformer (ViT) [4] into the encoder to capture contextual information and utilizes a convolutional decoder with multi-scale skip connections to learn local features together for 3D medical image segmentation. Similarly, SwinUNETR [7] employs the SwinTransformer [14] in its encoder to enable efficient extraction of features at multiple resolutions. Nonetheless, these Transformer-based methods imposes significant scalability constraints for high-resolution 3D medical imaging analysis. This computational bottleneck manifests particularly in memory-intensive segmentation tasks like biomedical image segmentation, resulting in suboptimal throughput rates during both training and inference phases despite recent hardware advancements.

Emerging from State Space Model, Mamba [5, 21, 24] presents an innovative approach to long-range dependency learning through selective mechanism and hardware-efficient algorithms. This paradigm achieves CNN-level training stability coupled with RNN-like inference efficiency [6], all maintained within linear computational complexity. Current advances demonstrate growing applications of Mambas in medical imaging analysis: In 2D contexts, U-Mamba [15] augments standard nnUNet frameworks [10] by integrating directional scanning modules into encoder pathways, significantly enhancing the performance in medical image segmentation. Moreover, Swin-U-Mamba [13] strategically incorporates large-scale ImageNet pretrained representations, thereby enhancing Mamba’s efficacy in clinical image segmentation scenarios. For 3D medical image analysis, Seg-Mamba [27] pioneers voxel-level contextual modeling through its tri-orientated Mamba (ToM) module for feature modeling and gated spatial convolution (GSC) module for the representation enhancement of spatial features. However, Existing 2D methodologies fail to address cross-slice correlations inherent in volumetric datasets effectively; meanwhile SegMamba exhibit suboptimal coordination between localized feature retention and global context integration when processing hierarchically partitioned image patches.

In this paper, we propose HybridMamba, a hierarchical architecture addressing two fundamental challenges in 3D medical image analysis: 1) multi-resolution contextual preservation across spatial scales; and 2) synergistic integration of frequency-spatial feature representations. To our best knowledge, this is the first method to deploy the feature extraction strategy in both the frequency domain and the spatial domain to facilitate more accurate and robust rep-

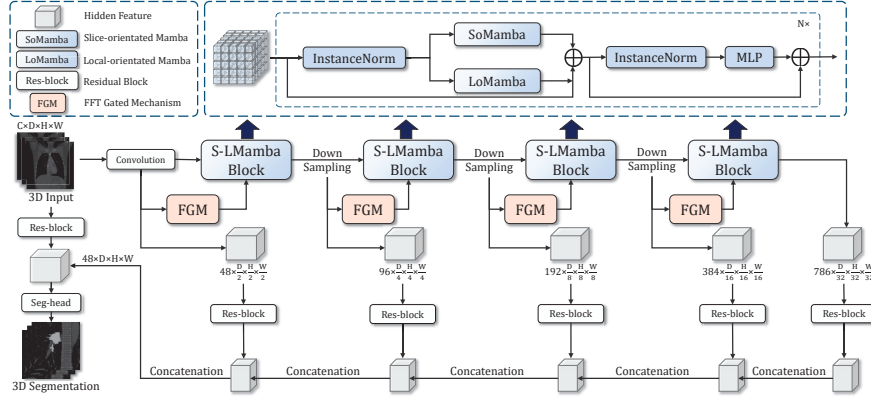


Fig. 1. The overview of the proposed HybridMamba. The encoder consists of a multiple S-LMamba blocks for balancing local and global features and an FFT Gate Mechanism (FGM) for merging the features from spatial and frequency domain dynamically according to the characteristics of different layers.

representations on multiple 3D medical image segmentation tasks. Consider the non-negligibility of relationships between global and local information, we design a fused ergodic mechanism named SoMamba (Slice-oriented Mamba) and LoMamba (Local-oriented Mamba) to increase the sequential modeling of 3D features. Following this, we further propose the gated mechanism that dynamically weights frequency-transformed features against spatially encoded patterns prior to cascaded Mamba processing stages. We verify the superior performance of the proposed HybridMamba on both public MRI dataset and our collected CT dataset. The results of experiments showcase the efficiency of our method.

2 Method

In this section, we introduce the implementation approach of HybridMamba, which is built upon SegMamba [27] with the particular improvements on encoder. We incorporate the frequency domain feature into encoder modeling by utilizing the gated mechanism before transforming from two kinds of traversing pathways to sequence. Fig. 1 demonstrates the overview of HybridMamba. The details of the encoder will be further described in this section.

2.1 Slice-Local Mamba (S-LMamba) Block

It is of great necessity to utilize Mamba to model long-range dependencies within 3D high-resolution biomedical image segmentation [20, 26, 28]. While transformer architectures [22, 23] effectively capture global information, they impose substantial computational costs when processing excessively long feature sequences. Accordingly, Mamba is utilized especially for high-resolution biomedical image

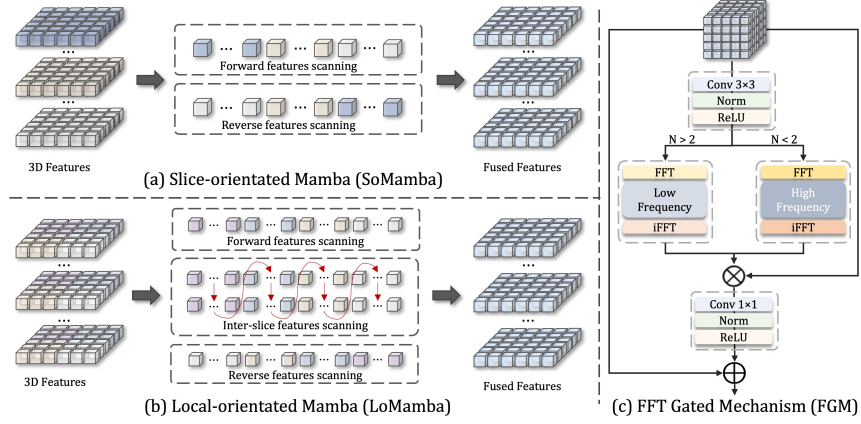


Fig. 2. The left side (a) and (b) shows the detailed implementation in SoMamba and LoMamaba. The right side (c) demonstrates the specific layers in the FGM.

field. It fully leverages long-distance modeling while being exponentially more computationally efficient than the Transformer framework. However, prioritizing long-distance dependencies overlooks the extraction of local segmentation information within and between slices, potentially leading to incoherent partitioning and inadequate semantic understanding in the model. Drawing inspiration from the scanning method presented in Local Mamba [9], we design an S-LMamba Block to partition all slices of the medical image, generating windows of a desired size according to different scales in each layer. We also recognize the importance of long-distance dependencies throughout the training process. Therefore, we introduce a slice-oriented traversing strategy to enhance the learning of contextual information within the medical image. This approach allows Mamba to extract more relevant segmentation information from the sequences.

The encoder illustrated in Fig. 1 is composed of multiple S-LMamba blocks and FGM modules. After the initial preprocessing of the input 3D volume $I \in \mathbb{R}^{48 \times D \times H \times W}$ using a large convolutional layer with a kernel size of $7 \times 7 \times 7$, the resulting 3D feature $x_0 \in \mathbb{R}^{48 \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}}$ is processed through the S-LMamba blocks and FGM modules, along with subsequent down-sampling layers. For the n^{th} S-LMamba Block, the computational operation can be defined as:

$$\tilde{x}_n^l = \text{MLP}(\text{IN}(\text{SoMamba}(\text{LN}(\hat{x}_n^l)) + \text{LoMamba}(\text{LN}(\hat{x}_n^l)) + \hat{x}_n^l)) + \tilde{x}_n^l, \quad (1)$$

where the SoMamba and LoMamba refer to the proposed Slice-oriented and Local-oriented Mamba module, respectively, which will be discussed next. $l \in \{0, 1, \dots, n-1\}$, LN refers to layer normalization, and IN refers to instance normalization. MLP refers to the multi-layer perceptron for enriching the feature representation.

Slice-oriented Mamba (SoMamba) and Local-oriented Mamba (Lo-Mamba). The Mamba layer captures feature dependencies by flattening the

3D features into a 1D sequence. Managing the order of this flattening process is crucial, as it directly impacts the model’s learning efficiency. To optimize the arrangement of the flattened 1D sequence, we introduce the Slice-orientated Mamba (SoMamba) and Local-orientated Mamba (LoMamba) module illustrated in part (a) and (b) of Fig. 2.

$$\text{SoMamba}(x) = \text{Mamba}(x_f) + \text{Mamba}(x_r), \quad (2)$$

$$\text{LoMamba}(x) = \text{Mamba}(x_{lf}) + \text{Mamba}(x_{lr}) + \text{Mamba}(x_{ls}), \quad (3)$$

where Mamba denotes the Mamba layer to model the global information within the sequence. f and r in Eq. 2 denote forward and reverse direction respectively. lf , lr , and ls in Eq.3 refer to local-window forward direction, local-window reverse direction, and local-window across slices, correspondingly.

For SoMamba, we design a traversing path that spans the entire slice in both forward and reverse directions to compute feature dependencies, aiming to effectively capture the global information inherent in high-dimensional features. Additionally, we observe the presence of short-distance dependencies within the overall medical image, particularly as lesion areas may constitute a smaller proportion of the total. To address this, we design a Local modeling strategy known as LoMamba to focus on the adjacent pixels with the same semantic region, which empower the model to better aggregate the features because of more compact physical position of key information after flattening into a 1D sequence.

Fig. 3 showcases the distinction between normal scanning and local scanning modes using the example of the local flattening sequence with a window size of three. In practice, the size of the local window is determined dynamically by the proportion and location of key information areas within the overall medical image. To be specific, the sequence through the standard flattening method exhibits a considerable distance between neighboring pixels that contain important segmentation information. In contrast, the continuous key segmentation pixels flattened by local windows of size three manifesting a strong distance coherence to a certain extent.

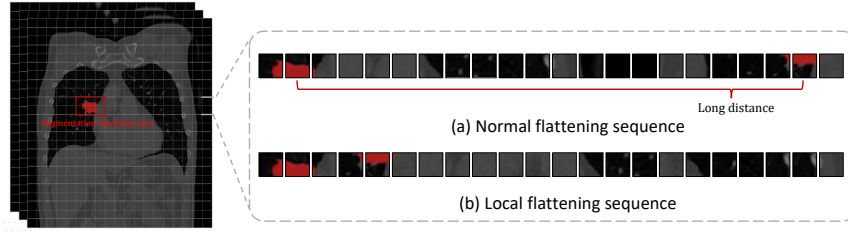


Fig. 3. Take 3×3 local window size around segmentation key area on one slice as an example. Sequence (a) showcases the normal flatten method, which traverse the whole image patches from beginning to end. Sequence (b) demonstrates the local flatten method.

2.2 FFT Gated Mechanism (FGM)

Fast Fourier Transform (FFT) [2] is leveraged to calculate the frequency values from spatial domain. In cases of CT images with poor contrast and high noise, as well as MRI images affected by artifacts, frequency information from high level and low level can correspondingly provide some boundary and shape cues to the model [30], which enables it to capture a wider range of feature representations, thus increasing the robustness. Furthermore, as highlighted in [29], the deeper layers of deep neural network tend to retain more low-frequency information. To capitalize on this, we design the FFT Gated Mechanism (GSM) to integrate the frequency and spatial features to enhance the performance of Mamba model. As illustrated in subgraph (c) of Fig. 2, the input 3D features are first fed into a convolution block, which contains a convolution, a normalization, and an activation layer. Then, the feature are transformed into the Fourier domain according to different layers, being extracted high and low frequency leveraging learnable thresholds within a filter. Subsequently, conducting inverse FFT to transform the required frequency feature back. Finally, a convolution block is used to further fuse the frequency features and spatial features after gated multiplication, while a residual connection is utilized to reuse the input features.

$$x_{\text{fre}} = \text{IFFT}(\text{Filter}(\text{FFT}(x))), \quad x_{\text{out}} = x_s * \text{gate} + x_{\text{fre}} * (1 - \text{gate}), \quad (4)$$

$$\text{Filter} = \begin{cases} x * (|x| < f_{\text{low}}), & X \in \text{low-level frequency}, \\ x * (|x| > f_{\text{high}}), & X \in \text{high-level frequency}, \end{cases} \quad (5)$$

where $\text{gate} = \text{Conv}_{3 \times 3 \times 3}(x_{\text{ft}}, x_s)$ is used to get the prior embedding to realize the coordination of features in spatial domain and frequency domain. $f_{\text{low}}, f_{\text{high}}$ denote the learnable thresholds of the filter with the initial values set to 0.1 and 0.9 respectively.

3 Experiments

3.1 Datasets and Implementation Details

BraTS2023 Dataset [1,11,16]. This dataset comprises 1,251 3D MRI volumes of the brain. Each volume is presented in four different imaging modalities: T1, T1Gd, T2, and T2-FLAIR. Three segmentation targets are identified within each volume: Whole Tumor (WT), Enhancing Tumor (ET), and Tumor Core (TC).

Lung Cancer (LC) Dataset. We gathered a collection of 828 3D chest CT scans from multiple centers, focusing on cases of central type lung carcinoma with small lesions. These cases pose inherent challenges in detection and analysis, as they appear as subtle anomalies within the imaging data. Each volume contains a single segmentation target, with the dominant lesion accurately annotated for each case. The visualization of this dataset is depicted in Fig 4.

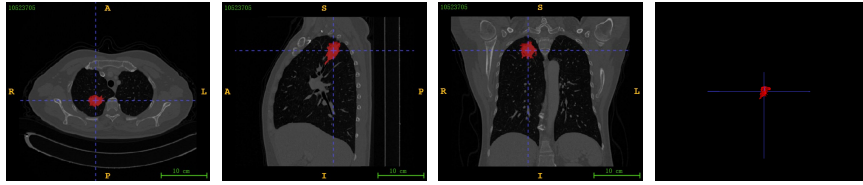


Fig. 4. The data visualization for LC dataset, which highlights the challenges posed by small lesions, which can be difficult to detect and analyze.

Evaluation and Metrics. In line with established approaches [27], we employ the Dice score (Dice) and Hausdorff Distance (HD95) to quantitatively assess our network’s performance and compare it with state-of-the-art (SOTA) methods.

Implementation Details. Our model is developed using PyTorch 2.0.1 with CUDA 11.8 and Monai 1.3.0. During training, we apply a random crop size of $128 \times 128 \times 128$ and utilize a batch size of 2 per GPU for each dataset. We employ cross-entropy loss for all experiments, using an SGD optimizer with a polynomial learning rate scheduler (initial learning rate set at $1e-4$ and a decay rate of $3e-5$). Each dataset undergoes 1000 training epochs, and we incorporate the following data augmentations: additive brightness, gamma correction, rotation, scaling, mirroring, and elastic deformation. All experiments are conducted on a cloud computing platform equipped with four NVIDIA GeForce RTX 4090 GPUs. For each dataset, we randomly assign 70% of the 3D volumes for training, 10% for validation, and the remaining 20% for testing.

3.2 Experimental Results

Compared Methods. We evaluate our network by comparing it to six state-of-the-art (SOTA) 3D image segmentation methods, which includes three CNN-based approaches (SegResNet [17], UX-Net [12], MedNeXt [18]), two transformer-

Table 1. Quantitative comparison on BraTS2023 dataset. The bold value denotes the best performance.

Methods	BraTS2023								LC	
	WT		TC		ET		Avg		Lung Cancer	
	Dice↑	HD95↓	Dice↑	HD95↓	Dice↑	HD95↓	Dice↑	HD95↓	Dice↑	HD95↓
SegresNet [17]	92.02	4.07	89.10	4.08	83.66	3.88	88.26	4.01	71.56	55.03
UX-Net [12]	93.13	4.56	90.03	5.68	85.91	4.19	89.69	4.81	72.63	59.06
MedNeXt [18]	92.41	4.98	87.75	4.67	83.96	4.51	88.04	4.72	57.76	111.47
UNETR [8]	92.19	6.17	86.39	5.29	84.48	5.03	87.68	5.49	65.12	101.7
SwinUNETR [7]	92.71	5.22	87.79	4.42	84.21	4.48	88.23	4.70	65.83	89.27
SegMamba [27]	93.61	3.37	92.65	3.85	87.71	3.48	91.32	3.56	72.36	60.33
Ours	94.10	3.78	92.84	3.30	88.83	3.35	91.92	3.48	75.34	44.52

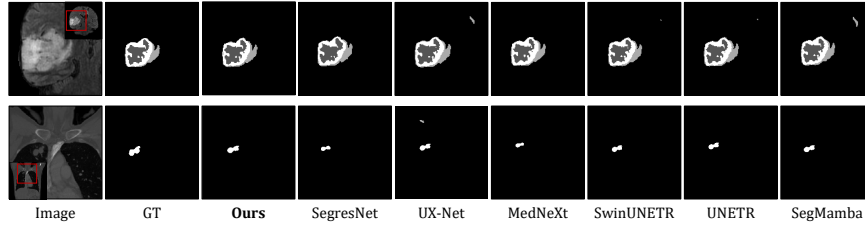


Fig. 5. Visual comparisons of proposed HybridMamba and state-of-the-art methods.

based methods (UNETR [8] and SwinUNETR [7]), and one Mamba-based technique (SegMamba [27]).

Quantitative Comparisons. Table 1 summarizes the Dice score and HD95 for each modalities on BraTS2023 and LC dataset and the total average scores. For BraTS2023 dataset, SegMamba, the Mamba-based method, achieves the best performance among the comparison methods, with an average Dice of 91.32% and an average HD95 of 3.56. In comparison, our HybridMamba achieves the highest Dice of 94.10%, 92.84%, and 88.83% on WT, TC, and ET, respectively, and the best HD95 with 3.30 on TC and 3.35 on ET except for 3.78 on WT. All in all, the total average scores shows the best segmentation robustness. In addition, our HybridMamba outperforms the SOTA method exceeding 2.98% for Dice and 15.81 for HD95 on LC dataset, getting 75.34% and 44.52% respectively. This proves the most effectiveness of HybridMamba compared to other approaches.

Visual Comparisons. We choose six comparative methods for visual assessment on two datasets to evaluate image segmentation performance. As shown in Fig. 5, our HybridMamba effectively delineates the boundary of each tumor region in the BraTS2023 dataset. Similarly, our approach successfully identifies cancerous areas in the LC dataset. The segmentation results demonstrate enhanced consistency compared to other state-of-the-art techniques.

Ablation Studies. Table 2 confirms the effectiveness of both S-LMamba (M1) and FGM (M2) modules on the LC dataset. S-LMamba improves SegMamba’s Dice by 1.47% to 73.83% and reduces HD95 by 9.45 to 69.78. FGM further boosts Dice to 74.98% and reduces HD95 by 18.35 to 51.43. HybridMamba, combining both modules, achieves optimal results with 75.34% Dice and 44.52 HD95.

Table 2. Ablation study for different modules on LC dataset.

Methods	Modules		Dice \uparrow	HD95 \downarrow
	S-LMamba	FGM		
SegMamba			72.36	69.78
M1	✓		73.83	60.33
M2		✓	74.98	51.43
Ours	✓	✓	75.34	44.52

4 Conclusion

In this work, we have developed HybridMamba to enhance the 3D biomedical segmentation task. Specifically, our network makes two primary contributions. First, we devise the S-LMamba block to effectively balance the modeling of global and local dependencies. Second, we aggregate frequency features with spatial features to enhance the representation of the model. Experimental results on two datasets demonstrate that our framework clearly outperforms SOTA methods in terms of 3D medical image segmentation task.

Acknowledgments. This research is supported by the Guangdong Science and Technology Department (No. 2024ZDZX2004), the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things(No.2023B1212010007), the National Natural Science Foundation of China (No. 82001903), the Natural Science Foundation of Shanghai (No. 25ZR1402077 and 21ZR1414200), and the Excellent Young Talents Project of Shanghai Public Health Three-year (2023-2025) Action Plan (No. GWVI-11.2-YQ48).

Disclosure of Interests. The authors declare that they have no competing interests.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* **4**(1), 1–13 (2017)
2. Bergland, G.D.: A guided tour of the fast fourier transform. *IEEE spectrum* **6**(7), 41–52 (1969)
3. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306* (2021)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020)
5. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752* (2023)
6. Gu, A., Dao, T., Ermon, S., Rudra, A., Ré, C.: Hippo: Recurrent memory with optimal polynomial projections. *Advances in neural information processing systems* **33**, 1474–1487 (2020)
7. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: *International MICCAI brainlesion workshop*. pp. 272–284. Springer (2021)

8. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
9. Huang, T., Pei, X., You, S., Wang, F., Qian, C., Xu, C.: Localmamba: Visual state space model with windowed selective scan. arXiv preprint arXiv:2403.09338 (2024)
10. 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)
11. Kazerooni, A.F., Khalili, N., Liu, X., Haldar, D., Jiang, Z., Anwar, S.M., Albrecht, J., Adewole, M., Anazodo, U., Anderson, H., et al.: The brain tumor segmentation (brats) challenge 2023: focus on pediatrics (cbtn-connect-dipgr-asnr-miccai brats-peds). ArXiv pp. arXiv-2305 (2024)
12. Lee, H.H., Bao, S., Huo, Y., Landman, B.A.: 3d ux-net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation. arXiv preprint arXiv:2209.15076 (2022)
13. Liu, J., Yang, H., Zhou, H.Y., Xi, Y., Yu, L., Li, C., Liang, Y., Shi, G., Yu, Y., Zhang, S., et al.: Swin-umamba: Mamba-based unet with imagenet-based pretraining. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 615–625. Springer (2024)
14. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
15. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
16. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* **34**(10), 1993–2024 (2014)
17. Myronenko, A.: 3d mri brain tumor segmentation using autoencoder regularization. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4. pp. 311–320. Springer (2019)
18. Roy, S., Koehler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., Jaeger, P.F., Maier-Hein, K.H.: Mednext: transformer-driven scaling of convnets for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 405–415. Springer (2023)
19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
20. Wang, H., Chen, J., Zhang, S., He, Y., Xu, J., Wu, M., He, J., Liao, W., Luo, X.: Dual-reference source-free active domain adaptation for nasopharyngeal carcinoma tumor segmentation across multiple hospitals. *IEEE Transactions on Medical Imaging* (2024)
21. Wang, H., Chen, Y., Chen, W., Xu, H., Zhao, H., Sheng, B., Fu, H., Yang, G., Zhu, L.: Serp-mamba: Advancing high-resolution retinal vessel segmentation with selective state-space model. arXiv preprint arXiv:2409.04356 (2024)

22. Wang, H., Yang, G., Zhang, S., Qin, J., Guo, Y., Xu, B., Jin, Y., Zhu, L.: Video-instrument synergistic network for referring video instrument segmentation in robotic surgery. *IEEE Transactions on Medical Imaging* (2024)
23. Wu, H., Yang, Y., Aviles-Rivero, A.I., Ren, J., Chen, S., Chen, H., Zhu, L.: Semi-supervised video desnowing network via temporal decoupling experts and distribution-driven contrastive regularization. In: *European Conference on Computer Vision*. pp. 70–89. Springer (2024)
24. Wu, H., Yang, Y., Xu, H., Wang, W., Zhou, J., Zhu, L.: Rainmamba: Enhanced locality learning with state space models for video deraining. In: *Proceedings of the 32nd ACM International Conference on Multimedia*. pp. 7881–7890 (2024)
25. Xing, Z., Wan, L., Fu, H., Yang, G., Yang, Y., Yu, L., Lei, B., Zhu, L.: Diff-unet: A diffusion embedded network for robust 3d medical image segmentation. *Medical Image Analysis* p. 103654 (2025)
26. Xing, Z., Wan, L., Fu, H., Yang, G., Zhu, L.: Diff-unet: A diffusion embedded network for volumetric segmentation. *arXiv preprint arXiv:2303.10326* (2023)
27. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 578–588. Springer (2024)
28. Xing, Z., Zhu, L., Yu, L., Xing, Z., Wan, L.: Hybrid masked image modeling for 3d medical image segmentation. *IEEE Journal of Biomedical and Health Informatics* (2024)
29. Xu, Z.Q.J., Zhang, Y., Luo, T., Xiao, Y., Ma, Z.: Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523* (2019)
30. Zhou, Y., Huang, J., Wang, C., Song, L., Yang, G.: Xnet: Wavelet-based low and high frequency fusion networks for fully-and semi-supervised semantic segmentation of biomedical images. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 21085–21096 (2023)