Adaptive Frequency-guided Parallel Mamba Network for Polyp Segmentation

Aiwen Jiang[1*], Hongqian Yu[1], Xue Li[1], Zhiqiang Xu[1], and Rui Cui[2]

[1] School of Artificial Intelligence, Jiangxi Normal University, Nanchang, China
jiangaiwen@jxnu.edu.cn
[2] East China Institute of Digital Medical Engineering, Shangrao, China

Abstract. Accurate polyp segmentation is crucial for early detection and timely prevention of cancer. Although existing methods have achieved remarkable progress, they still struggle in real-world clinical scenarios due to strong polyp heterogeneity, fuzzy boundaries, and the high risk of missing small polyps. To address the challenges, in this paper, we propose to jointly exploit spatial structural priors and frequency-domain cues, and construct an adaptive frequency-guided parallel mamba framework for polyp segmentation. Extensive experiments on five public benchmarks demonstrate that our method can achieve excellent balance among segmentation accuracy, efficiency, robustness, and cross-domain generalization. It achieves consistently leading performance compared with state-of-the-art methods, particularly in challenging cases with fuzzy boundaries, small polyps, and cross-domain shifts, while maintaining competitive computational cost.

Keywords: Polyp segmentation · Visual Mamba · State space model · Frequency-domain enhancement · Medical image segmentation.

1 Introduction

The precise segmentation of polyp regions during colonoscopy examinations is of great importance for early screening and decision making of cancer. However, automatic polyp segmentation remains challenging in real clinical settings, since polyps exhibit high heterogeneity in terms of size, shape, color, texture, boundary clarity, and illumination, and often co-occur with specular highlights, motion blur, intestinal folds, and complex background [1, 2]. As a result, many existing methods show significant performance drops when faced with the complex scenarios.

Recent work, such as VM-UNet [3], MSVM-UNet [4], and LKM-UNet [5] explored the potential to replace self-attention with visual Mamba [6] and highlighted the favorable trade-off between accuracy and efficiency. However, when directly applying 1D selective scan to 2D images, some intrinsic limitations have been revealed [7, 8]: (1) flattening 2D feature maps into 1D sequences inevitably

* Corresponding author: Aiwen Jiang (jiangaiwen@jxnu.edu.cn).

breaks natural spatial topology, degrading the modeling of boundaries, textures, and irregular shapes, particularly in complex polyp regions; (2) polyp boundaries are prone to oversmoothing and blurring, since visual mamba exhibits an inherent low-pass behavior.

Frequency-domain learning has recently shown its potential in polyp segmentation. Methods such as FTMF-Net [9] and FAGF-Net [10] have demonstrated that frequency-domain features help detect small objects and recover fine boundaries [11, 12]. However, most frequency-based methods adopted fixed frequency partitions or hand-crafted thresholds, ignoring sample-level spectral variations. The interaction between frequency-domain and spatial-domain features was usually implemented through simple concatenation or addition, which fails to fully utilize their complementarity.

Motivated by these observations, we propose an adaptive frequency-guided visual Mamba framework for polyp segmentation (Polyp-AFMamba), as shown in Fig. 1. It focuses on spatial structure preservation and frequency-aware enhancement rather than on extreme lightweight design. The contributions are summarized as follows:

- We have designed an adaptive parallel visual 2D-Mamba layer that can effectively preserve 2D spatial continuity while alleviating parameter explosion, thus balancing structural representation capacity and model efficiency.
- We have proposed a multi-scale high-frequency edge enhancement module and a low-frequency guided regional-prior adaptive module that can respectively refine polyp boundaries and enhance regional consistency in challenging scenes.
- Extensive experiments on five public benchmarks have demonstrated that the proposed Polyp-AFMamba achieves competitive state-of-the-art performance with competitive computational cost. It can serve as a promising solution for efficient and accurate computer-aided polyp screening.

2 Adaptive Frequency-guided Parallel Mamba Network for Polyp Segmentation

2.1 Adaptive parallel visual 2D-Mamba layer (APV-2DMamba)

The motivation for the APV-2DMamba is to maintain 2D spatial continuity while capturing long-range dependencies with linear complexity.

To achieve this goal, we adopt an adaptive multi-directional parallel subspace modeling strategy. Specifically, we uniformly split input feature map $\mathbf{X}_{\text{in}}$ along the channel dimension into four sub-tensors. Each sub-tensor is processed by an independent two-dimensional visual selective scanning module (2DVSS Block). The 2DVSS Block is similar to the block defined in [13]. To enhance directional diversity, different 2DVSS blocks adopt distinct scanning origins, starting respectively from the top-left (TL), bottom-right (BR), bottom-left (BL), and top-right (TR) corners, thereby performing row–column cascaded scanning from

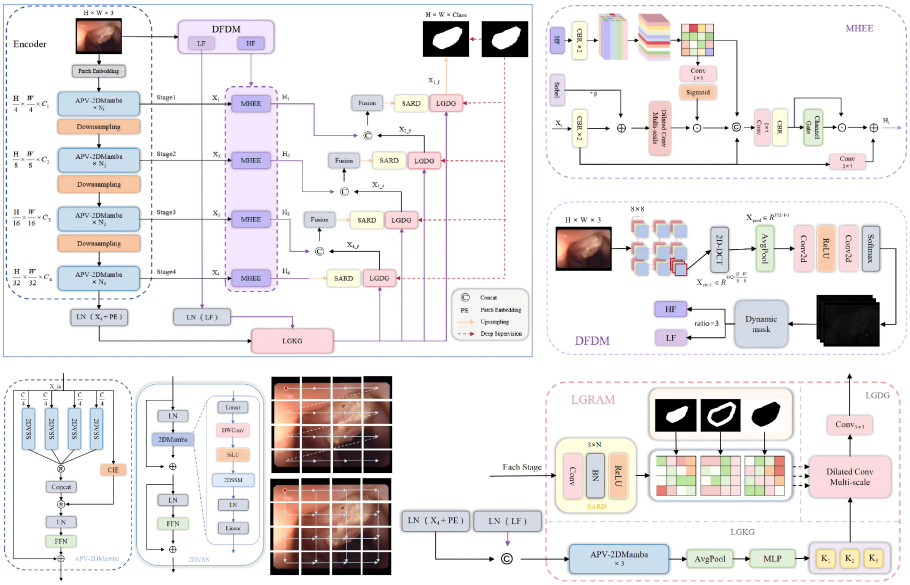

Fig. 1: Overview of the proposed Polyp-AFMamba framework.

multiple spatial directions. This multi-directional scanning strategy allows each subspace to capture contextual information from complementary spatial perspectives, alleviating the potential blind spots induced by single-direction scanning when handling complex anatomical structures. Besides, in parallel, channel-wise modulation weights w are learned in CIE path to explicitly model inter-channel dependencies among parallel subspaces.

2.2 Dynamic Frequency Decoupling Module

The dynamic frequency decoupling module performs block-wise DCT and learns a sample-adaptive frequency mask to decouple low- and high-frequency components in a bias-resistant manner, explicitly preserving boundary/detail information.

As shown in Fig 1, to dynamically identify frequency components that are most discriminative for polyp segmentation, we design a lightweight frequency attention branch that learns channel-wise importance weights over $\mathbf{X}_{\text{DCT}}$. For each sample, the top-K channels with the highest weights are selected as LF components, while the remaining channels are treated as HF components. The number of selected channels is determined by a ratio $top\_k\_ratio = 0.3$. Consequently, two binary masks $\mathbf{M}_L$ and $\mathbf{M}_H = 1 - \mathbf{M}_L$ are generated to indicate LF and HF channels, respectively.

To alleviate bias toward dominant frequency distributions in the training set, we introduce a stochastic perturbation mechanism during training. Specifically, 5% of the entries in $\mathbf{M}_L$ and $\mathbf{M}_H$ are randomly flipped using a Bernoulli process. This frequency-domain augmentation encourages the model to explore a broader

range of spectral patterns and improves generalization to unseen frequency distributions.

2.3 Low-frequency Guided Regional-prior Adaptive Module

Semantic-aware Regional Decoupling (SARD) To provide explicit region priors, we construct a lightweight auxiliary branch with three-channel head that in parallel predicts three masks $M_{\text{fg}}, M_{\text{bg}}, M_{\text{bd}}$ corresponding to foreground, boundary, and background, through a complementarity constraint to ensure $M_{\text{fg}} + M_{\text{bg}} + M_{\text{bd}} \approx 1$.

To adapt the boundary width to diverse polyp geometries, we introduce an adaptive *Boundary Ratio* (BR) that controls the relative thickness of the boundary region in a geometry-aware manner. The boundary ratio is calculated as Equation 1:

$$\text{BR} = \text{clip}\left(\text{BR}_0 \cdot (0.4\,\alpha_{\text{area}} + 0.4\,\alpha_{\text{bd}} + 0.2\,\alpha_{\text{shape}}), \frac{1}{8}, \frac{1}{2}\right) \tag{1}$$

where, α_{area}, α_{bd} and α_{shape} denote heuristic scaling factors triggered by area ratio, boundary complexity, and shape compactness, respectively.

Herein, the three geometry-related cues are defined as follows. (1) The area ratio is defined as $A_{\text{ratio}} = \frac{A_{\text{polyp}}}{A_{\text{image}}}$, which reflects the relative scale of the polyp. (2) The boundary complexity is measured by $C_{\text{bd}} = \text{Var}(\text{DT})$, where DT denotes the distance transform of the polyp mask. A larger variance indicates a more irregular and highly curved boundary. (3) The shape compactness is defined as $C_{\text{p}} = \frac{4\pi A_{\text{polyp}}}{P^2_{\text{polyp}}}$, where P_{polyp} denotes the perimeter of the binary polyp mask. Lower compactness values correspond to less compact and more irregular shapes.

Low-Frequency Guided Kernel Generator(LGKG) The LGKG synthesizes region-specific dynamic convolution kernels for all decoder stages by leveraging LF structural cues and deep encoder features.

Starting from the initialized state $\mathbf{H}_0$ which is projected from low-frequency feature, the kernel generator performs N stacked state transitions using the proposed APV-2DMamba layers, as illustrated in Equation 2:

$$\mathbf{H}_{n+1} = \text{APV-2DMamba}(\mathbf{H}_n,\, \mathbf{F}_B^{\text{pe}}), \quad n = 0, \ldots, N-1, \tag{2}$$

where, $\mathbf{F}_B^{\text{pe}}$ is augmented from the deepest encoder feature $\mathbf{X}_4$ with a two-dimensional positional encoding. This formulation preserves the intrinsic 2D topology while allowing low-frequency priors to continuously guide the state evolution.

We then apply global average pooling on the final hidden state $\mathbf{H}_N$ to obtain a compact representation for the generation of the region-specific dynamic convolution kernels $\mathcal{K}$.

Low-Frequency Guided Dynamic Decoding (LGDG) The LGDG performs region-adaptive and scale-aware feature refinement at each decoder stage, aiming to enhance boundary sharpness and regional consistency, particularly for polyps with large scale variations and irregular shapes.

Specifically, to inject region priors explicitly, the decoder feature $\mathbf{X}_t$ is first decomposed into region-specific branches through element-wise modulation, as done in Equation 3:

$$\mathbf{X}_t^{(r)} = \mathbf{X}_t \odot \mathbf{M}_r, \quad r \in \{\mathrm{fg}, \mathrm{bd}, \mathrm{bg}\}. \tag{3}$$

Given the dynamic kernel set $\mathcal{K}$ generated by LGKG, we parse it into four convolution groups corresponding to different dilation rates $d \in \{1, 3, 5, 7\}$. For each semantic region branch, multi-scale dynamic atrous convolutions are applied to obtain the semantic region features $\mathbf{Y}_t^{(r,d)}$. The final features $\mathbf{Y}_t$ are then obtained by aggregating and fusing different semantic region features of different dilation scales, as formulated in Equation 4:

$$\mathbf{Y}_t = \mathrm{Conv}_{3\times3}\left(\mathrm{Concat}_d\left(\sum_r \mathbf{Y}_t^{(r,d)}\right)\right). \tag{4}$$

2.4 Loss Function

The total loss combines segmentation loss L_{pred} and prior region-mask supervision L_{SARD}. L_{pred} consists of BCE and IoU losses, and L_{SARD} is an MSE loss over foreground, background, and boundary masks generated from the ground-truth annotation:

$$L_{\mathrm{total}} = L_{\mathrm{pred}} + \lambda L_{\mathrm{SARD}}, \quad L_{\mathrm{pred}} = L_{\mathrm{BCE}} + L_{\mathrm{IoU}}. \tag{5}$$

where $\lambda = 1$. Dynamic frequency masks have no additional mask label and are learned through L_{total} via the LF-guided regional decoding and HF-guided edge enhancement paths.

3 Experiment

3.1 Settings

We evaluate Polyp-AFMamba on five widely used colonoscopy polyp segmentation benchmarks: CVC-ClinicDB [14], Kvasir-SEG [15], CVC-ColonDB [16], ETIS-LaribPolypDB [17], and EndoScene [18]. These datasets cover diverse imaging conditions and acquisition centers. The training and test protocols are the same as the setting commonly done in [19, 20]. Mean Dice (mDice) and mean IoU (mIoU) are adopted as evaluation metrics, following prior polyp segmentation work [19, 21, 22].

The adaptive boundary ratio (BR) is implemented using deterministic rule-based heuristics derived from geometric properties of each polyp instance. Specifically, the area-related scaling factor is defined as $\alpha_{\mathrm{area}} = 0.7$ if $A_{\mathrm{ratio}} < 0.02$, and

Table 1: Quantitative comparison with state-of-the-art methods on five polyp segmentation datasets.

Method	ClinicDB (seen)		Kvasir (seen)		ColonDB (unseen)		ETIS (unseen)		EndoScene (unseen)	
	Dice	mIoU	Dice	mIoU	Dice	mIoU	Dice	mIoU	Dice	mIoU
U-Net [23]'15	0.823	0.755	0.818	0.746	0.512	0.444	0.398	0.335	0.710	0.627
PraNet [19]'20	0.899	0.849	0.898	0.840	0.709	0.640	0.628	0.567	0.871	0.797
TransFuse [24]'21	0.935	0.887	0.913	0.857	0.781	0.699	0.731	0.660	0.893	0.824
CFA-Net [25]'23	0.933	0.883	0.915	0.861	0.743	0.665	0.732	0.655	0.893	0.827
Polyp-PVT [20]'23	0.937	0.889	0.917	0.864	0.808	0.727	0.787	0.706	0.900	0.833
VM-UNet [3]'24	0.926	0.871	0.913	0.856	0.798	0.712	0.761	0.692	0.886	0.818
MEGANet [26]'24	0.938	0.894	0.913	0.863	0.793	0.714	0.739	0.665	0.899	0.834
CTNet [27]'24	0.936	0.887	0.917	0.863	0.813	0.734	0.810	0.734	0.908	0.844
PraNet-V2 [28]'25	0.931	0.881	0.915	0.861	0.758	0.680	0.764	0.687	0.899	0.831
PolyMamba [7]'25	0.940	0.894	0.919	0.866	0.815	0.738	**0.829**	**0.753**	0.904	0.843
Ours	**0.946**	**0.902**	**0.932**	**0.883**	**0.819**	**0.741**	0.825	0.748	**0.917**	**0.852**

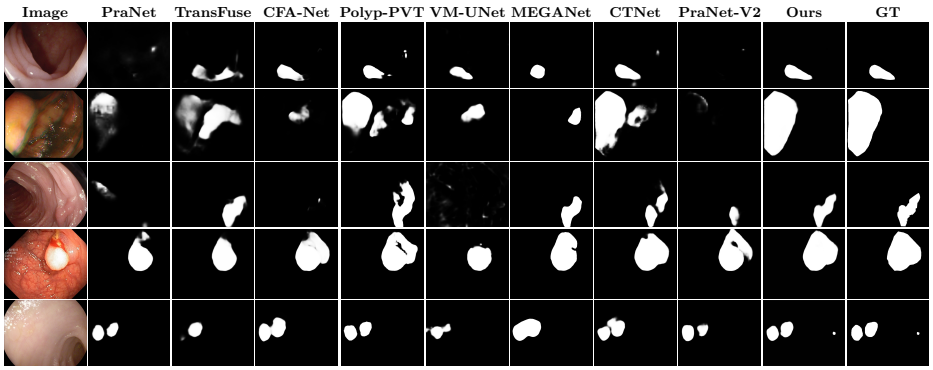

Fig. 2: Qualitative comparison of segmentation results by different methods on challenging cases. Polyp-AFMamba produces more coherent contours and fewer false positives than competing methods, consistent with Table 1.

$\alpha_{\text{area}} = 1.0$ otherwise, in order to prevent excessive boundary coverage for small polyps. The boundary-complexity factor is computed as $\alpha_{\text{bd}} = 1.0 + \min(0.5, 10 \cdot C_{\text{bd}})$, where the capped design avoids extreme boundary expansion caused by noisy or highly irregular masks. The shape-related factor is set to $\alpha_{\text{shape}} = 1.2$ if the compactness $C_{\text{p}} < 0.7$, and 1.0 otherwise, allowing larger boundary tolerance for irregular shapes. The resulting boundary ratio is finally clipped to the range $[1/8, 1/2]$ to ensure stable boundary supervision across different datasets.

3.2 Comparison with state-of-the-art methods

Table 1 reports Dice/mIoU on five benchmarks. We can observe that Polyp-AFMamba achieves consistently leading performance compared with state-of-the-art methods and remains highly competitive under cross-dataset shifts. In particular, it attains the highest Dice and mIoU on both seen datasets, indicating strong in-distribution accuracy. On unseen datasets, Polyp-AFMamba maintains

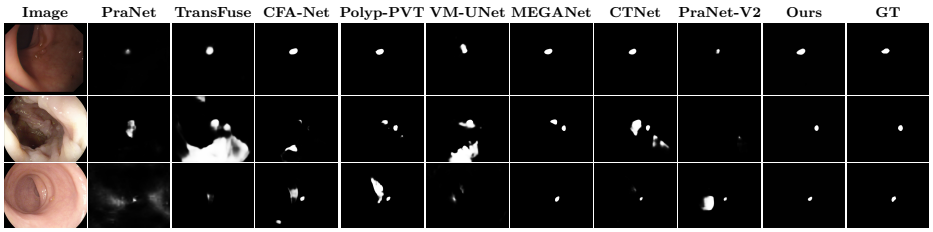

Fig. 3: Qualitative comparison for small polyps across different methods. Polyp-AFMamba is less prone to missing tiny targets and yields more compact predictions, indicating improved sensitivity to small foreground regions.

Table 2: Comparisons on model complexities. The inference speed is reported as the mean FPS with standard deviation over multiple runs on a single NVIDIA RTX 3090 GPU.

Method	Params (M)	GFLOPs	FPS
PraNet [19]'20	30.50	26.30	89.71 ± 0.16
Polyp-PVT [20]'23	25.11	20.04	106.63 ± 0.35
VM-UNet [3]'24	34.62	28.60	51.89 ± 0.22
PraNet-V2 [28]'25	32.55	27.16	74.16 ± 0.14
Ours	34.80	30.16	78.58 ± 0.26

stable performance compared with recent baselines. Although its ETIS results are slightly lower than those of PolyMamba, Polyp-AFMamba remains among the top methods overall with a balanced performance across all benchmarks.

From Table 2, we can observe that, compared with VM-UNet, Polyp-AFMamba achieves a substantially higher inference speed while maintaining a comparable number of parameters. Notably, although Polyp-AFMamba incorporates additional frequency-guided modeling and adaptive dynamic kernel generation, its overall computational cost remains moderate. This is mainly attributed to the proposed adaptive parallel 2D-Mamba design, which effectively improves hardware utilization and alleviates the efficiency bottleneck commonly observed in Mamba-based segmentation networks.

In contrast to CNN- or Transformer-based methods such as PraNet, Polyp-PVT and PraNet-V2, Polyp-AFMamba strikes a better balance between modeling capacity and inference efficiency. These results indicate that Polyp-AFMamba can deliver enhanced representation capability without incurring excessive computational overhead, suggesting its practical potential for polyp segmentation scenarios.

Fig. 2 visualizes representative challenging cases (e.g., blurry boundaries, specular highlights, and background textures similar to polyps). Compared with strong CNN/Transformer and Mamba baselines, Polyp-AFMamba better preserves boundary continuity and suppresses false positives. To further stress-test small targets, Fig. 3 reports representative small-polyp samples, where Polyp-

Table 3: Ablation study of M1, M2, and M3. Rows correspond to different combinations of enabled modules.

Modules			ClinicDB		Kvasir		ColonDB		ETIS		EndoScene	
M1	M2	M3	Dice	mIoU	Dice	mIoU	Dice	mIoU	Dice	mIoU	Dice	mIoU
			0.926	0.871	0.913	0.856	0.798	0.712	0.761	0.692	0.886	0.818
✓			0.933	0.882	0.921	0.866	0.806	0.722	0.772	0.706	0.893	0.827
	✓		0.932	0.879	0.919	0.864	0.804	0.720	0.778	0.710	0.893	0.826
		✓	0.936	0.892	0.926	0.870	0.810	0.728	0.801	0.727	0.905	0.839
	✓	✓	0.943	0.900	0.931	0.881	0.816	0.736	0.823	0.747	0.911	0.848
✓	✓	✓	**0.946**	**0.902**	**0.932**	**0.883**	**0.819**	**0.741**	**0.825**	**0.748**	**0.917**	**0.852**

AFMamba retains tiny foreground regions more reliably and avoids fragmented predictions in these examples.

4 Ablation Study and Discussion

We have systematically validated the contributions of each proposed component: the adaptive parallel visual 2D-Mamba layer (M1), the multi-scale high-frequency edge enhancement module (M2), and the low-frequency guided regional-prior adaptive module (M3).

The ablation results reveal that the proposed components are complementary rather than redundant. Each component contributes positively to the overall segmentation performance, leading to consistent performance gains across different evaluation metrics. The qualitative comparisons in Fig. 4 further support the conclusion. Baseline model tends to produce blurry boundaries or miss small polyp regions. In contrast, the full Polyp-AFMamba generates more accurate and coherent segmentation maps, particularly in challenging cases with ambiguous boundaries or low contrast. However, the current analysis mainly relies on standard Dice/mIoU scores and representative qualitative cases. A more targeted quantitative evaluation for small-polyp subsets and boundary quality, such as size-stratified Dice or boundary F-score, remains an important direction for future validation.

5 Conclusion

We have proposed an adaptive frequency-guided visual Mamba framework for polyp segmentation that jointly emphasizes spatial structure preservation and frequency-domain enhancement. Extensive experiments on five public polyp segmentation benchmarks have demonstrated that our method can effectively address the inherent limitations of visual SSMs in 2D topology preservation and high-frequency modeling. It achieves consistently leading performance compared with state-of-the-art methods, particularly in challenging cases with fuzzy boundaries, small polyps, and cross-domain shifts. Nevertheless, our evaluation still

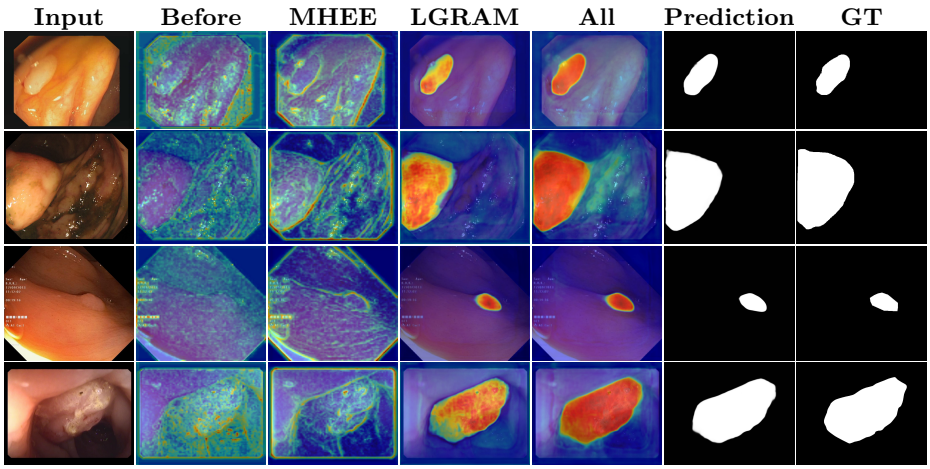

Fig. 4: Visualization of heat maps before and after MHEE and LGRAM. From left to right: input image, heat map before MHEE, heat map after MHEE, heat map after LGRAM, fused heat map after MHEE+LGRAM, model prediction, and ground truth.

lacks targeted quantitative analysis for small-polyp subsets and boundary quality, which will be important for more fine-grained clinical validation. Its reasonable computational cost also makes it promising for practical computer-aided diagnosis scenarios.

Acknowledgments

This work was supported by the Key Research and Development Program of Jiangxi Province (20261BCG330015).

Disclosure of Interests

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

References

1. Pooler, B.D., Kim, D.H., Matkowskyj, K.A., Newton, M.A., Halberg, R.B., Grady, W.M., Hassan, C., Pickhardt, P.J.: Growth rates and histopathological outcomes of small (6–9 mm) colorectal polyps based on ct colonography surveillance and endoscopic removal. GUT **72**(12), 2321–2328 (2023)
2. Qayoom, A., Xie, J., Ali, H.: Polyp segmentation in medical imaging: challenges, approaches and future directions. Artificial Intelligence Review **58**(6), 169 (2025)
3. Ruan, J., Li, J., Xiang, S.: Vm-unet: Vision mamba UNet for medical image segmentation. ACM Transactions on Multimedia Computing, Communications, and Applications (2024). https://doi.org/10.1145/3767748

4. Xu, H., et al.: Msvm-unet: Multi-scale vision mamba unet for medical image segmentation. arXiv preprint arXiv:2403.01357 (2024)
5. Wang, J., Chen, J., Chen, D., Wu, J.: Lkm-unet: Large kernel vision mamba unet for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 360–370. Springer Nature Switzerland (2024)
6. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
7. Fu, R., Hu, S., Zheng, X., Tang, C., Liu, X.: Polymamba: Spatial-prior guided mamba for polyp segmentation with high-frequency enhancement. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI) (2025), lecture Notes in Computer Science
8. Xiao, C., et al.: Spatial-mamba: Effective visual state space models via structure-aware state fusion. In: Proceedings of the International Conference on Learning Representations (2025)
9. Liu, G., Chen, Z., Liu, D., Chang, B., Dou, Z.: Ftmf-net: A fourier transform-multiscale feature fusion network for segmentation of small polyp objects. IEEE Transactions on Instrumentation and Measurement **72**, 1–15 (2023)
10. Li, Y., Zheng, Z., Ren, W., Nie, Y., Zhang, J., Jia, X.: Frequency aware and graph fusion network for polyp segmentation. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1586–1590 (2024)
11. Liu, Z., Deng, X., Jiang, P., Lv, C., Min, G., Wang, X.: Edge perception camouflaged object detection under frequency domain reconstruction. IEEE Transactions on Circuits and Systems for Video Technology **34**(10), 10194–10207 (2024)
12. Liu, Y., Tu, B., Liu, B., He, Y., Li, J., Plaza, A.: Spatial frequency domain transformation for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing (2025)
13. Zhang, J., Nguyen, A.T., Han, X., Trinh, V.Q.H., Qin, H., Samaras, D., Hosseini, M.S.: 2dmamba: Efficient state space model for image representation with applications on giga-pixel whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3583–3592 (2025)
14. 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**(2), 283–293 (2014)
15. 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: International conference on multimedia modeling. pp. 451–462. Springer (2019)
16. 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)
17. 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)
18. 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)

19. 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: Medical Image Computing and Computer-Assisted Intervention (MICCAI). Lecture Notes in Computer Science, vol. 12263, pp. 263–273. Springer (2020)
20. Dong, B., Wang, W., Fan, D.P., Li, J., Fu, H., Shao, L.: Polyp-pvt: Polyp segmentation with pyramid vision transformers. CAAI Artificial Intelligence Research **2**, 9150015 (2023)
21. Su, Y., Shen, Y., Ye, J., He, J., Cheng, J.: Revisiting feature propagation and aggregation in polyp segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). Lecture Notes in Computer Science, Springer (2023)
22. Shao, H., Zhang, Y., Hou, Q.: Polyper: Boundary sensitive polyp segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). vol. 38, pp. 4731–4739 (2024)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer (2015)
24. Zhang, Y., Liu, H., Hu, Q.: Transfuse: Fusing transformers and CNNs for medical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). Lecture Notes in Computer Science, vol. 12901, pp. 14–24. Springer (2021)
25. Zhou, T., Zhou, Y., He, K., Gong, C., Yang, J., Fu, H., Shen, D.: Cross-level feature aggregation network for polyp segmentation. Pattern Recognition **140**, 109555 (2023). https://doi.org/10.1016/j.patcog.2023.109555
26. Bui, N.T., Hoang, D.H., Nguyen, Q.T., Tran, M.T., Le, N.: Meganet: Multi-scale edge-guided attention network for weak boundary polyp segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 3592–3601 (2024)
27. Xiao, B., Hu, J., Li, W., Pun, C.M., Bi, X.: Ctnet: Contrastive transformer network for polyp segmentation. IEEE Transactions on Cybernetics **54**(9), 5040–5053 (2024)
28. Hu, X., Ji, G.P., Zhou, T., Fan, D.P., Shao, L.: Pranet-v2: Transformer-enhanced parallel reverse attention network for polyp segmentation. arXiv preprint arXiv:2501.01234 (2025)