



BoxPolyp: Boost Generalized Polyp Segmentation Using Extra Coarse Bounding Box Annotations

Jun Wei^{1,2,3}, Yiwen Hu^{1,2,3,6}, Guanbin Li⁷, Shuguang Cui^{1,2,3},
S. Kevin Zhou^{1,4,5}, and Zhen Li^{1,2,3(✉)}

¹ School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China

lizhen@cuhk.edu.cn

² Shenzhen Research Institute of Big Data, Shenzhen, China

³ The Future Network of Intelligence Institute, Shenzhen, China

⁴ School of Biomedical Engineering and Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, China

⁵ Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

⁶ Institute of Urology, The Third Affiliated Hospital of Shenzhen University (Luohu Hospital Group), Shenzhen, China

⁷ School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, China

Abstract. Accurate polyp segmentation is of great importance for colorectal cancer diagnosis and treatment. However, due to the high cost of producing accurate mask annotations, existing polyp segmentation methods suffer from severe data shortage and impaired model generalization. Reversely, coarse polyp bounding box annotations are more accessible. Thus, in this paper, we propose a boosted **BoxPolyp** model to make full use of both accurate mask and extra coarse box annotations. In practice, box annotations are applied to alleviate the over-fitting issue of previous polyp segmentation models, which generate fine-grained polyp area through the iterative boosted segmentation model. To achieve this goal, a fusion filter sampling (FFS) module is firstly proposed to generate pixel-wise pseudo labels from box annotations with less noise, leading to significant performance improvements. Besides, considering the appearance consistency of the same polyp, an image consistency (IC) loss is designed. Such IC loss explicitly narrows the distance between features extracted by two different networks, which improves the robustness of the model. Note that our BoxPolyp is a plug-and-play model, which can be merged into any appealing backbone. Quantitative and qualitative experimental results on five challenging benchmarks confirm that our proposed model outperforms previous state-of-the-art methods by a large margin.

Keywords: Polyp segmentation · Colonoscopy · Colorectal cancer

J. Wei and Y. Hu—Equal Contributions.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Wang et al. (Eds.): MICCAI 2022, LNCS 13433, pp. 67–77, 2022.

https://doi.org/10.1007/978-3-031-16437-8_7

1 Introduction

Colorectal Cancer (CRC) is one of the leading causes of malignant tumors death worldwide. As the precursor of CRC, colorectal polyps are the driver of CRC morbidity and mortality. Therefore, accurate polyp segmentation and diagnosis are of great significance to the survival of patients. Thanks to the evolution of computer technology, massive polyp segmentation models [1, 4, 6–8, 17, 20, 23, 24, 26, 29] have been proposed and achieved remarkable performance.

However, these models are always plagued by data shortages and suffer from severe over-fitting issue. Since the popularity of U-Net [17] and FCN [13], most of polyp segmentation models [7, 23] are based on convolutional neural networks (CNNs), which outperform traditional handcrafted ones but are data hungry. Unfortunately, accurate labeling of polyp masks is time-consuming and laborious, requiring pixel-by-pixel operation. Therefore, existing polyp segmentation datasets are relatively small. In particular, the widely adopted polyp training set [7, 23] contains only 1,451 images, far from enough to feed a large capacity CNN model. Thus, models trained on this dataset exhibit the unstable performance and are sensitive to noise, which hampers the practical clinical usage.

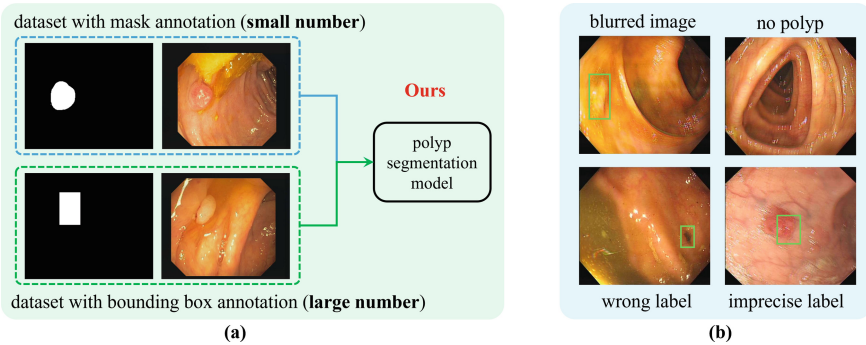


Fig. 1. (a) Our proposed polyp segmentation model using both accurate mask annotations and coarse bounding box ones. (b) Common annotation issues of the polyp detection dataset LDPolypVideo [14]. The second row shows the noisy annotations.

To provide effective clinical assistance, a generalized polyp segmentation model is urgently needed. In this paper, we struggle to achieve this goal using extra coarse bounding box annotations to expand the small segmentation dataset. Specifically, a large open-released polyp detection dataset LDPolypVideo [14] is adopted, which consists of 160 polyp video clips with 40,266 frames. Though all these images are labeled with only coarse bounding box annotations, they provide sufficient polyp appearance information and are much cheaper. Figure 1(a) shows the core idea of our method. Trained with a few images with mask annotations and a lot of images with bounding box annotations, a more generalized polyp segmentation model is achieved. However,

directly applying the bounding box annotations of LDPolypVideo is suboptimal. Because the bounding box area contains many background pixels. Taking the bounding box area as polyp mask will bring a lot of noise. Besides, as shown in Fig. 1(b), LDPolypVideo contains many blurred images, images with no polyps, wrong labels and imprecise labels, which also will mislead the model training.

To make the most of the good parts of LDPolypVideo annotations and reduce the bad parts, we propose the novel BoxPolyp model which mainly consists of two modules: fusion filter sampling and image consistency loss. In practice, fusion filter sampling (FFS) aims to generate pseudo labels for high-confidence regions of each image in LDPolypVideo. By combining the raw bounding box annotations and the predicted masks (derived from the model trained on a small polyp segmentation dataset), FFS efficiently produces pixel-wise pseudo masks for deterministic regions. For uncertain regions, pseudo masks are inaccurate and therefore discarded. However, these discarded regions also contain valuable information. To fully explore these regions, we propose the image consistency (IC) loss instead of generating pseudo masks. IC loss applies two different networks to extract features from the same image and explicitly reduces the distance between features of the uncertain regions. By forcing feature alignment, our model could learn robust polyp feature representations, requiring no mask annotations.

In summary, our contributions are three-folds: (1) We are the first to boost a generalized polyp segmentation model through extra bounding box annotations. (2) We propose the fusion filter sampling to generate pseudo masks with less noise and design the image consistency loss to enhance the feature robustness of uncertain regions. (3) Our proposed BoxPolyp is a plug-and-play model, which can largely enhance polyp segmentation performance using different backbones.

2 Related Work

Traditional polyp segmentation models [20, 28] are mostly based on low-level features (*i.e.*, color, texture and boundary). But limited by the poor semantics, these models fail when dealing with complex scenarios. Recently, fully convolutional networks (FCN) [13] have been widely adopted for polyp segmentation and make great progress. For example, U-Net [17], U-Net++ [29] and ResUNet++ [11] use the encoder-decoder architecture to handle the segmentation tasks, which has become the standard paradigm for subsequent works. However, the polyp boundaries are not well handled by these methods. Afterwards, PsiNet [15], LODNet [5], PraNet [7], MSNet [27] and SFANet [8] force the model to learn the feature differences, which greatly enhances the model’s perception for polyp boundaries and achieve the promising results.

Besides, ACSNet [24], HRENet [18] and CCBANet [16] pay more attention to context information. By adaptively aggregating multi-scale contexts, the ambiguity of local features will be reduced, thus leading to highly confident predictions. Unlike the above methods, SANet [23] deals with the polyp segmentation task in terms of data distribution. By eliminating the color bias of the image, SANet achieves robust performance gains in different scenarios. Furthermore,

with the success of transformer in image processing, researchers start working on the long-distance dependency. For instance, PNSNet [12] uses a self-attention block to mine the temporal and spatial relations in polyp videos. Polyp-Pvt [6] directly introduces a transformer encoder to replace the widely used CNN backbones. Differently, Transfuse [25] combines both CNN and transformer to extract spatial correlation and global context. All these methods have achieved remarkable performance. But limited by training set, these models suffer from the over-fitting issue. Therefore, we propose to use the cheap bounding box annotations to boost a generalized polyp segmentation model.

3 Method

Figure 2 depicts the whole framework of the proposed BoxPolyp segmentation model, consisting of two parts: fusion filter sampling and image consistency loss. Without special instructions, we use SANet [23] as our baseline model.

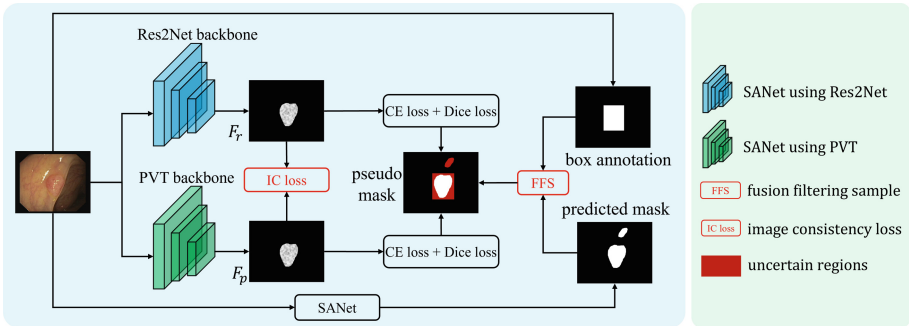


Fig. 2. The pipeline for our proposed BoxPolyp model. First, a SANet [23] trained on the small polyp segmentation dataset is used to predict the pixel-wise mask for each box-annotated image. Then, a FFS module combines the predicted mask and the box annotation to get the deterministic regions as pseudo labels. For regions of uncertainty, we propose the IC loss to reduce the distance between features extracted from two different backbones (*i.e.*, Res2Net [9] and PVT [22])

3.1 Fusion Filter Sampling

We integrate polyp detection dataset (*i.e.*, LDPolypVideo [14]) to enhance the polyp segmentation model. But LDPolypVideo is flawed in two ways. First, as shown in Fig. 1(b), there exists many wrongs and imprecise labels in LDPolypVideo, bringing noise for supervision. Second, bounding box annotations only provide coarse polyp contours and some background pixels are also included. Directly taking bounding box masks as pseudo labels will mislead the

model. To solve the above issues, we propose fusion filter sampling (FFS) to generate pseudo masks with less noise interference.

Specifically, FFS filters out noise through object-level bounding box annotations and pixel-level pseudo masks, where object-level annotations weed out mislabeled or hard images and pixel-level masks filter out the background pixels in bounding box regions. For the object-level operation, given an image I , we first convert its bounding box annotations into a binary mask B , as shown in Fig. 3(a). Meanwhile, a pre-trained SANet [23] model (trained on small polyp segmentation dataset) is applied to get a coarse prediction P for I . Intuitively, if there is a big difference between B and P , I may be a hard sample or a mislabeled sample. In either case, I will be filtered out and not involved in the model training. Thus, the issues shown in Fig. 1(b) will be alleviated. In practice, we choose Dice $d = \frac{2BP}{B+P}$ to measure the difference between B and P . Only images with $d > 0.7$ will be selected out to minimize the impact of object-level wrong annotations. For the pixel-level operation, we combine the complementarity of B and P to refine the pseudo masks. Specifically, we choose pixels where both B and P are equal to 1 as the foreground F . Namely, $F = B \cap P$. Similarly, only pixels where both B and P are equal to 0 will be regarded as the background K . Namely, $K = (1 - B) \cap (1 - P)$. Other pixels belong to the uncertain regions, as shown in Fig. 3(a). During training, only F and K are involved in supervision, while uncertain regions are dealt with IC loss (described in Sect. 3.2). Through FFS, we maximize the utilization of bounding box annotations and minimize the potential noise interference.

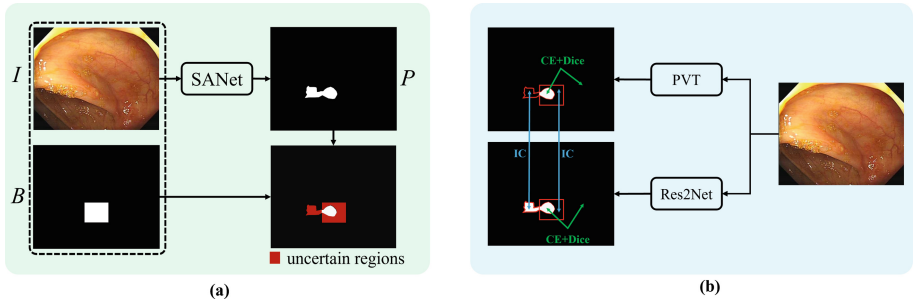


Fig. 3. (a) Refined pseudo mask generation using fusion filter sampling module, which consists of foreground, background and uncertain regions. (b) Different supervision for regions of certainty and regions of uncertainty.

3.2 Image Consistency (IC) Loss

By combining bounding box annotations and predicted masks, FFS module obtains deterministic foreground and background regions for supervision, as

shown in Fig. 3(a). However, the regions of uncertainty are not supervised during training. Because no matter the box mask or the predicted mask is used as the pseudo label, it will bring a lot of noise which is harmful to the model generalization. In view of this, we propose the image consistency loss which mines supervisory information from the relationship between images, instead of pseudo labels.

Specifically, for each polyp image, we send it to two SANet models but with different backbone networks (*i.e.*, Res2Net [9] and PVT [22]), as shown in Fig. 2. Due to the different architectures (*i.e.*, CNN and Transformer), the features F_r and F_p extracted by Res2Net and PVT present different characteristics. Meanwhile, F_r and F_p come from the same image. They should have similar appearance. To bring the supervision for regions of uncertainty, we propose the IC loss to explicitly reduce the distance between F_r and F_p , as shown in Eq. 1.

$$\mathcal{L}_{IC} = \frac{\sum_{i,j} (F_r^{i,j} - F_p^{i,j})^2 \cdot U^{i,j}}{\sum_{i,j} U^{i,j}} \quad (1)$$

where i and j are the pixel indexes of polyp regions, U represents the mask of uncertain regions. Thus, IC loss focuses on regions without labels. Supervised by the IC loss, our model outputs more consistent predictions and greatly reduces the over-fitting risk.

3.3 Loss Function

Following previous methods [7, 23], binary cross entropy \mathcal{L}_{BCE} and Dice loss \mathcal{L}_{Dice} are adopted. Besides, the proposed \mathcal{L}_{IC} is also involved in the total loss, as shown in Eq. 2.

$$\mathcal{L}_{total} = \mathcal{L}_{BCE} + \mathcal{L}_{Dice} + \mathcal{L}_{IC} \quad (2)$$

4 Experiments

4.1 Datasets and Training Settings

Five widely used polyp segmentation datasets are adopted to evaluate the model performance, including Kvasir [10], CVC-ClinicDB [2], CVC-ColonDB [3], EndoScene [21] and ETIS [19]. For the comparability, we follow the same dataset partition as [7]. Besides, nine state-of-the-art methods are used for comparison, namely U-Net [17], U-Net++ [29], ResUNet [26], ResUNet++ [11], SFA [8], PraNet [7], SANet [23], MSNet [27] and Polyp-Pvt [6]. Pytorch is used to implement our BoxPolyp model. All input images are uniformly resized to 352×352 . For data augmentation, random flip, random rotation and multi-scale training are adopted. The whole network is trained in an end-to-end way with a AdamW optimizer. Initial learning rate and batch size are set to $1e-4$ and 16, respectively. We train the entire model for 80 epochs.

Table 1. Performance comparison with different polyp segmentation models. The red column represents the weighted average (wAVG) performance of different testing datasets. Below the dataset name is the image number of each dataset.

Methods	ColonDB		Kvasir		ClinicDB		EndoScene		ETIS		wAVG	
	380		100		62		60		196		798	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
U-Net	.512	.444	.818	.746	.823	.750	.710	.627	.398	.335	.561	.493
U-Net++	.483	.410	.821	.743	.794	.729	.707	.624	.401	.344	.546	.476
ResUNet	-	-	.791	-	.779	-	-	-	-	-	-	-
ResUNet++	-	-	.813	.793	.796	.796	-	-	-	-	-	-
SFA	.469	.347	.723	.611	.700	.607	.467	.329	.297	.217	.476	.367
PraNet	.712	.640	.898	.840	.899	.849	.871	.797	.628	.567	.741	.675
MSNet	.751	.671	.905	.849	.918	.869	.865	.799	.723	.652	.785	.714
SANet	.753	.670	.904	.847	.916	.859	.888	.815	.750	.654	.794	.714
Ours-Res2Net	.820	.741	.910	.857	.904	.849	.903	.835	.829	.742	.846	.771
Polyp-Pvt	.808	.727	.917	.864	.937	.889	.900	.833	.787	.706	.833	.760
Ours-Pvt	.819	.739	.918	.868	.918	.868	.906	.840	.842	.755	.851	.776

4.2 Quantitative Comparison

To prove the effectiveness of the proposed BoxPolyp, nine state-of-the-art models are used for comparison, as shown in Table 1. BoxPolyp surpasses previous methods by a large margin on the weighted average (wAVG) performance of five datasets, demonstrating the superior performance of the proposed methods. In addition, Fig. 4 shows the Dice values of the above models under different thresholds (used to binarize the mask). From these curves, we observe that BoxPolyp consistently outperforms other models, which proves its good capability for polyp segmentation.

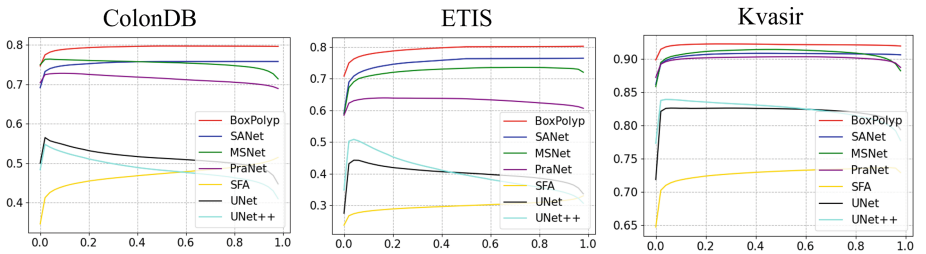


Fig. 4. Dice curves under different thresholds on three polyp datasets.

4.3 Visual Comparison

Figure 5 visualizes some predictions of different models. Compared with other counterparts, our method not only clearly highlights the polyp regions but also suppresses the background noise. Even for challenging scenarios, our model still handles well and generates accurate segmentation mask.

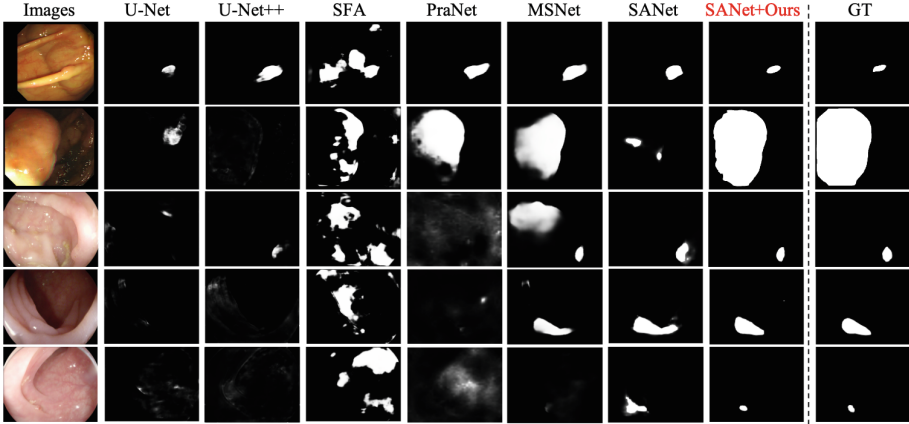


Fig. 5. Visual comparison between the proposed method and six state-of-the-art ones.

4.4 Ablation Study

To investigate the importance of each component in BoxPolyp, the weighted average (wAVG) performance is adopted. We evaluate the model on both Res2Net [9] and PVT [22] for ablation studies. As shown in Table 2, all proposed modules are beneficial for the final predictions. Combining all these modules, our model achieves the new state-of-the-art performance.

Table 2. Ablation studies for BoxPolyp with different backbone networks.

Settings	wAVG-Res2Net		wAVG-PVT	
	mDice	mIoU	mDice	mIoU
SANet	0.794	0.714	0.833	0.760
SANet+FFS	0.839	0.757	0.848	0.772
SANet+FFS+IC	0.846	0.771	0.851	0.776

5 Conclusion

Limited by the size of the dataset, existing polyp segmentation models are vulnerable to noise and suffer from over-fitting. For the first time, we leverage the cheap bounding box annotations to alleviate data shortage for a polyp segmentation task. Although coarse, these annotations can greatly improve the model generalization. It is achieved by the proposed FFS module and IC loss. In the future, we will explore the design of a weakly-supervised polyp segmentation model based on only bounding box annotations without masks.

Acknowledgement. This work is supported by the Guangdong Provincial Key Laboratory of Big Data Computing, The Chinese University of Hong Kong, Shenzhen, by NSFC-Youth 61902335, by Key Area R&D Program of Guangdong Province with grant No. 2018B030338001, by the National Key R&D Program of China with grant No. 2018YFB1800800, by Shenzhen Outstanding Talents Training Fund, by Guangdong Research Project No. 2017ZT07X152, by Guangdong Regional Joint Fund-Key Projects 2019B1515120039, by the NSFC 61931024 & 81922046, by helixon biotechnology company Fund and CCF-Tencent Open Fund.

References

1. Akbari, M., et al.: Polyp segmentation in colonoscopy images using fully convolutional network. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 69–72 (2018)
2. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilarino, F.: WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. *Comput. Med. Imaging Graph.* **43**, 99–111 (2015)
3. Bernal, J., Sánchez, J., Vilarino, F.: Towards automatic polyp detection with a polyp appearance model. *Pattern Recogn.* **45**(9), 3166–3182 (2012)
4. Brandao, P., et al.: Fully convolutional neural networks for polyp segmentation in colonoscopy. In: *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, p. 101340F (2017)
5. Cheng, M., Kong, Z., Song, G., Tian, Y., Liang, Y., Chen, J.: Learnable oriented-derivative network for polyp segmentation. In: de Bruijne, M., et al. (eds.) *MICCAI 2021*. LNCS, vol. 12901, pp. 720–730. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_68
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](https://arxiv.org/abs/2108.06932) (2021)
7. Fan, D.-P., et al.: PraNet: parallel reverse attention network for polyp segmentation. In: Martel, A.L., et al. (eds.) *MICCAI 2020*. LNCS, vol. 12266, pp. 263–273. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_26
8. Fang, Y., Chen, C., Yuan, Y., Tong, K.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: Shen, D., et al. (eds.) *MICCAI 2019*. LNCS, vol. 11764, pp. 302–310. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_34
9. Gao, S., Cheng, M., Zhao, K., Zhang, X., Yang, M., Torr, P.H.S.: Res2net: a new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(2), 652–662 (2021)

10. Jha, D., et al.: Kvasir-SEG: a segmented polyp dataset. In: Ro, Y.M., et al. (eds.) MMM 2020. LNCS, vol. 11962, pp. 451–462. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-37734-2_37
11. Jha, D., et al.: Resunet++: an advanced architecture for medical image segmentation. In: 2019 IEEE International Symposium on Multimedia (ISM), pp. 225–2255. IEEE (2019)
12. Ji, G.-P., et al.: Progressively normalized self-attention network for video polyp segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 142–152. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_14
13. 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)
14. Ma, Y., Chen, X., Cheng, K., Li, Y., Sun, B.: LDPolypVideo benchmark: a large-scale colonoscopy video dataset of diverse polyps. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 387–396. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_37
15. Murugesan, B., Sarveswaran, K., Shankaranarayana, S.M., Ram, K., Joseph, J., Sivaprakasam, M.: Psi-Net: shape and boundary aware joint multi-task deep network for medical image segmentation. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7223–7226 (2019)
16. Nguyen, T.-C., Nguyen, T.-P., Diep, G.-H., Tran-Dinh, A.-H., Nguyen, T.V., Tran, M.-T.: CCBANet: cascading context and balancing attention for polyp segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 633–643. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_60
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Shen, Y., Jia, X., Meng, M.Q.-H.: HRENet: a hard region enhancement network for polyp segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 559–568. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_53
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. *Int. J. Comput. Assist. Radiol. Surg.* **9**(2), 283–293 (2014). <https://doi.org/10.1007/s11548-013-0926-3>
20. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Trans. Med. Imaging* **35**(2), 630–644 (2015)
21. Vázquez, D., et al.: A benchmark for endoluminal scene segmentation of colonoscopy images. *J. Healthc. Eng.* **2017**, 1–9 (2017)
22. Wang, W., et al.: PVTv2: improved baselines with pyramid vision transformer. *Comput. Visual Media* **8**(3), 1–10 (2022). <https://doi.org/10.1007/s41095-022-0274-8>
23. Wei, J., Hu, Y., Zhang, R., Li, Z., Zhou, S.K., Cui, S.: Shallow attention network for polyp segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 699–708. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_66

24. Zhang, R., Li, G., Li, Z., Cui, S., Qian, D., Yu, Y.: Adaptive context selection for polyp segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 253–262. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_25
25. Zhang, Y., Liu, H., Hu, Q.: TransFuse: fusing transformers and CNNs for medical image segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 14–24. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_2
26. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **15**(5), 749–753 (2018)
27. Zhao, X., Zhang, L., Lu, H.: Automatic polyp segmentation via multi-scale subtraction network. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 120–130. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_12
28. Zhou, S., et al.: A review of deep learning in medical imaging: image traits, technology trends, case studies with progress highlights, and future promises. *Proc. IEEE* **109**(5), 820–838 (2020)
29. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1