Joint Polyp Detection and Segmentation with Heterogeneous Endoscopic Data

Wuyang LI[a], Chen YANG[a], Jie LIU[a], Xinyu LIU[a], Xiaoqing GUO[a] and Yixuan YUAN[a]

[a]City University of Hong Kong, 83 Tat Chee Ave, Kowloon Tong, 999077, Hong Kong SAR, China

Abstract

Endoscopy is commonly used for the early diagnosis of colorectal cancer. However, the endoscope images are usually obtained under different illumination conditions, at various sites of the digestive tract, and from multiple medical centers. The collected heterogeneous dataset is a challenging problem in developing automatic and accurate segmentation and detection models. To address these issues, we propose comprehensive polyp detection and segmentation in endoscopic scenarios with novel insights and strategies. For the detection task, we perform joint optimization of classification and regression with adaptive training sample selection strategies in order to deal with the heterogeneous problem. Our detection model achieves 1st place in both first and second rounds of EndoCV 2021 polyp detection challenge. Specifically, the proposed detection framework achieves full-scores (1.0) on AP_{large} and AP_{middle} in the $1st$ round, and 0.8986 ± 0.1920 of score-d on the $2nd$ round. For the segmentation task, we employ HRNet as our backbone and propose a low-rank module to enhance the generalization ability across multiple heterogeneous datasets. Our segmentation model achieves 0.7771 ± 0.0695 score and ranked 4th place in EndoCV 2021 polyp segmentation challenge.

1. Introduction

Colorectal cancer (CRC) is the second common cause of cancer-related deaths in the United States, with 53,200 estimated deaths in 2020. Fortunately, if an adenomatous polyp is detected and removed at its early stage, the deaths caused by CRC can be significantly reduced, and the survival rate is as high as 90%. Endoscopy is a commonly utilized clinical process to identify adenomatous polyps [1]. This process is usually performed manually by the clinician, which may suffer from human error and missed diagnosis of the polyp. Hence, there is a high demand for automatic polyp detection and segmentation models with satisfactory accuracy to facilitate the endoscopy procedures. Even though many methods [2, 3, 4, 5, 6] have been built for automatic detection and segmentation of polyps, existing models are mainly trained with homogeneous data collected from unique medical centers, and learning with highly heterogeneous dataset remains an open problem.

Polyp Detection Task. Most of the existing works [3, 2, 7] about polyp detection tend to perform model ensemble and blindly increase the scale of neural networks for heterogeneous datasets. However, this will lead to two potential inconsistencies, (1) Optimization Inconsistency (OI): The inconsistent optimization targets of classification and regression, and (2) Data Inconsistency (DI): The inconsistent standards of colonoscopy polyp annotations. To handle these

3rd International Workshop and Challenge on Computer Vision in Endoscopy (EndoCV2021) in conjunction with the 18th IEEE International Symposium on Biomedical Imaging ISBI2021, April 13th, 2021, Nice, France

Corresponding author: yxyuan.ee@cityu.edu.hk (Y. YUAN)

CEUR Workshop Proceedings (CEUR-WS.org)

two problems, we utilize GFL v2 [8] to jointly optimize classification and regression and use the regression offset distribution to relieve the influence of ambiguous annotations. To further improve the generalization ability of neural networks, we utilize Adaptive Training Sample Selection (ATSS) [9] strategy to select high-quality anchors with diverse spatial distributions.
Polyp Segmentation Task. Deep convolutional neural networks have achieved impressive progress in polyp segmentation task [6]. Most of the existing methods utilize existing networks, such as VGGNet, Unet, and Dilated ResNet, as the feature extractor. These models gradually reduce the feature resolution through convolution layers and pooling layers and recover the raw resolution through interpolation and convolution operation. This strategy will lead to intermediate low-resolution feature representations and lose a lot of critical detailed information, which is not an optimal solution for polyp semantic segmentation, i.e., pixel-wise classification. Thus, we employ HRNet [10] as our backbone for polyp segmentation in EndoCV2021 challenge[1]. Furthermore, considering this dataset heterogeneous property, we propose a low-rank module to enhance the generalization.

2. Proposed Methods

2.1. Method Details for Polyp Detection

As illustrated in Figure 1. Given an image, we first adopt ResNeXt-101-DCN with Feature Pyramid Network (FPN) [11] for feature extraction (§2.1.1). To relieve the aforementioned inconsistency (§2.1.2) , we perform joint optimization of classification and regression to bridge OI and use the regression offset distribution to relieve DI [8] in detection heads. To further improve the generalization ability (§2.1.3) of detection framework, ATSS [9] strategy is introduced to select high-quality training samples with diverse spatial distributions.

2.1.1. Feature Extraction

Due to the heterogeneous samples obtained from different medical centers, we found AP improves with the increase of model scale without over-fitting. Hence, we adopted ResNeXt101-DCN as our feature extractor. Specifically, we apply deformable convolutions from stage 3 to 5 of ResNeXt-101 and frozen parameters in stage 1. Besides, FPN is adopted in the backbone for multi-scale feature fusion.

2.1.2. Solutions for the Two Inconsistency

OI and DI are the major limitations for the performance of polyp detection in EndoCV 2021 challenges. For OI, classification and regression are optimized in two separated branches with inconsistent supervisions, which brings about the inconsistency during performing Non-Maximum Suppression (NMS) in inference. For DI, we found a large variance of bounding box coordinates caused by inconsistent standards of manual annotations. As shown in Figure 2 Left, the annotations on two sequential video frames should be similar but are different, obviously. Therefore, these ambiguous boxes will confuse neural networks and affect the detection performance significantly, especially in the case of small-scale endoscopic datasets.

[1] https://endocv2021.grand-challenge.org/EndoCV2021/

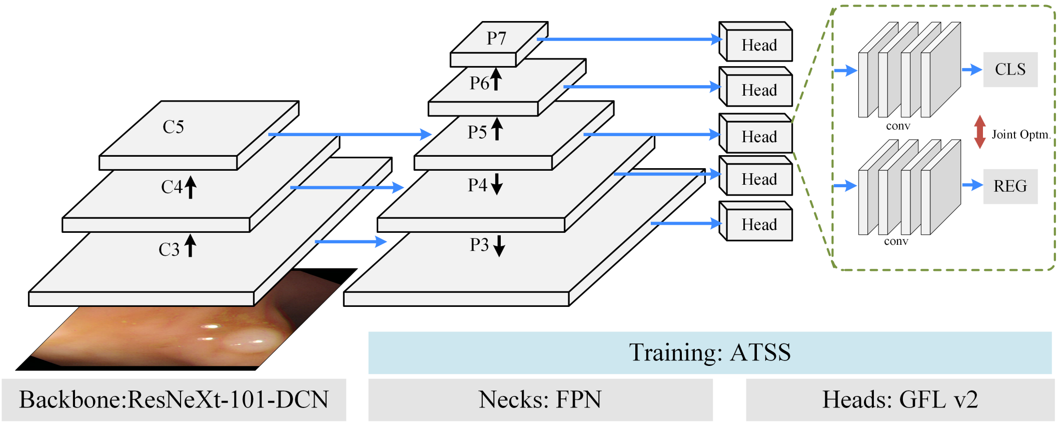

Figure 1: The overview of the detection framework. Specifically, ResNeXt-101-DCN with FPN necks are used as our feature extractors. Besides, GFL v2 [8] and ATSS [9] are these two key components in our model for joint optimizing classification and regression with diverse training samples.

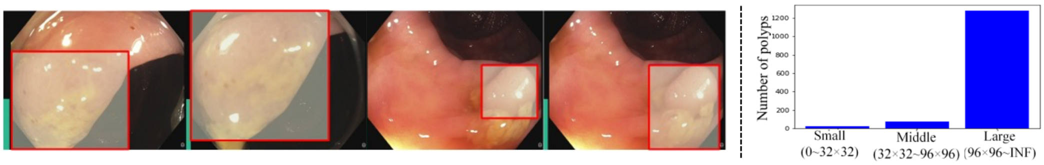

Figure 2: Left: Illustration of the phenomenon of DI. Ambiguous annotations may cause a severe performance drop in polyp detection; **Right:** The biased scale distribution of EndoCV 2021 polyp detection dataset. The area of most polyp instances is larger than size 96×96 pixels, which is defined as a large object in MS COCO matrix.

Most existing works tend to relieve OI by introducing localization quality estimation strategies, which are performed in the regression branch for comprehensive representations of detection results, such as the centerness in FCOS [12] and the Intersection of Union (IoU) scores in [13]. However, these methods may fail and lead to severe degeneration of localization estimation when the ground-truth of bounding boxes is ambiguous. To jointly handle these two problems, we adopt the well-designed strategies in [8] to learn the distributions of bounding boxes and use the statistics of regression offsets for the localization quality estimation. Then, Generalized Focal Loss [14] is used for the joint optimization of classification and regression, which eases the inconsistency skillfully and results in a significant improvement of detection accuracy.

2.1.3. Improving Generalization Ability

In addition to augmenting training data offline, improving the diversity of training samples in each image has great potential to promote the generalization ability of neural networks. Therefore, the allocation of training samples plays a decisive role in the model optimization for polyp detection, especially in the case of insufficient and high-variance EndoCV 2021 endoscopic

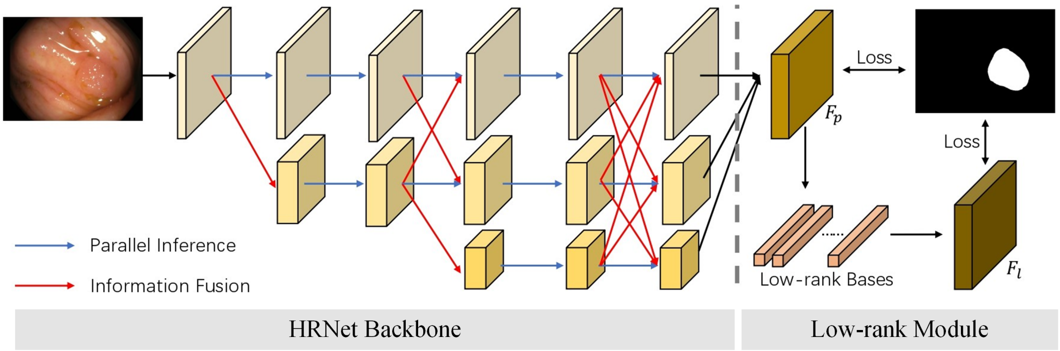

Figure 3: The overview of the segmentation model with HRNet backbone [10] and the proposed low-rank module.

datasets. Previous works tend to use hand-craft IoU thresholds (Faster RCNN [2], RetinaNet [15], etc.) and spatial constraints (FCOS [12], etc.) to select training samples, which may bring about the biased optimization for object detectors. Besides, in most FPN-based paradigms, training samples are allocated to different levels of feature pyramid layers according to their scales manually, leading to optimization difficulties. To relieve these problems, we utilize the novel ATSS [9] strategy to select high-quality training samples adaptively, which fully utilizes the statistics information of anchors. Nevertheless, the gap between endoscopic and natural scenes is still obvious and significant, which inspires us to adjust the number of selected positive samples to fit the endoscopic scenarios and improve the generalization ability of neural networks. Some key properties of endoscopic scenes for polyp detection are concluded as followed,

- Non-overlapping: the overlapping of polyps is extremely rare.
- Large-scale: the scales of polyps tend to be larger than 96×96 pixels shown in Figure 2 Right, which are defined as large objects using MS COCO evaluation matrix.
- Sparsity: the polyps in each image tend to be sparse.
- High-variance: high inter-sample variance is caused by different data sources.

To apply the strategy to the endoscopic scenarios and improve the generalization ability for endoscopic polyp detection, we enlarge the number of samples for each instance from k=9 to k=13 to increase the diversity of data in an online manner. The increasing number of samples won't generate many intolerable low-quality allocations thanks to the properties of sparsity, non-overlapping, and large-scale, but achieves the instance-level augmentation with feature representations in turn.

2.2. Method Details for Polyp Segmentation

We utilize HRNet as our backbone for polyp segmentation (§2.2.1), and then propose a low-rank module (§2.2.2) to enhance model generalization. Cross entropy and dice loss are utilized to optimize the whole model (§2.2.3). The whole framework is shown in Figure 3.

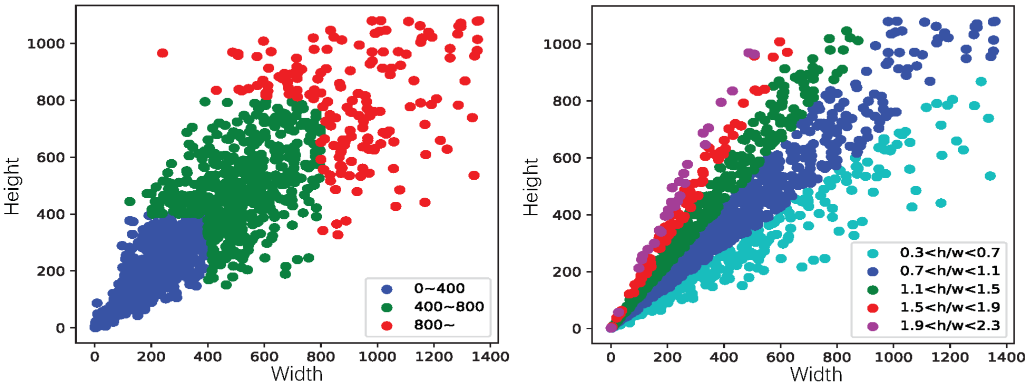

Figure 4: Illustration of the instance Resolution Distribution. **Left:** Instance scale statistic. **Right:** Instance ratio of height to width.

2.2.1. Backbone Selection

To choose a suitable solution for the polyp segmentation in EndoCV 2021 challenge, we conduct a detailed analysis of the dataset at the instance level. We regard the size in the range of 0-400 as small instances, 400-800 as middle instances, and above 800 as large instances. According to the polyp size statistics in Figure 4 Left, we find that small polyps are the majority ones. Figure 4Right analyzes the ratio of high to width for the polyps, showing that the ratio distributes widely. When feature representations become low-resolution inner the backbone, it's hard to recognize the small instances, especially with biased ratios. Considering these, we adopt HRNet [10] as our backbone network, which can maintain the high-resolution representations among the whole process. Two main components of this backbone are parallel inference and information fusion, as shown in Figure 3.

Parallel Inference The main idea of parallel inference is to perform convolution operations in three different resolutions, i.e., the blue lines in Figure 3. In this way, the high-resolution branch can keep the detailed information, and the low-resolution branch can grasp the semantic information over a wide range of regions.

Information fusion However, the high-resolution branch has difficulty in learning large patterns such as a pedunculated polyp, and the low-resolution branch can't learn detailed information such as hemorrhage polyp. Information fusion is utilized to address this problem, i.e., the red lines in Figure 3. Before each cross-branch information fusion, 2× down-sampling or 2× up-sampling is performed to ensure the strict resolution match. Then, concatenation is adopted to fuse multi-level information in each branch. At last, we resize the feature maps to the raw resolution and concatenate them, followed by 1×1 convolution to generate F_p.

2.2.2. Low-rank Module

In order to further eliminate noisy information in F_p and enhance model generalization, we propose a low-rank module to project the feature map F_p into a set of low-rank bases and reconstruct a low-rank feature map F_l to predict the final result. Specifically, we reshape the feature map F_p to $N \times C$, where $N = WH$ is the number of pixels and C is the channel number. To compress the semantic information, F_p is embedded into a low-dimension space of D free degree using an affine function $\phi(\cdot)$ with learned parameters, followed by a softmax function, $F_a = softmax(\phi(F_p)) \in R^{N \times D}$. Then, low-rank bases are calculated by $b = \frac{1}{\mathcal{N}} F_p^T F_a \in R^{C \times D}$, where $\mathcal{N}$ represents the normalized coefficient and D is a predefined number. Each base represents the concentrated semantic information of each degree in low-dimension space. In the end, the low-rank feature map F_l is reconstruct by $F_l = F_a b^T$. In general, the rank of F_p is $min\{N, C\}$, empirically 1k, and the rank of F_l is less than D. With this low-rank module, the feature map in the high dimensional space is redistributed to a low dimensional manifold, which removes unnecessary information and enhances the model generalization.

2.2.3. Optimization objective

We employ two supervised losses, cross entropy and dice loss, to supervise the learning of F_p and F_l. Cross-entropy loss is formulated as $L_{CE} = \sum_{i=1}^{N} y_i \log p_i$, where N is the pixel number of the whole image, y_i is the one-hot ground truth and p_i is the prediction probability. Dice-loss measures the overlap between the predicted region and ground truth, which is defined as $L_D = 1 - 2\frac{\sum_{i=1}^{N} y_i p_i}{\sum_{l \in L} \sum_{i=1}^{N} (y_i + p_i)}$.

3. Experiments

3.1. Experiments for Polyp Detection

Experiment Setting. To perform model selection and method verification, we conduct extensive experiments on the released training data [16] with 80% for training (1062 images) and 20% for validation (266 images) before offline data augmentation, while the final model is trained using all data. Pretrained on ImageNet, we further train our models with SGD using 2 NVIDIA V100 GPUs with a batch-size 8 for 24 epochs (2× training schedule). The learning rate is set 0.01 and decreased by 10 at epoch 16 and 22. The Average Precision (AP) is calculated with linear IoU thresholds from .5 to .95 with 0.05 interval. For the final model used in EndoCV 2021 polyp detection challenge, multi-scale training is performed by randomly re-scaling images from (1333, 480) to (1333, 960) with 120 intervals. We not only use the common online data augmentation strategies, e.g., randomly cropping and flipping with 50% probability, but also use some offline augmentation methods, such as random rotation (75% probability to rotate arbitrary angle), gamma contrast ($\gamma \in [0.5, 2.0]$), and brightness transformation with a random value from -10 to 10, etc. The NMS threshold is set 0.01, and the score-threshold is set 0.3 for lower AP and *dev* or 0 for higher AP and *dev*, which can be viewed as a trade-off between robustness and accuracy.

Baseline Selection. In some scenes, well-designed one-stage detectors [9, 13, 8] have achieved higher detection performance and shown more potential on inference speed. Instead of rashly choosing two-stage, cascade, or ensemble pipelines, we perform extensive experiments on the baseline selection shown in Table 1. As we expected, one-stage baselines show absolute advantages in polyp detection. This is because *RPN will degenerate into an inefficient single-stage detector when the number of categories is small.* Therefore, we choose one-stage detector

	Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Two-stage	FRCNN† [11]	46.8	68.1	52.6	12.5	28.2	48.7
	Cascade RCNN† [3]	49.3	69.0	55.4	10.4	23.9	51.6
	Double-Head-Ext [17]	48.0	69.2	52.4	12.5	25.2	50.3
	D2Det [7]	48.2	75.3	56.2	5.5	36.0	49.9
One-stage	RetinaNet† [15]	46.6	68.1	51.6	9.2	26.6	48.1
	FCOS† [12]	48.1	74.7	50.6	7.6	27.6	50.1
	PAA [13]	51.8	77.5	56.1	12.3	25.2	54.5
	ATSS [9]	49.9	72.1	56.1	10.4	28.4	52.0

Table 1

Comparison results (%) of two-stage and one-stage detection frameworks on validation set using ResNet-50 backbone with 2× training schedule. † represents the optional baseline models.

Backbone	Necks	DCN	MS_{train}	MS_{test}	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
ResNet-50	FPN				51.0	73.5	56.8	12.5	28.3	53.5
ResNet-50	FPN		✓		53.1	75.8	59.9	14.6	30.6	55.3
ResNet-50	FPN	✓	✓		53.8	76.6	57.8	12.3	27.3	56.0
ResNet-101	FPN	✓	✓		55.0	77.3	62.4	18.0	36.0	58.4
ResNeXt-101	FPN	✓	✓		**56.3**	76.9	**63.5**	12.5	**33.0**	**58.5**
ResNeXt-101	FPN	✓	✓	✓	52.1	73.1	60.9	14.2	33.0	53.1
ResNeXt-101	PAFPN [18]	✓	✓		53.3	76.1	60.6	12.5	32.2	55.2
ResNeXt-101	BiFPN [19]	✓	✓		55.8	77.0	62.2	15.8	32.9	57.2

Table 2

Comparison results (%) of different feature extractors in our model. MS_{train} and MS_{test} indicate multi-scale training and test strategies.

FCOS [12] as our baseline.

Investigation on Feature Extractors. For polyp detection in EndoCV2021 challenge, heavier backbones may not be suitable for small-scale datasets [16] due to the potential over-fitting of neural networks. Fortunately, we find a consistent improvement as increasing the scale of neural networks, as shown in Table 2, which demonstrates the high-variance of data distributions can reduce the possibility of over-fitting. On the contrary, utilizing stronger multi-scale feature fusion methods, e.g., PAFPN [18] and BiFPN [19], doesn't improve the performance due to the biased scale distribution. Besides, performing multi-scale inference leads to significantly AP drops, as demonstrated in Table 2.

Ablation Analysis. As shown in Table 3, we perform ablation analysis on each component using our validation set. Compared with the FCOS baseline, introducing ATSS [9] can achieve 1.8 AP improvement, which demonstrates the effectiveness of the sample selection strategy. After relieving the influence of OI and DI, a significant 2.9 AP improvement can be achieved by introducing both ATSS and GFL v2 [8] together with the comparison of our baseline. In addition to achieving state-of-the-art performance on endoscopic polyp detection, our model also has obvious advantages in the inference speed because of the one-stage detection pipeline.

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
FCOS (baseline) [12]	48.1	74.7	50.6	7.6	27.6	50.1
FCOS+ATSS [9]	49.9	72.1	56.1	10.4	28.4	52.0
FCOS+ATSS+GFLv2 [8]	**51.0**	**73.5**	**56.8**	**12.5**	**28.3**	**53.5**

Table 3
Ablation study results (%) on validation set using ResNet-50 backbone with 2× training schedule.

Method	Dice (%)	Spe (%)	Sen (%)	Acc (%)	IoU_p (%)	IoU_b (%)	mIoU (%)
UNet++ [4]	66.067	99.339	72.788	96.939	59.552	96.797	78.175
PraNet [5]	61.822	97.444	79.383	95.937	53.300	95.688	74.494
ACSNet [20]	72.708	98.903	77.753	97.463	66.691	97.285	81.988
ACFNet [21]	76.204	**99.369**	76.204	98.084	70.491	97.960	84.225
HRNet [10]	88.351	98.839	93.224	98.496	79.133	98.405	88.769
HRNet+Low-rank	**90.364**	99.007	**96.288**	**98.856**	**82.422**	**98.791**	**90.607**

Table 4
Comparison with state-of-the-art polyp segmentation methods.

3.2. Experiments for Polyp Segmentation

Experiment setting. To evaluate our method on the released training data [16], we first split them into 80% for training (1062 images) and 20% for testing (266 images). Images are resized to 512×512 pixels. We apply augmentation techniques upon images: random flipping and rotation with 50% probability, color shift (brightness, color, sharpness, and contrast), and Gaussian noise $\mathcal{N}$ (0.2, 0.3). At last, we normalize these images into [-1, 1]. The backbone of the segmentation model is HRNetV2 with parameters initialized on ImageNet. We utilize the SGD optimizer with the base learning rate of 0.01, the momentum of 0.9, and the weight decay of 0.0005. All experiments are implemented by the Pytorch framework and trained on four parallel Nvidia GeForce 2080Ti GPUs with a batch-size of 16 for 484 epochs. To evaluate the performance of polyp segmentation, seven common criteria including Dice Score (Dice), Sensitivity (Sen), Specificity (Spe), Accuracy (Acc), IoU of polyp regions (IoU_p), IoU of backgrounds (IoU_b) and Mean IoU (mIoU) are utilized.

Comparison with State-of-the-art Methods. To verify the effectiveness of our method, we perform a comprehensive comparison with state-of-the-art polyp segmentation methods, including UNet++ [4], PraNet [5], ACSNet [20] and ACFNet [21], as shown in Table 4. Specifically, our method achieves the best performance, with dice score of 90.364% and mIoU of 90.607%, demonstrating the superiority of our method over state-of-the-art polyp segmentation methods. In EndoCV 2021 polyp segmentation challenge, our segmentation model achieves 0.7771 ± 0.0695 score and ranked 4th place based on EndoCV metrics that included generalisation deviation scores between test sets [22].

4. Conclusion

Automatic polyp detection and segmentation are challenging due to the collected heterogeneous dataset. We find OI and DI as two major limitations for high-quality polyp detection for the polyp detection task. To handle these issues, we jointly optimize classification and regression to bridge OI and use the regression offset distribution to relieve DI. To further promote the generalization ability of neural networks, we utilize ATSS to improve the diversity of training samples in each image. For the polyp segmentation task, we find the small polyps make up the majority of the dataset. Hence we exploit HRNet as the backbone. To enhance the generalization of the model, we propose the low-rank module. Extensive experiments demonstrate the effectiveness of our methods. In the future, we aim to integrate the detection and segmentation framework for high-quality polyp instance segmentation.

References

[1] F. A. Haggar, R. P. Boushey, Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors, Clinics in colon and rectal surgery 22 (2009) 191.

[2] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: NeurIPS, 2015, pp. 91–99.

[3] Z. Cai, N. Vasconcelos, Cascade r-cnn: Delving into high quality object detection, in: CVPR, 2018, pp. 6154–6162.

[4] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: Redesigning skip connections to exploit multiscale features in image segmentation, IEEE Trans. Med. Imag. (2019).

[5] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Pranet: Parallel reverse attention network for polyp segmentation, in: MICCAI, Springer, 2020, pp. 263–273.

[6] X. Guo, C. Yang, Y. Liu, Y. Yuan, Learn to threshold: Thresholdnet with confidence-guided manifold mixup for polyp segmentation, IEEE Transactions on Medical Imaging (2020).

[7] J. Cao, H. Cholakkal, R. M. Anwer, F. S. Khan, Y. Pang, L. Shao, D2det: Towards high quality object detection and instance segmentation, in: CVPR, 2020, pp. 11485–11494.

[8] X. Li, W. Wang, X. Hu, J. Li, J. Tang, J. Yang, Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection, CVPR (2021).

[9] S. Zhang, C. Chi, Y. Yao, Z. Lei, S. Z. Li, Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection, in: CVPR, 2020, pp. 9759–9768.

[10] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, et al., Deep high-resolution representation learning for visual recognition, IEEE transactions on pattern analysis and machine intelligence (2020).

[11] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: CVPR, 2017, pp. 2117–2125.

[12] Z. Tian, C. Shen, H. Chen, T. He, Fcos: Fully convolutional one-stage object detection, in: ICCV, 2019, pp. 9627–9636.

[13] A. Hu, F. Cotter, N. Mohan, C. Gurau, A. Kendall, Probabilistic future prediction for video scene understanding, in: ECCV, 2020, pp. 767–785.

[14] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, J. Yang, Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection, NeurIPS (2020).

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: ICCV, 2017, pp. 2980–2988.

[16] S. Ali, D. Jha, N. Ghatwary, S. Realdon, R. Cannizzaro, M. A. Riegler, P. Halvorsen, C. Daul, J. Rittscher, O. E. Salem, D. Lamarque, T. de Lange, J. E. East, Polypgen: A multi-center polyp detection and segmentation dataset for generalisability assessment, arXiv (2021).

[17] Y. Wu, Y. Chen, L. Yuan, Z. Liu, L. Wang, H. Li, Y. Fu, Rethinking classification and localization for object detection, in: CVPR, 2020, pp. 10186–10195.

[18] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation, in: CVPR, 2018, pp. 8759–8768.

[19] M. Tan, R. Pang, Q. V. Le, Efficientdet: Scalable and efficient object detection, in: CVPR, 2020, pp. 10781–10790.

[20] R. Zhang, G. Li, Z. Li, S. Cui, D. Qian, Y. Yu, Adaptive context selection for polyp segmentation, in: MICCAI, Springer, 2020, pp. 253–262.

[21] F. Zhang, Y. Chen, Z. Li, Z. Hong, J. Liu, F. Ma, J. Han, E. Ding, Acfnet: Attentional class feature network for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6798–6807.

[22] S. Ali, F. Zhou, B. Braden, A. Bailey, S. Yang, G. Cheng, P. Zhang, X. Li, M. Kayser, R. D. Soberanis-Mukul, S. Albarqouni, X. Wang, C. Wang, S. Watanabe, I. Oksuz, Q. Ning, S. Yang, M. A. Khan, X. W. Gao, S. Realdon, M. Loshchenov, J. A. Schnabel, J. E. East, G. Wagnieres, V. B. Loschenov, E. Grisan, C. Daul, W. Blondel, J. Rittscher, An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy, Scientific Reports 10 (2020) 2748. doi:10.1038/s41598-020-59413-5.