FPN-in-FPN: A Nested Multi-Scale Aggregation Network for Polyp Segmentation

Jin Ye[1], Yanzhou Su[2], Yicheng Wu[1], Junjun He[3,4], Bohan Zhuang[1], Zhaolin Chen[1], and Jianfei Cai[1†]

[1] Department of Data Science & AI, Faculty of Information Technology, Monash University, Melbourne, VIC 3168, Australia
[†]Jianfei.Cai@monash.edu
[2] Fuzhou University, Fujian, China
[3] Shanghai Artificial Intelligence Laboratory, Shanghai, China
[4] Shanghai Innovation Institute, Shanghai, China

Abstract. Colorectal cancer is a leading cause of cancer-related deaths worldwide, and precise polyp segmentation plays a crucial role in its early detection. U-shaped architectures are widely used for polyp segmentation due to their ability to capture multi-scale contextual information effectively. However, it is suboptimal to solely use top-down or bottom-up fusion flow in traditional U-shaped architectures. Additionally, most existing methods only focus on improving the feature fusion module, often introducing more computational costs. In this work, we propose a novel and efficient nested multi-scale feature aggregation network that integrates high-level semantic information with low-level boundary details within skip connections, effectively handling the diverse shapes and sizes of polyp regions. Specifically, we introduce a bidirectional FPN-in-FPN module that fuses features across stages through both bottom-up and top-down pathways. This module adds only 0.12M extra parameters with minimal computational overhead while significantly enhancing segmentation performance in small networks. Extensive experiments on polyp segmentation datasets demonstrate that our network outperforms existing methods in both accuracy and efficiency. Code is available at https://github.com/Yejin0111/FPN-in-FPN

Keywords: Feature Aggregation · FPN · Polyp Segmentation.

1 Introduction

Colorectal cancer is a leading cause of worldwide cancer-related deaths, and early detection is crucial for improving prognosis [9]. Intestinal polyps, often detected during colonoscopy, serve as early indicators of the disease. Accurate and timely segmentation of these polyps is essential for aiding diagnosis, reducing missed detections, and enhancing overall diagnostic efficiency. However, polyp segmentation is challenging due to the variability in shape and size, and the subtle boundaries that often blend with surrounding tissues.

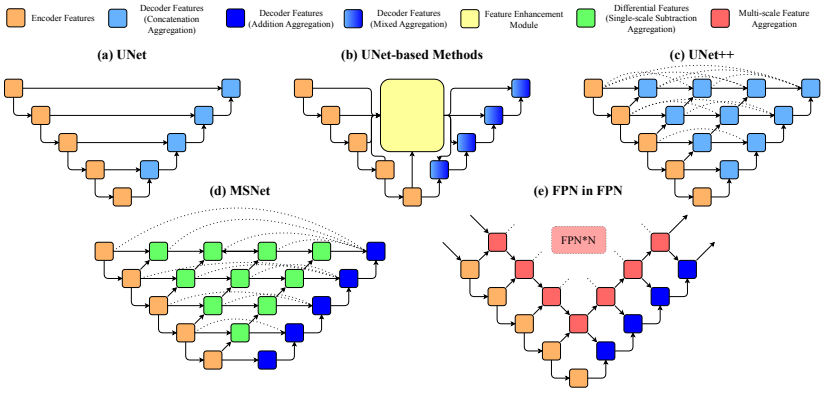

Fig. 1: Illustration of different feature aggregation methods in U-Shaped architecture. (a) Traditional UNet, (b) UNet variants, and (c) UNet++, which only focus on improving the feature fusion module, often making it more complex and computationally expensive. (d) MSNet. Previous approaches are suboptimal in feature fusion flow, as they consider only a single pathway, such as top-down or bottom-up. (e) FPN-in-FPN, which introduces an effective and efficient bidirectional feature fusion approach with minimal computational overhead.

Existing polyp segmentation methods are typically based on U-Net [10], a widely used encoder-decoder architecture [5,6] with Feature Pyramid Networks (FPN) [7] technique. The key challenge of the U-shaped structure is efficiently and effectively integrating multi-scale semantic information across stages to accurately identify diverse polyps. By now, several studies have explored multi-scale feature fusion from different perspectives. As shown in Fig 1(a), conventional U-Net fuses up-sampled feature maps from the decoder with feature maps skipped from the encoder using concatenation or addition operations, leading to two critical issues [18]: First, element-wise addition or concatenation may introduce redundancy or weaken level-specific features. Second, limiting the fusion to only two-level feature maps restricts the richness of feature representations. U-Net variants, as shown in Fig 1(b), address the first challenge with advanced feature fusion modules, such as attention mechanisms [2,5] and gate mechanisms [16,19], but at the cost of increased computational overhead. Meanwhile, U-Net++ [21,22] mitigates the second issue by introducing nested and dense skip connections, improving the feature propagation across different levels, as shown in Fig 1(c). Recently, an advanced multi-scale feature fusion method, MSNet [20] as shown in Fig 1(d), obtains rich multi-scale difference information through a multi-scale subtraction module, which extracts the difference features between adjacent encoder stages. Its improved version, M^2SNet [18] extends this module into an intra-layer multi-scale subtraction module, providing the decoder with both pixel-level and structure-level difference information. The essential draw-

back of the MSNet series is its subtraction module, which can unintentionally discard important information when both feature maps have similar details.

Therefore, in this paper, we propose FPN-in-FPN: A Nested Multi-Scale Aggregation Network, a novel architecture designed to enhance multi-scale feature fusion. Our method leverages bidirectional feature fusion (top-down and bottom-up paths) to effectively integrate high-level semantic-rich information with low-level boundary-rich details. Unlike previous approaches that rely solely on top-down or bottom-up fusion, our method enables more comprehensive feature aggregation at each stage, effectively integrating both local and global information for all feature maps. Specifically, as shown in Fig 1(e), the FPN-in-FPN module relies on a bidirectional architecture that first combines high-resolution boundary-rich features with low-resolution semantic-rich features via a bottom-up pathway, followed by a top-down fusion of low-resolution semantic-rich features with high-resolution boundary-rich features. To control additional computational costs, we incorporate lightweight convolutional blocks for FPN-in-FPN module, introducing only 0.12M extra parameters with minimal computational overhead. Additionally, we adopt deep supervision in the decoding stages, providing direct feedback at multiple levels of the decoder to enhance the segmentation of small or poorly defined polyps. In summary, the contributions of this paper can be summarized as follows:

- We propose a novel and efficient nested multi-scale feature aggregation network for polyp segmentation, which integrates high-level semantic information with low-level boundary details within skip connections, effectively handling the diverse shapes and sizes of polyp regions.
- A bidirectional FPN-in-FPN module is proposed to boost multi-scale feature fusion across stages through bottom-up and top-down pathways. This module adds minimal computational cost while significantly improving segmentation performance in small networks.
- Experiments on five polyp segmentation datasets, detailed in Table 1, showcase our method's efficiency and effectiveness. Our approach delivers competitive performance with just 2.55G MACs and 4.01M parameters, making it notably faster than state-of-the-art models.

2 Methodology

In this section, we detail our proposed FPN-in-FPN: A Nested Multi-Scale Aggregation Network, which is designed for effective and efficient polyp segmentation. The overall architecture consists of three main components: **1) Encoder**, which extracts features from the input image, **2) FPN-in-FPN module**, our core contribution, which fuses feature information across stages via both bottom-up and top-down pathways, and **3) Decoder**, which maps the feature maps to an output mask. In the following sections, we present the details of our proposed FPN-in-FPN method.

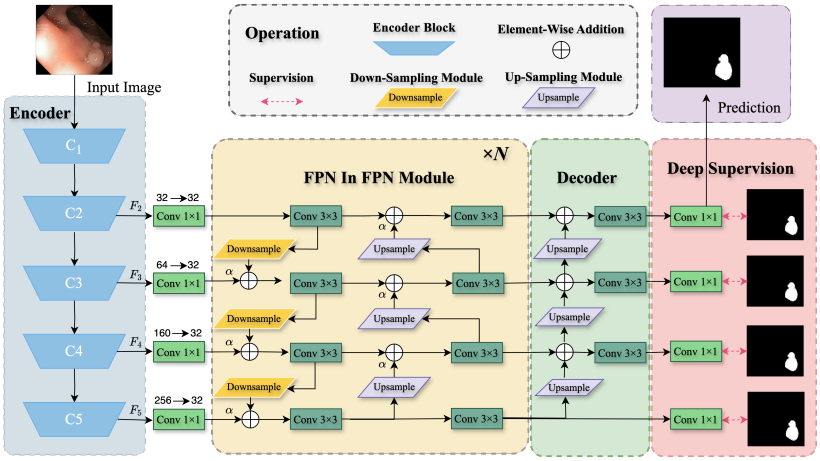

Fig. 2: Overview of U-Shaped architecture with FPN-in-FPN.

2.1 Overview of proposed framework

The framework overview is illustrated in Fig 2, given an input image $X \in \mathbb{R}^{3\times H\times W}$, a backbone CNN (*i.e.*, VAN [3]) serves as the encoder, extracting five feature maps $\{F_1, F_2, F_3, F_4, F_5\}$, where F_1 has the highest spatial resolution and is rich in boundary features, while F_5 has the lowest resolution and contains semantic-rich information. In our method, F_1 is not included in the feature fusion module. Due to the differing channel dimensions of these feature maps, we apply a series of 1×1 convolutional layers to align the channel dimensions of $\{F_2, F_3, F_4, F_5\}$. Formally,

$$\hat{F}_i = \mathrm{Conv}_{1\times1}(F_i), \quad i = 1, 2, 3, 4, \tag{1}$$

where each $\mathrm{Conv}_{1\times1}$ outputs a feature map of size (C, H_i, W_i) at each i stage, with $C = 32$ as the unified channel dimension used in our experiments. The channel dimension reduction is illustrated in Figure 2, e.g., "32->32". After fusing all encoder features with FPN-in-FPN module, the decoder further refines the multi-scale features through addition. Specifically, at each stage, the decoder takes two feature maps from different levels and produces an aggregated output:

$$\bar{F}_i = Conv_{3\times3}(F'_i + UpSample(F'_{i+1})), \tag{2}$$

where F'_i and $\bar{F}_i$ are the enhanced feature from the FPN-in-FPN module and the decoder feature used to generate the segmentation mask at stage i, respectively. An additional convolutional layer is applied for smoothing. Since polyps often have ambiguous boundaries, small sizes, or irregular shapes, we employ a multi-branch deep supervision strategy to enhance robust learning. During training, each branch is supervised by the groundtruth mask, providing gradients that help the network capture both high-level semantics and fine-grained boundary

details. The multi-scale feedback improves model generalization, particularly for challenging polyp cases.

2.2 FPN-in-FPN Module

After the channel translation, the aligned encoder features $\{\hat{F}_2, \hat{F}_3, \hat{F}_4, \hat{F}_5\}$ serve as inputs for the feature fusion. Instead of a traditional single top-down or bottom-up pass, our network employs multiple stacked FPN blocks, each performing both bottom-up and top-down aggregation. Specifically, FPN-in-FPN is a bidirectional architecture that first combines high-resolution boundary-rich features with low-resolution semantic-rich features via a bottom-up pathway, followed by a top-down fusion of low-resolution semantic-rich features with high-resolution boundary-rich features. To control additional computational costs, we incorporate lightweight convolutional blocks for FPN-in-FPN module, introducing only 0.12M extra parameters with minimal computational overhead.

Bottom-Up Path. In each FPN-in-FPN block, it first fuses high-resolution boundary-rich features into low-resolution semantic-rich features through downsampling, which can be formulated as:

$$F_3^{bu} = \alpha \times ConvBlock(F_3^{td}) + (1-\alpha) \times ConvBlock(DownSample(F_2^{td})), \quad (3)$$

$$F_4^{bu} = \alpha \times ConvBlock(F_4^{td}) + (1-\alpha) \times ConvBlock(DownSample(F_3^{td})), \quad (4)$$

$$F_5^{bu} = \alpha \times ConvBlock(F_5^{td}) + (1-\alpha) \times ConvBlock(DownSample(F_4^{td})) \quad (5)$$

where $ConvBlock(\cdot)$ is the **Conv-BatchNorm-ReLU** operation, we achieve $DownSample(\cdot)$ with bilinear interpolation to match the target spatial size, α is a learnable parameter. The input feature F_i^{td} for the FPN-in-FPN module is initially set to $\hat{F}_i$. After each fusion, a convolutional layer with 3×3 kernel refines the aggregated feature.

Top-Down Path. Following the bottom-up path, the fused bottom-up features are then passed through a top-down path that propagates low-resolution semantic-rich information back to high-resolution boundary-rich features:

$$F_4^{td} = \alpha \times ConvBlock(F_4^{bu}) + (1-\alpha) \times ConvBlock(UpSample(F_5^{bu})), \quad (6)$$

$$F_3^{td} = \alpha \times ConvBlock(F_3^{bu}) + (1-\alpha) \times ConvBlock(UpSample(F_4^{bu})), \quad (7)$$

$$F_2^{td} = \alpha \times ConvBlock(F_2^{bu}) + (1-\alpha) \times ConvBlock(UpSample(F_3^{bu})) \quad (8)$$

where $UpSample(\cdot)$ also denotes bilinear interpolation to match the target spatial resolution, followed by a refining convolution. After passing through n FPN-in-FPN blocks, the final fused feature F_i^{td} from the top-down path becomes F_i'.

Overall, this nested FPN design enables dual-direction, multi-scale feature aggregation that enhances high-level context while preserving fine-grained spatial details, making it particularly effective for accurate polyp segmentation across diverse lesion sizes and appearances.

3 Experiments

3.1 Datasets

We conducted experiments on five datasets, including CVC-T [13], Kvasir [4], CVC-ClinicDB [1], CVC-ColonDB [12], and ETIS [11]. Following the experimental setup of [2,6], we used a portion of images from CVC-ClinicDB and Kvasir for training, consisting of 550 and 900 images, respectively. Thus, the training set consisted of 1450 images, and the remaining data from these two datasets were used to test the model's learning ability. In addition, the remaining three datasets were used to evaluate the model's cross-dataset generalization performance, consisting of 60, 380, and 196 images, respectively.

3.2 Implementation Details

We implement the proposed framework in PyTorch and train it on an NVIDIA Tesla A100 GPU. Following a typical FPN design, we adopt VAN [3] (pretrained on ImageNet) as our encoder backbone, while the remaining layers are randomly initialized. The model is trained for 120 epochs with a batch size of 16, using the AdamW optimizer (weight decay $= 1 \times 10^{-4}$) and an initial learning rate of 1×10^{-4}. We further apply data augmentation techniques such as random flipping, random rotation, and multi-scale training. For evaluation, we use two widely adopted metrics, mean Dice (mDice) and mean Intersection over Union (mIoU), as primary quantitative measures. Additionally, Frames Per Second (FPS), Multiply ACcumulate operations (MACs) and parameter count are used to assess model efficiency.

Table 1: Comparison of SOTA approaches on five polyp segmentation datasets.

Method	Backbone	MACs	Params	FPS	CVC-ClinicDB		Kvasir		CVC-T		CVC-ColonDB		ETIS	
					mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
UNet[10]	-	-	-	-	0.823	0.755	0.818	0.746	0.710	0.627	0.512	0.444	0.398	0.335
PraNet[2]	Res2Net50	13.15	32.55	44.42	0.899	0.849	0.898	0.840	0.871	0.797	0.709	0.640	0.628	0.567
ResUNet++[5]	-	134.12	14.48	25.25	0.846	0.786	0.807	0.727	0.687	0.598	0.588	0.497	0.337	0.275
SANet[15]	Res2Net50	11.32	23.90	34.94	0.916	0.859	0.904	0.847	0.888	0.815	0.752	0.669	0.750	0.654
MSNet[20]	Res2Net50	17.03	29.74	44.34	0.921	0.879	0.907	0.862	0.869	0.807	0.755	0.678	0.719	0.664
M^2SNet[18]	Res2Net50	17.09	29.74	44.01	0.922	0.880	0.912	0.861	0.903	0.842	0.758	0.685	0.749	0.678
CaraNet[8]	Res2Net101	21.76	46.64	26.97	0.921	0.876	0.913	0.859	0.902	0.836	0.775	0.700	0.740	0.660
LDNet[17]	Res2Net50	66.57	33.38	21.45	0.923	0.872	0.912	0.855	0.893	0.826	0.794	0.715	0.778	0.707
SSFormer[14]	MiT-b2	19.10	29.57	52.38	0.916	0.873	0.925	0.878	0.887	0.821	0.772	0.697	0.767	0.698
FPN-in-FPN	VAN-B0	2.55	4.01	61.25	0.921	0.872	0.914	0.860	0.900	0.837	0.786	0.710	0.802	0.724

3.3 Comparisons with State-of-the-Art

As shown in Table 1, the proposed FPN-in-FPN method demonstrates competitive performance compared to state-of-the-art models across all five benchmark datasets, achieving superior or comparable results in both mDice and mIoU.

On the CVC-ClinicDB dataset, our method achieves an mDice of 0.921 and an mIoU of 0.872, surpassing models like U-Net, PraNet, and ResUNet++, while remaining competitive with the highest reported mDice of 0.923 and mIoU of 0.879. Similarly, on the Kvasir dataset, it attains a competitive mDice of 0.914 and an mIoU of 0.860, outperforming leading methods such as the MSNet series and CaraNet. The method also excels on the CVC-T dataset, achieving an mDice of 0.900 and an mIoU of 0.837, outperforming models like LDNet and SSFormer, which struggle with complex polyp structures. On CVC-ColonDB, our method achieves an mDice of 0.786 and an mIoU of 0.710, ranking second and very closely approaching the top-performing method in both metrics. Finally, on the difficult ETIS dataset, our approach achieves an mDice of 0.802 and an mIoU of 0.724, surpassing all existing state-of-the-art methods by a large margin, further demonstrating its robustness. Moreover, as shown in Table 1, our method achieves a favorable balance between accuracy and efficiency, with significantly lower computational costs in terms of MACs, parameters, and FPS, compared to state-of-the-art models. In summary, the FPN-in-FPN architecture provides a highly effective and efficient solution for polyp segmentation, offering improved multi-scale feature fusion and superior handling of challenges such as irregular polyp shapes and subtle boundaries compared to existing methods.

Table 2: Ablation study on Different Numbers of FPN-in-FPN modules.

Method	CVC-ClinicDB		Kvasir		CVC-T		CVC-ColonDB		ETIS	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
Baseline	0.911	0.859	0.901	0.847	0.877	0.809	0.774	0.695	0.764	0.676
N=1	0.915	0.862	0.913	0.859	0.89	0.822	0.765	0.694	0.774	0.686
N=2	0.921	0.872	0.914	0.860	0.900	0.837	0.786	0.710	0.802	0.724
N=3	0.919	0.874	0.911	0.858	0.897	0.834	0.791	0.714	0.781	0.707
N=4	0.912	0.863	0.910	0.859	0.899	0.834	0.785	0.708	0.763	0.686

3.4 Ablation on Different Numbers of FPN-in-FPN modules

In this ablation study, we analyze the impact of varying the number of FPN-in-FPN modules in the network. The results in Table 2 show that increasing the number of iterations from $N = 1$ to $N = 2$ improves performance across all datasets, with noticeable gains in both mDice and mIoU scores. Specifically, the performance improvement is particularly significant on the three challenging datasets: CVC-T (+0.10 mDice, +0.15 mIoU), CVC-ColonDB (+0.21 mDice, +0.16 mIoU), and ETIS (+0.28 mDice, +0.58 mIoU). However, beyond $N = 2$, the performance begins to decline, except for CVC-ColonDB, suggesting that excessive feature fusion does not consistently benefit the network. This indicates that only a few FPN-in-FPN modules are optimal for polyp segmentation, balancing feature representation effectiveness and computational efficiency. Based on these findings, we set $N = 2$ in our experiments to achieve the optimal trade-off between performance and computational cost for polyp segmentation.

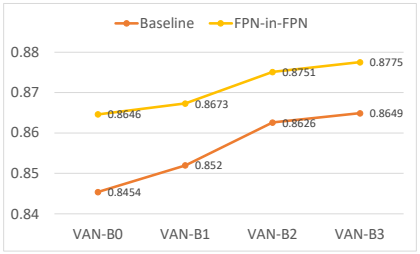

Fig. 3: Ablation results in mDice with different backbones.

Table 3: Ablation experiments on computational cost in terms of the number of parameters and MACs.

Method	Baseline		FPN-in-FPN	
	MACs (G)	Params (M)	MACs (G)	Params (M)
VAN-B0	2.31	3.89	2.55	4.01
VAN-B1	6.38	13.41	6.62	13.52
VAN-B2	12.61	26.13	12.85	26.24
VAN-B3	22.41	44.31	22.65	44.43

3.5 Ablation on Computational Cost of FPN-in-FPN module

In this ablation study, we analyze the impact of computational cost after integrating the FPN-in-FPN module into networks with different complexities. The average mDice across five datasets (Fig. 3) shows that the smallest vanilla VAN-B0 achieves only 0.8454 mDice, but this improves to 0.8646 after incorporating our FPN-in-FPN module, closely approaching the largest VAN-B3 baseline of 0.8649. Notably, the FPN-in-FPN module requires only an additional 0.24G MACs and 0.12M parameters as shown in Table 3, making it significantly more efficient than the VAN-B3 baseline. In addition, our FPN-in-FPN module improves all baselines with the same extra computational cost.

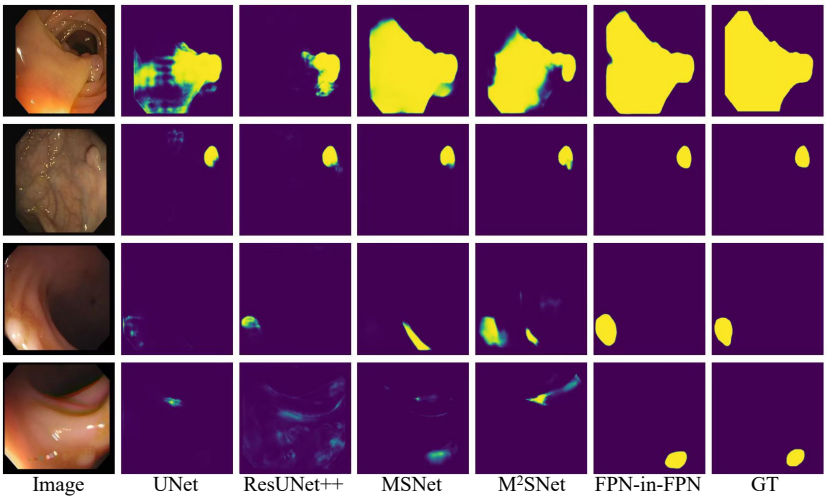

Fig. 4: Visual comparison of different state-of-the-art methods.

4 Conclusion

We proposed FPN-in-FPN: A Nested Multi-Scale Aggregation Network for polyp segmentation, which effectively fuses multi-scale features through top-down and bottom-up pathways. This design captures both global context and fine details, improving accuracy, especially for challenging polyp cases. Deep supervision further enhances boundary detection and small lesion segmentation. Extensive experiments show that our method in general outperforms state-of-the-art approaches, especially taking into account both accuracy and efficiency.

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. CMIG **43**, 99–111 (2015)
2. 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 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23. pp. 263–273. Springer (2020)
3. Guo, M.H., Lu, C.Z., Liu, Z.N., Cheng, M.M., Hu, S.M.: Visual attention network. Computational Visual Media **9**(4), 733–752 (2023)
4. 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: ICMM. pp. 451–462. Springer (2020)
5. 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: ISM. pp. 225–2255. IEEE (2019)
6. Kim, T., Lee, H., Kim, D.: Uacanet: Uncertainty augmented context attention for polyp segmentation. In: ACM MM. pp. 2167–2175 (2021)
7. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
8. Lou, A., Guan, S., Ko, H., Loew, M.H.: Caranet: context axial reverse attention network for segmentation of small medical objects. pp. 81 – 92. SPIE (2022)
9. 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)
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
11. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. IJCARS **9**(2), 283–293 (2014)
12. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. TMI **35**(2), 630–644 (2015)

13. 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. JHE (2017)
14. Wang, J., Huang, Q., Tang, F., Meng, J., Su, J., Song, S.: Stepwise feature fusion: Local guides global. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part III. pp. 110–120. Springer (2022)
15. Wei, J., Hu, Y., Zhang, R., Li, Z., Zhou, S.K., Cui, S.: Shallow attention network for polyp segmentation. In: MICCAI. pp. 699–708. Springer (2021)
16. Zhang, L., Dai, J., Lu, H., He, Y., Wang, G.: A bi-directional message passing model for salient object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1741–1750 (2018)
17. Zhang, R., Lai, P., Wan, X., Fan, D.J., Gao, F., Wu, X.J., Li, G.: Lesion-aware dynamic kernel for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part III. pp. 99–109. Springer (2022)
18. Zhao, X., Jia, H., Pang, Y., Lv, L., Tian, F., Zhang, L., Sun, W., Lu, H.: M^2snet: Multi-scale in multi-scale subtraction network for medical image segmentation. arXiv preprint arXiv:2303.10894 (2023)
19. Zhao, X., Pang, Y., Zhang, L., Lu, H., Zhang, L.: Suppress and balance: A simple gated network for salient object detection. In: Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, part II 16. pp. 35–51. Springer (2020)
20. Zhao, X., Zhang, L., Lu, H.: Automatic polyp segmentation via multi-scale subtraction network. In: MICCAI. pp. 120–130. Springer (2021)
21. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4. pp. 3–11. Springer (2018)
22. 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)