WeakPolyp: You only Look Bounding Box for Polyp Segmentation

Jun Wei[1,2], Yiwen Hu[1,2,5], Shuguang Cui[1,2], S. Kevin Zhou[3,4], and Zhen Li[1,2(✉)]

[1] FNii, CUHK-Shenzhen, Shenzhen, China
junwei@link.cuhk.edu.cn, lizhen@cuhk.edu.cn
[2] SSE, CUHK-Shenzhen, Shenzhen, China
[3] School of Biomedical Engineering and Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, China
[4] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[5] South China Hospital, Shenzhen University, Shenzhen, China

Abstract. Limited by expensive pixel-level labels, polyp segmentation models are plagued by data shortage and suffer from impaired generalization. In contrast, polyp bounding box annotations are much cheaper and more accessible. Thus, to reduce labeling cost, we propose to learn a weakly supervised polyp segmentation model (*i.e.,*WeakPolyp) completely based on bounding box annotations. However, coarse bounding boxes contain too much noise. To avoid interference, we introduce the mask-to-box (M2B) transformation. By supervising the outer box mask of the prediction instead of the prediction itself, M2B greatly mitigates the mismatch between the coarse label and the precise prediction. But, M2B only provides sparse supervision, leading to non-unique predictions. Therefore, we further propose a scale consistency (SC) loss for dense supervision. By explicitly aligning predictions across the same image at different scales, the SC loss largely reduces the variation of predictions. Note that our WeakPolyp is a plug-and-play model, which can be easily ported to other appealing backbones. Besides, the proposed modules are only used during training, bringing no computation cost to inference. Extensive experiments demonstrate the effectiveness of our proposed WeakPolyp, which surprisingly achieves a comparable performance with a fully supervised model, requiring no mask annotations at all. Codes are available at https://github.com/weijun88/WeakPolyp.

Keywords: Polyp segmentation · Weak Supervision · Colorectal cancer

1 Introduction

Colorectal Cancer (CRC) has become a major threat to health worldwide. Since most CRCs originate from colorectal polyps, early screening for polyps is necessary. Given its significance, automatic polyp segmentation models [5,8,16,18]

J. Wei and Y. Hu—Equal contributions.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14222, pp. 757–766, 2023.
https://doi.org/10.1007/978-3-031-43898-1_72

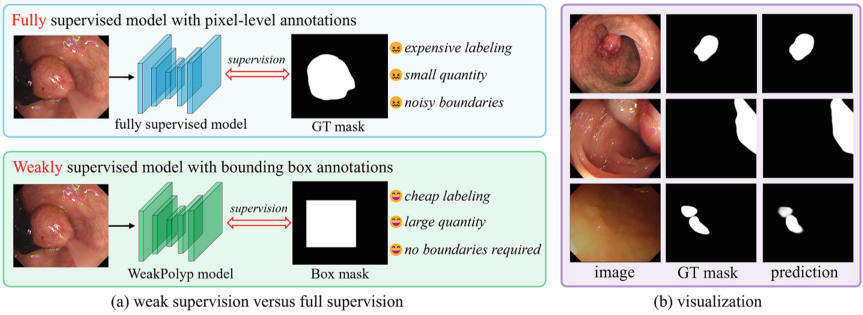

Fig. 1. (a) Comparison between the fully supervised model and our proposed WeakPolyp using box mask only. (b) Visualization of prediction from WeakPolyp.

have been designed to aid in screening. For example, ACSNet [21], HRENet [14], LDNet [20] and CCBANet [11] propose to use convolutional neural networks to extract multi-scale contexts for robust predictions. LODNet [2], PraNet [5], and MSNet [23] aim to improve the model's discrimination of polyp boundaries. SANet [19] eliminates the distribution gap between the training set and the testing set, thus improving the model generalization. Recently, TGANet [15] introduces text embeddings to enhance the model's discrimination. Furthermore, Transfuse [22], PPFormer [1], and Polyp-Pvt [3] introduce the Transformer [4] backbone to extract global contexts, achieving a significant performance gain.

All above models are fully supervised and require pixel-level annotations. However, pixel-by-pixel labeling is time-consuming and expensive, which hampers practical clinical usage. Besides, many polyps do not have well-defined boundaries. Pixel-level labeling inevitably introduces subjective noise. To address the above limitations, a generalized polyp segmentation model is urgently needed. In this paper, we achieve this goal by a weakly supervised polyp segmentation model (named **WeakPolyp**) that only uses coarse bounding box annotations. Figure 1(a) shows the differences between our WeakPolyp and fully supervised models. Compared with fully supervised ones, WeakPolyp requires only a bounding box for each polyp, thus dramatically reducing the labeling cost. More meaningfully, WeakPolyp can take existing large-scale polyp detection datasets to assist the polyp segmentation task. Finally, WeakPolyp does not require the labeling for polyp boundaries, avoiding the subjective noise at source. All these advantages make WeakPolyp more clinically practical.

However, bounding box annotations are much coarser than pixel-level ones, which can not describe the shape of polyps. Simply adopting these box annotations as supervision introduces too much background noise, thereby leading to suboptimal models. As a solution, BoxPolyp [18] only supervises the pixels with high certainty. However, it requires a fully supervised model to predict the uncertainty map. Unlike BoxPolyp, our WeakPolyp completely follows the weakly supervised form that requires no additional models or annotations. Surprisingly, just by redesigning the supervision loss without any changes to the

model structure, WeakPolyp achieves comparable performance to its fully supervised counterpart. Figure 1(b) visualizes some predicted results by WeakPolyp.

WeakPolyp is mainly enabled by two novel components: mask-to-box (M2B) transformation and scale consistency (SC) loss. In practice, M2B is applied to transform the predicted mask into a box-like mask by projection and back-projection. Then, this transformed mask is supervised by the bounding box annotation. This indirect supervision avoids the misleading of box-shape bias of annotations. However, many regions in the predicted mask are lost in the projection and therefore get no supervision. To fully explore these regions, we propose the SC loss to provide a pixel-level self-supervision while requiring no annotations at all. Specifically, the SC loss explicitly reduces the distance between predictions of the same image at different scales. By forcing feature alignment, it inhibits the excessive diversity of predictions, thus improving the model generalization.

In summary, our contributions are three-fold: (1) We build the WeakPolyp model completely based on bounding box annotations, which largely reduces the labeling cost and achieves a comparable performance to full supervision. (2) We propose the M2B transformation to mitigate the mismatch between the prediction and the supervision, and design the SC loss to improve the robustness of the model against the variability of the predictions. (3) Our proposed WeakPolyp is a plug-and-play option, which can boost the performances of polyp segmentation models under different backbones.

2 Method

Model Components. Fig. 2 depicts the components of WeakPolyp, including the segmentation phase and the supervision phase. For the segmentation phase, we adopt Res2Net [6] as the backbone. For input image $I \in R^{H \times W}$, Res2Net extracts four scales of features $\{f_i | i = 1, ..., 4\}$ with the resolutions $[\frac{H}{2^{i+1}}, \frac{W}{2^{i+1}}]$. Considering the computational cost, only f_2, f_3 and f_4 are utilized. To fuse them, we first apply a 1×1 convolutional layer to unify the channels of f_2, f_3, f_4 and then use the bilinear upsampling to unify their resolutions. After being transformed to the same size, f_2, f_3, f_4 are added together and fed into one 1×1 convolutional layer for final prediction. Instead of the segmentation phase, our contributions primarily lie in the supervision phase, including mask-to-box (M2B) transformation and scale consistency (SC) loss. Notably, both M2B and SC are independent of the specific model structure.

Model Pipeline. For each input image I, we first resize it into two different scales: $I_1 \in R^{s_1 \times s_1}$ and $I_2 \in R^{s_2 \times s_2}$. Then, I_1 and I_2 are sent to the segmentation model and get two predicted masks P_1 and P_2, both of which have been resized to the same size. Next, an SC loss is proposed to reduce the distance between P_1 and P_2, which helps suppress the variation of the prediction. Finally, to fit the bounding box annotations (B), P_1 and P_2 are sent to M2B and converted into box-like masks T_1 and T_2. With T_1/T_2 and B, we calculate the binary cross entropy (BCE) loss and Dice loss, without worrying about noise interference.

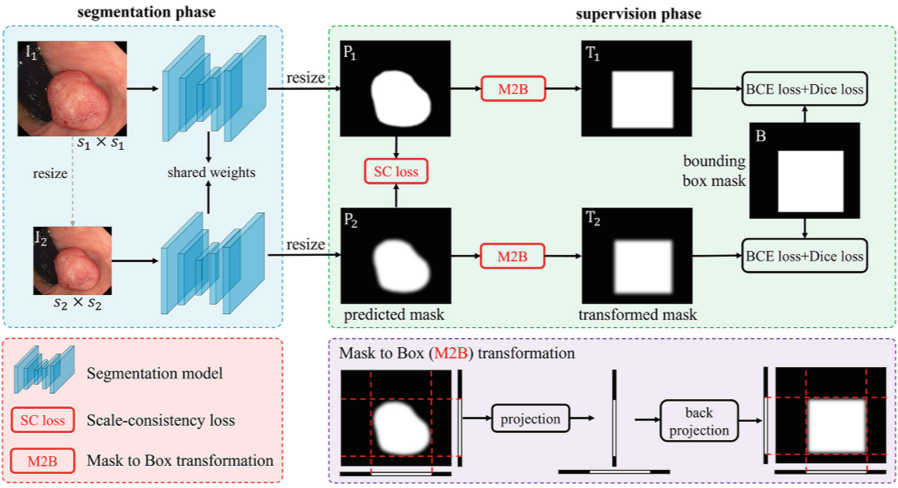

Fig. 2. The framework of our proposed WeakPolyp model, which consists of the segmentation phase and the supervision phase. The segmentation phase predicts the polyp mask for each input firstly, and the supervision phase uses the coarse box annotation to guide previous predicted mask. Note that our contributions mainly lie in the supervision phase, where the proposed M2B transformation converts the predicted mask into a box mask to accommodate the bounding box annotation. Besides, another proposed SC loss is introduced to provide dense supervision from multi-scales, which improves the consistency of predictions.

2.1 Mask-to-Box (M2B) Transformation

One naive method to achieve the weakly supervised polyp segmentation is to use the bounding box annotation B to supervise the predicted mask P_1/P_2. Unfortunately, models trained in this way show poor generalization. Because there is a strong box-shape bias in B. Training with this bias, the model is forced to predict the box-shape mask, unable to maintain the polyp's contours. To solve this, we innovatively use B to supervise the bounding box mask (*i.e.*,T_1/T_2) of P_1/P_2, rather than P_1/P_2 itself. This indirect supervision separates P_1/P_2 from B so that P_1/P_2 is not affected by the shape bias of B while obtaining the position and extent of polyps. But how to implement the transformation from P_1/P_2 to T_1/T_2? We design the M2B module, which consists of two steps: projection and back-projection, as shown in Fig. 2.

Projection. As shown in Eq. 1, given a predicted mask $P \in [0,1]^{H\times W}$, we project it horizontally and vertically into two vectors $P_w \in [0,1]^{1\times W}$ and $P_h \in [0,1]^{H\times 1}$. In this projection, instead of using mean pooling, we use max pooling to pick the maximum value for each row/column in P. Because max pooling can completely remove the shape information of the polyp. After projection, only the position and scope of the polyp are stored in P_w and P_h.

$$P_w = \max(P, \text{axis} = 0) \in [0,1]^{1\times W}, \quad P_h = \max(P, \text{axis} = 1) \in [0,1]^{H\times 1} \quad (1)$$

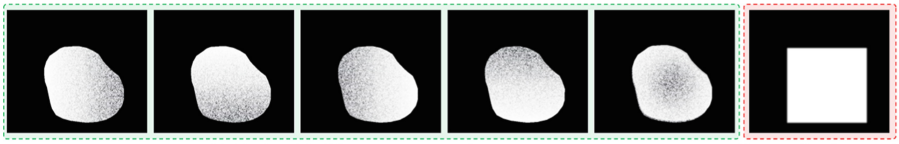

Fig. 3. Different predictions may correspond to the same bounding box mask.

Back-projection. Based on P_w and P_h, we construct the bounding box mask of the polyp by back-projection. As shown in Eq. 2, P_w and P_h are first repeated into $P_w^{'}$ and $P_h^{'}$ with the same size as P. Then, we element-wisely take the minimum of $P_w^{'}$ and $P_h^{'}$ to achieve the bounding box mask T. As shown in Fig. 2, T no longer contains the contours of the polyp.

$$\begin{aligned} P_w^{'} &= \text{repeat}(P_w, H, \text{axis} = 0) \in [0,1]^{H \times W} \\ P_h^{'} &= \text{repeat}(P_h, W, \text{axis} = 1) \in [0,1]^{H \times W} \\ T &= \min(P_w^{'}, P_h^{'}) \in [0,1]^{H \times W} \end{aligned} \tag{2}$$

Supervision. By M2B, P_1 and P_2 are transformed into T_1 and T_2, respectively. Because both T_1/T_2 and B are box-like masks, we directly calculate the supervision loss between them without worrying about the misguidance of box-shape bias. Specifically, we follow [5,19] to adopt BCE loss $\mathcal{L}_{BCE}$ and Dice loss $\mathcal{L}_{Dice}$ for model supervision, as shown in Eq. 3.

$$\mathcal{L}_{Sum} = \frac{\mathcal{L}_{BCE}(T_1, B) + \mathcal{L}_{BCE}(T_2, B)}{2} + \frac{\mathcal{L}_{Dice}(T_1, B) + \mathcal{L}_{Dice}(T_2, B)}{2} \tag{3}$$

Priority. By simple transformation, M2B turns the noisy supervision into a noise-free one, so that the predicted mask is able to preserve the contours of the polyp. Notably, M2B is differentiable, which can be easily implemented with PyTorch and plugged into the model to participate in gradient backpropagation.

2.2 Scale Consistency (SC) Loss

In M2B, most pixels in P are ignored in the projection, thus only a few pixels with high response values are involved in the supervision loss. This sparse supervision may lead to non-unique predictions. As shown in Fig. 3, after M2B projection, five predicted masks with different response values can be transformed into the same bounding box mask. Therefore, we consider introducing the SC loss to achieve dense supervision without annotations, which reduces the degree of freedom of predictions.

Method. As shown in Fig. 2, due to the non-uniqueness of the prediction and the scale difference between I_1 and I_2, P_1 and P_2 differ in response values. But

Table 1. Quantitative comparison between different baselines and our WeakPolyp, involving two datasets (SUN-SEG and POLYP-SEG) and two backbones (Res2Net-50 [6] and PVTv2-B2 [17]). The **gt** row is the performance upper bound. The **box** row is the performance lower bound. **'Bac.'** means backbone. **'Sup.'** means supervision. The highest and second-highest scores are marked in red and blue, respectively

Bac.	Sup.	SUN-SEG						POLYP-SEG			
		Easy Testing		Hard Testing		Training		Testing		Training	
		Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
Res.	gt	.772	.693	.798	.716	.931	.876	.761	.684	.936	.884
	grabcut	.595	.514	.617	.530	.706	.608	.660	.579	.778	.687
	box	.715	.601	.718	.599	.806	.685	.686	.566	.804	.683
	Ours	.792	.715	.807	.727	.899	.826	.760	.680	.909	.842
Pvt.	gt	.851	.780	.858	.784	.932	.878	.793	.715	.936	.883
	grabcut	.741	.648	.747	.649	.766	.670	.644	.559	.780	.683
	box	.769	.652	.770	.648	.804	.681	.734	.611	.824	.705
	Ours	.853	.781	.854	.777	.907	.839	.792	.707	.922	.859

P_1 and P_2 come from the same image I_1. They should be exactly the same. Given this, as shown in Eq. 4, we build the dense supervision $\mathcal{L}_{SC}$ by explicitly reducing the distance between P_1 and P_2, where (i, j) is the pixel coordinates. Note that only pixels inside bounding box are involved in $\mathcal{L}_{SC}$ to emphasize more on polyp regions. Despite its simplicity, $\mathcal{L}_{SC}$ brings pixel-level constraints to compensate for the sparsity of $\mathcal{L}_{Sum}$, thus reducing the variety of predictions.

$$\mathcal{L}_{SC} = \frac{\sum_{(i,j)\in box} |P_1^{i,j} - P_2^{i,j}|}{\sum_{(i,j)\in box} 1} \tag{4}$$

2.3 Total Loss

As shown in Eq. 5, combining $\mathcal{L}_{Sum}$ and $\mathcal{L}_{SC}$ together, we get WeakPolyp model. Note that WeakPolyp simply replaces the supervision loss without making any changes to the model structure. Therefore, it is general and can be ported to other models. Besides, $\mathcal{L}_{Sum}$ and $\mathcal{L}_{SC}$ are only used during training. In inference, they will be removed, thus having no effect on the speed of the model.

$$\mathcal{L}_{Total} = \mathcal{L}_{Sum} + \mathcal{L}_{SC} \tag{5}$$

3 Experiments

Datasets. Two large polyp datasets are adopted to evaluate the model performance, including SUN-SEG [9] and POLYP-SEG. SUN-SEG originates

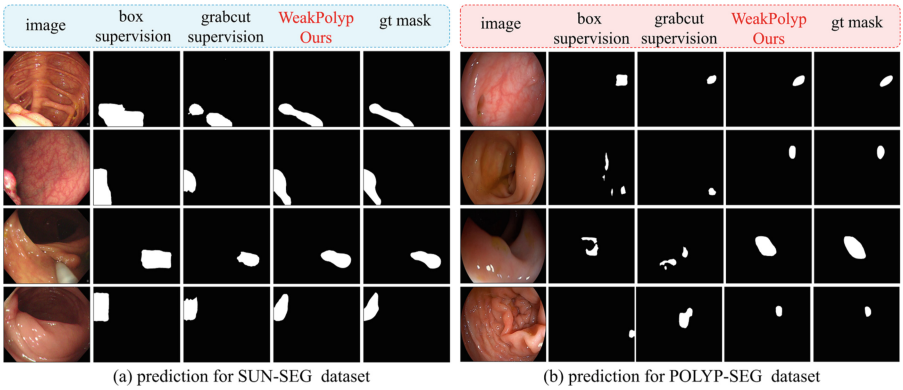

Fig. 4. Visualization comparison between predictions based on different supervisions.

Table 2. Ablation studies on the SUN-SEG testing set under different backbones.

Modules	**Res2Net-50**				**PVTv2-B2**			
	Easy Testing		Hard Testing		Easy Testing		Hard Testing	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
Base	.715	.601	.718	.599	.769	.652	.770	.648
Base+M2B	.748	.654	.768	.673	.822	.738	.822	.735
Base+M2B+SC	.792	.715	.807	.727	.853	.781	.854	.777

from [7,10], which consists of 19,544 training images, 17,070 easy tesing images, and 12,522 hard testing images. POLYP-SEG is our private polyp segmentation dataset, which contains 15,916 training images and 4,040 testing images. Note that, during training, only bounding box annotations are adopted in our WeakPolyp.

Training Settings. WeakPolyp is implemented using PyTorch. 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 an AdamW optimizer. Initial learning rate and batch size are set to 1e-4 and 16, respectively. We train the entire model for 16 epochs.

Quantitative Comparison. Table. 1 compares the model performance under different supervisions, backbones, and datasets. The overall performance order is *gt*>*WeakPolyp*>*box*>*grabcut*. The model supervised by grabcut [13] masks performs the worst, because the foreground and background of polyp images are similar. Grabcut can not well distinguish between them, resulting in poor masks. Our WeakPolyp predictably outperforms the model supervised by box masks because it is not affected by the box-shape bias of the annotations. Interestingly, WeakPolyp even surpasses the fully supervised model on SUN-SEG, which indicates that there is a lot of noise in the pixel-level annota-

Table 3. Performance comparison with previous fully supervised models on SUN-SEG.

Model	Conference	Backbone	Easy Testing		Hard Testing	
			Dice	IoU	Dice	IoU
PraNet [5]	MICCAI 2020	Res2Net-50	.689	.608	.660	.569
2/3D [12]	MICCAI 2020	ResNet-101	.755	.668	.737	.643
SANet [19]	MICCAI 2021	Res2Net-50	.693	.595	.640	.543
PNS+ [9]	MIR 2022	Res2Net-50	.787	.704	.770	.679
Ours		Res2Net-50	**.792**	**.715**	**.807**	**.727**
Ours		PVTv2-B2	**.853**	**.781**	**.854**	**.777**

tions. But WeakPolyp does not require pixel-level annotations so it avoids noise interference.

Visual Comparison. Fig. 4 visualizes some predictions based on different supervisions. Compared with other counterparts, WeakPolyp not only highlights the polyp shapes but also suppresses the background noise. Even for challenging scenarios, WeakPolyp still handles well and generates accurate masks.

Ablation Study. To investigate the importance of each component in WeakPolyp, we evaluate the model on both Res2Net-50 and PVTv2-B2 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 highest performance.

Compared with Fully Supervised Methods. Table. 3 shows our WeakPolyp is even superior to many previous fully supervised methods: PraNet [5], SANet [19], 2/3D [12] and PNS+ [9], which shows the excellent application prospect of weakly supervised learning in the polyp field.

4 Conclusion

Limited by expensive labeling cost, pixel-level annotations are not readily available, which hinders the development of the polyp segmentation field. In this paper, we propose the WeakPolyp model completely based on bounding box annotations. WeakPolyp requires no pixel-level annotations, thus avoiding the interference of subjective noise labels. More importantly, WeakPolyp even achieves a comparable performance to the fully supervised models, showing the great potential of weakly supervised learning in the polyp segmentation field. In future, we will introduce temporal information into weakly supervised polyp segmentation to further reduce the model's dependence on labeling.

Acknowledgement. This work was supported in part by Shenzhen General Program No.JCYJ20220530143600001, by the Basic Research Project No. HZQB-KCZYZ-2021067 of Hetao Shenzhen HK S&T Cooperation Zone, by Shenzhen-Hong Kong Joint Funding No. SGDX20211123112401002, by Shenzhen Outstanding Talents Training Fund, by Guangdong Research Project No. 2017ZT07X152 and No. 2019CX01X104, by the Guangdong Provincial Key Laboratory of Future Networks of Intelligence (Grant No. 2022B1212010001), by the Guangdong Provincial Key Laboratory of Big Data Computing, The Chinese University of Hong Kong, Shenzhen, by the NSFC 61931024&81922046, by zelixir biotechnology company Fund, by Tencent Open Fund.

References

1. Cai, L., et al.: Using guided self-attention with local information for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 629–638 (2022)
2. 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
3. Dong, B., Wang, W., Fan, D.P., Li, J., Fu, H., Shao, L.: Polyp-PVT: polyp segmentation with pyramid vision transformers. arXiv preprint arXiv:2108.06932 (2021)
4. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. In: ICLR (2021)
5. 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
6. 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)
7. Itoh, H., Misawa, M., Mori, Y., Oda, M., Kudo, S.E., Mori, K.: Sun colonoscopy video database. http://amed8k.sundatabase.org/ (2020)
8. Ji, G.P., Chou, Y.C., Fan, D.P., Chen, G., Jha, D., Fu, H., Shao, L.: Progressively normalized self-attention network for video polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2021)
9. Ji, G.P., et al.: Video polyp segmentation: a deep learning perspective. Mach. Intell. Res. (2022). https://doi.org/10.1007/s11633-022-1371-y
10. Misawa, M., et al.: Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). Gastrointest. Endosc. **93**(4), 960–967 (2021)
11. 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
12. Puyal, J.G.B., et al.: Endoscopic polyp segmentation using a hybrid 2D/3D CNN. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 295–305 (2020)
13. Rother, C., Kolmogorov, V., Blake, A.: GrabCut interactive foreground extraction using iterated graph cuts. ACM Trans. Graphics (TOG) **23**(3), 309–314 (2004)

14. 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
15. Tomar, N.K., Jha, D., Bagci, U., Ali, S.: TGANet: text-guided attention for improved polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 151–160 (2022). https://doi.org/10.1007/978-3-031-16437-8_15
16. Wang, J., Huang, Q., Tang, F., Meng, J., Su, J., Song, S.: Stepwise feature fusion: local guides global. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 110–120 (2022). https://doi.org/10.1007/978-3-031-16437-8_11
17. Wang, W., et al.: PVT v2: improved baselines with pyramid vision transformer. Comput. Visual Media **8**(3), 1–10 (2022)
18. Wei, J., Hu, Y., Li, G., Cui, S., Kevin Zhou, S., Li, Z.: BoxPolyp: boost generalized polyp segmentation using extra coarse bounding box annotations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 67–77 (2022). https://doi.org/10.1007/978-3-031-16437-8_7
19. 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
20. 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 (2022). https://doi.org/10.1007/978-3-031-16437-8_10
21. 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
22. Zhang, Y., Liu, H., Hu, Q.: Transfuse: fusing transformers and CNNs for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 14–24 (2021)
23. 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