Polyp Segmentation in Colonoscopy Images Using Ensembles of U-Nets with EfficientNet and Asymmetric Similarity Loss Function

Le Thi Thu Hong
Information Technology Institute, MIST
HaNoi, VietNam
lethithuhong1302@gmail.com

Nguyen Chi Thanh
Information Technology Institute, MIST
HaNoi, VietNam
thanhnc80@gamail.com

Tran Quoc Long
University of Enginineering and Technology,VNU
HaNoi, VietNam
tqlong@gmail.com

***Abstract*—Automatic polyp detection and segmentation are highly desirable for colon screening due to polyp miss rate by physicians during colonoscopy, which is about 25%. Diagnosis of polyps in colonoscopy videos is a challenging task due to variations in the size and shape of polyps. In this paper, we adapt U-net and evaluate its performance with different modern convolutional neural networks as its encoder for polyp segmentation. One of the major challenges in training networks for polyp segmentation raises when the data are unbalanced, polyp pixels are often much lower in numbers than non-polyp pixels. A trained network with unbalanced data may make predictions with high precision and low recall, being severely biased toward the non-polyp class which is particularly undesired because false negatives are more important than false positives. We propose an asymmetric similarity loss function to address this problem and achieve a much better tradeoff between precision and recall. Finally, we propose an ensemble method for further performance improvement. We evaluate the performance of well-known polyp datasets CVC-ColonDB and ETIS-Larib PolypDB. The best results are 89.13% dice, 79.77% IOU, 90.15% recall, and 86.28% precision. Our proposed method outperforms the state-of-the-art polyp segmentation methods.**

Keywords—Polyp Segmentation, Medical image analysis, transfer learning, deep learning

I. INTRODUCTION

Colorectal cancer is the third most common cause of cancer-related death in the world for both men and women, with 551,269 deaths (account for 5.8% of all cancer deaths) worldwide in 2018 [1]. Colorectal cancer usually arises from polyps abnormal growths inside the colon, although, polyps grow slowly and may take years to turn into cancer. While the advanced stages of colorectal cancer have a poor five-year survival rate of 10%, the early diagnosis has shown a more favorable five-year survival rate of 90%. Early diagnosis of colorectal cancer is achievable [2]. Colonoscopy is the primary method for screening and preventing polyps from becoming cancerous. However, colonoscopy is dependent on highly skilled endoscopists and a high level of eye-hand coordination, and recent clinical studies have shown that 22%–28% of polyps are missed in patients undergoing colonoscopy [3]. Segmenting out polyps from the normal mucosa can help endoscopists to improve their segmentation errors and subjectivity. The segmented polyps size directly has an impact on the miss rates in colonoscopy, because doctors usually cannot easily evaluate small polyps, which are tiny and difficult to see, yet they can later naturally become cancer tumors. Different methods have been proposed with the aim of accurate polyp segmentation. The existing research work in polyp segmentation can be roughly grouped into three main approaches. The first approach belongs to image processing based segmentations which do not use any learning methods. The second group of approaches belongs to methods that first extract features and then use classifiers for segmentation. The third group of approaches belongs to methods that use convolutional neuronal networks (CNN) and perform the segmentation.

In this work, we propose a novel polyp segmentation method based on CNNs. We adapt U-net [4] which is proposed for biomedical image segmentation in recent years, showing the state of art result, to segment polyp automatically. We aim to evaluate different CNN architectures (e.g. MobileNet [5], Resnet[6], and EfficentNets [7]) as the backbone of the U-net for polyp segmentation. We choose EfficentNet as the backbone of U-net for our segmentation polyp model because its performance is the highest. To deal with significantly unbalanced imaging data, we propose a novel loss function combining pixel-wise cross-entropy loss and an asymmetric loss function. By training models with the proposed loss function, we found that the network can achieve a considerably better Dice score and give a better prediction. Finally, we propose an ensemble method for further performance improvement. We evaluate our method using well-known public available datasets: ETIS-Larib [9] from the MICCAI 2015 polyp detection challenge [10], and CVC-ColonDB [11]. The main contributions of our work can be summarized as follows:

1) We present a transfer learning method based on U-net and EfficientNet for polyp segmentation. To the best of our knowledge, this is the first work to use U-net and EfficientNet for the task of polyp segmentation.

2) We present a novel loss function to address the unbalanced data problem and achieve better performance. The combination of the loss function and our model results in a better performance.

3) We present an ensemble method to combine the results of two U-net models with different encoder structures (EfficientNet B4 and EfficientNet B5) to get better performance.

4) We demonstrate that our proposed method outperforms state-of-the-art methods using datasets from the MICCAI 2015 polyp detection challenge.

The rest of this work is organized as follows. In Section 2, we review related research on polyp segmentation. In Section 3, we present our proposed method for polyp segmentation. The experimental results are presented in Section 4. Finally, in Section 5 we summarize and conclude this work.

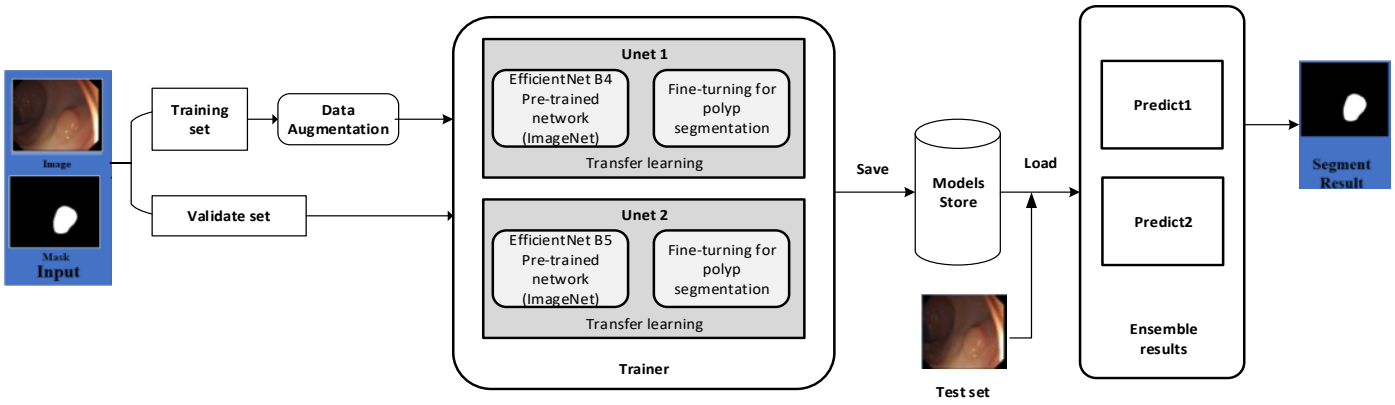


Fig.1. Overview of proposed method

II. RELATED WORK

The first approach for polyp segmentation is to use image processing segmentation methods. Many methods have been proposed to segment the polyps automatically. Bernal et al. [10] proposed a method using "depth of valleys" of an image to segment colorectal polyps. They use the watershed algorithm to segment images into polyp candidate regions and then classify each region into polyp and non-polyp, this classification is based on region information and "depth of valleys" in each region. Ganz et al. [12] propose a method based on Hough transform to detect the region of interest (ROI) and specular reflection suppression with an exemplar-based image in painting as a preprocessing method. Then, they use an algorithm called shape-UCM [13] for image segmentation, shape-UCM works based on image gradient contours and spectral clustering. After performing the shape-UCM algorithm, they use a scheme to improve edges resulted from the shape-UCM algorithm.

The second approach in polyp segmentation is feature extraction from image patches and labeling of patches as polyp and non-polyp based on extracted features. Tajbakhsh et al. [14] presented a method based on Canny edge detector in each of the three RGB channels. This is done to produce edge maps and then oriented patches for each pixel are extracted to classify them as polyp or non-polyp. Tajbakhsh et al. [15] also proposed a feature extraction method to extracts sub-patch with a 50% overlap and calculates their average vertically resulting in one- dimensional signal. After that, they use DCT coefficients as a feature for each extracted patch. Finally, they use a two-stage random forest classifier to label each patch.

The third approach for polyp segmentation is using Convolutional Neutral Networks (CNN). In the 2015 MICCAI sub-challenge on automatic polyp detection, most of the proposed methods were based on CNN, including the winner [16]. The author in [17] showed that fully convolution network (FCN) architectures could be refined and adapted to recognize polyp structures. Zhang et al. [18] used FCN-8S to segment polyp region candidates, and texton features computed from each region were used by a random forest classifier for the final decision. Shin et al. [19] showed that Faster R-CNN is a promising technique for polyp detection.

III. PROPOSED METHOD

In this section, we describe the methodology on which the proposed method is based on. First, we use the U-net architecture for polyps segmentation and evaluate the performance of U-nets with different CNN encoders. We selected U-net architectures with EfficentNet B4 and EfficentNet B5 encoder for our polyp segmentation framework. Second, we propose a novel loss function that can effectively boost the segmentation performance of our network. In the last step, we adapt an ensemble method to combine the results of two U-net models with different encoder structures (EfficientNet B4 and EfficientNet B5) for better performance. The overview of the proposed method can be seen in Fig.1. The proposed method consists of these components: 1) data augmentation, 2) two U-net with different encoder structures (EfficientNet B5 and EfficientNet B4), 3) the loss function that combined with the U-net model for better performance, and 4) an ensemble model to combine results of two U-net models with two different backbones to enhance the segmentation performance.

A. Data agumentation

One of the challenges in training polyp segmentation models is the insufficient number of data for training because access data is limited due to privacy concerns. Since the endoscopy procedures involving moving camera control, color calibration are not consistent, the appearance of endoscopy images significantly changes across different laboratories. The data augmentation step brings endoscopy images into an extended space that can cover all their variances. By augmenting training data, we can also reduce the over-fitting problem on training models. Fig.2. shows the examples of the data augmentation method applied to the original polyp image (Fig.2.a).

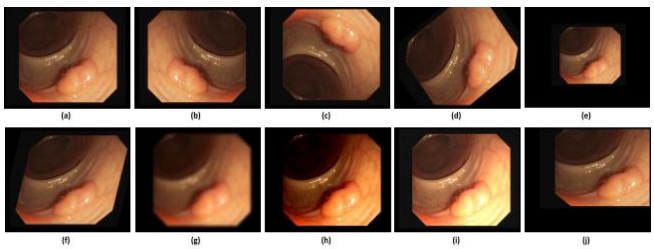


Fig.2. Examples of data augmentation

The methods of augmentation used in our work are: Vertical flipping, horizontal flipping, random rotation between -10 and 10 degrees, random scaling ranging from 0.5 to 1.5, random shearing between -5 and 5 degrees, random Gaussian blurring with a sigma of 3.0, random contrast normalization by a factor of 1 to 1.5, random brightness ranging from 1 to 1.5, and random cropping and padding by 0–5% of height and width.

B. *Encoder networks*

The U-net was developed by Olaf Ronneberger et al. for BioMedical Image Segmentation [4]. The architecture, shown in Fig.3., has two paths. First path is the contraction path (also called the encoder) which is used to capture the context in the image, consists of convolutional and max-pooling layers. The second path is the symmetric expanding path (also called the decoder) which is used to enable precise localization using transposed convolutions. Because the decoding process loses some of the higher-level features the encoder learned, the U-net has skip connections. That means that the outputs of the encoding layers are passed directly to the decoding layers so that all the important pieces of information can be preserved. For polyp segmentation, we adapt a transfer learning approach, we use U-net with a CNN model pre-trained on the ImageNet dataset as the encoder. In the first path of U-net, we need a convolution neural network as an encoder to extract features from the input image. The choice of the encoder is essential because the CNN architecture, the number of parameters and type of layers directly affect the speed, memory usage and most importantly the performance of the U-net. In this study, we select three architectures to compare and evaluate their performance in polyp segmentation: MobileNet, Resnet, and EfficientNet. MobileNet is a family of mobile-first computer vision models from Google. They are designed to effectively maximize accuracy while being mindful of the restricted resources for an on-device or embedded application. MobileNet has two different versions: MobileNet V1 and MobileNet V2 [5]. With MobileNetV2 as a backbone for feature extraction, state-of-the-art performances are also achieved for object detection and semantic segmentation. We choose MobileNetV2 as the encoder of U-net in our experiment. Resnet [6] is a residual learning framework to ease the training of deep networks by explicitly reformulating the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. With Resnet, we can benefit from deeper CNN networks to obtain even higher level of features which are essential for difficult tasks such as polyp segmentation. We use two Resnet backbone structures (ResNet50 and ResNet101) as encoders of U-net for polyp segmentation. EfficientNets [7] are the latest family of image classification models from Google, which achieves state-of-the-art accuracy on ImageNet.

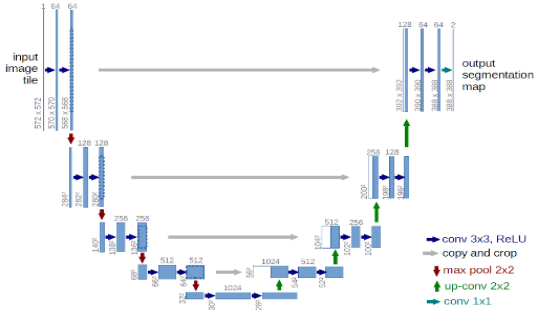


Fig.3. U-net architecture

The EfficientNets was developed by Mingxing Tan and Quoc V. Le, they developed EfficientNets based on AutoML and Compound Scaling. In particular, they use the AutoML MNAS Mobile framework to develop a mobile-size baseline network, named EfficientNet-B0. Then, they use the compound scaling method to scale up this baseline to obtain EfficientNet-B1 to EfficientNet-B7. Starting from the smallest EfficientNet configuration B0 to the largest B7, accuracies are steady increasing while maintaining a relatively small size. In our experiment, we select EfficentNet B4 and EfficentNet B5 as the encoder of U-net.

After experimenting and evaluating results of U-net with different CNN encoders, we selected U-net with EfficentNet B5 encoder (U-net1) and U-net with EfficentNet B4 encoder (U-net2) for our segmentation polyp model.

C. *Asymmetric similarity loss function*

To boost segmentation results, we propose a novel simple loss function that is a combination of basic loss functions with hyper-parameters to perform the segmentation: cross-entropy loss and asymmetric F_β loss function. Pixel-wise cross-entropy loss was used by Ronneberger et al. in [4] for the task of image segmentation. This loss simply verified each pixel individually, comparing the class predictions that are defined as a depth-wise pixel vector to the target vector. The cross-entropy loss function is defined as:

$$CE = -\sum_{i,j} g_{i,j} * \log(p_{i,j}) \quad (1)$$

where $p_{(i,j)}$ is the predicted binary segmentation volume and $g_{(i,j)}$ stands for the ground truth at image pixel (i,j). Because cross-entropy loss function asserts every single pixel and colonoscopy image usually have a low surface area, the segmentation network trained with a cross-entropy loss function is biased towards the background image rather than the object itself. Furthermore, as the foreground region is often missing or only partially detected, it is not easy for the model to see the object. In the medical community, the Dice score coefficient (DSC) is an overlap index that is widely used to asses segmentation maps. Let P and G be the set of predicted and ground truth binary labels, respectively. The Dice similarity coefficient D between P and G is defined as:

$$DSC(P,G) = \frac{2|PG|}{|P|+|G|} \quad (2)$$

Loss functions based on the Dice similarity coefficient have been proposed as alternatives to cross-entropy to improve training U-Net and other network architectures. However DSC, as the harmonic mean of precision and recall, weighs false positives (FPs) and false negatives (FNs) equally, forming a symmetric similarity loss function. To make a better adjustment of the weights of FPs and FNs (and achieve a better balance between precision and recall) in training fully convolutional deep networks for highly unbalanced data, where detecting small number of pixels in a class is important, we use an asymmetric similarity loss function [20] based on the F_β scores to replace Dice similarity coefficient. F_β scores is defined as:

$$F_\beta = (1 + \beta^2)\frac{precision*recall}{\beta^2*precision+recall} \quad (3)$$

By adjusting the hyperparameter β we can control the trade-off between precision and recall (FPs and FNs). Equation (3) can be written as:

$$F(P,G,\beta) = \frac{(1+\beta^2)|PG|}{(1+\beta^2)|PG|+\beta^2|G \setminus P|+|P \setminus G|} \quad (4)$$

where |P \ G| is the relative complement of G on P. To define the F_β loss function we use the following formulation:

$$F_\beta = \frac{(1+\beta^2)\sum p_{i,j} g_{i,j}}{(1+\beta^2)\sum p_{i,j} g_{i,j} + \beta^2 \sum (1-p_{i,j}) g_{i,j} + \sum p_{i,j}(1-g_{i,j})} \quad (5)$$

The asymmetric F_β loss function with the hyper-parameter β generalizes the Dice similarity coefficient and the Jaccard (IoU) index. More specifically, in the case of β =1 the score simplifies to be the Dice loss function (F1) while β = 2 generates the F2 score and β = 0 transforms the function to precision. Larger β weighs recall higher than precision (by placing more emphasis on false negatives).

We proposed a combination of cross-entropy loss and asymmetric F_β loss function to reduce the negative aspects of the former. This is because asymmetric F_β loss function can strongly measure the overlap between two objects, one is a prediction and the remaining is ground truth. The loss function is defined as:

$$L = \alpha * CE + DL \quad (6)$$

where CE is cross-entropy loss and DL=1-F_β is asymmetric F_β loss function, while hyperparameter α is used for balancing. Our experimental results prove that this loss function is more robust compared to the classical cross-entropy loss function and basic dice loss function. We trained our U-net2 with different hyper-parameters α,β values and used CVC-ColonDB for testing. Appropriate values of the hyper-parameters can be defined based on class imbalance ratios, the best results were obtained from training our U-net 2 model with α =0.4. and β = 1.6.

D. Ensemble models

In this work, we use two U-nets with different encoder structures (EfficientNet B5 and EfficientNet B4) for our polyp segmentation framework. The two CNN encoders compute different types of features due to differences in their number of layers and architectures. If U-net was initialized with different pre-trained backbone structure models, the network is therefore virtually guaranteed to converge to different solutions, although it uses the same training data, for example, U-net with EfficientNet B5 encoder produced better segmentation results than U-net with EfficientNet B4 encoder for some polyp images. Besides, a deeper CNN can compute a higher level of features from the input image while it loses some spatial information due to the contraction and pooling layers. Some polyps might be missed by one of the CNN models while it could be detected by another one. Based on these observations, we propose an ensemble method that combines the results of two U-nets for better performance. We use U-net with EfficientNet B5 (Unet1) encoder as the main model and its output is always relied on, and U-net with EfficientNet B4 encoder model (Unet2) as an auxiliary model to support the main model. We only take into account the outputs from the auxiliary model when the probability that pixel is polyp is > 0.96 (an optimized value using a validate dataset see section III-b)

IV. EXPERIMENTS AND RESULTS

A. Dataset

We use well-known datasets from the MICCAI 2015 polyp detection challenge in colorectal segmentation : CVC-ClinicDB[7], ETIS-Larib[8], and CVC-ColonDB[9]. The datasets are briefly described in the following paragraphs.

- CVC-ClinicDB contains 612 images, where all images show at least one polyp. The segmentation labels obtained from 31 colorectal video sequences were acquired from 23 patients.
- ETIS-LaribPolypDB contains 196 images, where all images show at least one polyp.
- CVC-ColonDB ontains 379 frames from 15 different colonoscopy sequences, where each sequence shows at least one polyp each.

The datasets were obtained with different imaging systems and contain binary masks as the ground truths to indicate the location of the polyps for each image. All ground truths of polyp regions for these datasets were annotated by expert video endoscopists from the corresponding associated clinical institutions. There are similar image frames within the same colonoscopy dataset. Therefore, for more reliable evaluation, we assign the above-mentioned different datasets into training and testing set separately as the recommendation of the MICCAI challenge guidelines: CVC-CLINIC for training and ETIS-Larib for testing. Furthermore, we also report results from another public dataset (CVC-ColonDB) as a testing set.

B. Evaluation metrics

For the evaluation of polyp segmentation, we use a common segmentation evaluation metric similarity score Dice coefficient as the main metric. Furthermore, to provide a general view of the effectiveness of our method, we also employed interception over union (IoU), recall (Re) which is also known as sensitivity, precision (Prec) metrics to evaluate the proposed method. We use these metrics to compare our prediction results (PR) with the ground truth (GT). If a pixel of polyp is correctly classified, it is counted as a true positive (TP). Every pixel segmented as a polyp pixel that falls outside of a polyp mask counts as a false positive (FP). Finally, every polyp pixel that has not been detected counts as a false negative (FN). The evaluation metrics are calculated as follows:

$$Dice = \frac{2PR \cap GT}{|PR| + |GT|} \quad (7)$$

$$IoU = \frac{PR \cap GT}{PR \cup GT} \quad (8)$$

$$Re = \frac{TP}{TP + FN} \quad (9)$$

$$Pre = \frac{TP}{TP + FP} \quad (10)$$

C. Training details

We use the CVC-CLINIC for training, this dataset contains 32 different polyps presented in 612 images. The training set is split into 80% for learning the weights and 20% for validating our model during the training step. We use the pre-trained weights of the backbone models on ImageNet dataset as the training begins. We unfreeze the backbone model and update the entire network via Adam optimizer, the learning rate of Adam is set to 10-4. The model generated at the epoch with the max dice score on the validation set is used as our final mode. Furthermore, all algorithms have been programmed/trained using Keras and Tensorflow backend on a PC with a GeForce GTX 1080 Ti GPU.

D. Performance evaluation on CNN pre-trained encoders

In this section, we reported the performance of U-net models for polyp segmentation with different pre-trained CNNs as encoders. In this experiment, we use the CVC-ClinicDB dataset for training the models, ETIS-Larib and CVC-ColonDB for testing. Table 1 presents our results using the ETIS-Larib dataset as the test set. Table 1 shows that U-net with EfficientNet B4 and U-net with EfficientNet B5 have the best performance among the models, the U-net with EfficientNet B4 achieves the highest in all evaluation metrics, with a Dice of 81.13% and IoU of 69.6%, recall (Re) of 80.8%, precision(Pre) of 83.4%. Table 2 presents the experimental results on the CVC-ColonDB dataset. The table also shows that U-net with EfficientNet B4 and U-net with EfficientNet B5 have the best performance among the models, but the U-net with EfficientNet B5 achieved the highest in all evaluation metrics, with a Dice of 87.69% and IoU 78.44%, recall of 88.07%, precision of 83.40%. Moreover, examples of different segmentations produced by the different U-net networks could be depicted in Fig.6. The figure describes that U-net with EfficientNet B4 and U-net with EfficientNet B5 can recognize the polyp mask as much as possible what others could not do.

E. The effect of proposed loss function

We evaluated the effect of our proposed loss function on performance of the model, compare it with basic loss functions in polyp segmentation. The improvement of performance metrics are reported in Table 3 and Figure 7 describes the comparison of effect to network learning progress between our proposed loss function and cross-entropy loss function. Table 3 demonstrates that our proposed loss function reduces the negative aspects of the cross-entropy, it makes a better balance between precision and recall so that the performance of models trained with our proposed loss function can improve. Comparing to using cross-entropy loss function for training model, using our proposed loss function with Unet1 (EfficientNetB4 encoder) could improve dice by 12.4% and IoU by 11%, and recall by 16.3% and with Unet2 (EfficientNetB5 encoder) could improve dice by 9.8% and IoU by 7%, and recall by 13.9%. Precision got decreased in both cases.

TABLE 1. COMPARISON OF U-NET MODELS ON THE ETIS-LARIB

Network	Dice(%)	IoU(%)	Re(%)	Pre(%)
U-net_MobileNetV2	70.06	54.58	76.85	65.22
U-net_Resnet50	68.16	52.26	69.03	68.55
U-net_Resnet101	74.56	61.60	77.60	74.50
U-net_EfficientNetB4	**81.30**	**69.60**	**80.80**	**83.40**
U-net_EfficientNetB5	78.69	65.68	79.44	79.07

TABLE 2. COMPARISON OF U-NET MODELS ON THE CVC-COLONDB

Network	Dice(%)	IoU(%)	Re(%)	Pre(%)
U-net_MobileNetV2	83.21	71.67	88.72	78.85
U-net_Resnet50	85.20	74.85	84.61	86.25
U-net_Resnet101	86.22	76.75	90.01	83.91
U-net_EfficientNetB4	86.96	77.26	87.55	86.80
U-net_EfficientNetB5	**87.69**	**78.44**	**88.07**	**87.78**

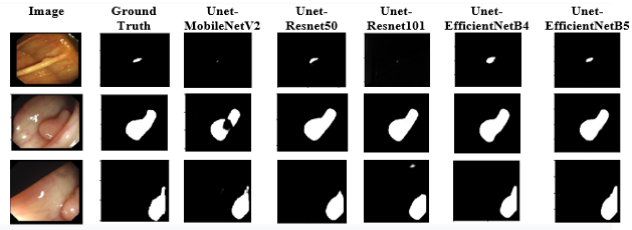
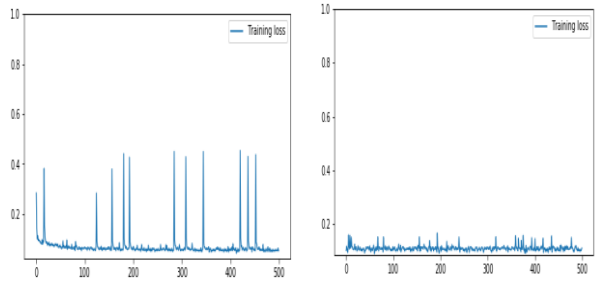


Fig.6. Example of different segmentations produced by the U-nets

TABLE 3. THE EFFECT OF THE PROPOSED LOSS FUNCTION ON THE ETIS-LARIB

Network	Dice(%)	IoU(%)	Re(%)	Pre(%)
Unet1 with bce loss	68.90	58.60	64.60	86.50
Unet1 with dice loss	73.69	58.72	64.94	86.29
Unet1 with F_β loss	78.10	64.29	71.50	86.62
Unet1 with proposed loss	**81.30**	**69.60**	**80.80**	**83.40**
Unet2 with bce loss	68.86	58.56	65.56	86.43
Unet2 with dice loss	74.41	59.61	75.74	73.86
Unet2 with F_β loss	69.45	53.58	68.88	70.94
Unet2 with proposed loss	**78.70**	**65.70**	**79.40**	**79.10**



a) Basic cross-entropy loss function c) Proposed loss function

Fig.7. The effect of proposed loss function to network learning progress on the same dataset by comparing to cross-entropy loss function.

F. Ensemble Results

Our experiments show that segmentation performance can be improved by combining the output results of U-net models using our ensemble method. We used the validation set to select a suitable probability threshold for the auxiliary model. Based on this optimization step, the output of the auxiliary model is only taken into account when the probability that pixel is polyp is >0.96. Table 4 shows the results of the ensemble on the CVC-ColonDB. Table 4 illustrates that the auxiliary model could add a small improvement in the performance of the main model. The ensemble could improve Dice by 1.44% and IoU by 1.33%.

G. Comparison with Other Methods

We evaluate our proposed segmentation method and compare it with the other competitor methods on the ETIS-Larib dataset of the MICCAI challenge.

TABLE 4. ENSEMBLE RESULTS OBTAINED ON THE CVC-COLONDB BY COMBINING THE RESULTS OF TWO U-NET MODELS

Network	Dice(%)	IoU(%)	Re(%)	Pre(%)
Unet1	86.54	76.76	89.61	84.24
Unet2	87.69	78.44	88.07	87.78
Ensemble	**89.13**	**79.77**	**90.15**	**86.28**

TABLE 5. COMPARISON OF THE PROPOSED METHOD WITH OTHER METHODS ON THE ETIS-LARIB

Criterion	Dice(%)	IoU(%)	Re(%)	Pre(%)
Qadir, Hemin Ali, et al[23]	70.40	61.20	72.60	80.00
Kang, Jaeyong, et al[22]		66.10	74.40	73.80
Proposed	**82.25**	**70.24**	**87.78**	**77.85**

TABLE 6. COMPARISON OF THE PROPOSED METHOD WITH OTHER METHODS ON THE CVC-COLONDB

Criterion	Dice(%)	IoU(%)	Re(%)	Pre(%)
Akbari, Mojtaba, et al[21]	81.00	-	75.70	88.30
Kang, Jaeyong, et al[22]	-	69.46	76.25	77.92
Proposed	**89.13**	**79.77**	**90.15**	**86.28**

Our results are presented in Table 5. The table shows that our proposed model outperforms previous methods in the segmentation of colorectal polyps on the ETIS-Larib dataset Moreover, we also evaluated our network's performance on the well-known dataset CVC-ColonDB, as shown in Table 6. Our proposed model achieves the highest in all metrics among the models.

V. CONCLUSION

In this paper we presented a transfer learning method based on U-net and EfficientNet model for colorectal polyp segmentation.We adapted and evaluated U-net with recent pre-trained CNN encoders i.e. MobileNetV2, Resnet50, Resnet101, EfficientNetB4 and EfficientNetB5 for polyp segmentation. We also presented a novel loss function to address unbalanced data problem and achieve better performance. Furthermore, we proposed an ensemble results method to improve the performance of the models. The proposed framework consists of elements: 1) data augmentation, 2) two U-net with different backbone structures (EfficientNetB4 and EfficientNetB5) pre-trained on the ImageNet, and 3) the ensemble method that combined results from two U-net. Our method is validated using well known datasets from MICCAI 2015 polyp detection challenge. Our experimental results show that the proposed method outperforms the state-of-the-art polyp segmentation methods. Our research is still flawed, but we hope to try to break through existing research results in a variety of ways. To improve segmentation performance, we plan to explore other semantic segmentation models combining with our proposed loss function. Besides, we also continue to find other ensemble methods to boost performance of models.

REFERENCES

[1] Bray, Freddie, et al. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: a cancer journal for clinicians 68.6 (2018): 394-424.

[2] M. Gschwantler, S. Kriwanek, E. Langner, B. Goritzer, C. SchrutkaKolbl, E. Brownstone, H. Feichtinger, and W. Weiss. "High-grade dysplasia and invasive carcinoma in colorectal adenomas: a multivariate analysis of the impact of adenoma and patient characteristics," European journal of gastroenterology hepatology, 14(2):183188, 2002.

[3] A. M. Leufkens, M. G. H. van Oijen, F. P. Vleggaar, and P. D. Siersema. "Factors influencing the miss rate of polyps in a back-to-back colonoscopy study," Endoscopy, 44(05):470475, 2012.

[4] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[5] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks. In 2018 IEEE." CVF Conference on Computer Vision and Pattern Recognition.

[6] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[7] Tan, Mingxing, and Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." arXiv preprint arXiv:1905.11946 (2019).

[8] Bernal, Jorge, et al. "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians." Computerized Medical Imaging and Graphics 43 (2015): 99-111.

[9] Bernal, Jorge, Javier Sánchez, and Fernando Vilarino. "Towards automatic polyp detection with a polyp appearance model." Pattern Recognition 45.9 (2012): 3166-3182.

[10] Silva, Juan, et al. "Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer." International Journal of Computer Assisted Radiology and Surgery 9.2 (2014): 283-293.

[11] Bernal, Jorge, Javier Sánchez, and Fernando Vilarino. "Towards automatic polyp detection with a polyp appearance model." Pattern Recognition 45.9 (2012): 3166-3182.

[12] Ganz, Melanie, Xiaoyun Yang, and Greg Slabaugh. "Automatic segmentation of polyps in colonoscopic narrow-band imaging data." IEEE Transactions on Biomedical Engineering 59.8 (2012): 2144-2151.

[13] Browet, Arnaud, P-A. Absil, and Paul Van Dooren. "Community detection for hierarchical image segmentation." International Workshop on Combinatorial Image Analysis. Springer, Berlin, Heidelberg, 2011.

[14] Tajbakhsh, Nima, Suryakanth R. Gurudu, and Jianming Liang. "Automated polyp detection in colonoscopy videos using shape and context information." IEEE transactions on medical imaging 35.2 (2015): 630-644.

[15] Tajbakhsh, Nima, Suryakanth R. Gurudu, and Jianming Liang. "A classification-enhanced vote accumulation scheme for detecting colonic polyps." International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging. Springer, Berlin, Heidelberg, 2013.

[16] Bernal, Jorge, et al. "Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge." IEEE transactions on medical imaging 36.6 (2017): 1231-1249.

[17] Brandao, Patrick, et al. "Fully convolutional neural networks for polyp segmentation in colonoscopy." Medical Imaging 2017: Computer-Aided Diagnosis. Vol. 10134. International Society for Optics and Photonics, 2017.

[18] Zhang, Lei, Sunil Dolwani, and Xujiong Ye. "Automated polyp segmentation in colonoscopy frames using fully convolutional neural network and textons." Annual Conference on Medical Image Understanding and Analysis. Springer, Cham, 2017.

[19] Shin, Younghak, et al. "Automatic colon polyp detection using region based deep cnn and post learning approaches." IEEE Access 6 (2018): 40950-40962.

[20] Hashemi, Seyed Raein, et al. "Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection." IEEE Access 7 (2018): 1721-1735.

[21] Akbari, Mojtaba, et al. "Polyp segmentation in colonoscopy images using fully convolutional network." 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018.

[22] Kang, Jaeyong, and Jeonghwan Gwak. "Ensemble of Instance Segmentation Models for Polyp Segmentation in Colonoscopy Images." IEEE Access 7 (2019): 26440-26447.

[23] Qadir, Hemin Ali, et al. "Polyp Detection and Segmentation using Mask R-CNN: Does a Deeper Feature Extractor CNN Always Perform Better?." 2019 13th International Symposium on Medical Information and Communication Technology (ISMICT). IEEE, 2019.