Efficient Supervision Net: Polyp Segmentation Using EfficientNet and Attention Unit

Sabari Nathan[1], Suganya Ramamoorthy[2]

[1]Couger Inc,Japan   [2]Thiagarajar college of engineering

ABSTRACT

Colorectal cancer is the third most common cause of cancer worldwide. In the era of medical Industry, identifying colorectal cancer in its early stages has been a challenging problem. Inspired by these issues, the main objective of this paper is to develop a Multi supervision net algorithm for segmenting polys on a comprehensive dataset. We have taken the Medico polyp challenge dataset, which consists of 1000 segmented polyp images from the gastrointestinal tract. We proposed an efficient Net B4 as a pre-trained architecture in multi-supervision net. The model is trained with multiple output layers. We present quantitative results on colorectal dataset to evaluate the performance and achieved good results in all the performance metrics. The experimental results proved that the proposed model is robust and provides a good level of accuracy in segmenting polyps on a comprehensive dataset for different metrics such as dice coefficient, recall, precision, and F2.

1 INTRODUCTION

The goal of *Medico automatic polyp segmentation challenge* [3] the benchmarking of polyp segmentation algorithms on new test images for automatic polyp segmentation that can detect and mask out polyps (including irregular, small, or flat polyps) with high accuracy. This work presents the abilities of a deep learning-based network for polyp segmentation in Medico automatic challenge task 1. We also evaluate our method and compare them with the results of state-of-the-art networks for automatic polyp segmentation. Our results show that deep neural networks can be used to segment polyps (including irregular, small or flat polyps) with good accuracy.

Our key contributions are summarized as follows:

(1) We created a novel Multi-Supervision Net architecture is proposed, i.e., the model is trained with multiple output layer.
(2) We evaluated the proposed architecture on a challenging polyp segmentation data set.
(3) We used EfficientNetB4 as a backbone for the proposed architecture.
(4) Achieved good experimental results in terms of accuracy, F1 score, and loss.

2 METHODS

To deal with the challenge of a robust Medico automatic polyp segmentation task, we have proposed a Multi-Supervision Net architecture. The detailed architecture is presented in Figure 3. In the following subsection, we will discuss the details of each module we used in the proposed architecture.
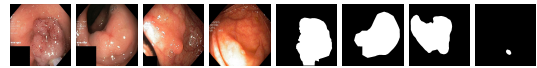
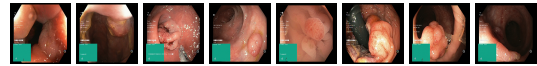**Figure 1: Polyps and corresponding masks taken from Kvasir-SEG.**

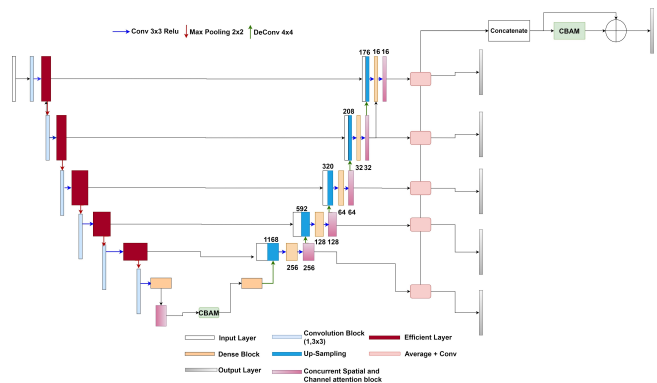**Figure 2: Examples polyps taken from the test data set.**

Figure 3: Proposed architecture diagram of proposed method.

2.1 Proposed Methodology

Our proposed polyp segmentation method uses a deep convolutional network to learn a connection between the input polyps' images. The overall block diagram of the proposed architecture is shown in Figure 3, which mainly consist of five layers (1) Convolutional layer, (2) Efficient layer [6], (3) Encoder block with efficient Net B4 (4) Decoder block combination of dense block and Concurrent Spatial and Channel Attention (CSCA) [5] blocks (5) Convolution Block Attention Module (CBAM) [7]. In this architecture, the input image pixels 332×487 is resized to 384×256 and divided by 255 are passed in to the convolutional layer. The proposed network was inspired by the multilevel hyper vision Net [1] and had properties of an encoder, decoder structure with supervision layers. Encoder block used efficient Net B4 as the backbone of the architecture. Decoder block consists of a combination of dense block [2] and Concurrent Spatial and Channel Attention (CSCA) block. Both the encoder and decoder are connected by Convolution block attention CBAM block. All the output of encoder are supervision, i.e., take individual decoder output and upsampled with output layer and supervised by the loss function, and also all upsampled output are concatenated and fed into CBAM. Totally six outputs are obtained from our proposed architecture. In the upsampling, we have used a convolution transpose layer. CBAM, an effective attention module for feed-forward convolutional neural networks. CBAM has two

sequential sub-modules: channel attention module and spatial attention module. The intermediate feature map is adaptively refined through CBAM at every convolutional block of deep networks. The CBAM block is utilized to generate spatial attention on the channel attention of the encoder output and the decoder output. The output of this attention is added with the input image. Equation 1 and Equation 2 shows the details of mathematical operations

$$C_A = f_a[w_1\left(w_0 \sum_n^i x_i\right) + w_1\left(max(x_i)\right))] \tag{1}$$

$$S_A = f_{con}[w_0 \sum_n^i c_{Ai} \|max(c_{Ai})\|] \tag{2}$$

2.2 Decoder part

The output of the CSCA block was passed into CBAM. The CBAM shares the processing results into another dense block. Now the algorithm starts from a bottom-up approach. Add the dense block and 5th efficient layer output results and forward to the consecutive dense and CSCA blcok. The CSCAB was connected to the average pooling and convolutional layer. The result of pooling and convolution is fed into the output layer. Now upsample the image or results of CSCA block and concatenate with 4th efficient layer, and remaining are same as the previous layer. After completing the operations up to 5 times, we concatenate every "average pooling + conv layer" result pass the output into CBAM and another concatenator. Finally, the concatenator sent output to the output layer. The output layer produces the final result of the given image. Loss function combination of categorical and dice loss

3 RESULTS

3.1 Dataset

The Kvaris-SEG [4] contains 1,000 polyp images and their corresponding ground truth mask, as shown in Figure 1. The dataset was collected from real routine clinical examinations at Baerum Hospital in Norway by expert gastroenterologists. The resolution of images varies from 332×487 to 1920×1072 pixels. Some of the images contain a green thumbnail in the lower-left corner of the images showing the scope position marking from the ScopeGuide (Olympus) refer to Figure 2. The Medico 2020 team annotate another separate dataset and use it as the test set to benchmark our proposed approach. Figure 2 shows some examples of test images used in the challenge.

3.2 Training

The proposed architecture has been trained and validated with the Medico automatic polyp segmentation challenges Task 1: Polyp Segmentation dataset. The proposed network is inspired by the Hyper Vision Net [4]. The shared dataset consists of 1000 segmented polyp images from the gastrointestinal tract images. We randomly split it into 70 percentage for training and 30 percentage for validation from the whole dataset. We used the Adam optimizer with an initial learning rate of 0.001. The learning rate was decreased to 0.00001 while we trained our models for 500 epochs. The proposed network was trained with the IntelCore i7 processor, GeForce GTX

Table 1: Results of the proposed model on Medico test data set.

mIoU	DICE	Recall	Precision	Accuracy	F2
0.777	0.850	0.916	0.839	0.957	0.879

1070 GPU, 8GB RAM, Platform Keras. The model is evaluated with Jaccard, dice coefficient, recall, precision, accuracy, and F2.

3.3 Data Augmentation

We divided the image by 255 to normalize it between 0-1. We applied data augmentation techniques such as HorizontalFlip, VerticalFlip, Blur (limit = 3), and Rotate(-10, 10) to increase the image count. A mixture of two different loss functions is used to supervise the network model outputs: categorical cross-entropy and dice loss. In this work, Adam optimizer is applied, which perfectly revise network weights in an iterative approach in training data. We used Adam optimizer with a learning rate of 0.001 to 0.00001 and 500 epochs for training the model. Our Loss function is defined by categorical Cross entropy and dice loss.

4 DISCUSSION

Currently, there is a growing interest in the development of computer-aided diagnosis (CADx) systems that could act as a second observer and digital assistant for endoscopists. Algorithmic benchmarking is an efficient approach to analyze the results of different methods. The task will use mean Intersection over Union (mIoU), or Jaccard index, as an evaluation metric, which is a standard metric for all medical segmentation task. Our proposed algorithm shows the Jaccard value as 0.777 in run 1 and run 2. The other evaluation metrics, such as dice coefficient, precision, recall, F2, and frame per second (FPS) for a comprehensive evaluation, also show effectiveness. In the challenge overview paper [3], the organizers will be calculating the metrics such as the Dice coefficient, mIoU, recall, precision, overlap, F2, FPS, the method submitted by each team and presented in Table 1.

5 CONCLUSION

We have presented a novel and unique Multi-supervision Net with EfficientNetB4 as the architecture's backbone to improve image segmentation accuracy under different factors. We accomplished this by training a multilevel attention network to take images from the Medico challenge 2020 polyp segmentation dataset. Moreover, we present a CSCA block in the decoder to improve image quality. As a major contribution, CBAM enhances the overall mechanism and utilizes significant features from the Encoder block. Many evaluations and comparisons to previous state-of-the-art approaches show that we can achieve good performance in qualitative and quantitative experimentation and proved the efficacy of the proposed architecture.

REFERENCES

[1] D.Sabarinathan, M.Parisa Beham, and S.M.Md.Mansoor Roomi. 2019. Hyper Vision Net: Kidney Tumor Segmentation Using Coordinate Convolutional Layer and Attention Unit. *arXiv preprint arXiv:1908.03339*.
[2] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. 2017. Densely Connected Convolutional Networks. *arXiv:1608.06993*.

[3] 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*.

[4] 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 *Proc. of International Conference on Multimedia Modeling*. 451–462.

[5] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. 2018. Concurrent Spatial and Channel Squeeze Excitation in Fully convolutional Networks. *arXiv:1803.02579*.

[6] Mingxing Tan and Quoc V. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv:1905.11946*.

[7] Woo and Sanghyun. 2018. CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*.