2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

APRNet: Alternative Prediction Refinement Network for Polyp Segmentation

Yutian Shen[1], Xiao Jia[2], Jin Pan[1] and Max Q.-H. Meng*, *Fellow, IEEE*

***Abstract*— Colorectal cancer has become the second leading cause of cancer-related death, attracting considerable interest for automatic polyp segmentation in polyp screening system. Accurate segmentation of polyps from colonoscopy is a challenging task as the polyps diverse in color, size and texture while the boundary between polyp and background is sometimes ambiguous. We propose a novel alternative prediction refinement network (APRNet) to more accurately segment polyps. Based on the UNet architecture, our APRNet aims at exploiting all-level features by alternatively leveraging features from encoder and decoder branch. Specifically, a series of prediction residual refinement modules (PRR) learn the residual and progressively refine the segmentation at various resolution. The proposed APRNet is evaluated on two benchmark datasets and achieves new state-of-the-art performance with a dice of 91.33% and an accuracy of 97.31% on the Kvasir-SEG dataset, and a dice of 86.33% and an accuracy of 97.12% on the EndoScene dataset.**

***Clinical relevance*— This work proposes an automatic and accurate polyp segmentation algorithm that achieves new state-of-the-art performance, which can potentially act as an observer pointing out polyps in colonoscopy procedure.**

I. INTRODUCTION

Colorectal cancer is the second highest prevalent cause of cancer-related death around the globe with a death toll of 915, 880 in 2020 [1]. Early-stage diagnosis and therapeutic treatment can greatly increase the likelihood of survival , where colonoscopy is the preferred method for analyzing inside the colon and removing colorectal polyps. To this end, automatic and accurate segmentation of polyps has become an active field of research for the past few decades.

In recent years, deep learning technologies have promoted automatic polyp segmentation, similar to other medical imaging applications. The fully convolutional neural networks (FCNs) [2][3] replaced the fully connected layers of the general convolutional neural networks (CNNs) with convolutional ones to form segmentation by pixel-wise classification, where skip connections could be adopted to combine multi-scale features. Brandao *et al.* [3] adopted the FCN with a pre-trained VGG model to identify and segment polyps from colonoscopy images. Later, UNet [4] was proposed and UNet-based architectures [5][6][7] became very popular in medical segmentation tasks where the expansive path is more or less symmetric to the contracting path yielding a U-shaped architecture and channel-wise concatenation is adopted for skip connect operation instead of summation in FCNs. Fang *et al.* [6] incorporated a sharing encoder branch and two decoder branches with upward concatenation and selective kernel module to fuse and learn multi-level features.

Although these networks have achieved high performance, there still exists some defects that the reconstructed mask was generated considering only the final semantic level. Moreover, deep semantic features from the decoder sub-network and shallow low-level features from the encoder sub-network were directly combined with skip connections, during which the semantic gap of features might cause ineffective recovery of fine details. Since different scales of reconstructed feature maps can be generated at various levels, it is expected that more efficient segmentation can be achieved by carefully exploiting multi-scale features.

In this paper, we propose the alternative prediction refinement network (APRNet) to more accurately segment the polyps by leveraging advantages of the residual learning and information encoded in multiple layers. Specifically, our APRNet is based on the UNet framework and the intent is to progressively refine and update the prediction map at each residual refinement step by alternatively taking in the semantic coarse-grained features from the decoder branch and the shallow fine-grained features from the encoder branch. To achieve this, we first design a global prediction generation (GPG) module to give an initial prediction seed map. Then we adopt a series of prediction residual refinement (PRR) modules that take in features from encoder-decoder architecture concatenated with the previous prediction as input to learn the residual and outputs a new prediction map that enhances polyp details and suppresses the background. Finally, we take the prediction map from the last PRR module as the output. The whole network is trained in an end-to-end manner.

To summarize, the contributions of this work mainly include: (1) An alternative prediction refinement network (APRNet) is proposed to progressively refine the segmentation by leveraging the features encoded in multiple layers of the U-shaped architecture. (2) A GPG module and a series of PRR modules are elaborated to generate a initial prediction seed and to effectively use multi-level features, respectively. (3) The proposed APRNet achieves a new state-of-the-art

[1]Y. Shen and J. Pan are with the Department of Electronic Engineering, The Chinese University of Hong Kong, N.T., Hong Kong SAR, China. yt.shen@link.cuhk.edu.hk, jpan@link.cuhk.edu.hk

[2]X. Jia is with the Department of Radiation Oncology, Stanford University, Stanford, CA, USA. jiaxiao@stanford.edu

*Max Q.-H. Meng is with the Department of Electronic and Electrical Engineering of the Southern University of Science and Technology in Shenzhen, China, on leave from the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong and also with the Shenzhen Research Institute of the Chinese University of Hong Kong in Shenzhen, China. max.meng@ieee.org

* Corresponding author.

performance on two benchmark datasets.

II. METHODS

The architecture of our APRNet is shown in Fig. 1, which can be regarded as an enhanced and deformed UNet. The encoder and decoder branch have five blocks each, where ResNet34 [9] is adopted as encoder and each decoder block consists of two Conv-BN-ReLu combinations and an upsample operation.

The global prediction generation (GPG) module is placed on top of the encoder branch, which captures the global context information and generates an initial prediction seed map. Then a series of prediction residual refinement (PRR) modules are utilized to learn the residual from the previous prediction map concatenated with encoder (or decoder) features, which is then used to refine and update the predicted segmentation. The previous prediction is upsampled for each PRR module exploiting features from the decoder branch. Meanwhile, the refined prediction map from each PRR module is supervised by the down-sampled ground truth with corresponding resolutions respectively. The prediction map generated by the last PRR module is taken as the final segmentation output of our network.

A. Global Prediction Generation Module (GPG)

We borrow the idea from atrous spatial pyramid pooling [10] and squeeze and excitation network [11] to build our GPG module, which is put on top of the encoder branch capturing global context to give an initial prediction map. Later, the initial prediction map is forwarded to the consecutive prediction residual refinement modules with features taken from decoder and encoder branches alternatively.

As shown in Fig. 2, GPG contains four atrous spatial pyramid pooling branches with different dilation rates and one image-level pooling branch to capture multi-scale information. Meanwhile, we incorporate a squeeze and excitation block consisting of global average pooling layer and two fully connected layers with activation to adaptively extract channel-wise features. Finally, a prediction out block with two convolutional layers and a dropout layer is adopted to generate the initial prediction map.

B. Prediction Residual Refinement Module (PRR)

Inspired by deep residual learning [9][12], we alternatively takes the features from decoder and encoder branch to learn the difference between the previous prediction map and the ground truth and then progressively refine and update the segmentation results.

As shown in Fig. 3, a PRR module is defined as:

$$P_{j+1} = P_j \oplus \Phi_j(Cat(P_j, F_j)) \tag{1}$$

where the take-in feature F_j is set as feature from decoder and encoder branch alternatively, which is then the concatenated (Cat) with the previous prediction map P_j from module PRR$_{j-1}$. Φ_j denotes function at current module consisting of three convolutional layers to learn the j^{th} residual. And $\oplus$ means element-wise summation.

C. Deep Supervision Loss (DSLoss)

As shown in Fig. 1, our network can output prediction maps at different resolutions. During the training process, we apply deep supervision mechanism to impose a deep supervision loss. We compute the combined Binary Cross Entropy (BCE) loss and Dice loss between each refined prediction map from each PRR module and down-sampled ground truth (supervision). Thus, the total deep supervision loss (DSLoss) is formulated as:

$$\mathcal{L}_{DS} = \mathcal{L}(GT, Pred_{10}) + \lambda \sum_{j=2}^{9} \mathcal{L}(GT_j, Pred_j) \tag{2}$$

where $\mathcal{L}$ is defined as $\mathcal{L} = \mathcal{L}_{BCE} + \mathcal{L}_{Dice}$, GT and GT_j denotes ground truth and down-sampled ground truth, while $Pred_{j=2-10}$ denotes refined prediction map from $j-1^{th}$ PRR module, repectively. λ is set as 0.25.

III. EXPERIMENTAL RESULTS

A. Datasets and Evaluation Metrics

We evaluate our proposed method on two benchmark colonoscopy image datasets, including Kvasir-SEG [13] with 1000 polyp images and EndoScene [14] with 912 images (612 images from CVC-ClinicDB and 300 images from CVC-ColonDB). For fair comparison, we refer to the settings of training set, validation set and test set as in previous literature[7], and the experiments of state-of-the-art methods are also conducted with same data settings. Specifically, for the first dataset Kvasir-SEG, we resize all images to 320×320 and randomly use 60% of the dataset as training set, 20% as validation set and 20% as test set. And for the second dataset Endoscene, we resize images to 288×384 and adopt the default setting with 574 images as training set, 183 images as validation set and 182 images as test set.

To quantitatively evaluate the segmentation performance of our proposed APRNet, we adopt eight metrics implemented in [6] for fair comparison, *Recall (Rec), Specificity (Spec), Precision (Prec), Dice Score (Dice), Intersection-over-Union for Polyp (IoUP), Intersection-over-Union for Background (IoUB), Mean Intersection-over-Union (mIoU) and Accuracy (Acc)*, among which Dice and Mean IoU are the preferred metrics suggested by Kvasir-SEG dataset.

B. Implementation Details

We use Pytorch to implement our algorithm. The batch size is set to be 4, the learning rate is initialized to be 0.001 and decreased by a factor $(\frac{epoch}{n_{epoch}=150})^{0.9}$. The SGD optimizer is used with a weight decay of 10^{-5} and momentum of 0.9. In training stage, data augmentation including random horizontal and vertical flips, shift and rotation is adopted to enlarge the training set as in [6].

C. Comparison with the State-of-the-arts

As listed in Table I, we compare our APRNet with five state-of-the-art algorithms: FCN [3], UNet [4], UNet++ [5], SFANet [6] and PraNet [8]. Meanwhile, some visualization results are shown in Fig. 4. As can be concluded from

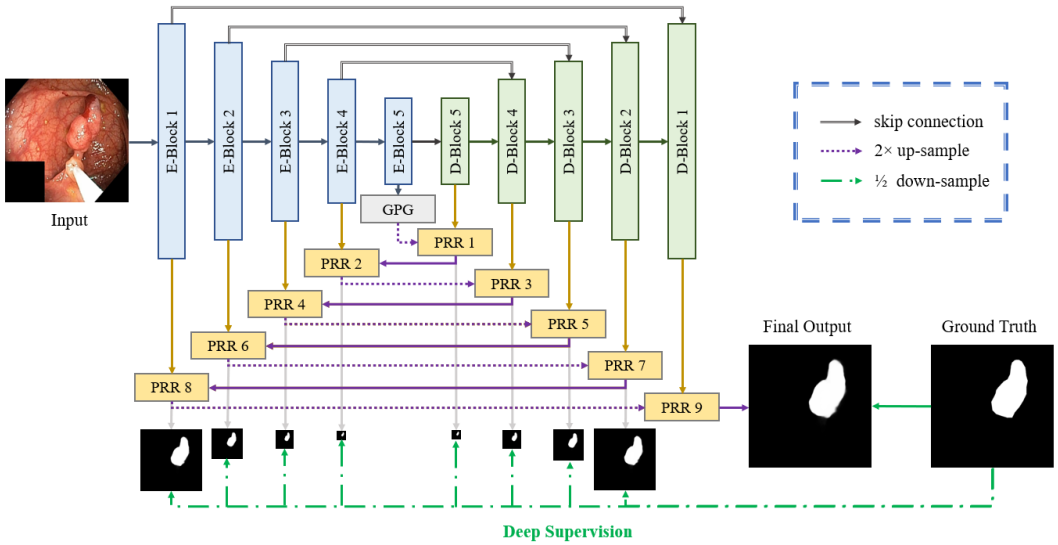

Fig. 1. Overview of our proposed APRNet. (Color figure online)

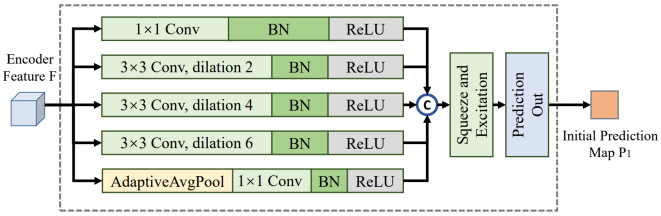

Fig. 2. Global Prediction Generation Module (GPG). (Color figure online)

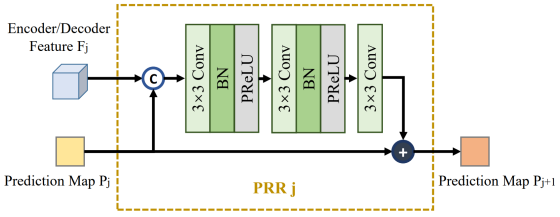

Fig. 3. Prediction Residual Refinement Module (PRR). (Color figure online)

both the quantitative and qualitative results, our APRNet consistently achieves the best segmentation performance on both benchmark dataset, showing effectiveness and also robustness of the proposed algorithm. Our APRNet achieves a mean Dice of 91.33%, a mean IoU of 91.23% and an accuracy of 97.32% on the Kvasir-SEG dataset, and a mean Dice of 86.33%, a mean IoU of 88.23% and an accuracy of 97.12% on the EndoScene dataset.

D. Ablation Study

The ablation experiments are done on the Kvasir-SEG dataset, and performance of our APRNet without each of proposed modules and deep supervision is also shown in Table I. Specifically, we replace the GPG module with the

TABLE I

COMPARISON WITH OTHER STATE-OF-THE-ART ALGORITHMS AND ABLATION STUDY

Dataset	Method	*Rec*	*Spec*	*Prec*	***Dice***[a]	*IoUP*	*IoUB*	***mIoU***[a]	*Acc*
Kvasir-SEG	FCN8s (EMBC'18) [3]	90.17	98.18	89.74	88.15	80.94	95.62	88.28	96.56
	UNet (MICCAI'15) [4]	87.91	97.30	87.27	84.70	76.77	94.26	85.52	95.33
	UNet++ (TMI'19) [5]	84.33	98.31	89.27	83.49	76.21	94.48	85.34	95.48
	SFANet (MICCAI'19) [6]	91.80	97.04	87.01	87.15	80.44	94.95	87.70	95.96
	PraNet (MICCAI'20) [8]	89.56	99.02	**93.04**	90.40	84.73	96.40	90.56	97.10
	APRNet (Ours)	**93.06**	98.32	91.86	**91.33**	**85.91**	**96.55**	**91.23**	**97.31**
	APRNet_w/o_PRR	89.71	96.77	86.97	85.38	77.34	94.63	85.99	95.69
	APRNet_w/o_GPG	92.13	98.10	91.26	89.95	84.39	95.98	90.19	96.77
	APRNet_w/o_DSLoss	92.67	**99.12**	90.47	90.01	84.11	96.22	90.17	97.00
EndoScene	FCN8s (EMBC'18) [3]	82.38	99.22	89.23	81.47	73.38	96.22	84.80	96.49
	UNet (MICCAI'15) [4]	68.44	98.59	**92.61**	72.19	63.72	95.20	79.46	95.44
	UNet++ (TMI'19) [5]	61.15	99.14	85.10	62.75	53.98	94.51	74.24	94.73
	SFANet (MICCAI'19) [6]	85.51	98.94	86.81	82.94	75.00	96.34	85.66	96.61
	PraNet (MICCAI'20) [8]	82.15	99.29	91.51	82.78	75.82	96.38	86.10	96.60
	APRNet (Ours)	**87.40**	**99.34**	91.34	**86.33**	**79.58**	**96.87**	**88.23**	**97.12**

[a] *Dice and mIoU are the preferred evaluation metrics suggested by the Kvasir dataset.*

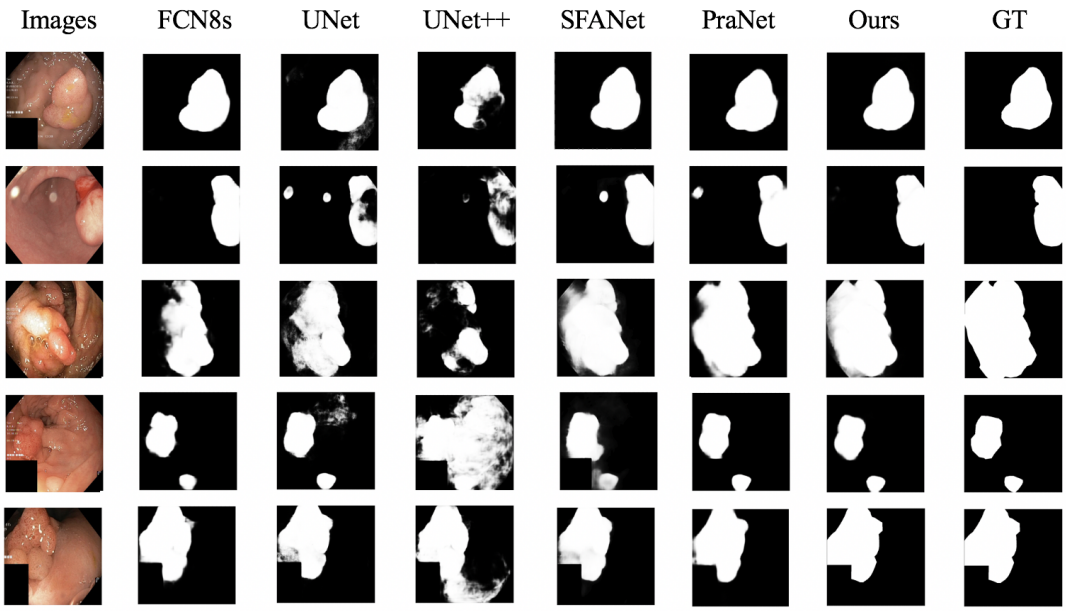

Fig. 4. Visualization of segmentation results on some samples. 1^{st} column: input images. 2^{nd} column to 7^{th} column: segmentation results of state-of-the-art methods and results of our APRNet, respectively. 8^{th} column: ground truths. Our proposed model efficiently learns to segment the polyps with more precise area and boundary, compared with other methods. (Color figure online)

prediction out block which contains two convolutional layers (see Fig. 2) to validate the effectiveness of GPG module. As for PRR modules, we directly use the original U-shaped encoder-decoder framework with same settings as APRNet and take output from the last decoder as final segmentation. Finally, the APRNet without deep supervision refers to a model adopting a BCEDice Loss supervising only the output segmentation map. As indicated from the results, all three developed parts help improve the segmentation performance, of which the series of PRR modules contributes most.

IV. CONCLUSIONS

In this paper, a novel UNet based segmentation algorithm, APRNet, is introduced to accurately segment polyps from colonoscopy images. Taking advantages of deep residual learning, we first generate an initial prediction seed with a proposed GPG module and then progressively refine and reproduce the prediction map via alternatively leveraging features from encoder and decoder branches of the UNet. Extensive experiments have demonstrated that our proposed APRNet outperforms state-of-the-art algorithms on two benchmark datasets.

ACKNOWLEDGMENT

This work was supported by National Key R&D program of China with Grant No.2019YFB1312400, Hong Kong RGC CRF grant C4063-18G and Hong Kong RGC GRF grant # 14211420.

REFERENCES

[1] SUNG, Hyuna, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 2021.

[2] LONG, Jonathan, et al. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2015. p. 3431-3440

[3] AKBARI, Mojtaba, 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). IEEE, 2018. p. 69-72.

[4] RONNEBERGER, Olaf, et al. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015. p. 234-241.

[5] ZHOU, Zongwei, et al. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging, 2019, 39.6: 1856-1867.

[6] FANG, Yuqi, et al. Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019. p. 302-310.

[7] ZHANG, Ruifei, et al. Adaptive Context Selection for Polyp Segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020. p. 253-262.

[8] FAN, Deng-Ping, et al. Pranet: Parallel reverse attention network for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020. p. 263-273.

[9] HE, Kaiming, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.

[10] HE, Kaiming, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 2015, 37.9: 1904-1916.

[11] HU, Jie, et al. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 7132-7141.

[12] DENG, Zijun, et al. R3net: Recurrent residual refinement network for saliency detection. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. AAAI Press, 2018. p. 684-690.

[13] JHA, Debesh, et al. Kvasir-seg: A segmented polyp dataset. In: International Conference on Multimedia Modeling. 2020. p. 451-462.

[14] VÁZQUEZ, David, et al. A benchmark for endoluminal scene segmentation of colonoscopy images. Journal of healthcare engineering, 2017, 2017.