HCMUS at Medico Automatic Polyp Segmentation Task 2020: PraNet and ResUnet++ for Polyps Segmentation

Tien-Phat Nguyen*[2,4], Tan-Cong Nguyen*[1,2,4], Gia-Han Diep*[2,4],
Minh-Quan Le[2,4], Hoang-Phuc Nguyen-Dinh[2,4], Hai-Dang Nguyen[2,4], Minh-Triet Tran[2,3,4]
[1]University of Social Sciences and Humanities, VNU-HCM
[2]University of Science, VNU-HCM
[3]John von Neumann Institute, VNU-HCM
[4]Vietnam National University, Ho Chi Minh city, Vietnam
ntcong@hcmussh.edu.vn,{ntphat,dghan,ndhphuc,lmquan,nhdang}@selab.hcmus.edu.vn,tmtriet@fit.hcmus.edu.vn

ABSTRACT

The Medico task, MediaEval 2020, explores the challenge of building accurate and high-performance algorithms to detect all types of polyps in endoscopic images. We proposed different approaches leveraging the advantages of either ResUnet++ or PraNet model to efficiently segment polyps in colonoscopy images, with modifications on the network structure, parameters, and training strategies to tackle various observed characteristics of the given dataset. Our methods outperform the other teams' methods, for both accuracy and efficiency. After the evaluation, we are at top 2 for task 1 (with Jaccard index of 0.777, best Precision and Accuracy scores) and top 1 for task 2 (with 67.52 FPS and Jaccard index of 0.658).

1 INTRODUCTION

Colorectal cancer is the third most frequently diagnosed cancer. Therefore, the prevention of colorectal cancer by early detecting and removing adenomas is critically important for the patients [8]. Detecting and segmenting various types of polyps is a challenging task, as the rate of overlooked polyps highly varied depending on types, shapes and sizes of the polyps [6]. There are previous approaches trying to accurately and automatically segment different types of polyps, such as an attraction propagation method based on a single interactable seed suggested by Ning Du et al. or the segmentation method leveraging patch-selection by Mojtaba Akbari et al. [1, 3]. Similarly, the goal of the Medico Task (MediaEval 2020) given Kvasir-SEG dataset aims to develop an efficient, accurate system to detect and segment multiple types of polyps to aid the clinicians or doctors by reducing overlooked polyps [5, 6].

We propose two approaches to solve the Medico task (results in section 3). For the first approach, we use PraNet [4], which is a parallel reverse attention network that helps to analyze, use the relationship between areas and boundary cues for accurate polyp segmentation. We use PraNet with Training Signal Annealing strategy to improve segmentation accuracy and effectively train from scratch on the given small dataset (section 2.1). For the second approach, we use ResUnet++ [7], a semantic segmentation neural network that takes advantage of residual blocks, squeeze and excitation blocks, atrous spatial pyramid pooling, and attention blocks. We modify the input path and integrate a guided mask layer to the original structure for a better segmentation accuracy (section 2.2).

*These authors contributed equally.

MediaEval'20, December 14-15 2020, Online

2 APPROACH

Given 1,000 polyp images with segmented masks for training and validation, we proposed two approaches for automatic polyps segmentation, both of which try to tackles the imbalanced characteristic of the polyps. The first one uses PraNet and takes advantages of the training signal annealing. The second one uses ResUnet++ and utilizes triple-path and weighted geodesic distance.

2.1 PraNet with Training Signal Annealing

This approach aims to learn the diversity of size, texture and the area-boundary in all type of polyps. To reach this goal, we retrain the PraNet [4] model from scratch with an effective training strategy and pre/post-processing methods.

2.1.1 Training strategy. The given training dataset majorly contains simple samples where polyps are discriminative or easy to recognize while those harder samples with small, flat or irregular polyps are few. As the result, we apply the Training Signal Annealing method proposed in [10] to prevent the models from overfitting on simple cases and force them to penalize on harder cases. Concretely, at each step in the training process, images with high segmentation score (dice coefficient) over a threshold are weighted less in the backpropagation process. The threshold formula that gives us the highest performance is:

$$\eta_t = \alpha_t * (1 - \frac{1}{K}) + \frac{1}{K} \tag{1}$$
$$\alpha_t = 1 - \exp(-\frac{t}{T} * 5)$$

where T is the total number of training step, η_t is the threshold at each training step t.

2.1.2 Pre/Post-processing method. We observe that the colonoscopy cameras may capture the polyps from any angle; Hence, for the preprocessing step, we focus on using random rotation as an augmentation method for training. We also apply test time augmentation with rotated inputs at inference time to improve the prediction accuracy. To enhance the polyp textures, we randomly use high-boost filtering to sharpen the input images when training.

2.2 ResUnet++ with Triple-Path and Geodesic Distance

In the second approach described in figure 1, we adapt the ResUnet++ architecture [7], modify it to aggregate three versions of the enhanced input image named the triple-path input. We also integrate a distance map layer using Geodesic Distance Transform as a guide mask to improve the accuracy of the original model.

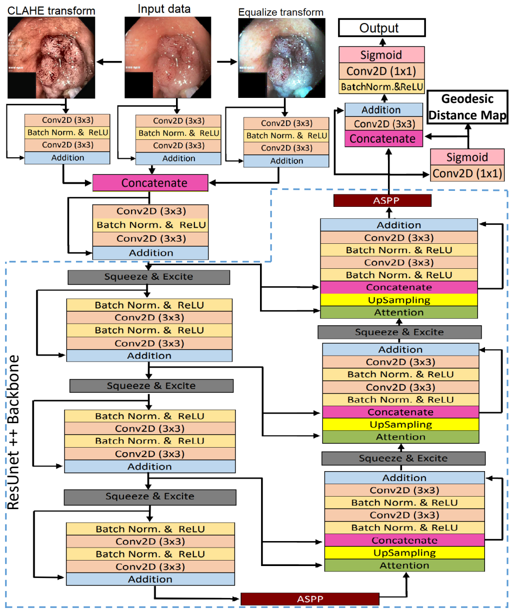

Figure 1: ResUnet++ with Triple-Path and Geodesic distance

2.2.1 Triple-Path Input. The size, shape and texture characteristics of the polyp are essential information to improve the accuracy of the model. So we use different enhancement methods to create two new versions of the original image (CLAHE transform [11] for the first and Equalize transform for the second one) and feed these two enhanced images together with the original one to the model. We use Albumentations [2] for this step. The enhanced images are put into separated convolution layers, then are concatenated together and passed into a sum convolution layer.

2.2.2 Guide map with Geodesic Distance. For medical segmentation problems, pixels closer the boundary should be treated differently from the pixels inside the polyps, depending on the importance of the pixel, especially for weak boundary problems. Hence, we smoothen the hard ground-truth masks by weighting each pixels differently based on the position and distance from the polyp center. So we integrate the geodesic distance map to increase the effect of pixel position and shape characteristics of polyps.

For each image in the training set, we use bounding boxes to compute the centre points of the polyps and apply GeodisTK [9] to calculate the geodesic distance map of the image based on those center points (figure 2). We add a new distance-map layer to the model and force the network to learn to predict the geodesic distance map. We then integrate the new layer to the original predict layer of the ResUnet++ architecture [7] for predicting the final result.

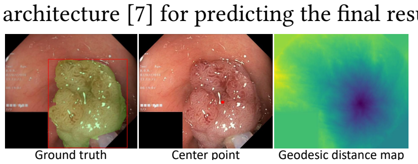

Figure 2: Compute geodesic distance map

3 EXPERIMENTS AND RESULTS

3.1 Experiments

In task 1, **for Run 1**, we use the PraNet model with the training, processing strategy discussed in Section 2.1 to train 3 models on 3 different train-val dataset split from the Kvasir-SEG dataset[6] (ratio 9:1). We average the results with a threshold of 0.6. **For Run 2**, with the dataset randomly divided into train and validation (ratio 9:1), we use the second approach in Section 2.2 to train 5 models and average them with a threshold of 0.5. **For Run 3**, we use all the trained models in Run 1 and Run 2, synthesize them in a similar way to Run 2. **For Run 4**, we continue training all trained models in Run 2 for several extra epoches with the full dataset that includes validation set. **Run 5** is the result of the single model (with highest validation score) in Run 2.

For the two runs in task 2, we choose the single model with the best validation score respectively from Run 1 and Run 2 in task 1.

3.2 Result

Table 1 shows our results in Task 1 (section 3.1). Run 1 is slightly better than run 2 in the Jaccard score. Run 2 is better than run 5 by using the ensemble method. Run 3 achieves a Jaccard score of 0.777. We received the 2^{nd} rank with this run as well as the best Precision and Accuracy scores for task 1. Because Run 3 is a combination of both of the approaches that are based on two completely different network architectures, implying that each proposed approach has its own strengths in recognizing types of polyp shapes.

Run ID	Jaccard	DSC	R	P	Accuracy	F2
Run 1	0.758	0.831	0.804	0.923	0.963	0.812
Run 2	0.753	0.831	0.845	0.862	0.955	0.832
Run 3	**0.777**	**0.848**	**0.850**	**0.890**	**0.964**	**0.845**
Run 4	0.754	0.831	0.840	0.876	0.958	0.830
Run 5	0.746	0.829	0.835	0.874	0.955	0.826

Table 1: Medico polyp segmentation task 1's result

Table 2 shows our results in Task 2 (section 3.1). Although our Run 2 is twice faster than Run 1, Run 1 has higher accuracy. Also we archived the best performance for this task with Run 1.

Run ID	FPS	Mean Time	Jac	DSC	Acc	F2
Run 1	33.28	0.030	**0.736**	**0.807**	**0.957**	0.806
Run 2	**67.52**	**0.015**	0.658	0.756	0.926	**0.811**

Table 2: Medico polyp segmentation task 2's result

4 CONCLUSION AND FUTURE WORKS

Medico task aims to develop semantic segmentation algorithms that can detect all types of polyps and we propose two approaches to solve it. We use PraNet with Training Signal Annealing strategy for the first approach and use ResUnet++ with Triple-Path and Geodesic Distance for the second. We also improve the training strategy to help those networks train more effectively. In the future, we plan to use other advanced ensemble methods instead of a simple average ensemble which may end up in an even higher score.

ACKNOWLEDGEMENTS

This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number DS2020-42-01 on "Artificial Intelligence and Extended Reality for Medical Diagnosis and Treatment Assistance".

Gia-Han Diep, VINIF.2020.ThS.JVN.04, was funded by Vingroup Joint Stock Company and supported by the Domestic Master/ Ph.D. Scholarship Programme of Vingroup Innovation Foundation (VINIF), Vingroup Big Data Institute (VinBigdata).

REFERENCES

[1] Mojtaba Akbari, Majid Mohrekesh, Ebrahim Nasr-Esfahani, S. M. Reza Soroushmehr, Nader Karimi, Shadrokh Samavi, and Kayvan Najarian. 2018. Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network. (2018). arXiv:eess.IV/1802.00368

[2] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. 2020. Albumentations: Fast and Flexible Image Augmentations. *Information* 11, 2 (2020). https://doi.org/10.3390/info11020125

[3] Ning Du, Xiaofei Wang, Jianhua Guo, and Meidong Xu. 2016. Attraction Propagation: A User-Friendly Interactive Approach for Polyp Segmentation in Colonoscopy Images. *PLOS ONE* 11, 5 (05 2016), 1–21. https://doi.org/10.1371/journal.pone.0155371

[4] 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. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 263–273.

[5] Debesh Jha, Steven A. Hicks, Krister Emanuelsen, Håvard 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 the MediaEval 2020 Workshop*.

[6] 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 *International Conference on Multimedia Modeling*. Springer, 451–462.

[7] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen. 2019. ResUNet++: An Advanced Architecture for Medical Image Segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*. 225–2255. https://doi.org/10.1109/ISM46123.2019.00049

[8] Lindsey A. Torre, Freddie Bray, Rebecca L. Siegel, Jacques Ferlay, Joannie Lortet-Tieulent, and Ahmedin Jemal. 2015. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* 65, 2 (2015), 87–108. https://doi.org/10.3322/caac.21262 arXiv:https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21262

[9] G. Wang, M. A. Zuluaga, W. Li, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren. 2019. DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 7 (2019), 1559–1572. https://doi.org/10.1109/TPAMI.2018.2840695

[10] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V Le. 2019. Unsupervised data augmentation for consistency training. *arXiv preprint arXiv:1904.12848* (2019).

[11] G. Yadav, S. Maheshwari, and A. Agarwal. 2014. Contrast limited adaptive histogram equalization based enhancement for real time video system. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. 2392–2397. https://doi.org/10.1109/ICACCI.2014.6968381