Automatic Polyp Segmentation via Parallel Reverse A□ention Network

Ge-Peng Ji[1,2], Deng-Ping Fan[1,*], Tao Zhou[1], Geng Chen[1], Huazhu Fu[1], Ling Shao[1]
[1]Inception Institute of Arti□cial Intelligence (IIAI), Abu Dhabi, UAE.
[2]School of Computer Science, Wuhan University, Hubei, China.
https://github.com/GewelsJI/MediaEval2020-IIAI-Med

ABSTRACT

In this paper, we present a novel deep neural network, termed **P**arallel **R**everse **A**ttention **Net**work (*PraNet*), for the task of automatic polyp segmentation at MediaEval 2020. Speci□cally, we □rst aggregate the features in high-level layers using a parallel partial decoder (PPD). Based on the combined feature, we then generate a global map as the initial *guidance area* for the following components. In addition, we mine the *boundary cues* using the reverse attention (RA) module, which is able to establish the relationship between areas and boundary cues. Thanks to the recurrent cooperation mechanism between areas and boundaries, our *PraNet* is capable of calibrating misaligned predictions, improving the segmentation accuracy and achieving real-time e□ciency (~**30fps**) on a single NVIDIA GeForce GTX 1080 GPU.

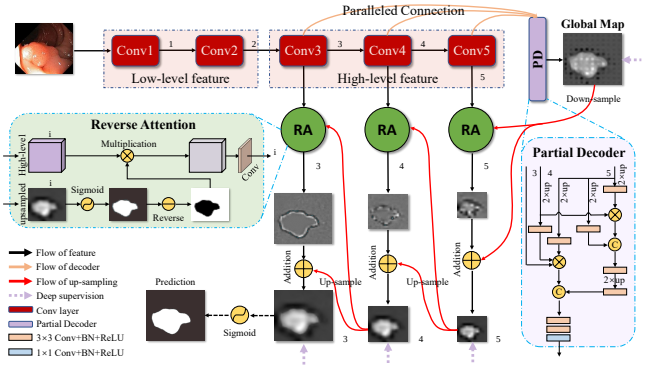

Figure 1: Pipeline of our *PraNet*, which consists of three reverse attention (RA) modules with a parallel partial decoder (PPD) connection. Please refer to § 2 for more details.

1 INTRODUCTION

Aiming at developing computer-aided diagnosis systems for automatic polyp segmentation, and detecting all types of polyps (*i.e.*, irregular polyps, smaller or □at polyps) with high e□ciency and accuracy, Medico Automatic Polyp Segmentation Challenge 2020[1] [10] benchmarks semantic segmentation methods for segmenting polyp regions in colonoscopy images on a publicly available dataset, emphasizing robustness, speed, and generalization. Following the protocols of this challenge, we participate two required sub-tasks including (i) Polyp segmentation task and (ii) Algorithm e□ciency task, more task descriptions refer to the challenge guidelines.

Recent years have witnessed promising progress in addressing the task of automatic polyp segmentation using traditional [12, 16] and deep learning based [1, 2, 7, 13, 20, 21] methods. However, there are three core challenges in this □eld, including (a) the polyps often vary in appearance, *e.g.*, size, color and texture, even if they are of the same type; (b) in colonoscopy images, the boundary between a polyp and its surrounding mucosa is usually blurred and lacks the intense contrast required for segmentation approaches; (c) the practical applications of existing algorithms are hindered by their low performance and e□ciency.

Based on these observations, we develop a real-time and accurate framework, termed **P**arallel **R**everse **A**ttention **Net**work (*PraNet*[2]), for the automatic polyp segmentation task. As can be seen from Fig. 1, *PraNet* utilizes a parallel partial decoder (see § 2.1) to generate a high-level semantic global map and a set of reverse attention modules (see § 2.2) for accurate polyp segmentation from the colonoscopy images. All components and implementation details will be elaborated as follows.

2 APPROACH

2.1 Parallel Partial Decoder (PPD)

Current popular medical image segmentation networks usually rely on a U-Net [15] or a U-Net shaped network (*e.g.*, U-Net++ [26], ResUNet [25], *etc*). These models are essentially encoder-decoder frameworks, which typically aggregate *all* multi-level features extracted from convolutional neural networks (CNNs). As demonstrated by Wu *et al.* [19], compared with high-level features, low-level features demand more computational resources due to their larger spatial resolutions, but contribute less to performance. Motivated by this observation, we propose to aggregate high-level features with a *parallel partial decoder* component. More specifically, for an input polyp image I with size $h \times w$, □ve levels of features $\{\mathbf{f}_i, i = 1, ..., 5\}$ with resolution $[h/2^{i-1}, w/2^{i-1}]$ can be extracted from a Res2Net-based [8] backbone network. Then, we divide $\mathbf{f}_i$ features into low-level features $\{\mathbf{f}_i, i = 1, 2\}$ and high-level features $\{\mathbf{f}_i, i = 3, 4, 5\}$. We introduce the partial decoder $p_3(\cdot)$ [19], a new state-of-the-art (SOTA) decoder component, to aggregate the high-level features with a paralleled connection. As shown in Fig. 1, the partial decoder feature is computed by $\mathbf{PD} = p_3(f_3, f_4, f_5)$, and we can obtain a global map $\mathbf{S}_6$.

[1]https://multimediaeval.github.io/editions/2020/tasks/medico/
*Corresponding Author: Deng-Ping Fan (Email: dengpfan@gmail.com)
Work was done while Ge-Peng Ji was an intern mentored by Deng-Ping Fan.
[2]This work is based on our paper [5] published at MICCAI-2020.

MediaEval'20, December 14-15 2020, Online

2.2 Reverse Attention (RA)

In a clinical setting, doctors first roughly locate the polyp region, and then carefully inspect local tissues to accurately label the polyp. As discussed in § 2.1, our global map $\mathbf{S}_6$ is derived from the deepest CNN layer, which can only capture a relatively rough location of the polyp tissues, without structural details (see Fig. 1). To address this issue, we propose a principle strategy to progressively mine discriminative polyp regions by erasing foreground objects [3, 18]. Instead of aggregating features from all levels like in [9, 22, 23], we propose to adaptively learn the *reverse attention* in three parallel high-level features. In other words, our architecture can sequentially mine complementary regions and details by erasing the existing estimated polyp regions from high-level side-output features, where the existing estimation is up-sampled from the deeper layer.

Specifically, we obtain the output reverse attention features R_k by multiplying (element-wise $\odot$) the high-level side-output feature $\{f_k, i = 3, 4, 5\}$ by a reverse attention weight A_k, as follows:

$$R_k = f_k \odot A_k. \tag{1}$$

The reverse attention weight A_k is de-facto for salient object detection in the computer vision community [3, 24], and can be formulated as:

$$A_k = \ominus(\sigma(\mathcal{P}(S_{k+1}))), \tag{2}$$

where $\mathcal{P}(\cdot)$ denotes an up-sampling operation, $\sigma(\cdot)$ is the Sigmoid function, and $\ominus(\cdot)$ is a reverse operation subtracting the input from matrix E, in which all the elements are 1. Fig. 1 (RA) shows the details of this process. It is worth noting that the erasing strategy driven by the reverse attention can eventually refine the imprecise and coarse estimation into an accurate and complete prediction map.

3 EXPERIMENTS

3.1 Learning Strategies

We use a hybrid loss function in the training process, which is defined as $\mathcal{L} = \mathcal{L}^F_{>*} + \mathcal{L}^F$, where $\mathcal{L}^F_{>*}$ and $\mathcal{L}^F$ represent the weighted Intersection over Union (IoU) loss and binary cross entropy (BCE) loss for the global restriction and local (pixel-level) restriction. The definitions of these losses are the same as in [14, 17] and their effectiveness has been validated in the field of salient object detection. Here, we adopt deep supervision for the three side-outputs (*i.e.*, S_3, S_4, and S_4) and the global map S_6. Each map is up-sampled (*e.g.*, $S_3^{D?}$) to the same size as the ground-truth map G. Thus the total loss for the proposed *PraNet* can be formulated as: $\mathcal{L}_{\mathcal{C}>\mathcal{C}O;} = \mathcal{L}(G, S_6^{D?}) + \sum_{k=3}^{k=5} \mathcal{L}(G, S_k^{D?})$.

3.2 Evaluation Metrics

We employ the metrics widely used in the medical segmentation field, including mean IoU (mIoU or Jaccard index), Dice coefficient, recall, precision, acccuracy and frame per second (FPS) for a comprehensive evaluation.

3.3 Implementation Details and Datasets

We randomly split the whole Kvaris-SEG[3] [11] into a training set (900 images) and validation set (100 images). Note that we do not use any extra data in this challenge. All the inputs of our model are uniformly resized to 352×352 and we augment all the training images using multiple strategies, including random horizontal flipping, rotating, color enhancement and border cropping. Parameters of the Res2Net-50 [8] backbone are initialized from the model pre-trained on ImageNet [4]. Other parameters are initialized using the default PyTorch settings. The Adam algorithm is used to optimize our model, and it is accelerated by an NVIDIA TITAN RTX GPU. We set the initial learning rate is 1e-4 and divide it by 10 every 50 epochs. It takes about 40 minutes to train the model with a mini-batch size of 26 over 100 epochs. Our final prediction map $S_?$ is generated by S_3 after a *Sigmoid* function. The testing dataset consists of 160 polyp images provided by organisers. Our code can be found at https://github.com/GewelsJI/MediaEval2020-IIAI-Med.

[3] https://datasets.simula.no/kvasir-seg/

Table 1: Quantitative results for both the polyp segmentation (task 1) and algorithm efficiency (task 2) on a single NVIDIA GeForce GTX 1080 GPU of Medico Automatic Polyp Segmentation Challenge 2020.

Team Name	Jaccard	DSC	Recall	Precision	Accuracy	F2	FPS
IIAI-Med	0.761	0.839	0.830	0.901	0.960	0.828	29.87

3.4 Results and Analysis

Without any bells and whistles, such as extra training data or a model ensemble, we introduce a new training scheme for addressing the polyp segmentation challenge based on the previous work [5]. For more hyper-parameters and data augmentation settings refer to § 3.3. In the submission phase, we train our model on the Kvasir-SEG dataset [11] and submit the inference results only once. Tab. 1 reports the quantitative results of our approach on sub-task 1, which achieve a very high performance (Precision=0.901 and Accuracy=0.960 on test set). Meantime, *PraNet* also runs at ~30fps on a single NVIDIA GeForce GTX 1080 GPU, demonstrating its simplicity and effectiveness. As a robust, general, and real-time framework, *PraNet* can help facilitate future academic research and computer-aided diagnosis for automatic polyp segmentation.

4 CONCLUSION

Automatic polyp segmentation is a challenging problem because of the diversity in appearance at polyps, complex similar environments, and require high accuracy and inference speed. We have presented a novel architecture, *PraNet*, for automatically segmenting polyps from colonoscopy images. *PraNet* efficiently integrates a cascaded mechanism and a reverse attention module with a parallel connection, which can be trained in an end-to-end manner. Another advantage is that *PraNet* is universal and flexible, meaning that more effective modules can be added to further improve the accuracy. We hope this study will offer the community an opportunity to explore more powerful models for related topics such as lung infection segmentation [6], or even on upstream tasks such as video-based understanding.

REFERENCES

[1] Mojtaba Akbari, Majid Mohrekesh, Ebrahim Nasr-Esfahani, SM Reza Soroushmehr, Nader Karimi, Shadrokh Samavi, and Kayvan Najarian. 2018. Polyp segmentation in colonoscopy images using fully convolutional network. In *IEEE EMBC*. 69–72.

[2] Patrick Brandao, Evangelos Mazomenos, Gastone Ciuti, Renato Caliò, Federico Bianchi, Arianna Menciassi, Paolo Dario, Anastasios Koulaouzidis, Alberto Arezzo, and Danail Stoyanov. 2017. Fully convolutional neural networks for polyp segmentation in colonoscopy. In *Medical Imaging 2017: Computer-Aided Diagnosis*, Vol. 10134. 101340F.

[3] Shuhan Chen, Xiuli Tan, Ben Wang, and Xuelong Hu. 2018. Reverse attention for salient object detection. In *ECCV*. 234–250.

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *IEEE CVPR*. 248–255.

[5] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. 2020. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. *MICCAI* (2020).

[6] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. 2020. Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. *IEEE TMI* (2020).

[7] Yuqi Fang, Cheng Chen, Yixuan Yuan, and Kai-yu Tong. 2019. Selective feature aggregation network with area-boundary constraints for polyp segmentation. In *MICCAI*. Springer, 302–310.

[8] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. 2020. Res2Net: A New Multi-scale Backbone Architecture. *IEEE TPAMI* (2020), 1–1. https://doi.org/10.1109/TPAMI.2019.2938758

[9] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. 2019. CE-Net: Context encoder network for 2d medical image segmentation. *IEEE TMI* 38, 10 (2019), 2281–2292.

[10] Debesh Jha, Steven A. Hicks, Krister Emanuelsen, Håvard D. Johansen, Dag Johansen, Thomas de Lange, Michael A. Riegler, and Pål Halvorsen. 2020. Medico Multimedia Task at MediaEval 2020: Automatic Polyp Segmentation. In *Proc. of MediaEval 2020 CEUR Workshop*.

[11] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. 2020. Kvasir-SEG: A Segmented Polyp Dataset. In *MMM*. 451–462.

[12] Alexander V Mamonov, Isabel N Figueiredo, Pedro N Figueiredo, and Yen-Hsi Richard Tsai. 2014. Automated polyp detection in colon capsule endoscopy. *IEEE TMI* 33, 7 (2014), 1488–1502.

[13] Balamurali Murugesan, Kaushik Sarveswaran, Sharath M Shankaranarayana, Keerthi Ram, Jayaraj Joseph, and Mohanasankar Sivaprakasam. 2019. Psi-Net: Shape and boundary aware joint multi-task deep network for medical image segmentation. In *IEEE EMBC*. 7223–7226.

[14] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. 2019. Basnet: Boundary-aware salient object detection. In *IEEE CVPR*. 7479–7489.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*. Springer, 234–241.

[16] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. 2015. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE TMI* 35, 2 (2015), 630–644.

[17] Jun Wei, Shuhui Wang, and Qingming Huang. 2020. F3Net: Fusion, Feedback and Focus for Salient Object Detection. In *AAAI*.

[18] Yunchao Wei, Jiashi Feng, Xiaodan Liang, Ming-Ming Cheng, Yao Zhao, and Shuicheng Yan. 2017. Object region mining with adversarial erasing: A simple classi□cation to semantic segmentation approach. In *IEEE CVPR*. 1568–1576.

[19] Zhe Wu, Li Su, and Qingming Huang. 2019. Cascaded partial decoder for fast and accurate salient object detection. In *IEEE CVPR*. 3907–3916.

[20] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Pheng Ann Heng. 2016. Integrating online and o□ine three-dimensional deep learning for automated polyp detection in colonoscopy videos. *IEEE JBHI* 21, 1 (2016), 65–75.

[21] Ruikai Zhang, Yali Zheng, Carmen CY Poon, Dinggang Shen, and James YW Lau. 2018. Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognition* 83 (2018), 209–219.

[22] Shihao Zhang, Huazhu Fu, Yuguang Yan, Yubing Zhang, Qingyao Wu, Ming Yang, Mingkui Tan, and Yanwu Xu. 2019. Attention Guided Network for Retinal Image Segmentation. In *MICCAI*. 797–805.

[23] Zhijie Zhang, Huazhu Fu, Hang Dai, Jianbing Shen, Yanwei Pang, and Ling Shao. 2019. ET-Net: A generic edge-attention guidance network for medical image segmentation. In *MICCAI*. Springer, 442–450.

[24] Zhao Zhang, Zheng Lin, Jun Xu, Wenda Jin, Shao-Ping Lu, and Deng-Ping Fan. 2020. Bilateral attention network for rgb-d salient object detection. *arXiv preprint arXiv:2004.14582* (2020).

[25] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. 2018. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters* 15, 5 (2018), 749–753.

[26] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. 2019. Unet++: A nested u-net architecture for medical image segmentation. *IEEE TMI* (2019), 3–11.