SD-SAM: Synergizing SAM and DINOv3 for Parameter-Efficient Polyp Segmentation and a Challenging Benchmark Dataset

Haoran Yang[1], Jiachuan Hou[2], Pu Huang[3], and Jianlei Liu[1]✉

[1] Qufu Normal University, Qufu, China
jianleiliu@qfnu.edu.cn
[2] Jining Rencheng District People's Hospital, Jining, China
[3] Shandong Key Laboratory of Medical Physics and Image Processing, School of Communication and Electronic Engineering, Shandong Normal University, Jinan, China

Abstract. Accurate polyp segmentation is critical for the early diagnosis of colorectal cancer, a task where the Segment Anything Model (SAM) has shown great potential. However, adapting SAM to this domain presents two major challenges. First, because SAM requires explicit spatial prompts, automating this process typically relies on parameter-heavy auxiliary networks, which often generate inaccurate prompts. Second, fine-tuning SAM introduces a strict trade-off between minimizing trainable parameters and maintaining segmentation accuracy. To address these issues, we propose SD-SAM, a parameter-efficient two-stage architecture. In the first stage, instead of directly delineating complete lesion boundaries, we model high-confidence polyp centers via Gaussian regression, circumventing heavy auxiliary networks to generate high-quality prompts with minimal parameters. In the second stage, we implement a hierarchical fine-tuning strategy that applies tailored adaptation to SAM's encoder layers based on their distinct representational roles, minimizing parameters while guaranteeing high accuracy. Furthermore, we open-source a new benchmark dataset characterized by predominantly tiny, highly dispersed polyps. Extensive experiments demonstrate that SD-SAM achieves superior performance with fewer than 3.5M trainable parameters. The code and dataset are available at `https://github.com/xxxy0523/SD-SAM`.

Keywords: Polyp Segmentation · Polyp Benchmark · SAM.

1 Introduction

Colorectal cancer remains one of the leading causes of cancer-related mortality, making the early detection of polyps a critical clinical measure [3]. In recent years, vision foundation models have revolutionized polyp detection. Notably, the Segment Anything Model (SAM) [9] enables flexible polyp segmentation through prompt-guided interactions, while DINOv3 [21] delivers robust, domain-agnostic

semantic features via advanced self-supervised learning, offering immense potential for precise polyp boundary delineation.

However, applying SAM to medical segmentation faces two critical challenges. The first challenge lies in the accuracy and efficiency of automatic prompt generation. Since SAM requires explicit prompts (e.g., points or bounding boxes) to initiate segmentation, methods like Yolo-SAM2 [17], SelfPrompt SAM [24], and APG-SAM [27] utilize parameter-heavy auxiliary networks to parse the entire lesion region to generate explicit prompts. This requirement to delineate complete boundaries is not only computationally expensive but also highly error-prone; any parsing failure directly yields inaccurate prompts that misguide SAM. The second challenge is the parameter-accuracy trade-off during fine-tuning. To adapt SAM to medical images, current approaches embed dense adapters (e.g., SAM-Adapter [4], Moe-SAM [12]) or external branches (e.g., ASPS [11], SAMUS [13]), resulting in a massive surge in trainable parameters, where any attempt to reduce this parameter count inevitably degrades segmentation accuracy.

Furthermore, mainstream benchmarks (e.g., Kvasir-SEG, CVC-ClinicDB) of polyps segmentation suffer from two primary limitations. First, they exhibit a severe "center bias", where polyps are predominantly clustered in the central image regions. Second, these datasets are heavily dominated by large polyps. These biases cause models to overfit to specific spatial locations and object scales, leading to failures when identifying the tiny, spatially dispersed polyps typical of real-world endoscopy.

To address these challenges, we propose SD-SAM, a parameter-efficient, two-stage framework for accurate polyp segmentation. SD-SAM involves two tightly coupled stages. First, rather than requiring the auxiliary network to parse the complete lesion region, we model the high-confidence polyp center via Gaussian heatmap regression. By explicitly optimizing for the "hit rate", this approach generates highly accurate prompts with a remarkably low parameter count. Second, recognizing the distinct representational roles of different SAM encoder layers [23], we propose a DINOv3-guided hierarchical fine-tuning strategy. Specifically, in the shallow layers of SAM's encoder, which primarily focus on contour information, we apply lightweight cosine similarity distillation to transfer DINOv3's superior boundary-capturing capabilities and enhance SAM's contour awareness. In contrast, for the deep layers that encode abstract semantics crucial for the final segmentation, we utilize a QKV-based feature fusion mechanism. This layer-specific strategy maximizes parameter reduction while guaranteeing high segmentation accuracy.

Furthermore, we introduce and open-source CPD1K, a clinically reliable dataset of 1,000 colonoscopy images designed to address the severe center bias and the distinct lack of minute targets in existing benchmarks. Annotated by medical professionals with formal ethical approval, CPD1K features an unprecedented proportion of micro-polyps. Crucially, its highly dispersed spatial distribution across the entire image space effectively mitigates center bias, preventing models from erroneously overfitting to location priors and providing a rigorous benchmark strictly aligned with complex clinical reality.

2 Challenging Polyp Dataset-CPD1K

To mitigate the center bias and over-representation of large polyps in existing benchmarks like Kvasir-SEG and CVC-ClinicDB, we present and open-source CPD1K. To ensure high clinical reliability, all endoscopic images were collected by medical professionals across multiple hospitals using diverse equipment, with every corresponding mask meticulously annotated by expert doctors. Our dataset has been granted formal ethical approval (ethics application number: 2026010). The dataset sample is shown in Fig 2.

Medical imaging models often overfit to location priors, erroneously favoring the central region as foreground. To quantify this "center bias", we introduce the maximum overlap probability (P_{max}). Specifically, we superimpose all binarized Ground Truth masks within a dataset and normalize the pixel-wise sum by the total number of images. This generates the spatial occurrence heatmap shown in Fig. 1, where P_{max} represents the highest probability of a polyp appearing at any single pixel location. Mainstream datasets exhibit significant center clustering, yielding P_{max} values of 0.515 for Kvasir-SEG [8], 0.342 for ClinicDB [1], 0.300 for CVC-300 [22], and 0.261 for ColonDB [2]. In contrast, CPD1K is highly dispersed, achieving the lowest P_{max} among all benchmarks at 0.223. By distributing polyps across the entire image space, including edges and corners, our dataset prevents models from relying on location shortcuts. This forces the learning of robust semantic features, providing a benchmark more closely aligned with clinical reality.

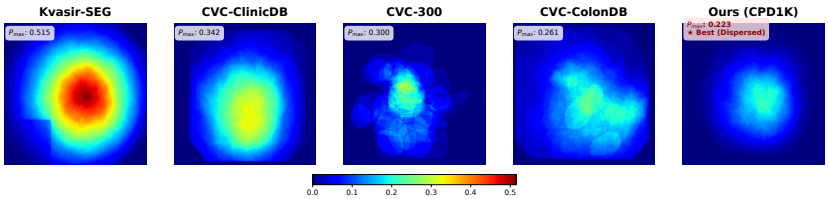

Fig. 1. Spatial distribution comparison between CPD1K and four public datasets

To evaluate dataset difficulty for fine segmentation, we quantified "micro-polyps" (area $< 2\%$) [15]. Our dataset contains 315 micro-polyp images, an order of magnitude higher than Kvasir-SEG (31) and significantly exceeding CVC-ColonDB (97) and CVC-300 (13). This high volume of minute targets presents a major challenge for computer vision, as small objects are prone to severe spatial information loss during encoder downsampling. By filling the gap in existing benchmarks for minute lesions, this dataset provides a rigorous platform for evaluating a model's boundary detection and feature recovery capabilities on extremely small targets.

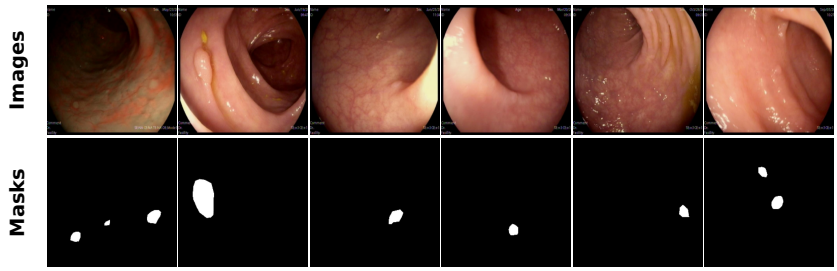

Fig. 2. Samples from the CPD1K dataset

3 Method

3.1 The Overall Architecture of SD-SAM

We propose SD-SAM, a two-stage, parameter-efficient framework for fully automatic segmentation. In the first stage, a lightweight Gaussian heatmap regression network locates lesion centers to generate high-confidence prompts, bypassing the computational burden and inaccurate prompts associated with full lesion parsing. In the second stage, we adapt the SAM encoder via a DINOv3-guided hierarchical strategy: applying cosine similarity distillation in shallow layers to enhance edge localization, and utilizing QKV-based feature fusion in deep layers for injecting robust semantic priors. The overall framework is shown in Fig 3.

3.2 Stage 1: Lightweight Gaussian Prompt Regression Network

To formulate the polyp localization task, given the ground truth binary mask M corresponding to the input image, we first extract the set of coordinates P for all foreground pixels. We then compute the geometric centroid of this set to serve as the polyp's target center $C(X, Y)$. Based on this calculated centroid, we generate a non-normalized 2D Gaussian distribution $G(x, y)$ for every spatial location (x, y) within the image, which is expressed as follows:

$$G_{(x,y)} = \exp\left(-\frac{(x-c_x)^2 + (y-c_y)^2}{2\sigma^2}\right) \cdot \mathbb{I}((x,y) \in \mathcal{P}) \tag{1}$$

We develop a 1M-parameter lightweight decoder using Depthwise Separable Convolution (DSConv) to efficiently fuse DINOv3 multiscale features through a top-down Feature Pyramid. Combined with Gaussian heatmap regression, this design significantly enhances polyp recognition and localization precision while maintaining minimal computational overhead.

During the training phase, we construct a hybrid loss function to supervise the network. The total optimization objective, L_{total}, is defined as the sum of the regression loss L_{mse} and the classification loss L_{bce}. which is expressed as follows:

$$\mathcal{L}_{stage1} = \mathcal{L}_{MSE}(\sigma(\hat{Y}), G) + \mathcal{L}_{BCE}(\hat{Y}, M) \tag{2}$$

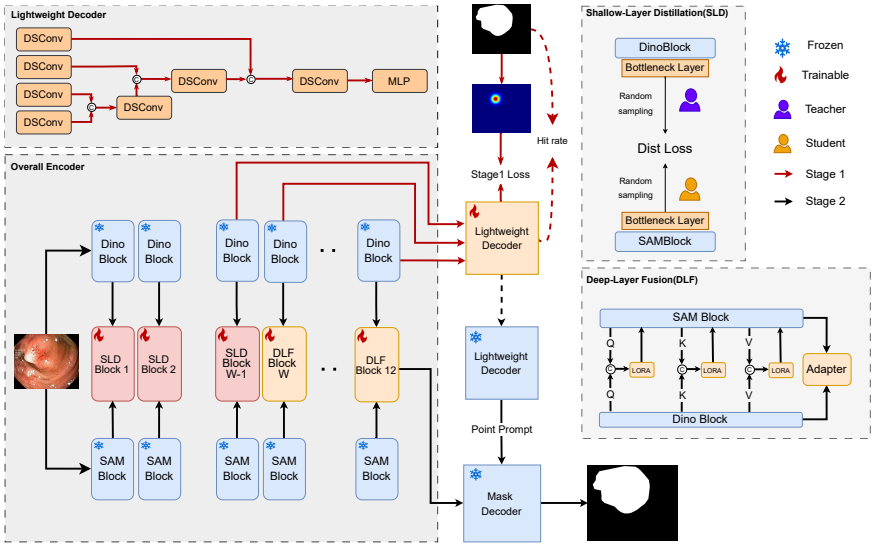

Fig. 3. The overall architecture of SD-SAM

Specifically, $\hat{Y}$ denotes the predicted logits, while G and M represent the ground truth Gaussian heatmap and binary mask, respectively. Due to the inconsistent sizes of polyps, we select only a single foreground point as the prompt input for SAM during prompt generation.

To ensure the reliability of this point prompt, we adopt the hit rate of the most confident foreground point within the actual foreground region as the evaluation metric, which is expressed as follows:

$$HitRate = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(p_i \in \mathcal{M}_i) \tag{3}$$

where N denotes the total number of images in the dataset, p_i represents the coordinate of the most confident foreground point for the i-th image, and $\mathcal{M}_i$ refers to the corresponding ground truth mask.

3.3 Stage 2: DINOv3-Guided Hierarchical Fine-Tuning for SAM

Shallow-Layer Distillation To overcome the limited contour perception capabilities of SAM's shallow encoder, we distill the robust boundary-awareness of DINOv3 into SAM. Specifically, in the shallow stages ($l < W$, $W \in [1, 13]$), DINOv3 serves as a teacher network, explicitly transferring its powerful edge-sensitive representations to enhance SAM's geometric perception. We first introduce a bottleneck projection layer to map the student features and teacher

features of the l-th layer into a unified metric space, which is expressed as follows:

$$\hat{F}_l^S = \phi_S(F_l^S), \quad \hat{F}_l^T = \phi_T(F_l^T) \tag{4}$$

Subsequently, to mitigate the computational overhead of high-resolution features and preserve the integrity of SAM's original semantic representations, we employ a spatial random sampling strategy to select a set of pixel indices $\mathcal{K}$. By maximizing the cosine similarity to align feature directions, the single-layer distillation loss is defined as:

$$\mathcal{L}_{\text{dist}} = 1 - \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \text{Cosine}\left(\frac{\hat{v}_{l,k}^S}{\|\hat{v}_{l,k}^S\|_2}, \frac{\hat{v}_{l,k}^T}{\|\hat{v}_{l,k}^T\|_2} \right) \tag{5}$$

Where $\hat{v}_{l,k}^s$ and $\hat{v}_{l,k}^t$ denote the projected feature vectors of the student and teacher networks, respectively, at spatial location k.

Deep-Layer Fusion To endow the deep layers of the SAM encoder with robust semantic understanding, we introduce a feature fusion mechanism based on QKV subspace injection. In the deep stages of the network ($l \geq W$, $W \in [1, 13]$), we perform explicit semantic enhancement on the Query (Q), Key (K), and Value (V) vectors within the attention blocks. Specifically, for each attention head at the l-th layer, we concatenate the SAM-generated features $(Q_{sam}, K_{sam}, V_{sam})$ with the corresponding features $(Q_{dino}, K_{dino}, V_{dino})$ extracted from the DINOv3 backbone along the channel dimension. Subsequently, employing a LoRA fine-tuning strategy, we use a linear projection layer to map the concatenated features back to their original dimensions. The calculation for the refined Q vector is defined as follows (with K and V calculated analogously):

$$\hat{Q}_{sam} = \text{Linear}(\text{Concat}(Q_{sam}, \text{Align}(Q_{dino}))) + Q_{sam} \tag{6}$$

Beyond the micro-level fusion of attention mechanisms, we also introduce a macro-level feature adaptation module at the end of each Transformer module. To integrate these features while maintaining efficiency, we employ a lightweight bottleneck adapter comprising three stages: dimension reduction, nonlinear activation, and dimension recovery. The workflow of this module is as follows:

$$X_{out} = \text{Linear}_{up}(\sigma(\text{Linear}_{down}(\text{Concat}(X_{sam}, F_{dino})))) + X_{sam} \tag{7}$$

4 Experiments

4.1 Datasets and Implementation Details

We evaluated our model on Kvasir [8], ClinicDB [1], ColonDB [2], CVC-300 [22], and our new CPD1K dataset. Training involved 900 images from Kvasir and 550 from ClinicDB, with ColonDB and CVC-300 as independent test sets [16]. The CPD1K dataset (1,000 pairs) was split 8:2 for further validation. Consistent random seeds were used throughout to ensure reproducible data splits.

Table 1. Quantitative comparison of our proposed method with state-of-the-art methods on five public polyp segmentation datasets.

Methods	Trainable Params	Kvasir		ClinicDB		ColonDB		CVC-300		CPD1K	
		Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
Unet[19]	31.0M	82.1	73.4	88.5	81.9	51.4	43.8	73.2	70.3	73.0	58.4
PraNet[7]	32.6M	89.8	83.8	91.9	87.6	73.5	66.2	88.5	81.0	81.2	71.1
MSNet[31]	29.7M	92.3	87.4	92.0	86.9	74.5	65.2	86.6	78.7	82.8	72.5
LDNet[29]	33.4M	91.7	86.7	93.0	88.5	76.3	66.9	87.8	80.2	81.8	72.0
Polyp-PVT[6]	25.1M	91.2	86.1	94.2	89.4	78.1	69.8	89.6	82.7	83.0	72.7
IMFF[14]	34.7M	82.4	70.6	91.1	83.7	61.5	47.9	70.4	55.2	69.6	56.2
ConDSeg[10]	44.6M	91.0	84.0	94.4	89.4	71.5	62.0	84.7	75.6	84.5	73.6
PolypMoE[30]	>20M	92.7	87.8	93.8	89.2	**83.4**	**75.5**	90.8	84.4	-	-
MSBP-Net[18]	25.5M	91.9	86.8	94.0	89.2	81.0	73.1	89.7	82.8	-	-
SAM-Adapter[4]	6.4M	84.7	76.3	77.4	67.3	67.1	56.8	81.5	72.5	83.2	73.2
AutoSAM[20]	20.7M	82.2	74.0	80.1	71.0	59.7	51.3	84.1	76.4	73.2	62.3
ASPS[11]	52.8M	92.0	87.1	94.4	89.5	78.8	70.8	90.6	83.9	81.0	70.6
Efficient-SAM[25]	>9M	91.4	84.9	94.2	89.1	78.2	68.0	90.1	81.9	-	-
PGP-SAM[26]	15.3M	90.7	85.0	84.1	77.4	74.5	66.2	85.3	77.7	-	-
Moe-SAM[12]	62.5M	65.6	54.7	77.0	68.4	49.5	38.9	65.7	54.6	50.8	39.5
SAMUS[13]	39.7M	75.5	69.3	90.4	83.3	54.6	48.5	70.4	63.3	37.6	30.8
SurgicalSAM[28]	4.7M	75.3	65.7	72.8	62.8	54.5	44.6	79.4	70.3	55.3	42.9
UN-SAM[5]	132.5M	84.8	77.0	81.5	72.4	61.2	51.3	76.2	66.3	70.4	59.8
Our	**3.4M|3.2M**	**94.5**	**89.8**	**94.5**	**89.9**	78.1	70.7	**91.3**	**84.6**	**84.9**	**78.2**

We implemented our method in PyTorch and conducted experiments on a single NVIDIA RTX 5090 GPU using the ViT-B version of SAM. Both DINOv3 and SAM branches utilize a 1024×1024 input resolution. The model was trained for 50 epochs with a batch size of 1, using the AdamW optimizer with a 1×10^{-4} learning rate. To ensure reproducibility, we maintained a consistent random seed across both stages, ensuring identical data splits.

4.2 Results and Analysis

Comparison with Other Methods As demonstrated in Table 1, we compared our method against SOTA polyp segmentation models and SAM-based frameworks. Our approach achieved significant performance gains, reaching Dice scores of 94.5%, 94.5%, and 91.3% on the Kvasir, ClinicDB, and CVC-300 datasets, respectively, surpassing traditional methods such as Polyp-PVT, ASPS, and PolypMoE. Notably, on our newly constructed CPD1K dataset, we achieved a peak Dice score of 84.9%, validating the effectiveness of the DINOv3-guided strategy. Crucially, SD-SAM is extremely parameter-efficient, with trainable parameters ranging from only 3.2M ($W = 8$) to 3.4M ($W = 6$). This is substantially lower than heavy frameworks like UN-SAM (132.5M) and Moe-SAM (62.5M). More importantly, our model is less than half the size of existing lightweight alternatives, such as SAM-Adapter (6.4M),Efficient-SAM (9M) and Surgical-SAM (4.7M). This demonstrates that our "hierarchical low-rank adaptation and Gaussian prior modeling" strategy achieves an extraordinary balance between performance and parameter efficiency.

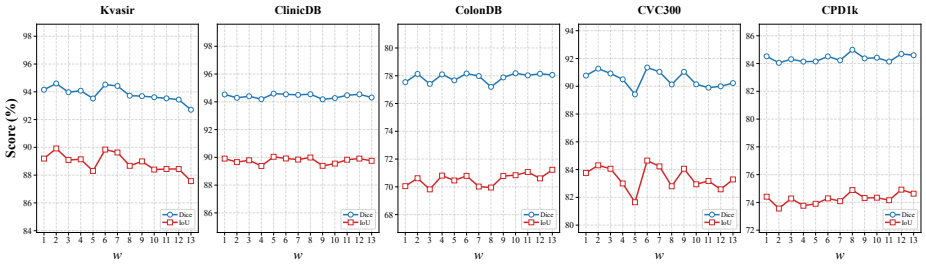

Fig. 4. Selection of Hyperparameter W across Five Datasets

Table 2. Ablation study on components (SAM, SLD, and DLF)

SAM	SLD	DLF	Kvasir		ClinicDB		ColonDB		CVC-300		CPD1K	
			Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
✓			62.6	52.0	47.7	38.5	48.5	40.3	60.4	52.6	31.5	23.5
✓	✓		90.3	83.9	93.0	87.8	75.0	67.0	85.5	78.2	83.2	72.5
✓	✓	✓	**94.5**	**89.8**	**94.5**	**89.9**	**78.1**	**70.7**	**91.3**	**84.6**	**84.9**	**78.2**

Hyperparameter Selection for W and Ablation Study As shown in Fig 4, we evaluated the layer-split hyperparameter $W \in [1, 13]$ to determine the optimal boundary between shallow distillation (SLD) and deep fusion (DLF), where the extremes $W = 1$ and $W = 13$ denote applying exclusively DLF or SLD across all layers, respectively. The model achieved its highest Dice scores at $W = 6$ for the four public datasets (89.6% average) and at $W = 8$ for our CPD1K dataset (84.9%).

To validate our SLD and DLF strategies (Table 2), we conducted ablation studies while retaining the pre-trained Stage 1 prompt generator. For the layer division threshold, we explicitly set $W = 6$ for public datasets and $W = 8$ for CPD1K. Under this configuration, sequentially incorporating SLD and DLF demonstrated a substantial, progressive leap over the Zero-shot SAM baseline, confirming both modules are essential for optimal segmentation.

5 Conclusion

We propose SD-SAM, a parameter-efficient, fully automatic polyp segmentation framework that addresses the issues of inaccurate automatic prompt generation, heavy auxiliary networks, and high fine-tuning costs. By utilizing Gaussian heatmap regression for prompt generation and a DINOv3-guided strategy of 'shallow distillation and deep fusion,' SD-SAM significantly enhances SAM's discriminative power with less than 3.5M trainable parameters. Furthermore, we open-source a dataset, CPD1K, featuring a highly dispersed spatial distribution and a large proportion of diminutive polyps to mitigate center bias and large-object dominance. Extensive experiments confirm that SD-SAM achieves state-of-the-art performance with minimal computational overhead.

Acknowledgments. This work was supported by the project of the Design and Development of Optical Detection Scheme (2021370102131997), project ZR2024MF025 supported by Shandong Provincial Natural Science Foundation.

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. Bernal, J., Sánchez, J., Vilarino, F.: Towards automatic polyp detection with a polyp appearance model. Pattern Recognition **45**(9), 3166–3182 (2012)
3. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., et al.: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca Cancer J Clin **68**(6), 394–424 (2018)
4. Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., Li, Z., Sun, L., Zang, Y., Mao, P.: Sam-adapter: Adapting segment anything in underperformed scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3367–3375 (2023)
5. Chen, Z., Xu, Q., Liu, X., Yuan, Y.: Un-sam: Universal prompt-free segmentation for generalized nuclei images. arXiv preprint arXiv:2402.16663 (2024)
6. 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)
7. 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)
8. 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)
9. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
10. Lei, M., Wu, H., Lv, X., Wang, X.: Condseg: A general medical image segmentation framework via contrast-driven feature enhancement. In: Proceedings of the AAAI conference on artificial intelligence. vol. 39, pp. 4571–4579 (2025)
11. Li, H., Zhang, D., Yao, J., Han, L., Li, Z., Han, J.: Asps: Augmented segment anything model for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 118–128. Springer (2024)
12. Li, R., Wu, L., Gu, J., Xu, Q., Chen, W., Cai, X., Bu, J.: Moe-sam: Enhancing sam for medical image segmentation with mixture-of-experts. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 367–377. Springer (2025)
13. Lin, X., Xiang, Y., Yu, L., Yan, Z.: Beyond adapting sam: Towards end-to-end ultrasound image segmentation via auto prompting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 24–34. Springer (2024)

14. Liu, M., Wang, Y., Wang, L., Hu, S., Wang, X., Ge, Q.: Imff-net: An integrated multi-scale feature fusion network for accurate retinal vessel segmentation from fundus images. Biomedical Signal Processing and Control **91**, 105980 (2024)
15. Lou, A., Guan, S., Loew, M.: Caranet: context axial reverse attention network for segmentation of small medical objects. Journal of Medical Imaging **10**(1), 014005–014005 (2023)
16. Luo, C., Wang, Y., Deng, Z., Lou, Q., Zhao, Z., Ge, Y., Hu, S.: Colonic polyp segmentation based on transformer-convolutional neural networks fusion. Pattern Recognition p. 112116 (2025)
17. Mansoori, M., Shahabodini, S., Abouei, J., Plataniotis, K.N., Mohammadi, A.: Self-prompting polyp segmentation in colonoscopy using hybrid yolo-sam2 model. In: ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2025)
18. Pan, X.L., Ding, J.R., Li, X., Liu, S., Wang, J., Hua, B., Tang, G.Z., Zhong, C.H.: Msbp-net: a multi-scale boundary prediction network for automated polyp segmentation. Pattern Recognition **170**, 112101 (2026)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
20. Shaharabany, T., Dahan, A., Giryes, R., Wolf, L.: Autosam: Adapting sam to medical images by overloading the prompt encoder. arXiv preprint arXiv:2306.06370 (2023)
21. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
22. 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)
23. Wu, S., Zhang, J., Zeng, P., Gao, L., Song, J., Shen, H.T.: Skip tuning: Pre-trained vision-language models are effective and efficient adapters themselves. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 14723–14732 (2025)
24. Xie, B., Tang, H., Cai, D., Yan, Y., Agam, G.: Self-prompt sam: Medical image segmentation via automatic prompt sam adaptation. arXiv preprint arXiv:2502.00630 (2025)
25. Xiong, Y., Varadarajan, B., Wu, L., Xiang, X., Xiao, F., Zhu, C., Dai, X., Wang, D., Sun, F., Iandola, F., et al.: Efficientsam: Leveraged masked image pretraining for efficient segment anything. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16111–16121 (2024)
26. Yan, Z., Yin, Z., Lin, T., Zeng, X., Liang, K., Ma, Z.: Pgp-sam: Prototype-guided prompt learning for efficient few-shot medical image segmentation. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2025)
27. Yin, D., Zheng, Q., Chen, L., Hu, Y., Wang, Q.: Apg-sam: Automatic prompt generation for sam-based breast lesion segmentation with boundary-aware optimization. Expert Systems with Applications **276**, 127048 (2025)
28. Yue, W., Zhang, J., Hu, K., Xia, Y., Luo, J., Wang, Z.: Surgicalsam: Efficient class promptable surgical instrument segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 6890–6898 (2024)

29. Zhang, R., Lai, P., Wan, X., Fan, D.J., Gao, F., Wu, X.J., Li, G.: Lesion-aware dynamic kernel for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 99–109. Springer (2022)
30. Zhang, W., Ye, W., Yu, Z., Ju, J.: A colon polyp segmentation network via collaborative decision-making of mixture of experts. Expert Systems with Applications p. 131117 (2026)
31. Zhao, X., Zhang, L., Lu, H.: Automatic polyp segmentation via multi-scale subtraction network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 120–130. Springer (2021)