Mamba Guided Boundary Prior Matters: A New Perspective for Generalized Polyp Segmentation

Tapas K. Dutta[1], Snehashis Majhi[2], Deepak Ranjan Nayak[3], and Debesh Jha[4]

[1] University of Surrey, United Kingdom
[2] Côte d'Azur University, France
[3] Malaviya National Institute of Technology Jaipur, India
[4] University of South Dakota, USA
drnayak.cse@mnit.ac.in
https://github.com/deepak1113/SAM-MaGuP

Abstract. Polyp segmentation in colonoscopy images is crucial for early detection and diagnosis of colorectal cancer. However, this task remains a significant challenge due to the substantial variations in polyp shape, size, and color, as well as the high similarity between polyps and surrounding tissues, often compounded by indistinct boundaries. While existing encoder-decoder CNN and transformer-based approaches have shown promising results, they struggle with stable segmentation performance on polyps with weak or blurry boundaries. These methods exhibit limited abilities to distinguish between polyps and non-polyps and capture essential boundary cues. Moreover, their generalizability still falls short of meeting the demands of real-time clinical applications. To address these limitations, we propose **SAM-MaGuP**, a groundbreaking approach for robust polyp segmentation. By incorporating a boundary distillation module and a 1D-2D Mamba adapter within the Segment Anything Model (SAM), SAM-MaGuP excels at resolving weak boundary challenges and amplifies feature learning through enriched global contextual interactions. Extensive evaluations across five diverse datasets reveal that SAM-MaGuP outperforms state-of-the-art methods, achieving unmatched segmentation accuracy and robustness. Our key innovations—a Mamba-guided boundary prior and a 1D-2D Mamba block—set a new benchmark in the field, pushing the boundaries of polyp segmentation to new heights.

Keywords: Colorectal cancer · polyp segmentation · Mamba-guided boundary prior.

1 Introduction

Colorectal Cancer (CRC) is one of the leading causes of cancer-related death across the globe, ranking among the third most frequently diagnosed cancer [3, 18]. It starts from certain growths or polyps in the inner lining of the colon and often leads to CRC if not treated timely. The incidence of CRC can be reduced significantly with early diagnosis and on-time treatment, saving many lives. Colonoscopy is the gold standard screening procedure for identifying and removing polyps. However, this detection procedure is manual, time-consuming, and highly reliant on the clinician's skill and expertise. Additionally, the miss rate of polyp detection is considerably high (6-27%) [1]. Therefore, it is of utmost significance to design automated polyp segmentation approaches to detect the polyp and provide clinicians with precise location and boundary information, reducing the miss rate.

There has been remarkable progress in the development of polyp segmentation methods based on deep learning in the past years, and the results shown have been awe-inspiring [17, 28, 9]. In early efforts, U-Net and its variants, such as UNet++ and ResUNet++ [12], have been adopted as a model of choice. However, these models fall short of capturing boundary details. Following these works, various notable architectures such as FCN [10], PraNet [9], MSNet [25], CFA-Net [27], and MEGANet [4] have been proposed. While such architectures exhibit their capability to capture the boundary cues and handle variability in the size of polyps, they encounter challenges in extracting global contextual information, which is crucial for polyp identification. Later, inspired by the remarkable success of transformers, a few architectures such as PVT-Cascade [16] and CTNet [23] has recently been introduced, which allows modeling global feature dependencies and learning prominent features and have demonstrated impressive segmentation results.

Despite the significant progress, segmenting polyps has remained a daunting challenge due to the camouflage characteristics of polyps, i.e., intrinsic resemblance among polyps and the surrounding mucosa, considerable variability in polyp's shape and size, and absence of sharp boundaries. This results in unstable segmentation performance and poor localization of polyps. In addition, the existing models have shown limited capacity to learn distinguishable features and detailed boundary cues from polyp regions. Segment Anything Model (SAM), a foundational model, has recently garnered the interest of researchers for segmenting images with exceptional generalization capabilities [13]. However, its performance is limited when applied to medical images because of inadequate domain-specific knowledge [26]. Additionally, it leads to high computational overhead and memory requirements [22]. To address these issues, an adapter module integrated SAM architecture has recently been proposed by leveraging Mamba [6], an efficient approach known for its excellent capability to model long-range interactions with linear computational complexity [29]. Although this approach has obtained superior generalization performance with exceptional computational benefits, it does not take into account boundary information, leading to suboptimal polyp segmentation performance. Additionally, the generalizability needs to be improved further to meet real-time clinical requirements.

To address the persistent challenges in polyp segmentation, we introduce **SAM-MaGuP**, a novel framework that leverages boundary distillation and the Mamba adapter to enhance SAM's performance. The key building block **MaGuP** module revolutionizes general-adapter-based [5] fine-tuning by refining SAM's feature representations to capture subtle boundary details often overlooked by conventional models. The module is composed of two transformative components, **1D-2D Mamba:** A unique feature fusion technique that merges channel (1D) and spatial (2D) information, empowering SAM to learn global contexts at multiple scales, unlocking richer feature representations. **Boundary Distillation Component (BDC):** A dynamic unit that sharpens the model's ability to detect exact boundary locations, even in ambiguous or weakly defined regions. Extensive experiments across five diverse datasets demonstrate that SAM-MaGuP effectively tackles blurry boundary issues, setting a new standard by outshining current state-of-the-art (SoTA) methods. In summary, our major contributions are:

– We introduce a Mamba-guided boundary prior in SAM, the first of its kind, to effectively tackle the weak-boundary challenge in polyp segmentation.
– We propose a boundary distillation component that empowers SAM to accurately detect polyps in weak-boundary scenarios, enhanced by a 1D-2D Mamba block that optimizes feature interactions across spatial and channel dimensions.
– We demonstrate the outstanding performance of SAM-MaGuP on five diverse datasets, consistently surpassing existing SoTA polyp segmentation techniques.

2 Proposed SAM-MaGuP

In this section, we present the **SAM-MaGuP** framework, specifically tailored to tackle the challenging task of weak-boundary polyp segmentation. By leveraging the versatility of the SAM backbone, we elevate its capabilities to address complex segmentation tasks. We first introduce SAM, a powerful model designed for general-purpose object segmentation, before diving into the core of **SAM-MaGuP**, which enhances SAM's performance in polyp segmentation through the innovative **MaGuP Module**. This module brings unprecedented precision to weak-boundary scenarios, ensuring more effective segmentation.

2.1 SAM Overview and Limitations for Polyp Segmentation

SAM comprises three key components, **Image Encoder**: Built on the ViT-H/16 architecture and pre-trained with Masked Autoencoders (MAE), this encoder uses windowed attention and global attention blocks to output down-sampled image embeddings. **Prompt Encoder**: This module processes both sparse and dense prompts, tailored with positional encoding and learned embeddings. **Mask Decoder**: A modified transformer decoder with dynamic mask prediction, refining the interaction between image and prompt embeddings to output prediction.

While SAM excels in general object segmentation, its performance falters in weak-boundary polyp segmentation due to two major limitations: **Inferior Transfer Learning**: SAM's traditional full fine-tuning approach results in overfitting or feature degradation when trained on small or limited datasets. **Weak-Boundary Challenge**: SAM is ill-equipped to handle ambiguous, low-contrast polyp regions that blend with the background, making it difficult to generate

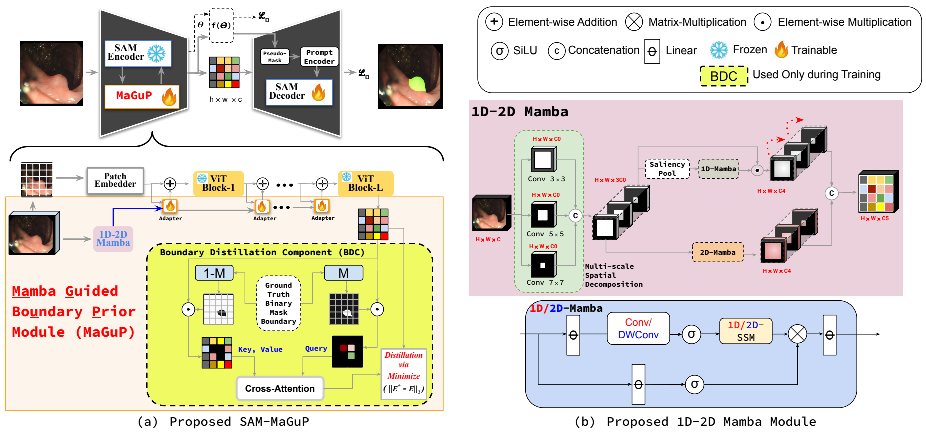

Fig. 1. Overall framework of the SAM-MaGuP for weak-boundary polyp segmentation. It incorporates the MaGuP Adapter-based fine-tuning into the SAM backbone to enhance its representation ability for weak-boundary polyp segmentation. In MaGuP, the BDC is only used during training to refine SAM's feature learning ability pertaining to diverse polyp shapes and sizes.

accurate prompts or define subtle boundaries. Thus, using the original SAM for segmenting weak-boundary polyps remains a challenge.

2.2 Mamba Guided Boundary Prior (MaGuP) Module

The **MaGuP Module** introduces a groundbreaking approach to general-adapter-based fine-tuning, improving SAM's ability to tackle polyp segmentation. It acts as a plug-in general-adapter, refining SAM's feature representations to capture intricate boundary details that are often missed in traditional models. The module consists of two key components: **1D-2D Mamba**: A feature transformation that bridges the gap between channel (1D) and spatial (2D) representations, enabling SAM to better fuse channel and spatial information in global contexts for more accurate segmentation. **BDC**: A unit focused on sharpening boundary representations by transferring knowledge of precise edge localization, even in weak or ambiguous regions.

1D-2D Mamba: To enhance SAM's pre-trained features for polyp segmentation, we introduce the **1D-2D Mamba** (Fig. 1b). This component incorporates two powerful strategies: **Multi-scale Spatial Decomposition (MSD)**: Decomposes the input image into multiple scales through parallel convolutions with varying receptive fields, creating a multi-scale pyramid with coarse-to-fine features. **Mamba Spatial and Channel Interaction**: Captures long-range dependencies across channel and spatial dimensions by 1D and 2D Mamba layers, enhancing the ability to inject multi-scale polyp-specific cues into SAM's encoder for improved segmentation accuracy.

Multi-scale Spatial Decomposition (MSD): An input image $I \in \mathbb{R}^{H \times W \times C}$ is processed through parallel convolutions with varying receptive fields ($k \in \{3, 5, 7\}$) to capture features at multiple scales. The resulting maps F_k are padded and stacked along the channel dimension to form a multi-scale feature pyramid $F^* \in \mathbb{R}^{H \times W \times 3C_0}$, with coarse-grained features (F_7) at the top and fine-grained features (F_3) at the bottom. This decomposition allows the model to analyze the polyp region with fine-to-coarse spatial semantics.

Mamba Spatial and Channel Interaction: Mamba efficiently captures intra-pixel interactions of salient (F^S) and contextual (F^C) maps using two parallel 1D and 2D Mamba layers [29, 14], each encoding dependencies in the multi-scale channel and spatial distributions. A gated mechanism refines feature representations, where the outputs are computed via a linear layer ($\phi(.)$), SiLU activation (σ), and matrix multiplication ($\otimes$). To retain fine-grained cues of channels, the original multi-scale feature maps are combined with the gated outputs using a skip connection, forming a domain-specific embedded feature map $F^D \in \mathbb{R}^{H \times W \times C_5}$. However, the skip connection is discarded for the spatial stream to avoid information redundancy. These interaction encodings enhance the model's potential to capture critical polyp cues across scales.

Boundary Distillation Component (BDC): The **BDC**, as shown in Fig. 1a, refines SAM's ability in to understand polyp shapes and structures. It uses the ground truth mask to clearly define polyp (E_P) and non-polyp (E_{NP}) regions, enhancing segmentation precision. By starting with the latent feature embeddings from the ViT Block-L (E), we compute two activation regions:

$$E_P = E \odot M, \quad E_{NP} = E \odot (1 - M) \tag{1}$$

where $E \in \mathbb{R}^{h \times w \times c}$ is the feature embedding, M is the ground truth mask, and $1-M$ represents the complement of the ground truth mask (M), highlighting non-polyp regions.

To further refine the boundary details, we employ a **cross-attention mechanism** where E_P serves as the `query`, and E_{NP} as the `key` and `value`. The resulting embedding, E^*, captures the critical polyp boundaries while suppressing non-polyp regions. This refined boundary information is then integrated into SAM's representation space using a **knowledge distillation framework**, aligning the original embedding E with the boundary-enhanced embedding E^*. This is done by minimizing their representational difference:

$$\mathcal{L}_{distill} = \frac{1}{h * w * c} \sum_{i=1}^{h} \sum_{j=1}^{w} \sum_{k=1}^{c} \left(E_{i,j,k} - E^*_{i,j,k} \right)^2 \tag{2}$$

By minimizing this loss, the **BDC** can be discarded during inference, eliminating the need for explicit ground truth while still maintaining high boundary precision in polyp segmentation. This makes the process more efficient and accurate, even without relying on ground truth data.

2.3 SAM Decoder

The architecture of the mask decoder is adopted from [13], which leverages various prompts like bounding boxes, masks, or points for improved segmentation. However, this requires explicit prompts. To address this, **SAM-MaGuP** generates a pseudo mask via $f(\theta)$, trained with **MaGuP**, optimizing the model using $\mathcal{L}_D$ (discussed in Section 2.4). This pseudo mask subsequently serves as a prompt to refine the final segmented output during training, involving the image encoder and mask decoder.

2.4 Training Objective

The SAM-MaGuP model, including its components like general-adapter, MaGuP, $f(\theta)$, and mask decoder, is jointly trained with a loss $\mathcal{L}_D = \mathcal{L}_w^{\text{Dice}} + \mathcal{L}_w^{\text{BCE}}$. For a predicted segmentation mask (P) and the ground truth mask (M), $\mathcal{L}_w^{\text{Dice}} = 1 - \frac{2 \cdot \sum (P \cdot M)}{\sum P + \sum M + \epsilon}$, and $\mathcal{L}_w^{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} [M_i \cdot \log(P_i) + (1 - M_i) \cdot \log(1 - P_i)]$, where N indicates number of pixels.

We follow a two-stage training approach:

Stage I: Train adapters in the image encoder with deep supervision, using up-sampled outputs O_{Enc}^{Up} and ground-truth M, with loss: $\mathcal{L}_{\text{Stage-I}} = \mathcal{L}_D(M, O_{Enc}^{\text{Up}})$.
Stage II: Train both the image encoder's adapter and mask decoder with full supervision, using the outputs O_{Dec} and O_{Enc}^{up}, with loss: $\mathcal{L}_{\text{Stage-II}} = \mathcal{L}_D(M, O_{Dec})$.

3 Experiments and Results

This section presents the quantitative and qualitative comparisons with SoTA methods, ablation study, and implementation details.

3.1 Datasets, Evaluation Measures, and Implementation Details

Datasets:

We conduct a series of experiments on five widely used and challenging benchmark datasets: CVC-ClinicDB [2], ETIS [19], CVC-ColonDB [20], CVC-300 (EndoScene) [21], and Kvasir-SEG [11]. For a fair comparison, we follow the training settings of the previous study [9], where a mixture of images (1450 images) from Kvasir-SEG (900 images) and CVC-ClinicDB (550 images) datasets has been taken into account to train the model. For testing, the remaining 62 images from CVC-ClinicDB and 100 from Kvasir-SEG as well as 196, 380, and 60 images from ETIS, CVC-ColonDB, and CVC-300, are considered.

Evaluation Measures:

To evaluate the performance of our proposed model as well as SoTA methods, we adopt six evaluation measures: F-measure (F_β^w) [15], S-measure (S_α) [7], E-measure ($E_\phi^{\max}$) [8], mean absolute error ($\mathcal{M}$), mean Dice (mDice), and mean IoU (mIoU), following the previous works [9, 25].

Implementation Details: We implement our proposed SAM-MaGuP in PyTorch and conduct all experiments using an NVIDIA A100 GPU. To derive a fair comparison, we uniformly resize the input images to 352×352 pixels. To augment the data, we adopt a multi-scale training strategy with scaling factors {0.75, 1, 1.25}. We train the model for 200 epochs using the Adam optimizer. The initial learning rate and batch size are set to 1×10^{-5} and 8, respectively.

3.2 Quantitative Comparison

From the SoTA comparison (Table 1), **SAM-MaGuP** consistently outperforms all existing methods across **seen** and **unseen** datasets. On **seen datasets**, SAM-MaGuP achieves the highest scores across all metrics. Notably, on **Kvasir-SEG**, it yields an mDice of 94.7%, mIoU of 89.0%, and S_α of 95.1%, significantly outperforming Polyp-Mamba and SAM-Mamba. Its S_α (96.5%) and $E_\phi^{\max}$ (98.3%) reflect exceptional boundary localization. Similarly, on **ClinicDB**, it delivers an mDice of 95.3% and mIoU of 91.3%, reaffirming its robustness on challenging datasets. For **unseen datasets** (Table 2), SAM-MaGuP demonstrates unparalleled generalization. It achieves the best scores across all metrics, effectively handling weak and ambiguous boundaries. On the challenging ETIS dataset, SAM-MaGuP sets a new benchmark with an mDice of 85.4%, F_β^w of 86.2%, and the lowest MAE of 1.0, showcasing its capability to segment complex polyp structures. These results highlight SAM-MaGuP's robustness, adaptability, and SoTA performance. Further, compared to SAM-Mamba, SAM-MaGuP offers improved segmentation accuracy with minimal additional cost: 106M vs. 103M trainable parameters, identical inference-time parameters (103M), and a slight increase in GFLOPs (431 vs. 423).

In a nutshell, SAM-MaGuP not only excels on seen cases but also demonstrates exceptional generalization capabilities on unseen cases, consistently achieving top performance across all metrics. This highlights its robustness, superior boundary localization, and efficiency compared to existing SoTA methods.

Table 1. Comparison of quantitative results with SoTA approaches on **seen datasets**. 'Green' and 'Blue' color fonts represent the best and second-best results.

Methods	Kvasir-SEG (Seen)						ClinicDB (Seen)					
	mDice	mIoU	F_β^w	S_α	$E_\phi^{\max}$	$\mathcal{M}$	mDice	mIoU	F_β^w	S_α	$E_\phi^{\max}$	$\mathcal{M}$
U-Net (MICCAI'15) [17]	81.8	74.6	79.4	85.8	89.3	5.5	82.3	75.5	81.1	88.9	95.4	1.9
U-Net++ (TMI'19) [28]	82.1	74.3	80.8	86.2	91.0	4.8	79.4	72.9	78.5	87.3	93.1	2.2
SFA (MICCAI'19) [10]	72.3	61.1	67.0	78.2	84.9	7.5	70.0	60.7	64.7	79.3	88.5	4.2
PraNet (MICCAI'20) [9]	89.8	84.0	88.5	91.5	94.8	3.0	89.9	84.9	89.6	93.6	97.9	0.9
MSNet (MICCAI'21) [25]	90.7	86.2	89.3	92.2	94.4	2.8	92.1	87.9	91.4	94.1	97.2	0.8
CFA-Net (PR'23) [27]	91.5	86.1	90.3	92.4	96.2	2.3	93.3	88.3	92.4	95.0	98.9	0.7
PVT-Cascade (WACV'23) [16]	91.1	86.3	90.6	91.9	96.1	2.5	91.9	87.2	91.8	93.6	96.9	1.3
CTNet (TCYB'24) [23]	91.7	86.3	91.0	92.8	95.9	2.3	93.6	88.7	93.4	95.2	98.3	0.6
MEGANet (WACV'24) [4]	91.3	86.3	90.7	91.8	95.9	2.5	93.8	89.4	94.0	95.0	98.6	0.6
Polyp-Mamba (MICCAI'24) [24]	94.0	88.1	94.2	93.5	98.3	1.6	94.9	90.7	95.2	96.5	99.3	0.5
SAM-Mamba (WACV'25) [6]	92.4	87.3	94.2	93.6	96.1	2.5	94.2	88.7	94.3	95.5	98.2	0.6
SAM-MaGuP (Ours)	94.7	89.0	95.1	94.2	98.1	1.6	95.3	91.3	95.8	96.3	98.8	0.5

3.3 Qualitative Comparison

Qualitative evaluations (Fig. 2) demonstrate that our model excels in capturing intricate polyp boundaries across diverse datasets. Leveraging multi-scale boundary refinement, it precisely delineates polyp contours, even in low-contrast or ambiguous edge cases. On the ClinicDB dataset, SAM-MaGuP effectively distinguishes subtle boundary transitions from adjacent tissues, minimizing leakage and ensuring clear separation. In the challenging ETIS dataset, the model progressively sharpens coarse boundary cues, enhancing segmentation accuracy

Table 2. Comparison of quantitative results with SoTA approaches on **unseen datasets**. 'Green' and 'Blue' color fonts represent the best and second-best results.

Methods	CVC-300 (Unseen)				ColonDB (Unseen)				ETIS (Unseen)			
	mDice	mIoU	F_β^w	$\mathcal{M}$	mDice	mIoU	F_β^w	$\mathcal{M}$	mDice	mIoU	F_β^w	$\mathcal{M}$
U-Net (MICCAI'15) [17]	71.0	62.7	68.4	2.2	51.2	44.4	49.8	6.1	39.8	33.5	36.6	3.6
U-Net++ (TMI'19) [28]	70.7	62.4	68.7	1.8	48.3	41.0	46.7	6.4	40.1	34.4	39.0	3.5
SFA (MICCAI'19) [10]	46.7	32.9	34.1	6.5	46.9	34.7	37.9	9.4	29.7	21.7	23.1	10.9
PraNet (MICCAI'20) [9]	87.1	79.7	84.3	1.0	70.9	64.0	69.6	4.5	62.8	56.7	60.0	3.1
MSNet (MICCAI'21) [25]	86.9	80.7	84.9	1.0	75.5	67.8	73.7	4.1	71.9	66.4	67.8	2.0
CFA-Net (PR'23) [27]	89.3	82.7	93.8	0.8	74.3	66.5	72.8	3.9	73.2	65.5	69.3	1.4
PVT-Cascade (WACV'23) [16]	89.2	82.4	87.3	0.9	78.1	71.0	77.9	3.1	78.6	71.2	75.9	1.3
CTNet (TCYB'24) [23]	90.8	84.4	89.4	0.6	81.3	73.4	80.1	2.7	81.0	73.4	77.6	1.4
MEGANet (WACV'23) [4]	89.9	83.4	88.2	0.7	79.3	71.4	77.9	4.0	73.9	66.5	70.2	3.7
Polyp-Mamba (MICCAI'24) [24]	92.1	87.5	89.5	0.5	82.9	74.3	81.5	2.7	82.5	74.7	76.6	1.2
SAM-Mamba (WACV'25) [6]	92.0	86.1	88.8	0.6	85.3	77.1	85.6	1.7	84.8	78.2	85.5	1.0
SAM-MaGuP (Ours)	92.7	88.0	90.1	0.5	85.9	78.5	86.2	1.7	85.4	78.9	86.2	1.0

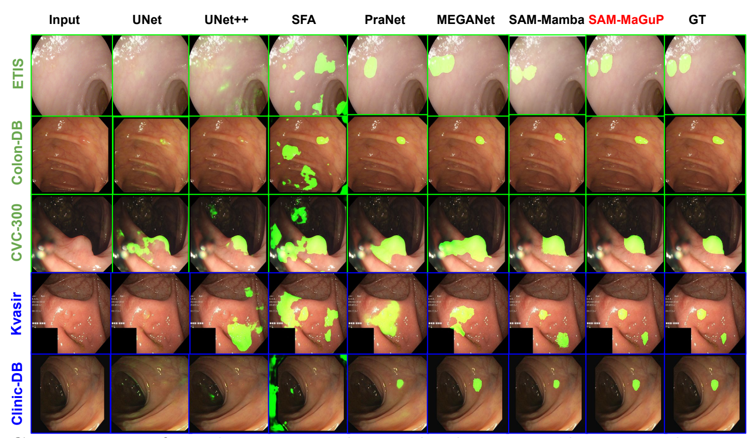

Fig. 2. Comparison of qualitative results on both seen and unseen datasets demonstrates the SAM-MaGuP's learning ability to segment polyps with shape, size, texture variations and intricate boundaries, and showcases strong generalization.

and reducing false positives by accurately localizing transitions between polyps and the background. These results underscore the critical role of our boundary-focused modules in delivering robust and precise polyp segmentation.

3.4 Ablation Study and Discussion

Our ablation study (Table 3) highlights the incremental contributions of each module to the overall performance of SAM-MaGuP. Across all datasets, the baseline Adapter achieves a reasonable mDice score. Subsequently, the addition of the MSD module significantly enhances multi-scale feature learning, boosting performance by an acceptable margin. Integrating the 1D-Mamba module further improves channel-wise feature discrimination, leading to an additional performance increase. Further, the 2D-Mamba module refines spatial context and local feature interactions, contributing a smaller but meaningful gain. Finally, the inclusion of the BDC sharpens boundary localization and aligns complementary features, resulting in an additional improvement. These results demonstrate the synergistic effect of each module, underscoring their critical role in achieving state-of-the-art performance.

Table 3. Results of ablation study on the components of the proposed MaGuP module.

Configuration					Seen		Unseen		
Adapter	MSD	1D-Mamba	2D-Mamba	BDC	Kvasir	Clinic-DB	CVC-300	ColonDB	ETIS
✓	-	-	-	-	89.9	89.9	80.9	80.1	80.6
✓	✓	-	-	-	90.9	91.3	90.3	80.8	81.2
✓	✓	✓	-	-	92.4	94.2	92.0	85.3	84.8
✓	✓	✓	✓	-	93.1	94.7	92.2	85.5	84.9
✓	✓	✓	✓	✓	94.7	95.3	92.7	85.9	85.4

4 Conclusion

In summary, SAM-MaGuP represents a significant leap forward in polyp segmentation, addressing longstanding challenges related to boundary ambiguity and feature generalization. By introducing a novel BDC and a 1D-2D Mamba adapter into the SAM, SAM-MaGuP achieves precise segmentation even for polyps with weak or indistinct boundaries. Its ability to capture global contextual interactions and refine boundary cues ensures superior segmentation accuracy, setting a new standard in the field. Extensive experiments on diverse datasets confirm its robustness and clinical relevance, demonstrating that SAM-MaGuP is not only a powerful innovation in medical imaging but also a practical solution for real-world diagnostic needs.

Acknowledgments. This work is supported by the Anusandhan National Research Foundation (ANRF), Government of India (grant number CRG/2023/007397).

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

References

1. Ahn, S.B., Han, D.S., et al.: The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. Gut and Liver **6**(1), 64 (2012)
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. Bernal, J., Sánchez, J., Vilarino, F.: Towards automatic polyp detection with a polyp appearance model. Pattern Recognition **45**(9), 3166–3182 (2012)
4. 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. pp. 7985–7994 (2024)
5. Chen, Z., Duan, Y., Wang, W., He, J., Lu, T., Dai, J., Qiao, Y.: Vision transformer adapter for dense predictions. arXiv preprint arXiv:2205.08534 (2022)
6. Dutta, T.K., Majhi, S., Nayak, D.R., Jha, D.: Sam-mamba: Mamba guided sam architecture for generalized zero-shot polyp segmentation. In: Proceedings of the Winter Conference on Applications of Computer Vision (WACV). pp. 4655–4664 (February 2025)

7. Fan, D.P., Cheng, M.M., Liu, Y., Li, T., Borji, A.: Structure-measure: A new way to evaluate foreground maps. In: Proceedings of IEEE International Conference on Computer Vision. pp. 4548–4557 (2017)
8. Fan, D.P., Gong, C., Cao, Y., Ren, B., Cheng, M.M., Borji, A.: Enhanced-alignment measure for binary foreground map evaluation. arXiv preprint arXiv:1805.10421 (2018)
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: Medical Image Computing and Computer Assisted Intervention (MICCAI). pp. 263–273. Springer (2020)
10. Fang, Y., Chen, C., Yuan, Y., Tong, K.y.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: 22nd International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2019). pp. 302–310. Springer (2019)
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. Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., De Lange, T., Halvorsen, P., Johansen, H.D.: Resunet++: An advanced architecture for medical image segmentation. In: 2019 IEEE International Symposium on Multimedia (ISM). pp. 225–2255. IEEE (2019)
13. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. arXiv:2304.02643 (2023)
14. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: Vmamba: Visual state space model. In: Advances in Neural Information Processing systems. pp. 103031–103063 (2025)
15. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps? In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2014)
16. Rahman, M.M., Marculescu, R.: Medical image segmentation via cascaded attention decoding. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6222–6231 (2023)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: 18th international conference on Medical image computing and computer-assisted intervention (MICCAI 2015). pp. 234–241. Springer (2015)
18. Siegel, R.L., Wagle, N.S., Cercek, A., Smith, R.A., Jemal, A.: Colorectal cancer statistics, 2023. CA: A Cancer Journal for Clinicians **73**(3), 233–254 (2023)
19. 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)
20. 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)
21. 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)

22. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)
23. Xiao, B., Hu, J., Li, W., Pun, C.M., Bi, X.: Ctnet: Contrastive transformer network for polyp segmentation. IEEE Transactions on Cybernetics (2024)
24. 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)
25. Zhao, X., Zhang, L., Lu, H.: Automatic polyp segmentation via multi-scale subtraction network. In: 24th International Conference on Medical Image Computing and Computer Assisted Intervention–MICCAI 2021. pp. 120–130. Springer (2021)
26. Zhou, T., Zhang, Y., Zhou, Y., Wu, Y., Gong, C.: Can sam segment polyps? arXiv preprint arXiv:2304.07583 (2023)
27. 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)
28. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging **39**(6), 1856–1867 (2019)
29. 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)