This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

LSSNet: A Method for Colon Polyp Segmentation Based on Local Feature Supplementation and Shallow Feature Supplementation

Wei Wang, Huiying Sun, and Xin Wang(✉)

Changsha University of Science and Technology, Changsha, China
wangxin@csust.edu.cn

Abstract. Accurate polyp segmentation methods are essential for colon polyp screening and colorectal cancer diagnosis. However, polyp segmentation faces the following challenges: (1) Small-sized polyps are easily lost during the identification process. (2) The boundaries separating the polyp from its surroundings are fuzzy. (3) Additional distracting information is introduced during the colonoscopy procedure, resulting in noise in the colonoscopy image and influencing the segmentation outcomes. To cope with these three challenges, a method for colon polyp segmentation based on local feature supplementation and shallow feature supplementation (LSSNet) is proposed by incorporating feature supplementation structures in the encoder-decoder structure. The multiscale feature extraction (MFE) module is designed to extract local features, the interlayer attention fusion (IAF) module is designed to fuse supplementary features with the current layer features, and the semantic gap reduction (SGR) module is designed to reduce the semantic gaps between the layers, which together form the local feature supplementation structure. The shallow feature supplementation (SFS) module is designed to supplement the features in the fuzzy areas. Based on these four modules LSSNet is proposed. LSSNet is evaluated on five datasets: ClinicDB, KvasirSEG, ETIS, ColonDB, and EndoScene. The results show that mDice scores are improved by 1.33%, 0.74%, 2.65%, 1.08%, and 0.62% respectively over the compared state-of-the-art methods. The codes are available at https://github.com/heyeying/LSSNet.

Keywords: Polyp · Local features · Shallow features · Supplementation.

1 Introduction

The second most common cause of cancer-related mortality globally is colorectal cancer (CRC) [12]. Colonoscopy is a commonly used screening method that can help reduce morbidity and mortality from colorectal cancer [11]. Colonoscopy can lower colorectal cancer mortality by approximately 60% and incidence by roughly 40%, according to the research [14]. However, this inspection process takes a lot of time for the doctor and there are cases of missed detections. The use of segmentation algorithms to recognize polyps in colonoscopy images can improve detection efficiency and assist in polyp diagnosis.

Methods for medical image segmentation have been developing quickly in recent years due to the rapid advancement of deep learning and the application of convolutional neural networks. FCN [7] was the first to propose end-to-end fully convolutional networks. Fully convolutional versions of existing networks could accept inputs of arbitrary size for semantic segmentation. UNet [9] was based on the FCN architecture and expanded it by proposing a symmetric U-shaped architecture, which was made up of a contracting path and an expanding path. UNet++ [26] proposed an encoder-decoder network using dense skip paths to reduce the semantic gap between the encoder and decoder. Given the success of Transformer's application in computer vision, PHCU-Net [21] added Transformer to UNet and proposed a dual-branch attention structure to extract local features and global features.

Apart from the medical image segmentation networks previously mentioned, specific networks for polyp segmentation were designed to tackle the difficulties associated with polyp segmentation. PraNet [2] introduced a parallel reverse attention network for the aggregation of high-level features and the generation of a global map. The global map guided the network to reversely extract boundary features. However, its processing was not robust enough in the presence of noise. SANet [19] introduced the shallow attention network to decouple image color and content while utilizing the shallow attention mechanism to suppress noise. Nonetheless, the network did not give careful thought to ways to enhance the capacity of the model for feature learning to enhance the model's performance. LDNet [23] introduced a lesion-aware dynamic network, which was capable of detecting hidden polyps by using dynamic kernel adaptive learning based on the input image to improve segmentation capability. However, the network is more prone to over-detection. CASCADE [8] introduced a cascaded attention decoder, using the pyramid transformer as the backbone, and using the convolutional attention method to build the encoder. However, interlayer feature fusion using only the information from the layers after backbone downsampling resulted in localized information loss.

Considering the drawbacks of the above methods and challenges such as loss of small-sized polyps, fuzzy boundaries, and noise, designing a method that enables the network to pay more attention to the local features of the polyps and boundary features is very necessary. Therefore, a method for colon polyp segmentation based on local feature supplementation and shallow feature supplementation (LSSNet) is proposed. The local feature supplementation structure mainly integrates local features from previous layers for supplementation while reducing semantic gaps between layers. Subsequently, multiscale feature extraction is applied to the supplemented features to prevent the loss of small-sized polyps. The shallow feature supplementation structure can focus on the fuzzy boundary features of the polyp, and the fuzzy boundary can be clarified by injecting shallow features. The main contributions of this work can be summarized as follows: (1) The LSSNet is proposed to improve the precision of polyp segmentation by incorporating two feature supplementation structures in the U-shaped encoder-decoder structure. (2) The local feature supplementation structure consisting of

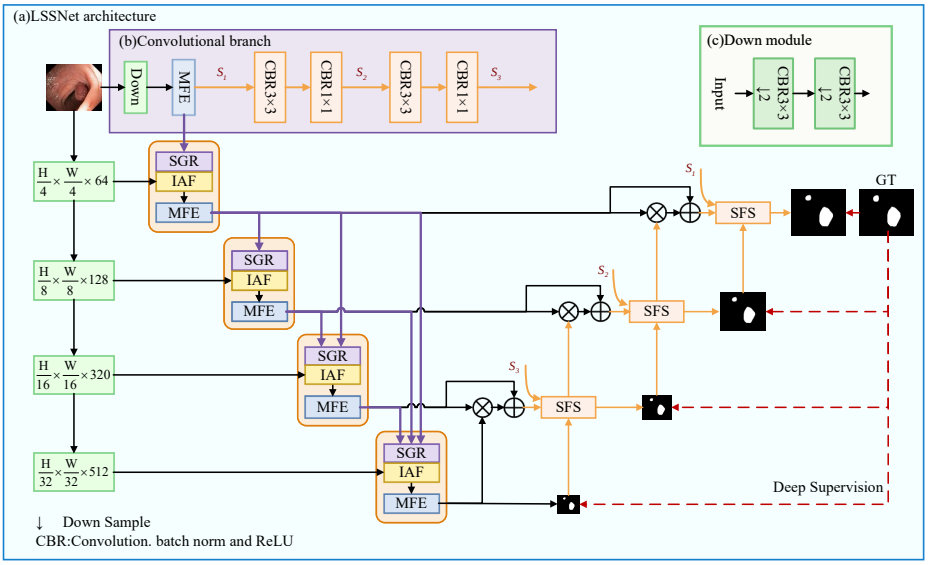

Fig. 1. (a) The overall architecture of LSSNet. (b) The structure of the convolution branch. (c) Illustration of Down module.

a multiscale feature extraction (MFE) module, a semantic gap reduction (SGR) module, and an interlayer attention fusion (IAF) module is developed. (3) The shallow feature supplementation structure consisting of a shallow feature supplementation (SFS) module and convolutional branch is proposed. (4) The proposed method is evaluated on five challenging polyp datasets, and the experimental results showcase its superiority over five other SOTA methods.

2 Method

The architecture of LSSNet is shown in Fig. 1(a). The local feature supplementation structure and shallow feature supplementation structure are added to the encoder-decoder structure. Specifically, the pre-trained PVT v2 [18] is used as the encoder. The local feature supplementation structure is composed of the SGR module, IAF module, and MFE module. The feature maps of the first few layers are aggregated by the SGR module. The IAF module uses an attention mechanism to fuse the encoder output feature maps with those of the SGR module, effectively suppressing interfering information such as noise. The MFE module extracts the local features using parallel multiscale convolutions to reduce the loss of small-sized polyps. The shallow feature supplementation structure consists of the convolution branch and the SFS module. The convolutional branch extracts shallow features with rich boundary information. The shallow features are injected into the SFS module during the fusion process in the decoder stage

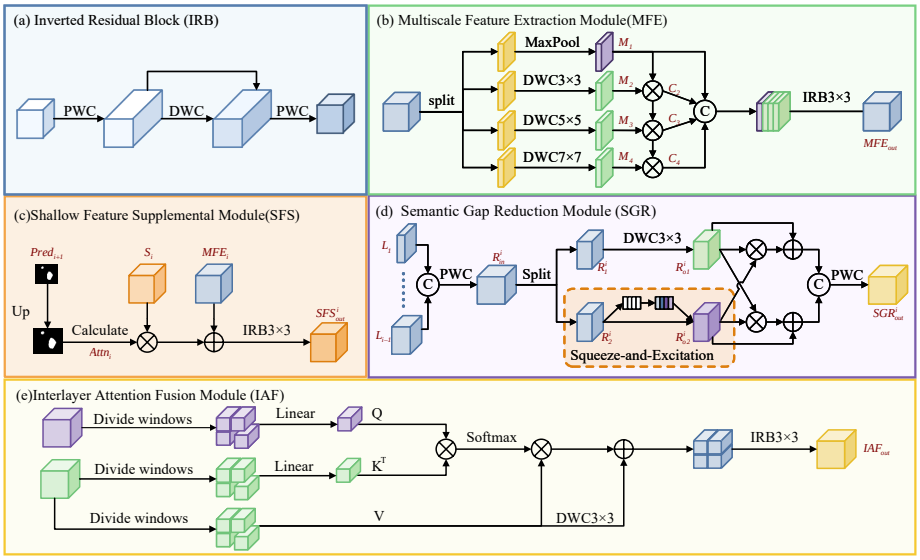

Fig. 2. (a) Illustration of IRB. (b) Illustration of MFE. (c) Illustration of SFS. (d) Illustration of SGR. (e) Illustration of IAF.

to complement the areas with fuzzy boundaries, and the output feature maps of the SFS module are used for prediction.

2.1 Convolutional Branch

Considering the advantage of CNN in extracting local features, LSSNet adds a convolutional branch to extract shallow features. The shallow features can be used as supplementary information for fuzzy areas. The convolutional branch is shown in Fig. 1(b). Specifically, the input feature maps are first convolutionally downsampled by the Down module, which is shown in Fig. 1(c). Then, multiscale local features are extracted by the MFE module. Convolution operations with kernel sizes of 3×3 and 1×1 extract the features and align the channels. Finally, the shallow feature maps $S_i, i \in 1, 2, 3$ are obtained.

2.2 Multiscale Feature Extraction Module

To address the problem that small-sized polyps are easily lost during encoder downsampling, the multiscale feature extraction (MFE) module is intended to improve the network's capability for extracting local features. As shown in Fig. 2(b), MFE consists of max pooling and depthwise convolutions in parallel with kernel sizes of 3×3, 5×5, and 7×7, respectively, to obtain outputs $M_i, i \in 1, 2, 3, 4$. The use of max pooling can emphasize the local texture in the features [13]. The strong coupling relation [20] is established by multiplication

operation during branch fusion, which can be expressed as $C_i = M_i \times M_{i-1}, i \in 2, 3, 4$. Then, the results of the four branches are concatenated. Finally, the features are enhanced after the inverted residual block(IRB) [10]. IRB is shown in Fig. 2(a). The MFE module can be represented as the formula:

$$MFE_{out} = IRB\left(Concat\left(M_1, C_2, C_3, C_4\right)\right) \tag{1}$$

2.3 Semantic Gap Reduction Module

Since there are semantic gaps between different layers, the semantic gap reduction (SGR) module is designed to aggregate features from previous layers to reduce the semantic gaps during feature fusion. SGR is depicted in Fig. 2(d). Firstly, the feature maps from the previous layers passing through the MFE module $\{L_1, \ldots, L_{i-1}\}$ are concatenated and the dimensions of the channels are adjusted by the convolutional layer to obtain R^i_{in}. The formula is expressed as:

$$R^i_{in} = PWConv\left(Concat\left(L_1, \ldots, L_{i-1}\right)\right) \tag{2}$$

where $PWConv$ is the pointwise convolution.

Then, R^i_{in} is split into two parallel branches according to the channel dimensions R^i_1, R^i_2. The R^i_1 branch uses depthwise convolution to extract spatial features R^i_{o1}. The R^i_2 branch extracts the channel features R^i_{o2} using the Squeeze-and-Excitation (SE) module [4]. The formula is expressed as:

$$R^i_{o1} = \ DWConv_{3\times3}\left(R^i_1\right), \quad R^i_{o2} \ = SE\left(R^i_2\right) \tag{3}$$

where $DWConv_{3\times3}$ is the depthwise convolution with kernel size of 3×3.

Finally, the feature maps of the two branches are multiplied for feature interaction and then via skip connections. Following the concatenation of the output feature maps from these two branches, SGR^i_{out} is obtained by fusing the features between the channels using pointwise convolution. The formula is expressed as:

$$SGR^i_{out} = \ PWConv\left(Concat\left(\left(R^i_{o1} \times R^i_{o2} + R^i_{o1}\right), \left(R^i_{o2} \times R^i_{o1} + R^i_{o2}\right)\right)\right) \tag{4}$$

2.4 Interlayer Attention Fusion Module

After aggregating local features, how to achieve feature supplementation is particularly important. For this reason, the interlayer attention fusion (IAF) module is designed to use the attention mechanism to focus on the features that are jointly important among the layers, and then carry out feature supplementation. The IAF module can be seen in Fig. 2(e). Specifically, the feature maps used for supplementation and the current layer feature maps go through the weight matrix W_q and W_k, respectively, to obtain Query(Q) and Key(K). Value(V) remains consistent with the input. The attention mechanism uses Expanded Window MHSA(EW-MHSA) [22]. Q, K, and V are divided into windows and the attention matrix is calculated. V undergoes depthwise convolution to increase the diversity of features [3]. Then, the windows are restored to their original input size. Finally, the representativeness of the features is strengthened by IRB. The formula is expressed as:

$$IAF_{out} = IRB\left(EW - MHSA\left(Q, K, V\right) + DWConv_{3\times3}\left(V\right)\right) \tag{5}$$

2.5 Shallow Feature Supplementation Module

The shallow feature supplementation (SFS) module is designed to address the challenge of polyps with ambiguous boundaries with normal regions. It can use shallow features to supplement the ambiguous regions and re-extract features. The SFS module is depicted in Fig. 2(c). Firstly, for layer i, the prediction map $Pred_{i+1}$ obtained from the subsequent layer is used to calculate the fuzzy areas with the formula $Attn_i = 1 - \frac{|Pred_{i+1} - 0.5|}{0.5}$ [24], and the fuzzy areas $Attn_i$ are used as the guidance information to multiply the elements with the shallow features S_i from the convolutional branch. Afterward, the features extracted from the current layer MFE_i are combined through elementwise summation to fuse the information. Finally the enhanced feature maps SFS^i_{out} are obtained after IRB. The formula is expressed as:

$$SFS^i_{out} = IRB\left(Attn_i \times S_i + MFE_i\right) \tag{6}$$

3 Experiments

3.1 Datasets and Evaluation Metrics

Five publicly available polyp segmentation datasets are used to evaluate LSSNet: CVC-ClinicDB [1], KvasirSEG [5], ETIS-LaribPolypDB [15], CVC-ColonDB [16], and EndoScene [17]. The dataset partitioning method remains consistent with that proposed in [2,25]. The training dataset comprises a total of 1450 images, including 900 images from KvasirSEG and 550 images from CVC-ClinicDB. The remaining images from KvasirSEG and CVC-ClinicDB are used as visible test sets to validate the learning ability of the model, while the other three datasets are used as invisible test sets to validate the generalization performance of the network.

Six metrics are used to evaluate the performance of LSSNet: mean Dice score (mDice), mean IoU score (mIoU), mean absolute error (MAE), weighted F-measure (F^w_β), E-measure (E_ξ), and S-measure (S_α).

3.2 Implementation Details

The overall network is implemented based on the PyTorch framework, and the experiments are carried out using the NVIDIA A800 GPU. The AdamW optimizer is used to train the model. The learning rate is set to 1e-4 and the batch size is set to 8. The input image size is resized to 352×352. The model is trained for 150 epochs, and the decay rate of the learning rate is 0.5 [6]. Data enhancement strategies including random rotation, random flip, and color jitter are used to increase the diversity of the training data and improve the robustness and generalization of the model.

Table 1. Comparison of experimental results across the five polyp datasets, with best results indicated in bold.

Datasets	Methods	mDice	mIoU	F_β^w	S_α	E_ξ	MAE
ClinicDB	UNet(MICCAI'15)	87.16	81.02	85.55	91.24	94.25	1.76
	PraNet(MICCAI'20)	91.90	87.04	90.18	94.50	97.42	0.90
	SANet(MICCAI'21)	92.41	87.40	90.85	94.70	97.61	0.90
	LDNet(MICCAI'22)	92.34	87.77	91.57	94.31	97.16	0.98
	CASCADE(WACV'23)	93.15	88.62	92.10	95.41	97.77	0.74
	LSSNet(Ours)	**94.48**	**90.04**	**93.94**	**95.83**	**98.84**	**0.59**
Kvasir	UNet(MICCAI'15)	86.34	79.70	82.64	88.12	92.28	4.59
	PraNet(MICCAI'20)	90.81	85.99	89.17	91.91	95.47	2.68
	SANet(MICCAI'21)	91.03	85.65	88.92	91.90	95.65	2.98
	LDNet(MICCAI'22)	90.81	85.85	89.56	91.59	94.71	3.37
	CASCADE(WACV'23)	91.87	86.97	90.78	92.67	**96.37**	2.34
	LSSNet(Ours)	**92.61**	**87.80**	**91.56**	**93.00**	96.34	**2.24**
ETIS	UNet(MICCAI'15)	63.65	54.82	55.12	77.90	81.73	2.66
	PraNet(MICCAI'20)	76.54	68.38	69.88	86.05	87.99	1.81
	SANet(MICCAI'21)	77.46	69.54	71.34	85.94	89.39	2.35
	LDNet(MICCAI'22)	74.13	66.59	69.66	84.38	87.16	2.44
	CASCADE(WACV'23)	81.99	74.90	78.16	89.16	92.34	1.47
	LSSNet(Ours)	**84.64**	**77.54**	**81.64**	**90.79**	**93.82**	**1.00**
ColonDB	UNet(MICCAI'15)	70.16	61.29	66.13	80.36	83.87	4.59
	PraNet(MICCAI'20)	76.16	69.07	74.13	84.85	87.68	3.90
	SANet(MICCAI'21)	76.75	68.76	74.03	84.51	88.18	4.06
	LDNet(MICCAI'22)	79.68	71.77	76.79	85.71	89.87	3.52
	CASCADE(WACV'23)	80.95	73.19	78.97	86.75	90.90	3.28
	LSSNet(Ours)	**82.03**	**74.13**	**80.31**	**87.27**	**91.68**	**2.80**
EndoScene	UNet(MICCAI'15)	86.35	78.27	82.02	91.33	94.44	1.20
	PraNet(MICCAI'20)	88.55	81.32	85.21	92.94	96.02	0.83
	SANet(MICCAI'21)	87.35	80.48	84.46	92.50	94.71	0.90
	LDNet(MICCAI'22)	88.76	81.87	85.92	92.87	95.82	0.85
	CASCADE(WACV'23)	87.69	80.69	84.67	92.56	94.93	0.96
	LSSNet(Ours)	**89.38**	**82.21**	**86.32**	**93.03**	**96.66**	**0.67**

3.3 Results

LSSNet is compared with five SOTA methods, including UNet[9], PraNet[2], SANet[19], LDNet[23], and CASCADE[8], on five publicly available polyp datasets. To ensure fair comparisons, the codes for the comparison networks are re-run with the same training strategies. The comparison results between LSSNet and other methods are shown in Table 1. LSSNet achieves the best mDice, mIoU on all five datasets, Specifically, a significant improvement on ClinicDB and ETIS, with 1.33% and 2.65% improvement in mDice, respectively, when compared to the best results among the other methods. It should be noted that in the results on Kvasir, the E_ξ score of LSSNet is slightly lower than CASCADE by 0.03%, which is almost the same, but the other scores are better than CASCADE. The visualization results of different methods are displayed in Fig. 3.

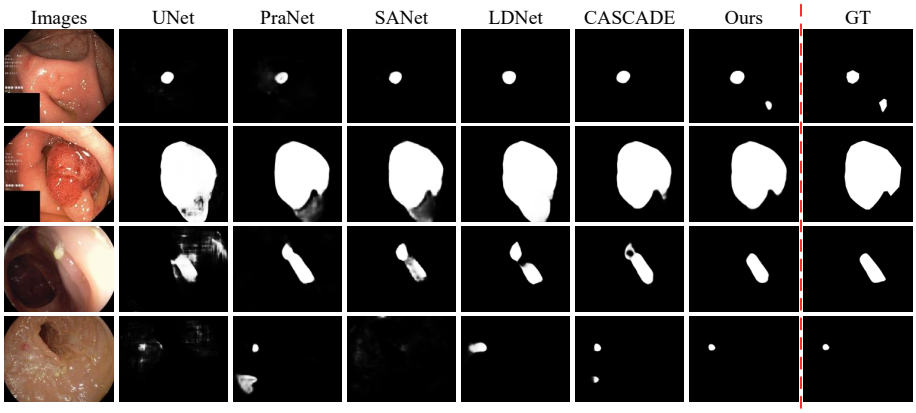

Fig. 3. Qualitative results of different methods.

Table 2. Ablation studies of LSSNet on CVC-ClinicDB and ETIS datasets.

MFE	IAF	SGR	SFS	ClinicDB			ETIS		
				mDice	mIoU	MAE	mDice	mIoU	MAE
✗	✓	✓	✓	92.90	88.23	0.78	80.74	73.45	1.20
✓	✗	✗	✓	93.15	88.68	0.69	82.60	75.54	1.57
✓	✓	✓	✗	92.92	88.38	0.76	81.57	73.93	1.42
✓	✓	✓	✓	**94.48**	**90.04**	**0.59**	**84.64**	**77.54**	**1.00**

3.4 Ablation Experiments

A series of ablation experiments are carried out on the CVC-ClinicDB and ETIS-LaribPolypDB datasets to confirm the importance and effectiveness of the LSSNet core modules. Specifically, the MFE module, IAF and SGR module, and SFS module are deleted on the basis of LSSNet, respectively. The results are shown in Table 2. On the CVC-ClinicDB dataset, deleting the MFE module decreases the mDice and mIoU scores the most, 1.58% and 1.81%. Since the SGR module is designed to assist the IAF module, the IAF and SGR modules are deleted together, and the mDice and mIoU scores decrease by 1.33% and 1.36%. Deleting the SFS module decreases the mDice and mIoU by 1.56% and 1.66%. The experimental results show that the core modules of LSSNet are all effective.

4 Conclusion

A high-precision network for colon polyp segmentation based on local feature supplementation and shallow feature supplementation is proposed to solve the problems of small-size polyp loss, fuzzy boundaries, and noise interference in colonoscopy images in polyp segmentation. The LSSNet adds a local feature supplementation structure to the encoder-decoder structure. In addition, rich

boundary features are added to the network through shallow feature supplementation, which makes the network prediction more accurate. Experiments have demonstrated the effectiveness of LSSNet, which can be used to assist in the diagnosis of colonic polyps, making the diagnosis more precise and automated.

Acknowledgments. This work is supported by the National Science Innovation Special Zone Project(2019XXX00701), the National Key Basic Research Program Project (62402010206), Key research and development projects of Hunan Province(2020SK2134), and the Natural Science Foundation of Hunan Province (2022JJ30625).

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. 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)
3. Han, D., Pan, X., Han, Y., Song, S., Huang, G.: Flatten transformer: Vision transformer using focused linear attention. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5961–5971 (2023)
4. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
5. 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)
6. Liu, G., Yao, S., Liu, D., Chang, B., Chen, Z., Wang, J., Wei, J.: Cafe-net: Cross-attention and feature exploration network for polyp segmentation. Expert Systems with Applications **238**, 121754 (2024)
7. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
8. 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)
9. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
10. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520 (2018)

11. Screening, P., Board, P.E.: Colorectal cancer screening (pdq®). In: PDQ Cancer Information Summaries [Internet]. National Cancer Institute (US) (2023)
12. Sedlak, J.C., Yilmaz, Ö.H., Roper, J.: Metabolism and colorectal cancer. Annual Review of Pathology: Mechanisms of Disease **18**, 467–492 (2023)
13. Si, C., Yu, W., Zhou, P., Zhou, Y., Wang, X., Yan, S.: Inception transformer. Advances in Neural Information Processing Systems **35**, 23495–23509 (2022)
14. 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)
15. 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)
16. 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)
17. Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., Drozdzal, M., Courville, A., et al.: A benchmark for endoluminal scene segmentation of colonoscopy images. Journal of healthcare engineering **2017** (2017)
18. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pvt v2: Improved baselines with pyramid vision transformer. Computational Visual Media **8**(3), 415–424 (2022)
19. Wei, J., Hu, Y., Zhang, R., Li, Z., Zhou, S.K., Cui, S.: Shallow attention network for polyp segmentation. 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. 699–708. Springer (2021)
20. Xia, Y., Yun, H., Liu, Y.: Mfefnet: Multi-scale feature enhancement and fusion network for polyp segmentation. Computers in Biology and Medicine **157**, 106735 (2023)
21. Xu, J., Wang, X., Wang, W., Huang, W.: Phcu-net: A parallel hierarchical cascade u-net for skin lesion segmentation. Biomedical Signal Processing and Control **86**, 105262 (2023)
22. Zhang, J., Li, X., Li, J., Liu, L., Xue, Z., Zhang, B., Jiang, Z., Huang, T., Wang, Y., Wang, C.: Rethinking mobile block for efficient attention-based models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1389–1400 (2023)
23. 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)
24. 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)
25. 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)
26. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Im-

age Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11. Springer (2018)