PolyMamba: Spatial-prior Guided Mamba for Polyp Segmentation with High-Frequency Enhancement

Renyu Fu[1], Shurui Hu[1], Xiao Zheng[2], Chang Tang[3⋆], and Xinwang Liu[4]

[1] School of Computer, China University of Geosciences, Wuhan 430074, China
https://github.com/Parker-rfu/PolyMamba
[2] School of Computer Science, Hubei University of Technology, Wuhan 430068, China
[3] School of Software Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
tangchang@hust.edu.cn
[4] School of Computer, National University of Defense Technology, Changsha 410073, China.

Abstract. Accurate polyp segmentation during colonoscopy is crucial for the early detection and timely intervention of colorectal cancer. Recently, Mamba, a State Space Model, has gained significant attention in polyp segmentation due to its remarkable ability to model long-range dependencies with linear computational complexity. However, Mamba-based methods face two key challenges: (1) their fixed scanning pattern limits the capture of dynamic spatial context, impairing the precise localization of irregular polyps; (2) during the calculation process, the high-frequency information that is crucial to local details is weakened, and the blurred mid-frequency information becomes dominant, thereby reducing the boundary accuracy. To overcome these limitations, we propose PolyMamba, a novel framework that integrates spatial priors while enhancing high-frequency information for more accurate polyp segmentation. Specifically, our framework introduces a **Spatial-Prior Guided** module, which leverages explicit spatial priors extracted from Transformer-based methods to counteract the local perception bias caused by Mamba's fixed scanning pattern. Additionally, we design a **Dual-Gate Frequency Enhancement** module, which applies two Gaussian filters to generate spectra with different high-frequency thresholds, and uses the difference between them as an attention map to selectively enhance high-frequency features, thereby refining the polyp boundaries. Comprehensive experiments on five widely used polyp segmentation datasets demonstrate that PolyMamba not only surpasses existing state-of-the-art techniques but also provides a novel frequency-domain perspective, offering new insights into improving segmentation performance.

Keywords: Polyp Segmentation · Mamba · Spatial Prior · Frequency Domain.

⋆ Corresponding author

1 Introduction

Colon polyps are abnormal tissues that grow on the inner wall of the colon. Accurate segmentation of polyps in colonoscopy images is crucial for the early detection and prevention of colon cancer [3,20]. Recently, a promising state-space model, Mamba, has garnered significant attention in the field of polyp segmentation. Unlike traditional CNN-based models [13,18], which have limited receptive fields, and Transformer-based models [26,23,5,10], which incur computational costs that increase quadratically with image size [6,24], Mamba can capture long-range dependencies while maintaining linear computational complexity. Its derivative variants[12,31,29,33,15,32,19] have been applied to various vision tasks, including U-Mamba [15], Segmamba [31], and VM-UNet [19]. However, to date, no study has analyzed Mamba from a frequency domain perspective in the context of polyp segmentation.

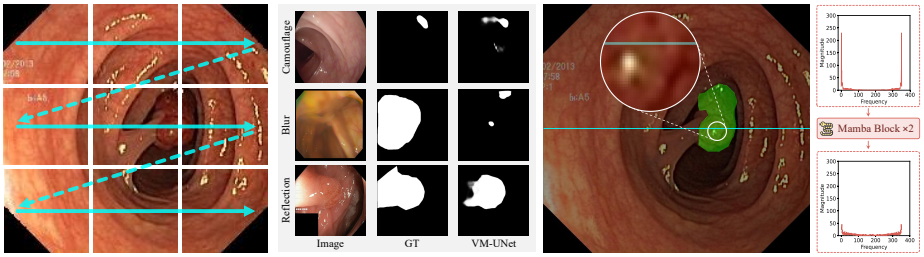


Fig. 1. Visualization of the limitations of the Mamba method.

We identify two main limitations in current Mamba-based polyp segmentation methods: 1) As shown in Fig. 1(a) and (b), the fixed spatial scanning pattern restricts the perceptual range, which may lead to errors in polyp localization, particularly in complex scenarios. 2) As illustrated in Fig. 1(c), during the Mamba calculation process, high-frequency components (ranging from 322 to 352), which are critical for capturing fine local details, gradually diminish, while blurred mid-frequency components (ranging from 31 to 321) increase. This ultimately compromises the accuracy of the segmentation.

To address these limitations, we propose the PolyMamba framework, which consists of two key components: the Spatial Prior Guidance (SG) module and the Dual-Gated Frequency Enhancement (DFE) module. This framework not only uses spatial priors to expand the perceptual range but also enhances high-frequency features from a frequency domain perspective to achieve accurate polyp segmentation. Specifically, the SG module integrates explicit spatial priors, extracted using a Transformer-based approach, to mitigate the local perception bias introduced by Mamba's fixed scanning pattern. Meanwhile, the DFE module applies two Gaussian filters to generate spectra with distinct high-frequency

thresholds. The difference between these spectra is used as an attention map to enhance high-frequency features, effectively mitigating the impact of blurred mid-frequency components and ultimately refining the polyp boundary. In summary, our contributions are threefold:

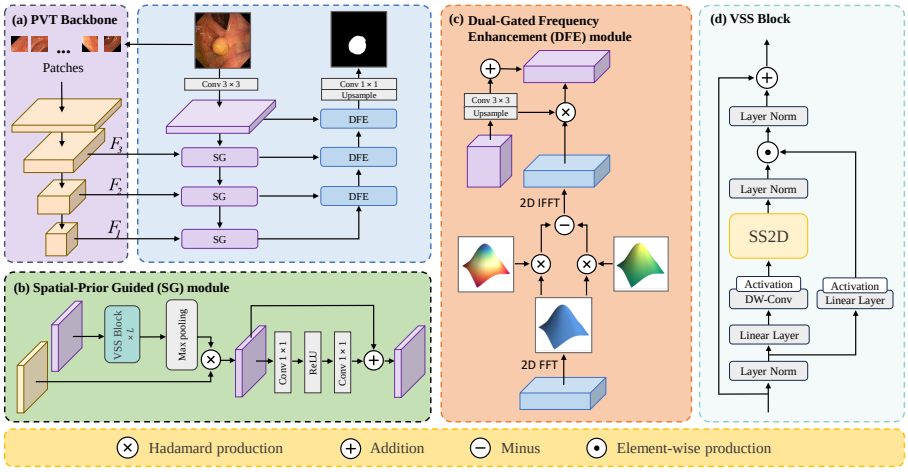


Fig. 2. Overview of the proposed PolyMamba, which consists of (a) PVT Backbone (b) Spatial-Prior Guided Module, and (c) Dual-Gated Frequency Enhancement Module.

1 Building on the observation that Mamba may suffer from local perception bias due to its scanning order and that it weakens high-frequency components while enhancing mid-frequency components during computation, our proposed PolyMamba framework achieves accurate polyp segmentation by integrating spatial priors and enhancing high-frequency features.
2 We introduced the SG and DFE modules to address the limitations of pure Mamba. The SG module incorporates explicit spatial priors to alleviate the local perception bias caused by the scanning order, while the DFE module enhances high-frequency information to refine the polyp boundaries. Together, these modules improve the model's ability to recognize polyps of various shapes from both spatial and frequency domain perspectives.
3 Extensive quantitative and qualitative experiments on five public colonoscopy image datasets show that our model outperforms most state-of-the-art methods in accurate polyp detection, highlighting its superior performance.

2 Method

2.1 Overall Architecture

The architecture of the proposed PolyMamba is sketched in Fig. 2. Given a colonoscopy image $I \in \mathbb{R}^{H \times W \times 3}$, we employ the Pyramid Vision Transformer

(PVT)[27,28] encoder network as the backbone to extract explicit spatial-prior features $F_i \in \mathbb{R}^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i}$, where $C_i \in \{512, 320, 128, 64\}$ and $i \in \{1, 2, 3, 4\}$. To mitigate the local perception bias in pure Mamba caused by the fixed scanning order, F_1, F_2, and F_3 are fed into the three Spatial-Prior Guided (SG) modules in the encoder as spatial-prior attention maps. These maps guide the encoder to expand its perception range and progressively learn high-level semantic information of polyps across multiple scales. Note that the first stage of the encoder is a convolutional layer, which captures rich high-frequency object details. Mirroring the encoder structure, the decoder comprises three Dual-Gated Frequency Enhancement (DFE) modules. The DFE module first applies a Fourier transform to the lower-layer features, then performs an inverse Fourier transform on the spectral difference obtained from two Gaussian filters to generate a high-frequency attention map. This map is subsequently fused with the output of the SG module to enhance the high-frequency components. After passing through the decoder, a final projection layer upsamples the features to restore their original height and width, and adjusts the number of channels to 1 to match the segmentation target. Further details of these modules are provided in the following sections.

2.2 Spatial-Prior Guided Module

In this module, we leverage the spatial prior features output by the PVT backbone as a guided attention map, directing the Mamba module to progressively learn high-level semantics across multiple perception ranges. We believe that the Transformer provides explicit spatial priors, while Mamba enables implicit global reasoning. Together, these two models are synergistic and complementary in capturing the semantics of polyps. As shown in Fig. 2 (b) and (d), each SG module comprises L VSS [33,1,7] blocks, each containing a 2D-Selective-Scan (SS2D) module, a linear layer, and a residual connection. To extract initial semantic information, the input feature f_C is first passed through an L-layer VSS stack following a fixed scanning pattern. It is then further refined by maximum pooling (Max). The corresponding Transformer feature f_T in PVT, which provides a richer perceptual space, is used to supply the prior for f_M:

$$f_M = Max(VSS_{\times L}(f_C)) \otimes f_T \tag{1}$$

Then, f_M is nonlinearly enhanced through a 1×1 convolution ($C_{1\times1}$), ReLU activation (Re), and a residual connection to obtain the final enhanced convolution feature f_S, which also serves as the input feature f_C for the next SG module in the encoder:

$$f_S = f_M + C_{1\times1}(Re(C_{1\times1}(f_M)) \tag{2}$$

By leveraging the SG module, PolyMamba extracts robust polyp semantic features from the rich perceptual space, allowing the DFE module to further refine these features from the frequency domain perspective, enhancing their discriminative power for more accurate recognition.

2.3 Dual-Gated Frequency Enhancement Module

Based on our observations during the Mamba calculation process, we found that high-frequency information gradually diminishes, while mid-frequency information is enhanced. Since high-frequency information is crucial for capturing local details essential for accurate polyp segmentation, the goal of this module is to boost the useful high-frequency components in the output of the SG module, where the middle-frequency components that include camouflaged background details are dominant. Specifically, the feature f_L from the low-layer is first transformed using a fast Fourier transform (FFT) and then processed with two Gaussian filters to generate two spectra with distinct high-frequency boundaries. The difference between these spectra is then restored to spatial features through an inverse fast Fourier transform (IFFT), which serve as attention maps, guiding the output f_{S_i} from the SG module to enhance high-frequency information. This ultimately yields the refined feature $f_{S_{i+1}}$. This process can be expressed as follows:

$$f_{S_{i+1}} = I((G_a - G_b) \otimes F(f_L)) \otimes Up(f_{S_i}) \tag{3}$$

Where F, I, and Up represent the FFT, IFFT, and upsampling operations respectively; G_a and G_b are the two distinct Gaussian filters; $\otimes$ denotes the Hadamard product. The average values of both Gaussian filters are set to the width and height of the polyp image.

2.4 Loss Function

To optimize our model, we consider the weighted intersection-over-union (IoU) loss and weighted binary cross-entropy (BCE) loss. Unlike the standard BCE loss, which assigns equal importance to all pixels, the weighted IoU loss and weighted BCE loss assign higher weights to more challenging pixels. This allows the model to constrain the prediction map from both the global structure (object-level) and local details (pixel-level) perspectives. Assuming that the prediction and ground truth are denoted as P and G, respectively, our total loss function can be expressed as:

$$\mathcal{L}_{total} = \mathcal{L}^{\omega}_{IoU}(P, G) + \mathcal{L}^{\omega}_{BCE}(P, G) \tag{4}$$

where $\mathcal{L}^{\omega}_{BCE}(\cdot)$ and $\mathcal{L}^{\omega}_{IoU}(\cdot)$ are the weighted IoU loss and weighted BCE loss.

3 Experiments and Results

Datasets and Metrics. To evaluate our proposed approach, we conducted extensive experiments on five widely used polyp segmentation datasets, including: Kvasir-SEG [11], ColonDB [22], Endoscene [25], ClinicDB [2], and ETIS [21]. The training set consists of 900 images from Kvasir-SEG and 550 images from ClinicDB. The test sets comprise 100 images from Kvasir-SEG, 62 images from CVC-ClincDB, 380 images from CVC-ColonDB, 60 images from Endoscene, and

196 images from ETIS. Six metrics are used to evaluate the performance of PolyMamba: mean Dice score (mDice) [17], mean IoU score (mIoU), mean absolute error (MAE), weighted F-measure (F_β^ω) [16], max E-measure (E_ϕ^{max}) [8], and S-measure (S_α) [4].

Table 1. Comparison of experimental results across the five polyp datasets, with best results indicated in bold. A total of five models were compared: U-Net [18], PraNet [9], SANet [30], Polyp-PVT [5], and VM-UNet [19].

Datasets	Methods	mDice	mIoU	F_β^ω	S_α	E_ϕ^{max}	MAE
ClinicDB	UNet(MICCAI'15)	0.823	0.755	0.811	0.889	0.954	0.019
	PraNet(MICCAI'20)	0.899	0.849	0.896	0.936	0.979	0.009
	SANet(MICCAI'21)	0.916	0.859	0.909	0.939	0.976	0.012
	Polyp-PVT(AIR'23)	0.937	0.889	0.936	0.949	0.989	0.006
	VM-UNet(arxir'24)	0.926	0.871	0.927	0.933	0.971	0.009
	PolyMamba(Ours)	**0.940**	**0.894**	**0.940**	**0.953**	**0.993**	**0.006**
Kvasir	UNet(MICCAI'15)	0.818	0.746	0.794	0.858	0.893	0.055
	PraNet(MICCAI'20)	0.898	0.840	0.885	0.915	0.948	0.030
	SANet(MICCAI'21)	0.904	0.847	0.892	0.915	0.953	0.028
	Polyp-PVT(AIR'23)	0.917	0.864	0.911	0.925	0.962	0.023
	VM-UNet(arxir'24)	0.913	0.856	0.902	0.918	0.958	0.027
	PolyMamba(Ours)	**0.919**	**0.866**	**0.911**	**0.924**	**0.967**	**0.023**
ETIS	UNet(MICCAI'15)	0.398	0.335	0.366	0.684	0.740	0.036
	PraNet(MICCAI'20)	0.628	0.567	0.600	0.794	0.841	0.031
	SANet(MICCAI'21)	0.750	0.654	0.685	0.849	0.897	0.015
	Polyp-PVT(AIR'23)	0.787	0.706	0.750	0.871	0.910	0.013
	VM-UNet(arxir'24)	0.761	0.692	0.743	0.869	0.900	0.015
	PolyMamba(Ours)	**0.829**	**0.753**	**0.794**	**0.902**	**0.934**	**0.012**
ColonDB	UNet(MICCAI'15)	0.512	0.444	0.498	0.712	0.776	0.061
	PraNet(MICCAI'20)	0.712	0.640	0.699	0.820	0.872	0.043
	SANet(MICCAI'21)	0.753	0.670	0.726	0.837	0.878	0.043
	Polyp-PVT(AIR'23)	0.808	0.727	0.795	0.865	0.919	0.031
	VM-UNet(arxir'24)	0.798	0.712	0.782	0.861	0.904	0.036
	PolyMamba(Ours)	**0.815**	**0.738**	**0.800**	**0.871**	**0.921**	**0.029**
EndoScene	UNet(MICCAI'15)	0.710	0.627	0.684	0.843	0.875	0.022
	PraNet(MICCAI'20)	0.871	0.797	0.843	0.925	0.972	0.010
	SANet(MICCAI'21)	0.888	0.815	0.859	0.928	0.972	0.008
	Polyp-PVT(AIR'23)	0.900	0.833	0.884	0.935	0.981	0.007
	VM-UNet(arxir'24)	0.886	0.818	0.849	0.921	0.968	0.009
	PolyMamba(Ours)	**0.904**	**0.843**	**0.888**	**0.945**	**0.985**	**0.007**

Implementation Details. We implement our method using PyTorch and conduct experiments on a single NVIDIA RTX 3090 GPU. The AdamW optimizer [14] is used for training over 100 epochs, with a learning rate of 1e-4, a weight decay of 1e-4 , and a batch size of 8. Data augmentation techniques, including random rotation, random flipping, and color jitter, are applied to enhance the diversity of the training data, thereby improving the model's robustness and generalization.

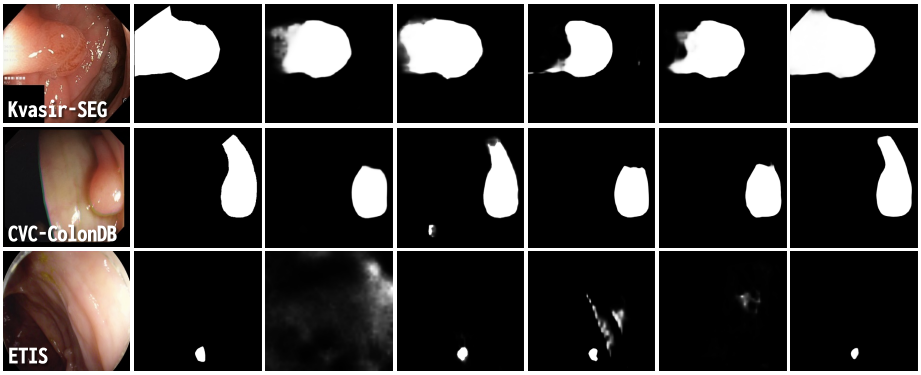
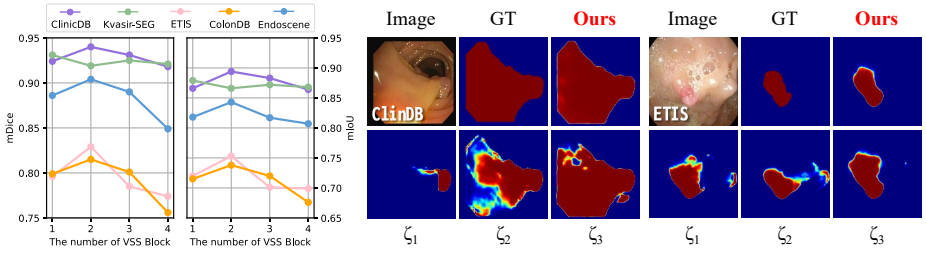


Fig. 3. Qualitative results of different methods.

Learning Ability and Generalization Ability. We conducted two experiments on two seen datasets, Kvasir and CVC-ClinicDB, to evaluate the learning ability of our model. To assess the generalization performance, we tested our model on three unseen datasets from different medical centers: ETIS, ColonDB, and EndoScene. As shown in Tab. 1 and Fig. 3, PolyMamba outperforms existing methods on both the seen and unseen datasets, demonstrating the model's superior ability to learn high-level semantic information for accurate polyp segmentation and its capacity to maintain exceptional performance even in unfamiliar and complex scenarios.

Table 2. Ablation studies on CVC-ClinicDB (seen) and ETIS (unseen) datasets.

Variation	Prior	Gauss	ClinicDB		ETIS	
			mDice	mIoU	mDice	mIoU
ζ_1	-	-	0.916	0.861	0.786	0.705
ζ_2	-	✓	0.919	0.872	0.797	0.721
ζ_3	✓	-	0.926	0.888	0.812	0.733
Ours	✓	✓	**0.940**	**0.894**	**0.829**	**0.753**



Analysis of the sensitivity of PolyMamba to different VSS Block numbers
(a)

Qualitative results on ClinicDB and ETIS datasets using different components, corresponding to settings in Tab. 2
(b)

Fig. 4. Visualization results of ablation experiments

Ablation Study. In this section, we conduct ablation experiments to evaluate the effectiveness of the proposed modules: the spatial priors extracted via the Transformer method in the SG module, and the Gaussian filter used to enhance high-frequency components in the DFE module. These experiments aim to verify whether the proposed modules can address the limitations of current Mamba-based polyp segmentation methods. Additionally, we investigate the sensitivity of our model to the number of VSS Blocks. As shown in Tab. 2, the results clearly demonstrate that each module contributes significantly to the overall performance improvement, with their combination achieving the best results. Fig. 4(a) shows that using too many VSS Blocks can lead to overfitting. This also suggests that if Mamba's local deviation is too strong, it may not be fully corrected even with the introduction of the prior. Through the visual ablation experiment in Fig. 4(b), we confirm that: 1) the introduction of spatial priors effectively enhances the model's spatial perception, aiding in the precise localization of polyps; 2) enhancing high-frequency information improves the accuracy of polyp boundary refinement, eliminates false positives, and enhances overall segmentation performance.

4 Conclusion

Building on the fixed spatial scanning mode of Mamba and the observed phenomena of high-frequency attenuation and middle-frequency amplification during the Mamba calculation process, we propose a novel polyp segmentation method, PolyMamba. This approach addresses two critical challenges: first, it mitigates the local perception bias caused by the scanning order in Mamba, and second, it enhances high-frequency information from the frequency domain to refine the polyp boundary, offering a fresh perspective for achieving accurate polyp segmentation. The SG module enhances the perception capabilities of the Mamba module by incorporating explicit spatial priors extracted via the Transformer, which helps in accurately locating polyps and extracting high-level semantic features. Additionally, the DFE module utilizes the spectral difference between two

distinct Gaussian filters as an attention map to enhance high-frequency information, progressively refining the polyp boundary. Experiments on five widely used polyp datasets demonstrate the effectiveness and robustness of PolyMamba. The results show that PolyMamba significantly supports the diagnosis of colon polyps, improving both the accuracy and automation of the diagnostic process.

Acknowledgments. The work was supported in part by the National Natural Science Foundation of China under grant 62476258, and in part by the Natural Science Foundation of Hubei Province under grant 2025AFA113.

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

References

1. Ba, J.L.: Layer normalization. arXiv preprint arXiv:1607.06450 (2016)
2. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized medical imaging and graphics **43**, 99–111 (2015)
3. Board, P.C.G.E.: Brca1 and brca2: Cancer risks and management (pdq®). In: PDQ Cancer Information Summaries [Internet]. National Cancer Institute (US) (2023)
4. Cheng, M.M., Fan, D.P.: Structure-measure: A new way to evaluate foreground maps. International Journal of Computer Vision **129**, 2622–2638 (2021)
5. Dong, B., Wang, W., Fan, D.P., Li, J., Fu, H., Shao, L.: Polyp-pvt: Polyp segmentation with pyramid vision transformers. arXiv preprint arXiv:2108.06932 (2021)
6. Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
7. Elfwing, S., Uchibe, E., Doya, K.: Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural networks **107**, 3–11 (2018)
8. Fan, D.P., Ji, G.P., Qin, X., Cheng, M.M.: Cognitive vision inspired object segmentation metric and loss function. Scientia Sinica Informationis **6**(6), 5 (2021)
9. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 263–273. Springer (2020)
10. Fang, X., Shi, Y., Guo, Q., Wang, L., Liu, Z.: Sub-band based attention for robust polyp segmentation. In: Elkind, E. (ed.) Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23. pp. 736–744 (8 2023), main Track
11. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., De Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26. pp. 451–462. Springer (2020)
12. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: Vmamba: Visual state space model. Advances in neural information processing systems **37**, 103031–103063 (2025)

13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
14. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
15. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
16. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps? In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 248–255 (2014)
17. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
19. Ruan, J., Li, J., Xiang, S.: Vm-unet: Vision mamba unet for medical image segmentation. arXiv preprint arXiv:2402.02491 (2024)
20. Sedlak, J.C., Yilmaz, Ö.H., Roper, J.: Metabolism and colorectal cancer. Annual Review of Pathology: Mechanisms of Disease **18**(1), 467–492 (2023)
21. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. International journal of computer assisted radiology and surgery **9**, 283–293 (2014)
22. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. IEEE transactions on medical imaging **35**(2), 630–644 (2015)
23. Tang, F., Xu, Z., Huang, Q., Wang, J., Hou, X., Su, J., Liu, J.: Duat: Dual-aggregation transformer network for medical image segmentation. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 343–356. Springer (2023)
24. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)
25. Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., Drozdzal, M., Courville, A.: A benchmark for endoluminal scene segmentation of colonoscopy images. Journal of healthcare engineering **2017**(1), 4037190 (2017)
26. Wang, J., Huang, Q., Tang, F., Meng, J., Su, J., Song, S.: Stepwise feature fusion: Local guides global. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 110–120. Springer (2022)
27. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 568–578 (2021)
28. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pvt v2: Improved baselines with pyramid vision transformer. Computational Visual Media **8**(3), 415–424 (2022)
29. 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)

30. Wei, J., Hu, Y., Zhang, R., Li, Z., Zhou, S.K., Cui, S.: Shallow attention network for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 699–708. Springer (2021)
31. 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)
32. Xu, Z., Tang, F., Chen, Z., Zhou, Z., Wu, W., Yang, Y., Liang, Y., Jiang, J., Cai, X., Su, J.: Polyp-mamba: Polyp segmentation with visual mamba. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 510–521. Springer (2024)
33. 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)