HCMUS-Juniors at Medico Polyp Segmentation Task 2021: Efficient U-Net for Polyps Segmentation

Quoc-Huy Trinh[1,4], Trong-Hieu Nguyen Mau[1,4],Minh-Van Nguyen[1,4],
Van-Son Ho [1,4], Tan-Cong Nguyen [2,3,4], Hai-Dang Nguyen[1,3,4], Minh-Triet Tran[1,3,4]

[1]Faculty of Information Technology, University of Science, VNU-HCM
[2] University of Social Sciences and Humanities, VNU-HCM
[3]John von Neumann Institute, VNU-HCM
[4]Vietnam National University, Ho Chi Minh city, Vietnam
{20120013,20127094,20120081,20120021}@student.hcmus.edu.vn,ntcong@hcmussh.edu.vn,nhdang@selab.hcmus.edu.vn,tmtriet@fit.hcmus.edu.vn

1 ABSTRACT

Medico task in the Mediaeval with the target to segment the Polyps in the endoscopic images. In this paper, we propose methods that use Efficient Unet and propose the Multiscale Efficient Unet to deal with this task. In the experiment, we also benchmark our method with others previous methods.

2 INTRODUCTION

With the developing of bio-medical and the information technology, medical images now are stored on a digital database. Moreover, with the increase of cases that have abnormal findings and symptoms on the digestive system, it is necessary to have a system to help the doctor accurately diagnose and detect the position of the abnormal in the medical images. That is why many methods have been proposed to help diagnose the polyps or the abnormal in the digestive system through endoscopic images.

On the other hand, the improvement of the Convolutional Neural Network architecture leads to improving the task for the segmentation of the medical images, and several architectures have been proposed such as U-Net[7], PSP-Net[9], PraNet[3], etc. However, there are many drawbacks in each method and need to be improved and many challenges to the researchers to improve the performance of their methods.[8]

The goal of the Medico automatic polyp segmentation challenge is to evaluate various methods for automatic polyp segmentation that can be used to detect and mask out various types of polyps (including irregular, small or flat polyps) with high accuracy[4]. In this challenge, our goals are to segment the mask of all types of polyps in the dataset.[5]

3 DATASET

To evaluate our proposed method, we use the Hyper Kavsir dataset proposed in 2020. This open dataset includes a comprehensive multi-class image and video dataset for gastrointestinal endoscopy, including the ground truth with mask and the bounding boxes value for the multi-task on the endoscopic images.[2]

In this task, we use the segmentation part of the Hyper Kvasir dataset. This dataset consists of 1000 Ground Truth images with masks to experiment on the segmentation tasks. With the test dataset, we evaluate our on the test dataset of Medico task organizer of Mediaeval.

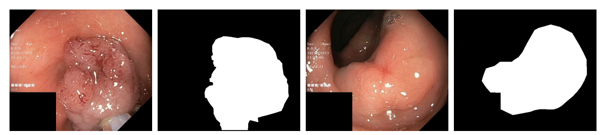

Figure 1: Polyps and corresponding masks from Hyper Kvasir Segmented.

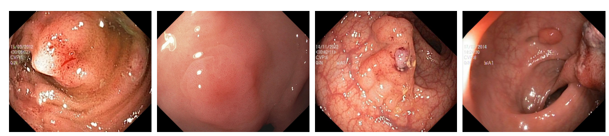

Figure 2: Examples polyps from the test images.

4 METHODS

We consider five solutions corresponding to our five submitted runs. To evaluate the performance of the proposed method, we also compare the results with those from other methods, such as are ResUNet and PraNet.

4.1 Propose Architecture

In our proposed methods, we propose the architecture that uses the Efficient Net for the encoder block. Moreover, we propose using the low scale feature to get a better mask of the output and improve the model's performance, which is Multiscale Efficient U-Net. This architecture includes three main blocks Multiscale Block (MC Block), EfficientNet Encoder block (EEN Block), and Decoder Block.

Initially, the input with the shape (w, h, c) passes through the MC Block; this block includes 3 Max Pooling layers, 2 Convolution 2D layers, and 1 Batch Normalization layer. MC layers to create the new low-scale feature for the model[6]. The parameter of the layers can be changed to adapt to the feature representation of images.

MediaEval'21, December 13-15 2021, Online

From the output of the MC block, there are two types of features: low scale and high scale features. Then these features pass the encoder block created by the Efficient Encoder Block, with high scale features, they map to the decoder block while low scale features pass the encoder block.
After passing the EEN Blocks, the feature will be concatenated at the block Encoder 4, then continue to the Decoder Block.

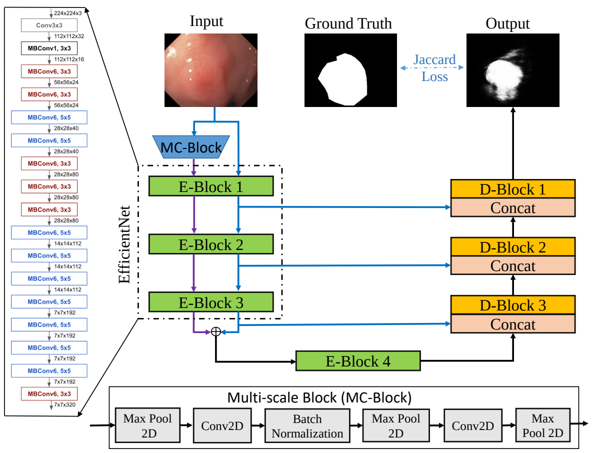

Figure 3: Multiscale Efficient UNet presentation

By using MC Block, the model can use the low-scale feature to enrich the feature in the learning process, particularly when the model has to adapt to the small dataset.
There is a limitation for this architecture because there are two types of features to the encoder block; this is why this architecture costs more computing resources than the traditional U-Net.

4.2 Loss Function

To use the proposed architecture, we propose using the Jaccard Loss Function with the following formula [1]:

$$JaccardLoss(y, \hat{y}) = \alpha * (1 - \frac{\alpha + \sum_c^C y_c * \hat{y}_c}{\alpha + \sum_c^C y_c + \hat{y}^c - y_c * \hat{y}_c}) \quad (1)$$

This loss function enable the segmentation process better and can control the performance of model on the pitch of the tissues.

4.3 Data Augmentation

To enrich the dataset, we propose some augmentation methods. We use Center Crop, Random Rotate, GridDistortion, Horizontal, and Vertical Flip to improve the quantity of the dataset.
Following is the sample of the data after augmentation:

Figure 4: Data Augmentation

5 RESULTS

Evaluation is on the test set of Mediaeval- Medico task, which includes 200 images of Polyps in endoscopic images. The Benchmark table shows that the Efficient Unet model with low features performs better than the original Efficient Unet and PraNet overall. However, if we compare Precision or Recall metrics, the model gets a lower performance than the other two models.

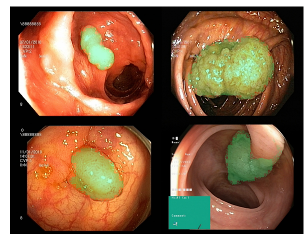

Figure 5: Visualization result

The performance of the proposed architecture is positive. The mask can cover almost all the tissue on the images, and it can cover cases with difficult shapes.

Method	Jaccard	Dice	Recall	Precision	Accuracy	F1
EfficientUnet	0.6572	0.7425	0.7264	0.8442	0.9529	0.7425
MCEU	**0.7059**	**0.7961**	**0.8167**	**0.8295**	**0.9565**	**0.7961**
PraNet	0.6929	0.7774	0.8204	0.8160	0.9511	0.7774
ResUnet	0.6739	0.7737	0.8371	0.7766	0.9495	0.7737

Table 1: MediaEval 2021 challenge's result base on the team method.

With the benchmark table, the Efficient Unet that uses the low feature achieves the high score. The reason is the data for the training and validation is the limitation, and augmentation can be used to enrich the quantity of data. However, some features can be as similar as the original sample. That is why the lower scale feature can help the model adapt better to the low quantity of sample dataset.

6 CONCLUSION

In general, we propose the Multiscale Efficient U-Net to deal with the segmentation task. MCEU has the merit that can enrich the feature for the training process. Moreover, this architecture can help normalize the high-scale feature to help the model adapt to the small dataset; however, some limitations exist. Regarding the evaluation of the experiment, the result we achieved is quite positive, compared to the PraNet, Res-UNet, and Efficient-UNet, our model achieves better performance. This positive impact can help the later architecture have another approach to deal with this task.

ACKNOWLEDGMENT

This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number DS2020-42-01.

REFERENCES

[1] Jeroen Bertels, Tom Eelbode, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, and Matthew B. Blaschko. 2019. Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019* (2019), 92–100. https://doi.org/10.1007/978-3-030-32245-8_11

[2] Hanna Borgli, Vajira Thambawita, Pia H Smedsrud, Steven Hicks, Debesh Jha, Eskeland Sigrun L, Kristin Ranheim Rand l, Konstantin Pogorelov, Mathias Lux, Duc Tien Dang Nguyen, Dag Johansen, Carsten Griwodz, Stensland Håkon K, Enrique Garcia-Ceja, Peter T Schmidt, Hugo L Hammer, Michael A Riegler, Pål Halvorsen, and Thomas de Lange. 2020. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data* 7, 1 (2020), 283. https://doi.org/10.1038/s41597-020-00622-y

[3] 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 *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz (Eds.). Springer International Publishing, Cham, 263–273.

[4] Steven Hicks, Debesh Jha, Vajira Thambawita, Hugo Hammer, Thomas de Lange, Sravanthi Parasa, Michael Riegler, and Pål Halvorsen. 2021. Medico Multimedia Task at MediaEval 2021: Transparency in Medical Image Segmentation. In *Proceedings of MediaEval 2021 CEUR Workshop*.

[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] Pardha Saradhi Mittapalli and Thanikaiselvan V. 2021. Multiscale CNN with compound fusions for false positive reduction in lung nodule detection. *Artificial Intelligence in Medicine* 113 (2021), 102017. https://doi.org/10.1016/j.artmed.2021.102017

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *LNCS* 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi (Eds.). Springer International Publishing, Cham, 234–241.

[9] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. 2016. Pyramid Scene Parsing Network. *CoRR* abs/1612.01105 (2016). arXiv:1612.01105 http://arxiv.org/abs/1612.01105