Learning Diverse and Realistic Polyp Data via Style-Fused Diffusion Models

Longfei Han[1,2], Ying Zhai[1,3], and Haisheng Li[1*]

[1] School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing, China lihsh@btbu.edu.cn
[2] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China
[3] Research and Development Department, Hefei WiseOmni Technology Company Ltd, Hefei, China

Abstract. Polyp segmentation is critical for early colorectal cancer screening, yet its performance is limited by the quantity and diversity of annotated data. Existing synthesis methods typically rely on masks from public datasets, inheriting structural biases and limiting morphological diversity, while often neglecting style realism, producing synthetic images with inconsistent color and texture. To address these issues, we propose *SFD-Polyp*, a style-fused conditional diffusion framework that conditions generation on ground-truth masks and injects appearance cues via a *Style Integration Module* (SIM), enhancing texture and style consistency while preserving mask-aligned semantics. We further introduce *Morphology-Preserving Mask Synthesis* (MPMS) to generate multiple anatomically plausible mask variants at inference, expanding morphological diversity. Experiments on five public polyp datasets demonstrate that training with the synthesized data consistently improves segmentation across multiple baseline models, highlighting the potential of diffusion-based data generation to enhance generalization in data-limited medical imaging scenarios. Our code is avaliable at: https://github.com/zhaiying0925/SFD-polyp.

Keywords: Data Generation · Style Fusion · Diffusion Models · Polyp Segmentation

1 Introduction

Colorectal cancer typically develops progressively from polyps, and accurate polyp segmentation in endoscopy is crucial for early screening and clinical decision-making. Although deep learning methods (e.g., PraNet[8], SANet[24], and Polyp-PVT[5]) have improved segmentation performance, their generalization remains limited due to scarce and narrowly covered annotated data. High-quality pixel-level labels require substantial clinical effort, making it difficult to obtain diverse samples.

To alleviate data scarcity, previous studies have explored data augmentation and synthesis methods. Early augmentation techniques generate samples

guided by model predictions [4], apply color inversion to expand appearance variations [24], or use meta-learning to refine uncertain supervision into more reliable labels [9]. In addition, GAN-based synthesis has also been widely studied to expand training data, e.g., CycleGAN-style translation [17], [22] proposes a single-image-based approach for generating synthetic training data, enabling the creation of diverse and high-fidelity medical images from a single annotated example, and [14] proposes a GAN framework based on a cyclic structure and illumination constraints, which enhances the brightness and contrast of medical images while preserving fine details, thereby generating high-quality augmented training data. Despite progress, these methods often struggle to simultaneously achieve high diversity and high realism, which limits their effectiveness for improving segmentation generalization [18].

Diffusion models [10] have recently emerged as a promising paradigm for high-fidelity medical image synthesis and have been explored for endoscopic data generation [6, 7]. However, two gaps remain for data-limited polyp segmentation. First, most conditional pipelines rely on masks sampled from existing datasets as generation priors. When the available masks cover limited shapes and scales, the synthesized data largely inherits the same structural distribution, leading to insufficient morphological diversity for robust generalization[15]. Second, style realism is often under-modeled: synthetic samples may exhibit inconsistent color and texture statistics across endoscopic domains, which reduces clinical fidelity and weakens downstream training benefits[16].

To address these issues, we propose *SFD-Polyp*, a style-fused conditional diffusion framework that models both morphological diversity and style realism of polyp images. As shown in Fig. 1, SFD-Polyp conditions generation on ground-truth (GT) masks and employs a *Style Integration Module (SIM)* to inject appearance cues from real endoscopic images during diffusion training, enhancing texture and color consistency. To further improve morphological diversity without introducing anatomically implausible lesion shapes, we introduce *Morphology-Preserving Mask Synthesis (MPMS)*, which generates conservative mask variants through controlled local contour and sub-region perturbations. These variants serve as conditional inputs to the trained diffusion model for synthesizing diverse yet plausible image-mask pairs. In summary, our main contributions are as follows:

- We propose **SFD-Polyp**, a style-fused conditional diffusion framework. Unlike conventional methods that rely solely on mask-based conditions, SFD-Polyp incorporates training-only real endoscopic appearance cues as style conditions during diffusion modeling, effectively reducing color and texture discrepancies and preserving mask-guided lesion structures;
- We propose **MPMS**, a anatomy-constrained mask-condition augmentation strategy that derives morphology-preserving variants from existing annotations without extra manual labeling. By combining superpixel-based connected component expansion with IoU constraints, MPMS introduces controlled local contour and sub-region variations, thereby enhancing the conditional inputs for polyp synthesis.

– We validate the synthesized data on five public benchmarks and multiple segmentation backbones, demonstrating consistent improvements in segmentation performance and generalization.

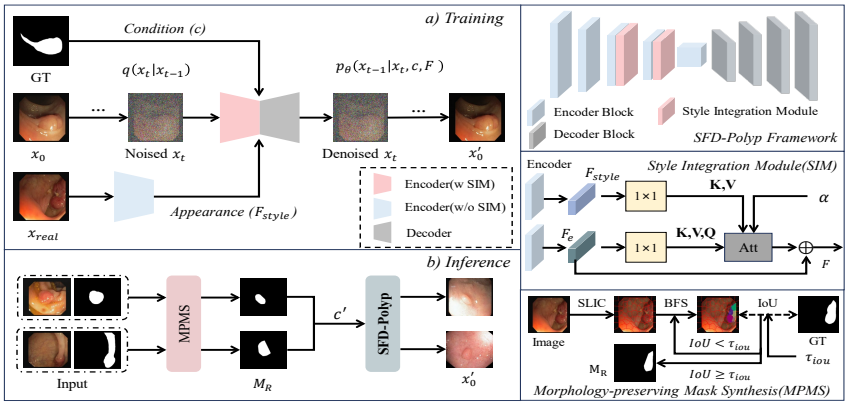

Fig. 1: SFD-Polyp architecture: (a) Training: SIM injects real image style into the denoising pathway, enhancing style realism while preserving mask-aligned semantics; (b) Inference: MPMS generates plausible mask variants as conditions for synthesizing structurally diverse and realistic polyp images.

2 Method

Overview. Our goal is to synthesize realistic and structurally diverse polyp images to augment training data for polyp segmentation. We learn a mask-conditioned diffusion generator that models the conditional distribution $p_\theta(x_0 \mid c)$, where c is a binary polyp mask. Given c, the generator produces a polyp image whose lesion geometry is aligned with the provided mask.

Diffusion training. In the forward diffusion process, a clean image x_0 is gradually perturbed by Gaussian noise over timesteps $t = 1, \ldots, T$:

$$q(x_t \mid x_{t-1}) = \mathcal{N}\big(x_t;\ \sqrt{1-\beta_t}\, x_{t-1},\ \beta_t \mathbf{I}\big). \tag{1}$$

where β_t denotes a variance schedule. This Markov process admits a closed form:

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \quad \bar{\alpha}_t = \prod_{i=1}^{t} (1-\beta_i). \tag{2}$$

In the reverse process, we learn a U-Net denoiser conditioned on the mask c and the style-fused feature F, F preserves the structure while injecting style, serving for noise prediction and decoding. Its probability distribution is given by:

$$p_\theta(x_{t-1} \mid x_t, c, F) = \mathcal{N}\big(x_{t-1};\, \mu_\theta(x_t, t, c, F),\, \sigma_t^2 \mathbf{I}\big). \tag{3}$$

Following the standard diffusion training paradigm, according to the forward diffusion process (Eq. (2)), a clean sample x_0 can be expressed in terms of the noisy observation x_t. By applying Bayes' theorem, it can be derived that the predicted mean is a linear function of the added noise:

$$\mu_\theta(x_t, t, c, F) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t, c, F) \right). \tag{4}$$

Therefore, the denoising learning process can be reformulated as a noise prediction task, where the predicted noise $\epsilon_\theta(x_t, t, c, F)$ by the denoising network is optimized:

$$\mathcal{L}_{\text{diff}}(\theta) = \mathbb{E}_{t, x_0, \epsilon} \left[\|\epsilon - \epsilon_\theta(x_t, t, c, F)\|_2^2 \right]. \tag{5}$$

Diffusion inference. During the synthesis stage, we start from $x_T \sim \mathcal{N}(0, \mathbf{I})$ and iteratively sample x_{t-1} for $t = T, \ldots, 1$, ultimately obtaining the synthesized image x_0'. The sampling formula is given as follows:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\, \epsilon_\theta(x_t, t, c', F) \right) + \sigma_t z, \quad z \sim \mathcal{N}(0, \mathbf{I}). \tag{6}$$

MPMS is applied prior to sampling to generate diverse mask variants c, while SIM is activated within the denoising network. Here, we construct a style bank from all training images and randomly assign training style F to mask variants c'(obtained from Eq.(11)) to generate images at inference stage. The style features provide global color and texture cues, while lesion geometry remains constrained by the mask throughout denoising. Overall, during training, images and masks are fed into the diffusion network, with SIM applied in the deeper layers of the U-Net to fuse real image styles and enhance visual realism; during inference, MPMS generates diverse masks as conditions for producing morphologically rich synthetic images.

2.1 Style Integration Module

In mask-based endoscopic image generation, a major challenge is to simultaneously preserve the mask-aligned structural semantics and enhance visual realism, while purely mask-driven generation may lead to cross-domain visual inconsistencies. To address this, we propose the **Style Integration Module (SIM)**, which injects real image style into the generated features via a weighted mechanism, enabling color and texture transfer while maintaining spatial structure.

Specifically, given a real image $x_{real} \in \mathbb{R}^{C \times H \times W}$, we extract its features using the encoder of the original denoising network, denoted as $F_{style} \in \mathbb{R}^{C \times H_j \times W_j}$

$(j = 1, \ldots, 4)$, which serve as the style condition of each encoder in the reverse diffusion process. Let $F_e \in \mathbb{R}^{C \times H_j \times W_j}$ denote the features of the generated image at the i-th downsampling stage of the UNet encoder. The SIM module fuses style information via an attention mechanism as follows:

$$\begin{aligned} Q &= \mathcal{W}_Q(F_e), \\ K &= \alpha\, \mathcal{W}_K(F_{style}) + (1-\alpha)\, \mathcal{W}_K(F_e), \\ V &= \alpha\, \mathcal{W}_V(F_{style}) + (1-\alpha)\, \mathcal{W}_V(F_e). \end{aligned} \tag{7}$$

where $\mathcal{W}_Q$, $\mathcal{W}_K$, and $\mathcal{W}_V$ are 1×1 convolutional projections, and α denotes the style fusion ratio. The final output feature is obtained via a residual connection:

$$F = F_e + \text{Proj}\left(\text{softmax}\left(\frac{QK^\top}{\sqrt{C}}\right)V\right). \tag{8}$$

where Proj denotes a 1×1 convolution that maps the attention output back to the original feature dimension. Keeping the Query unchanged preserves semantic and spatial consistency, while fusing Key and Value only modifies feature content, enabling structure-style decoupling. The resulting F combines the original structure with real style and is used for subsequent decoding. Using Eq. (4), the predicted noise $\epsilon_\theta(x_t, t, c, F)$ can be obtained to compute the predicted loss. Here, the SIM injects style only at the third and fourth downsampling stages to avoid structural disruption, making the generated images closer to the real color and texture.

2.2 Morphology-preserving Mask Synthesis

During inference, to increase mask diversity while preserving anatomical plausibility, we propose **MPMS**, which generates diverse masks as conditional inputs and, combined with the SFD-Polyp model $\epsilon_\theta(x_t, t, c, F)$, produces morphologically rich synthetic images. Given an image $x_0 \in \mathbb{R}^{H \times W \times 3}$ and its binary mask $M \in \{0,1\}^{H \times W}$, MPMS performs superpixel-guided connected expansion with IoU control, introducing boundary variations without disrupting the topology.

First, we apply SLIC algorithm[1] to obtain a superpixel label map $S \in \{0, 1, \ldots, k-1\}^{H \times W}$, which partitions the image into k compact regions. For each superpixel region:

$$\Omega_k = \{(m, n) \mid S(m, n) = k\}. \tag{9}$$

where m, n denotes the pixel in x_0, then we compute its lesion occupancy ratio as:

$$r_k = \frac{\sum_{(m,n)\in\Omega_k} M(m, n)}{|\Omega_k|}. \tag{10}$$

Here, $|\Omega_k|$ denotes the total number of pixels in the region, and superpixels with coverage $r_k \geq 0.5$ are considered high-confidence lesion regions. During mask generation, MPMS performs connected expansion based on breadth-first

search (BFS): a seed superpixel is randomly selected and progressively expanded, adding neighboring superpixels that meet the coverage threshold and have not yet been selected. The resulting expanded region set is denoted as $\mathcal{R}$, and the synthesized mask is given by:

$$M_{\mathcal{R}}(m,n) = \begin{cases} 1, & S(m,n) \in \mathcal{R} \\ 0, & \text{otherwise} \end{cases}. \tag{11}$$

where S denotes the superpixel label map. To avoid deviating from the original annotation, we monitor the IoU between the expanded mask $M_{\mathcal{R}}$ and the original mask M, stopping the expansion when IoU $\geq \tau_{\text{iou}}$. The hyperparameter τ_{iou} balances diversity and fidelity. A convex hull is then applied to smooth boundaries, followed by intersection with the original mask, producing a boundary-consistent and anatomically plausible mask c'. Using Eq. (6), reverse diffusion is iterated to $t = 1$, yielding the final image x'_0, which conforms to the generated mask morphology while preserving the real image style, achieving both morphological and visual diversity.

3 Experiments

Dataset and experimental settings. The experiments were conducted on five publicly available polyp datasets: Kvasir-SEG[11], CVC-ClinicDB[2], CVC-ColonDB[21], CVC-300[23], and ETIS-LaribPolypDB[19]. The dataset split follows PraNet[8], with 900 Kvasir-SEG images and 550 CVC-ClinicDB images used for training, and the remaining images used for testing. All images were resized to a uniform resolution of 256×256. Based on MPMS masks, we use SFD-Polyp to synthesize additional training samples and, following the results in [16], train the segmentation model with a 1:1 mix of synthetic and real data. Using the same baseline optimization settings, this strategy significantly improves robustness to variations in appearance and morphology. Training was performed with a batch size of 16 for 100,000 iterations, a learning rate of 1×10^{-4}, 250 diffusion steps, and a style fusion weight $\alpha = 0.5$. Our work focuses on task-oriented data synthesis for polyp segmentation, aiming to assess whether generated data can improve downstream performance under limited annotations. Therefore, we adopt Dice and IoU as evaluation metrics. Experiments were conducted on a workstation with an NVIDIA V100 GPU (32GB memory), taking approximately one day.

Results and Analysis. We use SFD-Polyp–generated data to train Polyp-PVT and compare it with state-of-the-art methods. As shown in Table 1, our method achieves Dice scores of 0.904 and 0.817 on CVC-300 and CVC-ColonDB, respectively, outperforming the baseline and notably improving segmentation of small polyps. On the small-sample CVC-ColonDB and low-contrast ETIS datasets, our approach also surpasses most existing methods, demonstrating that the synthetic data enhances adaptability to diverse morphological and visual variations. Fig. 2 shows representative results, with polyps of varying scales and

Table 1: Quantitative comparison (Dice/IoU) on five polyp segmentation benchmarks. Best results are in bold and second-best are underlined.

Methods	Pub.'Year	CVC-300		CVC-ClinicDB		Kvasir-SEG		CVC-ColonDB		ETIS	
		Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
ACSNet[27]	MICCAI'2020	0.863	0.787	0.882	0.826	0.898	0.838	0.716	0.649	0.578	0.509
PraNet[8]	MICCAI'2020	0.871	0.797	0.899	0.849	0.898	0.840	0.712	0.640	0.628	0.567
MSNet[28]	MICCAI'2021	0.869	0.807	0.921	0.879	0.907	0.862	0.755	0.678	0.719	0.664
SANet[24]	MICCAI'2021	0.888	0.815	0.916	0.859	0.904	0.847	0.753	0.670	0.750	0.654
BLENet[20]	ACM'2022	0.879	0.805	0.926	0.878	0.905	0.854	0.731	0.658	0.673	0.594
DCRNet[25]	ISBI'2022	0.856	0.788	0.896	0.844	0.886	0.825	0.704	0.631	0.556	0.496
CaraNet[13]	SPIE'2023	0.903	**0.838**	0.936	0.887	0.918	**0.865**	0.773	0.689	0.747	0.672
Polyp-PVT[5]	CAAI'2023	0.900	0.833	0.937	0.889	0.917	0.864	0.808	0.727	0.787	0.706
MEGANet[3]	WACV'2024	0.899	0.834	**0.938**	**0.894**	0.913	0.863	0.793	0.713	0.739	0.665
EFANet[30]	MIR'2025	0.894	0.830	0.919	0.871	0.914	0.861	0.774	0.696	0.749	0.670
SAFE-Net[26]	BSPC'2025	0.895	0.828	0.937	0.889	0.917	0.865	0.781	0.704	**0.794**	**0.713**
WeakPolyp-SAM[29]	KBS'2025	0.886	0.813	0.882	0.805	0.869	0.797	0.700	0.621	0.691	0.603
Ours	-	**0.904**	0.836	0.937	0.889	**0.918**	0.864	**0.817**	**0.740**	0.772	0.691

Table 2: Comparison of data synthesis strategies for training segmentation models on five benchmark datasets using two backbone networks (Dice/IoU).

Methods	CVC-300		CVC-ClinicDB		Kvasir-SEG		CVC-ColonDB		ETIS	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
SANet[24]	0.888	0.815	0.916	0.859	0.904	0.847	0.753	0.670	0.750	0.654
+Copy-Paste[12]	0.897	0.830	0.902	0.851	0.903	0.848	0.777	0.700	0.774	0.688
+ArSDM[7]	0.902	0.832	0.914	0.861	0.911	0.856	0.777	0.700	**0.780**	0.695
+PolypDDPM[6]	0.909	0.841	0.921	0.866	**0.912**	**0.859**	0.764	0.686	0.777	0.694
+Ours	**0.910**	**0.845**	**0.923**	**0.869**	0.902	0.844	**0.799**	**0.721**	0.777	**0.697**
Polyp-PVT[5]	0.900	0.833	0.937	0.889	0.917	0.864	0.808	0.727	0.787	0.706
+Copy-Paste[12]	0.880	0.809	0.934	0.887	0.917	**0.871**	0.798	0.718	0.792	0.713
+ArSDM[7]	0.882	0.812	0.922	0.875	0.915	0.863	0.817	0.738	**0.806**	**0.729**
+PolypDDPM[6]	0.896	0.826	0.913	0.862	0.918	0.868	0.800	0.717	0.787	0.708
+Ours	**0.904**	**0.836**	**0.937**	**0.889**	**0.918**	0.864	**0.817**	**0.740**	0.772	0.691

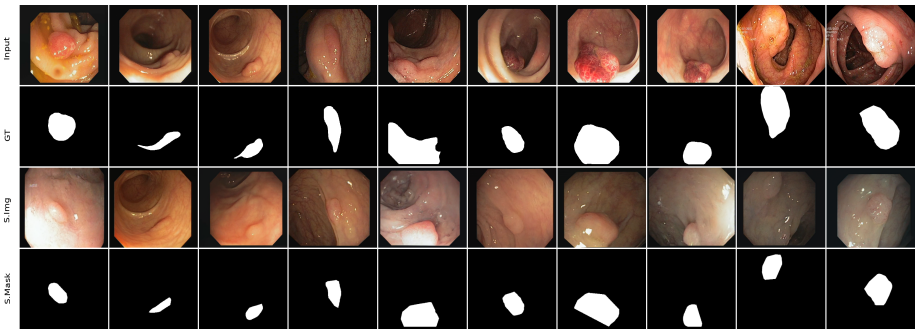

Fig. 2: Visualization of the method: GT is the ground-truth mask, Input is the original image, S.Img is the synthesized image, and S.Mask is the generated mask.

boundaries while maintaining coherent textures and natural color transitions, validating the effectiveness of SIM in improving realism and usability of the synthesized samples.

We compare SFD-Polyp with representative image synthesis methods (Table 2) and observe improvements across both backbones and multiple datasets.

Table 3: Ablation study of MPMS and SIM using SANet as the segmentation backbone on five benchmarks (Dice/IoU).

Methods			CVC-300		CVC-ClinicDB		Kvasir-SEG		CVC-ColonDB		ETIS	
Base	MPMS	SIM	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
✓			0.888	0.815	0.916	0.859	0.904	0.847	0.753	0.670	0.750	0.654
✓	✓		0.880	0.811	0.917	0.867	**0.915**	**0.860**	0.776	0.695	0.756	0.673
✓		✓	0.895	0.822	0.918	0.867	0.905	0.850	0.760	0.686	0.765	0.685
✓	✓	✓	**0.910**	**0.845**	**0.923**	**0.869**	0.902	0.844	**0.799**	**0.721**	**0.777**	**0.697**

Table 4: Performance comparison under different IoU thresholds τ_{iou} and style fusion weights α based on the SANet backbone.

IoU Threshold τ_{iou}			Style Fusion Weight α		
Value	mDice	mIoU	Value	mDice	mIoU
0	0.850	0.784	0	0.837	0.769
0.9	0.852	0.787	1.0	0.845	0.780
0.5	**0.862**	**0.795**	0.5	**0.862**	**0.795**

With SANet, Dice on CVC-300 and CVC-ColonDB increases by 2.5% and 6.1%, showing that MPMS mask variants enhance handling of irregular boundaries and scale variations. With Polyp-PVT, IoU on cross-domain CVC-ColonDB improves by 1.8%, indicating better generalization. These gains stem from MPMS providing morphological diversity and SIM preserving style realism, enabling generated data to balance structural plausibility and visual consistency, supporting more robust segmentation training.

Module Ablation Study. We conduct ablation studies to assess the contributions of MPMS and SIM using SANet as the segmentation backbone (Table 3). Each component improves performance over the baseline, and combining both yields the best overall results. MPMS provides stronger gains on unseen datasets such as ETIS, indicating that expanding mask structures beyond dataset priors benefits generalization. SIM brings notable improvements on datasets with larger appearance variation (e.g., CVC-ClinicDB and CVC-ColonDB), suggesting that style-consistent synthesis improves the usefulness of generated samples for downstream learning. The combination of MPMS and SIM achieves the most stable improvements across five benchmarks, demonstrating their complementarity in modeling structural diversity and appearance realism.

Parameter Sensitivity Analysis. Table 4 shows the impact of key parameters on segmentation performance. A low IoU threshold generates many small masks, biasing the model toward minor lesions, while a high threshold slightly improves accuracy but limits mask diversity. An intermediate value, $\tau_{\text{iou}} = 0.5$, achieves the best balance between lesion contour preservation and mask diversity. Regarding the style fusion weight, $\alpha = 0$ uses only real images, limiting the utilization of synthesized features, whereas $\alpha = 1$ relies entirely on generated

styles, increasing diversity but deviating from real styles and reducing accuracy. $\alpha = 0.5$ achieves the highest mDice and mIoU, indicating that moderate style fusion improves the model's adaptability to variations in polyp appearance.

4 Conclusion

This work proposes the style-fused diffusion framework SFD-Polyp, which enhances the style realism of generated images via SIM and expands mask morphological diversity through MPMS, enriching the conditional space for data generation. Experiments show that the synthesized data consistently improve the cross-dataset generalization of various segmentation models. Future work will explore hyperparameter sensitivity, additional diffusion-based baselines, and complementary image-level evaluations to further assess the diversity and realism of synthesized data.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (No. U24A20341, 62277001) and Anhui Provincial Key R&D Programmes (2023s07020001).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence **34**(11), 2274–2282 (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. 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)
4. Chaitanya, K., Karani, N., Baumgartner, C.F., Erdil, E., Becker, A., Donati, O., Konukoglu, E.: Semi-supervised task-driven data augmentation for medical image segmentation. Medical Image Analysis **68**, 101934 (2021)
5. 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** (2023)
6. Dorjsembe, Z., Pao, H.K., Xiao, F.: Polyp-ddpm: Diffusion-based semantic polyp synthesis for enhanced segmentation. In: 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 1–7. IEEE (2024)
7. Du, Y., Jiang, Y., Tan, S., Wu, X., Dou, Q., Li, Z., Li, G., Wan, X.: Arsdm: colonoscopy images synthesis with adaptive refinement semantic diffusion models. In: International conference on medical image computing and computer-assisted intervention. pp. 339–349. Springer (2023)

8. 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)
9. Guo, X., Chen, Z., Liu, J., Yuan, Y.: Non-equivalent images and pixels: Confidence-aware resampling with meta-learning mixup for polyp segmentation. Medical image analysis **78**, 102394 (2022)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
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. Lin, D., Cao, Y., Zhu, W., Li, Y.: Few-shot defect segmentation leveraging abundant defect-free training samples through normal background regularization and crop-and-paste operation. In: 2021 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE Computer Society (2021)
13. 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)
14. Ma, Y., Liu, Y., Cheng, J., Zheng, Y., Ghahremani, M., Chen, H., Liu, J., Zhao, Y.: Cycle structure and illumination constrained gan for medical image enhancement. In: International conference on medical image computing and computer-assisted intervention. pp. 667–677. Springer (2020)
15. Qiu, K., Gao, Z., Zhou, Z., Sun, M., Guo, Y.: Noise-consistent siamese-diffusion for medical image synthesis and segmentation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 15672–15681 (2025)
16. Qiu, R., Xia, K., Gao, F., Yang, S., Cai, D., Wang, J., Chen, Y., Wang, L.: Accurate boundary alignment and realism enhancement for colonoscopic polyp image-mask pair generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 34–44. Springer (2025)
17. Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. Scientific reports **9**(1), 16884 (2019)
18. Sharma, V., Kumar, A., Jha, D., Bhuyan, M.K., Das, P.K., Bagci, U.: Controlpolypnet: Towards controlled colon polyp synthesis for improved polyp segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2325–2334 (2024)
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. Ta, N., Chen, H., Lyu, Y., Wu, T.: Ble-net: boundary learning and enhancement network for polyp segmentation. Multimedia Systems **29**(5), 3041–3054 (2023)
21. 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)
22. Thambawita, V., Salehi, P., Sheshkal, S.A., Hicks, S.A., Hammer, H.L., Parasa, S., Lange, T.d., Halvorsen, P., Riegler, M.A.: Singan-seg: Synthetic training data generation for medical image segmentation. PloS one **17**(5), e0267976 (2022)

23. 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)
24. Wei, J., Hu, Y., Zhang, R., Li, Z., Zhou, S.K., Cui, S.: Shallow attention network for polyp segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 699–708. Springer (2021)
25. Yin, Z., Liang, K., Ma, Z., Guo, J.: Duplex contextual relation network for polyp segmentation. In: 2022 IEEE 19th international symposium on biomedical imaging (ISBI). pp. 1–5. IEEE (2022)
26. Yu, J., Qi, L.: Safe-net: Shape-aware and feature enhancement network for polyp segmentation. Biomedical Signal Processing and Control **99**, 106906 (2025)
27. Zhang, R., Li, G., Li, Z., Cui, S., Qian, D., Yu, Y.: Adaptive context selection for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23. pp. 253–262. Springer (2020)
28. Zhao, X., Zhang, L., Lu, H.: Automatic polyp segmentation via multi-scale subtraction network. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 120–130. Springer (2021)
29. Zhao, Y., Zhou, T., Gu, Y., Zhou, Y., Zhang, Y., Wu, Y., Fu, H.: Weakpolyp-sam: Segment anything model-driven weakly-supervised polyp segmentation. Knowledge-Based Systems p. 113701 (2025)
30. 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)