Hybrid Graph Mamba: Unlocking Non-Euclidean Potential for Accurate Polyp Segmentation

Yueyue Zhu[1], Haolin Lv[1], Geng Chen✉[1], Zhonghao Zhang[1], Haotian Jiang[1], and Yong Xia[1]

National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, China

Abstract. Colorectal polyp segmentation can assist doctors in screening colonoscopy images, which is crucial for the prevention of colorectal cancer. Although deep learning has significantly advanced polyp segmentation, three issues remain: (1) Most polyp segmentation methods only extract Euclidean features such as shape and texture, while neglecting non-Euclidean features, such as the geometric topology between the polyp and its surrounding tissue; (2) Non-Euclidean features vary across different regions, but most feature fusion methods overlook both the non-Euclidean topological structures and the differences between internal, edge, and background regions. (3) Low-level features are not fully exploited, and the differences between low- and high-level features are not effectively addressed. To resolve these issues, we propose Hybrid Graph Mamba (HGM) based on Mamba and Graph Convolutional Network (GCN). Our model first uses the pyramid vision transformer to extract features at different levels. Next, we propose hybrid graph Mamba modules to process low-level features from multiple directions using quad-directional Mamba and extract non-Euclidean features with GCN. A boundary discrimination fusion module is also designed to handle high-level features, extracting semantic information for the interior, edges, and background to improve the fusion of low- and high-level features. Finally, a bidirectional Mamba decoder combines bidirectional Mamba and dilated convolutions to aggregate multi-scale features, minimizing information loss and producing the final prediction. Extensive experiments on five benchmark datasets demonstrate that HGM significantly outperforms eight State-Of-The-Art models. Our code is publicly available at https://github.com/YueyueZhu/HGM.

Keywords: Mamba, GCN, Polyp Segmentation

Equal contribution: Yueyue Zhu and Haolin Lv

Corresponding author: Geng Chen (Email: geng.chen@ieee.org)

This work was supported in part by the National Natural Science Foundation of China (No. 61540047).

1 Introduction

Colorectal Cancer (CRC) is the third leading cause of cancer-related deaths worldwide, with most cases developing from polyps. Early detection of polyps through colonoscopy is crucial for preventing CRC. However, manual observation is time-consuming, labor-intensive, and subjective, leading to missed detections. Automated polyp segmentation techniques have improved accuracy, aiding doctors in diagnosis and treatment planning. Despite this, accurate segmentation remains challenging due to the varying shapes, sizes, and ambiguous boundaries of the polyps with surrounding tissues.

Early automated polyp segmentation relied on handcrafted features, but these were limited in expressiveness, leading to high false detection rates. Deep learning can learn more comprehensive and accurate features. Recently, a large number of deep learning-based methods have been applied to polyp segmentation [7,10]. PraNet [5] improves the accuracy of edge segmentation using reverse attention to focus on the segmentation of background regions. Polyp-PVT [3] extracts and processes both low- and high-level features and then fuses them to obtain more comprehensive features. CAFE [9] supplements details through the Feature Selection and Enhancement Module (FSEM) and retains lower-level features with cross-attention, improving the accuracy of small polyp segmentation. Polyper [12] enhances polyp segmentation using morphological operators and a boundary-sensitive attention module. VMUNetV2 [23] introduces Visual State Space (VSS) blocks to capture more contextual information and uses Semantic and Detail Injection (SDI) mechanisms to improve feature fusion. G-CASCADE [11] employs a graph convolutional decoder and an attention module to enhance feature map optimization and segmentation performance. VANet [2] uses viewpoint classification to localize polyps and designs VAFormer to reduce the interference of surrounding tissues with attention, thus obtaining better polyp representations. HSNet [24] employs a Cross-Scale Attention (CSA) to link encoder–decoder features and a dual-branch Hybrid Scale Context (HSC) combining Transformers [15] and CNNs [8] for both global context and local detail recovery. FusionMamba [21] proposes a Mamba-based fusion framework that adaptively boosts intra- and inter-modal representations, balancing CNN efficiency with ViT global modeling to excel in multimodal fusion tasks with lower complexity. EFA-Net [25] integrates an Edge-aware Guidance Module (EGM), Scale-aware Convolution Module (SCM), and Cross-level Fusion Module (CFM) for multi-scale, cross-level feature fusion, sharpening polyp boundaries and improving segmentation across datasets.

While existing methods improve the accuracy of polyp segmentation, several issues remain: (1) Colonoscopy images contain not only Euclidean features, such as shape and texture, but also many non-Euclidean features, such as the geometric and topological structures formed by a polyp and its surrounding tissues. Existing methods primarily focus on extracting Euclidean features of polyps while neglecting the rich non-Euclidean features. (2) Non-Euclidean features contain a wealth of information and vary across different regions. When integrating high- and low-level features, existing methods treat the entire feature set uniformly,

without considering both the non-Euclidean topological structure and the differences between internal, edge, and background regions. (3) There is a lack of further exploration of low-level feature information, and existing methods fail to effectively address the gap between low- and high-level features.

To this end, we propose a novel deep learning model, called Hybrid Graph Mamba (HGM), which achieves accurate polyp segmentation with a mixture of Mamba and Graph Convolutional Network (GCN). Specifically, we use a pyramid vision transformer [19] to extract features at different levels. For low-level features, we design the Hybrid Graph Mamba Module (HGMM), which focuses on detailed features using Quad-directional Mamba (QM) and extracts non-Euclidean features using GCN [22]. For high-level features, we use the Cascaded Fusion Module (CFM) [3] to extract semantic and positional information of polyps. The processed low- and high-level features are then fused in the Boundary Discrimination Fusion Module (BDFM), ensuring that the final feature map contains positional information while also focusing on edge details. Finally, the outputs of CFM and BDFM are passed through a HGMM to further extract non-Euclidean features and to prevent the loss of prior features. These outputs are then input into the Bidirectional Mamba Decoder (BMD), which uses multi-size receptive fields to extract various features and fuse multi-scale features, ultimately obtaining the segmentation result. Experimental results show that our model achieves superior segmentation results and outperforms State-Of-The-Art methods on five benchmark datasets.

2 Method

2.1 Overall Architecture

As shown in Fig. 1, our model consists of five components. The pyramid vision transformer is employed to extract features at multiple levels from the original image. The CFM aggregates high-level features from $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, and $\boldsymbol{X}_4$, which provide global semantic information. Through the CFM, the intermediate feature $\boldsymbol{T}_1$ is obtained. For the low-level feature $\boldsymbol{X}_1$, the HGMM transforms it into a one-dimensional vector and extracts features from four directions using Mamba. These features are then fused across different channels using GCN, which helps extract non-Euclidean features. These non-Euclidean features capture rich boundary details, which assist in distinguishing the target from the background. The BDFM is then applied to fuse the high-level features processed by the CFM and the low-level features processed by HGMM, ultimately producing the output $\boldsymbol{T}_2$. Mathematically, the whole procedure is defined as follows:

$$\boldsymbol{T}_1 = \mathrm{CFM}(\boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4), \quad \boldsymbol{T}_2 = \mathrm{BDFM}(\boldsymbol{T}_1, \mathrm{HGMM}(\boldsymbol{X}_1)). \tag{1}$$

To fuse the features and further extract non-Euclidean features, both the outputs of BDFM and the outputs of CFM are passed through a HGMM. These processed features are then fused with the original features that have been processed by

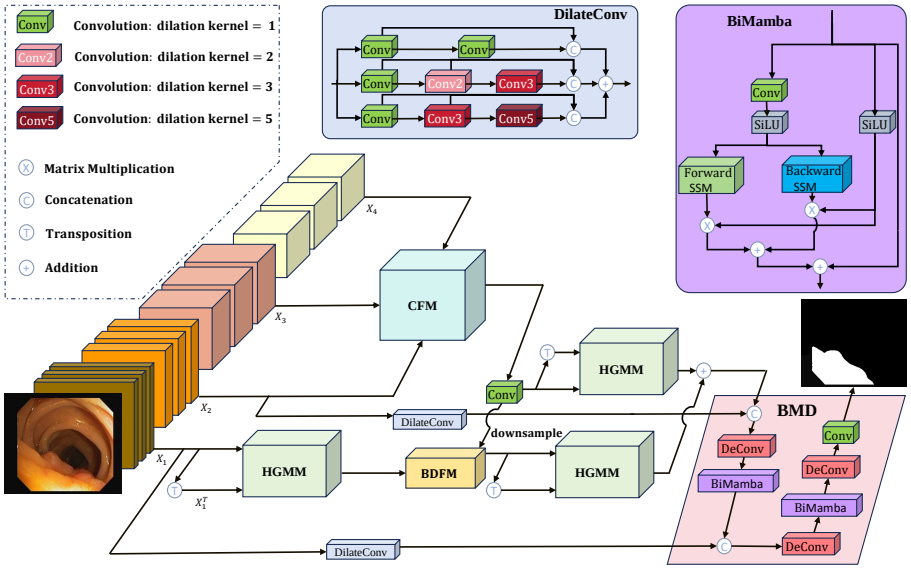

Fig. 1. Overall architecture of HGM. Our model consists of a pyramid vision transformer, a CFM, three HGMMs, a BDFM, and a BMD.

dilated convolutions, in the BMD. This fusion process ultimately generates the final segmentation mask $\boldsymbol{Y}$:

$$\boldsymbol{Y} = \mathrm{BMD}(\mathrm{HGMM}(\boldsymbol{T}_1), \mathrm{HGMM}(\boldsymbol{T}_2), \mathrm{DilateConv}(\boldsymbol{X}_1, \boldsymbol{X}_2)), \tag{2}$$

where $\mathrm{DilateConv}(\cdot)$ represents the dilated convolution component, as shown at the top of Fig. 1.

2.2 Hybrid Graph Mamba Module

As shown in Fig. 2, in HGMM we construct the QM for feature extraction using the BiMamba [18]. We use four BiMamba blocks, connected in pairs. The input consists of $\boldsymbol{X}$ and its transpose $\boldsymbol{X}^\top$. For a single BiMamba cascade, the first block is called the pre-BiMamba, which has a single output. The second block is called the post-BiMamba, which, in addition to the output of the pre-BiMamba, produces two additional direction-specific immediate features, $\boldsymbol{X}_\mathrm{F}$ and $\boldsymbol{X}_\mathrm{B}$. Mathematically, this is defined as:

$$(\boldsymbol{X}_\mathrm{F}, \boldsymbol{X}_\mathrm{M}, \boldsymbol{X}_\mathrm{B}) = \mathrm{BiMamba}_\mathrm{post}(\mathrm{BiMamba}_\mathrm{pre}(\boldsymbol{X})), \tag{3}$$

where $\boldsymbol{X}_\mathrm{M}$ represents the output feature of the post-BiMamba, and $\mathrm{BiMamba}_\mathrm{pre}(\cdot)$ and $\mathrm{BiMamba}_\mathrm{post}(\cdot)$ represent the pre- and post-BiMamba. BiMamba inputs the data from both the forward and backward directions into the SSM and is defined as:

$$\mathrm{BiMamba}(\boldsymbol{x}) = \mathrm{RS}(\boldsymbol{x} + \boldsymbol{x}'\mathrm{SSM}_\mathrm{F}(\boldsymbol{x}'') + \boldsymbol{x}'\mathrm{SSM}_\mathrm{B}(\boldsymbol{x}'')), \tag{4}$$

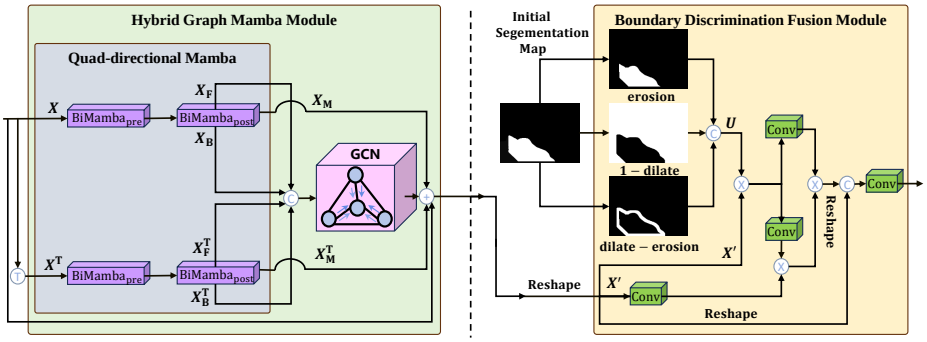

Fig. 2. Illustrations of two proposed modules.

where $\mathrm{RS}(\cdot)$ denotes a reshape operation that transforms $\boldsymbol{X}$ into $\boldsymbol{x}$, i.e., $\boldsymbol{x} = \mathrm{RS}(X)$. The $\mathrm{SSM_F}(\cdot)$ and $\mathrm{SSM_B}(\cdot)$ represent the forward and backward SSM, respectively. $\boldsymbol{x}'$ and $\boldsymbol{x}''$ are defined as $\boldsymbol{x}' = \mathrm{SiLU}(\boldsymbol{x})$ and $\boldsymbol{x}'' = \mathrm{SiLU}(\mathrm{Conv}(\boldsymbol{x}))$ with $\mathrm{SiLU}(\cdot)$ denoting SiLU activation function [4].

As shown in Fig. 2, the four intermediate features from BiMamba blocks, denoted as $\boldsymbol{X}_\mathrm{F}$, $\boldsymbol{X}_\mathrm{B}$, $\boldsymbol{X}_\mathrm{F}^\top$, and $\boldsymbol{X}_\mathrm{B}^\top$, are then concatenated along the channels and fed into the GCN [22]. This fuses the features from different channels and extracts non-Euclidean features from the image. Finally, we add the extracted features to the features $\boldsymbol{X}_\mathrm{M}$ and $\boldsymbol{X}_\mathrm{M}^\top$ obtained from the QM to produce the final output of the HGMM, as shown in the following formula:

$$\mathrm{HGMM}(\boldsymbol{X}) = \mathrm{GCN}([\boldsymbol{X}_\mathrm{F}, \boldsymbol{X}_\mathrm{B}, \boldsymbol{X}_\mathrm{F}^\top, \boldsymbol{X}_\mathrm{B}^\top], \boldsymbol{A}) + \boldsymbol{X}_\mathrm{M} + \boldsymbol{X}_\mathrm{M}^\top + \boldsymbol{X}, \tag{5}$$

where $[\cdot]$ denotes the concatenation operation and $\boldsymbol{A}$ is the adjacency matrix. To reduce computation, we set the value of specific positions in the adjacency matrix to one, while the rest are set to zero. Specifically, every 32 units along each axis is set to one, and the values along the axis of symmetry are also set to one to further enhance the matrix's structure.

2.3 Boundary Discrimination Fusion Module

The structure of BDFM is shown in Fig. 2. First, we generate an initial segmentation map by applying a downsampling operation and a ReLU activation function to the high-level features, which are the output of the CFM. This initial segmentation map is subsequently processed to derive three distinct feature maps corresponding to the internal, background, and boundary regions of the image. The separate processing of different regions helps capture the unique properties of each region in a more detailed manner. Next, to simplify the processing steps and make the information more compact and easier to handle, we flatten these three feature maps into one-dimensional vectors, forming the tensor $\boldsymbol{U}$.

Next, to allow effective fusion with the features output by the HGMM. We flatten each channel in each batch of the HGMM output features into a one-dimensional vector, resulting in $\boldsymbol{X}'$. The subsequent operations involve a series

Table 1. Quantitative comparison of HGM and other methods. The top of these results are shown in red and the second best in blue.

Module	CVC-300		ClinicDB		Kvasir		ColonDB		ETIS		All Datasets	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
UMNet	0.816	0.726	0.885	0.820	0.841	0.763	0.673	0.590	0.554	0.456	0.754	0.671
VMUNetV2	0.857	0.780	0.897	0.841	0.895	0.834	0.737	0.655	0.727	0.632	0.822	0.748
G-CASCADE	0.883	0.807	0.881	0.820	0.913	0.858	0.764	0.680	0.720	0.628	0.832	0.759
CFATransUnet	0.900	0.829	0.907	0.850	0.906	0.847	0.791	0.706	0.727	0.641	0.846	0.775
VANet	0.889	0.819	0.915	0.860	0.920	0.868	0.796	0.715	0.792	0.712	0.862	0.794
HSNet	0.877	0.807	0.925	0.882	0.926	0.883	0.819	0.742	0.790	0.713	0.867	0.805
Polyp-PVT	0.903	0.837	0.937	0.889	0.919	0.869	0.811	0.730	0.790	0.709	0.872	0.806
CAFE	0.897	0.835	0.932	0.885	0.926	0.878	0.828	0.749	0.803	0.725	0.877	0.814
HGM	0.910	0.846	0.939	0.896	0.928	0.879	0.837	0.759	0.820	0.743	0.887	0.825

of convolutional steps, as shown in the following formulas:

$$\boldsymbol{X}_{\mathrm{BDFM}} = \mathrm{Conv}([\mathrm{RS}(\mathrm{Conv}(\boldsymbol{U}\boldsymbol{X}')(\mathrm{Conv}(\boldsymbol{U}\boldsymbol{X}')\mathrm{Conv}(\boldsymbol{X}'))), \mathrm{RS}(\boldsymbol{X}')]). \quad (6)$$

2.4 Bidirectional Mamba Decoder and Loss Function

After passing the outputs of CFM and BDFM through HGMM, we fuse them with the original features processed by the dilated convolution module and input this fused feature map into the BMD for further decoding. We use two sets of deconvolution layers and BiMamba blocks for fusion. Leveraging the excellent feature extraction capabilities of the BiMamba block, the final fused features can effectively represent the characteristics of each individual feature. This approach ensures that while focusing on global information, it does not lose detailed features. Finally, we obtain the final output through a deconvolution layer and a fully connected layer.

Our model is trained with a mixture loss consisting of a weighted Binary Cross-Entropy (wBCE) loss and a weighted Intersection-over-Union (wIoU) loss.

3 Experiments

3.1 Experimental Settings

Datasets. In this study, based on the experimental setup of PraNet, we selected five challenging public datasets to validate the effectiveness of our framework. The datasets include: CVC-300 [16] containing 60 images, ClinicDB [1] with 612 images, Kvasir [6] comprising 1000 images, ColonDB [14] consisting of 380 images, and ETIS [13] containing 196 images. These datasets, derived from diverse sources, encompass various polyp morphologies and different colonic environments. This comprehensive collection enables thorough evaluation of our framework's performance in handling diverse complex scenarios, thereby establishing a solid data foundation for verifying the framework's effectiveness.

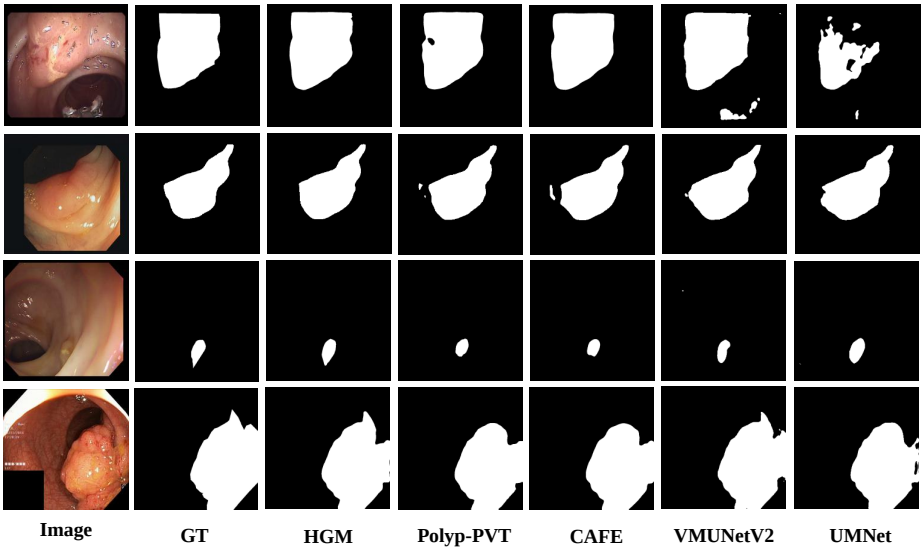

Fig. 3. Visualized segmentation results. In the five datasets mentioned in the previous experiment, three images are selected to compare the segmentation performance of our model with that of other models.

Implementation Details. In this experiment, we build and train the model using the PyTorch deep learning framework. During training, the hyperparameters are set as follows: the initial learning rate is set to 1e-4, the batch size is set to 16, and the input image resolution is set to 352×352. All experiments are conducted on a computing platform equipped with two NVIDIA RTX 3090 GPUs. To evaluate the performance of the model in the polyp segmentation task comprehensively and objectively, we use two widely applied evaluation metrics in this field: the Dice coefficient and IoU.

3.2 Comparison with State-Of-The-Art Models

Quantitative Results. To further verify the performance of HGM, under the same experimental environment and parameter settings, a performance comparison is conducted on five datasets with currently mainstream segmentation models, namely UMNet [20], VMUNetV2 [23], G-CASCADE [11], CFATransUnet [17], VANet [2], HSNet [24], Polyp-PVT [3], and CAFE [9]. The experimental results, shown in Table 1, indicate that HGM achieves the best or second-best performance in terms of Dice and IoU across all five datasets, with particularly significant advantages on ETIS (Dice: 0.820, IoU: 0.743). Furthermore, HGM achieves the best performance (Dice: 0.887, IoU: 0.825) on all datasets, outperforming the cutting-edge model, CAFE (Dice: 0.877, IoU: 0.814). While other methods like Polyp-PVT and CAFE excel in specific datasets (e.g., Polyp-PVT achieves a Dice of 0.937 on ClinicDB), HGM demonstrates superior consistency

Table 2. Ablation results of different modules. The top of these results are shown in red and the second best in blue.

Model					ClinicDB		CVC-300		ETIS	
Base	BMD	QM	GCN	BDFM	Dice	IoU	Dice	IoU	Dice	IoU
✓					0.917	0.869	0.883	0.818	0.787	0.705
✓	✓				0.924	0.876	0.899	0.829	0.801	0.725
✓	✓	✓			0.929	0.882	0.900	0.831	0.792	0.718
✓	✓	✓	✓		0.934	0.900	0.900	0.837	0.811	0.733
✓	✓	✓	✓	✓	0.939	0.896	0.911	0.847	0.820	0.743

and generalization. The findings suggest that HGM effectively improves segmentation accuracy, especially in complex scenarios (e.g., ETIS), validating its overall superiority.

Qualitative Results. Fig. 3 shows a qualitative comparison of the HGM with other segmentation models. Visually, HGM accurately segments polyps of varying sizes and achieves good segmentation results. Moreover, HGM demonstrates outstanding performance in boundary detection. As shown in rows 2, 3, and 4, our method effectively segments the edges of the polyps, reducing segmentation errors. This is due to the ability of the approach to extract non-Euclidean features and fuse multi-scale features, preserving positional, detail, and boundary information while minimizing feature loss during processing.

3.3 Ablation Study

The results of the ablation experiments in Table 2 show that each component makes an important contribution to the overall performance of the model. The first row of the Table 2 represents the baseline model, which uses convolutional layers to replace the corresponding modules. When the BMD is added, two metrics increase on the three datasets. Taking the CVC-300 dataset as an example, the Dice coefficient is increased to 0.899, and the IoU is increased to 0.829. This is attributed to the fact that this module can aggregate multi-scale features and use receptive fields of different sizes, reducing the loss of original features. When the QM component is added continuously, the performance of the model is further improved. On the ClinicDB dataset, the Dice coefficient reaches 0.929, and the IoU is 0.882. This is because its extraction of multi-directional features enhances the ability to recognize lesions. When the GCN component is added, the performance improvement of the model on each dataset is more prominent. On the ETIS dataset, the Dice coefficient is increased to 0.811, and the IoU is increased to 0.733, indicating that the extraction of non-Euclidean features by the GCN component has a great impact on improving the segmentation effect. Finally, when the BDFM is added, it improves the fusion effect between low-level features and high-level features, retains more information, enhances the robustness of polyp segmentation, and the model performance reaches the best.

4 Conclusion

In this paper, we proposed HGM, a novel deep learning model for accurate polyp segmentation, which employs a multi-level feature extraction and fusion architecture. It combines a pyramid vision transformer for feature extraction at different levels, the HGMM for low-level features, and the CFM for high-level features. The low- and high-level features are then fused using the BDFM. Finally, the BMD extracts and fuses multi-scale features using different-sized receptive fields, reducing feature loss to ensure more comprehensive feature information. Our model outperforms State-Of-The-Art methods on five benchmark datasets, achieving superior segmentation results.

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

References

1. 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)
2. Cai, L., Chen, L., Huang, J., Wang, Y., Zhang, Y.: Know your orientation: A viewpoint-aware framework for polyp segmentation. Medical Image Analysis **97**, 103288 (2024)
3. 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)
4. Elfwing, S., Uchibe, E., Doya, K.: Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Networks **107**, 3–11 (2018)
5. 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)
6. 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)
7. Ji, G.P., Xiao, G., Chou, Y.C., Fan, D.P., Zhao, K., Chen, G., Van Gool, L.: Video polyp segmentation: A deep learning perspective. Machine Intelligence Research **19**(6), 531–549 (2022)
8. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
9. Liu, G., Yao, S., Liu, D., Chang, B., Chen, Z., Wang, J., Wei, J.: CAFE-Net: Cross-attention and feature exploration network for polyp segmentation. Expert Systems with Applications **238**, 121754 (2024)
10. Mei, J., Zhou, T., Huang, K., Zhang, Y., Zhou, Y., Wu, Y., Fu, H.: A survey on deep learning for polyp segmentation: Techniques, Challenges and Future Trends. Visual Intelligence **3**(1), 1 (2025)

11. Rahman, M.M., Marculescu, R.: G-CASCADE: Efficient cascaded graph convolutional decoding for 2d medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7728–7737 (2024)
12. Shao, H., Zhang, Y., Hou, Q.: Polyper: Boundary Sensitive Polyp Segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 4731–4739 (2024)
13. 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)
14. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. IEEE Transactions on Medical maging **35**(2), 630–644 (2015)
15. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)
16. 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)
17. Wang, C., Wang, L., Wang, N., Wei, X., Feng, T., Wu, M., Yao, Q., Zhang, R.: CFATransUnet: Channel-wise cross fusion attention and transformer for 2d medical image segmentation. Computers in Biology and Medicine p. 107803 (2024)
18. Wang, J., Chen, J., Chen, D., Wu, J.: LKM-UNet: Large kernel vision mamba unet for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 360–370. Springer (2024)
19. 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)
20. Wang, Y., Zhang, W., Wang, L., Liu, T., Lu, H.: Multi-Source Uncertainty Mining for Deep Unsupervised Saliency Detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11727–11736 (2022)
21. Xie, X., Cui, Y., Tan, T., Zheng, X., Yu, Z.: FusionMamba: Dynamic feature enhancement for multimodal image fusion with mamba. Visual Intelligence **2**(1), 37 (2024)
22. Yao, L., Mao, C., Luo, Y.: Graph convolutional networks for text classification. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 7370–7377 (2019)
23. Zhang, M., Yu, Y., Jin, S., Gu, L., Ling, T., Tao, X.: VM-UNET-V2: rethinking vision mamba unet for medical image segmentation. In: International Symposium on Bioinformatics Research and Applications. pp. 335–346. Springer (2024)
24. Zhang, W., Fu, C., Zheng, Y., Zhang, F., Zhao, Y., Sham, C.W.: HSNet: A hybrid semantic network for polyp segmentation. Computers in Biology and Medicine **150**, 106173 (2022)
25. Zhou, T., Zhang, Y., Chen, G., Zhou, Y., Wu, Y., Fan, D.P.: Edge-aware feature aggregation network for polyp segmentation. Machine Intelligence Research **22**(1), 101–116 (2025)