Polyp segmentation on colonoscopy image using improved Unet and transfer learning

Le Thi Thu Hong[1*], Nguyen Sinh Huy[1], Nguyen Duc Hanh[1], Trinh Tien Luong[1],
Ngo Duy Do[1], Le Huu Nhuong[2], Le Anh Dung[2]

[1]Military Information Technology Institute, Academy of Military Science and Technology;
[2]Military Medical Hospital 354/General Department of Logistics.
[*]Corresponding author: lethithuhong1302@gmail.com

Received 14 Sep 2022; Revised 7 Dec 2022; Accepted 15 Dec 2022; Published 30 Dec 2022.
DOI: https://doi.org/10.54939/1859-1043.j.mst.CSCE6.2022.41-55

ABSTRACT

Colorectal cancer is among the most common malignancies and can develop from high-risk colon polyps. Colonoscopy remains the gold-standard investigation for colorectal cancer screening. The procedure could benefit greatly from using AI models for automatic polyp segmentation, which provide valuable insights for improving colon polyp dection. Additionally, it will support gastroenterologists during image analysation to correctly choose the treatment with less time. In this paper, the framework of polyp image segmentation is developed by a deep learning approach, especially a convolutional neural network. The proposed framework is based on improved Unet architecture to obtain the segmented polyp image. We also propose to use the transfer learning method to transfer the knowledge learned from the ImageNet general image dataset to the endoscopic image field. This framework used the Kvasir-SEG database, which contains 1000 GI polyp images and corresponding segmentation masks according to annotation by medical experts. The results confirmed that our proposed method outperforms the state-of-the-art polyp segmentation methods with 94.79% dice, 90.08% IOU, 98.68% recall, and 92.07% precision.

Keywords: Artificial Intelligence; Colonoscopy; Polyp Segmentation; Transfer Learning; Unet.

1. INTRODUCTION

Colorectal cancer (CRC) is one of the most common causes of cancer-related death in the world for both men and women, with 576,858 deaths (accounting for 5.8% of all cancer deaths) worldwide in 2020 [1]. Colorectal polyps are irregular cell growth from the mucous membrane in the gastrointestinal (GI) tract that are forerunners of colorectal cancer. According to anatomical findings, the structure of polyps is distinguished from normal mucosa by color, size, and surface type. The surface of polyps can be flat, elevated, or pedunculated based on a change in the gastrointestinal tract [2]. Colonoscopy is the primary method for colorectal cancer screening. However, colonoscopy suffers from human errors and failure to fully recognize polyps [3]. Automatic polyp detection is highly desirable for colon screening due to the polyp miss rate by physicians during colonoscopy.

The computerized algorithms for polyp detection are divided into the classification of polyps against non-polyp and pixel-polyp segmentation. Segmentation of polyps on colonoscopy images is an image semantic segmentation task in which image pixels are binary classified, either into polyp class pixels or non-polyp class pixels. Figure 1 is an illustration of the polyp segmentation. The segmentation of colonoscopy images is an effective modality to obtain regions of interest (ROIs) that contain a polyp. The ROI detection in each image is based on pixel distributions for improving the polyp diagnosis with less time. Over the past years, researchers have made several efforts to develop Computer-Aided Diagnosis(CADx) prototypes for automated polyp segmentation. Most of the prior polyp segmentation approaches were based on analyzing polyp color, texture, shape, or edge information to segment polyp regions. More recently, deep neural networks have been widely used to solve medical image segmentation problems, including polyp segmentation. The CADx system for automatically segmenting out polyps from normal mucosa on colonoscopy images can be an effective clinical tool that helps endoscopists for faster screening and higher accuracy.

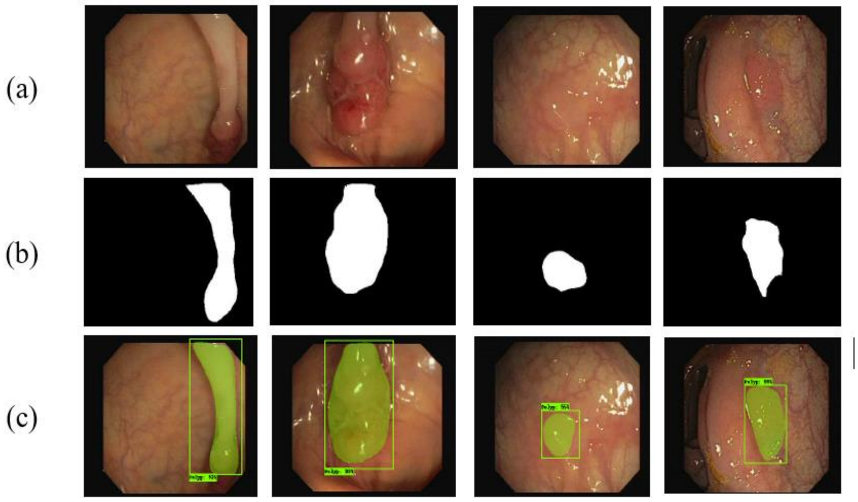


Figure 1. *Polyp segmentation: (a) Input image, (b) Results of polyp segmentation, (c): Visual display of polyp segmentation.*

Among various deep learning models, UNet [4] and its variants have demonstrated impressive performance in biomedical image segmentation. Motivated by the success of UNet, in this work, we propose a novel polyp segmentation method based on the UNet architecture. We aim to evaluate different CNN architectures (e.g. MobileNet [5], Resnet[6], and EfficentNets [7]) as the encoder 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. We also use the transfer learning method to transfer the knowledge learned from the ImageNet general image dataset to the endoscopic image field. We perform experiments using recent public datasets for polyp segmentation: Kvasir-SEG [8] for training our model and CVC-ColonDB [9], EITS-Larib [10] for testing. Finally, we evaluate our proposed method and compare it with state-of-the-art (SOTA) approaches.

The rest of the article is organized as section 2 reviews related research. In section 3, we describe our proposed method of polyp segmentation using Unet in detail. Section 4 outlines our experiment settings, experimental results, and discussion. Finally, in section 5, we summarize and conclude this work.

2. RELATED WORKS

The deep learning-based approach for polyp segmentation has gained much attention in recent years due to the automatic feature extraction process to segment polyp regions with unprecedented precision. Qadir et al. [11] proposed using Mask-RCNN incorporated with traditional CNN-based feature extractors to provide bounding boxes of the polyp regions. Kang and Gwak [12] used Mask-RCNN, which relies on ResNet50 and ResNet101, as a backbone structure for automatic polyp detection and segmentation. Akbari et al. [13] applied FCN network to polyp segmentation and combined Otsu thresholding to select the largest connected region. Sun et al. [14] utilized Unet architecture for polyp segmentation and further introduced a dilated convolution to learn high-level semantic features without resolution reduction. Zhou et al. [15] proposed UNet++ to redesign skip pathways and achieve better performance in polyp segmentation. Jha et al. [16] also propose ResUNet++, which takes advantage of residual blocks, squeeze and excitation units, ASPP, and the attention mechanism. Wang et al. [17] used the SegNet architecture to detect polyps in real-time and with high sensitivity and specificity. Afify et al. [18] presented an improved framework for polyp segmentation based on image preprocessing and two types of SegNet architecture. Despite the significant progress made by these methods, the performance of polyp segmentation is still limited by the small size of polyp databases, which require expensive and time-consuming manual labelling.

3. PROPOSED METHOD

3.1. Overview of the proposed method

The overall proposed method, which adapts U-net to segment polyp automatically, is depicted in figure 2.

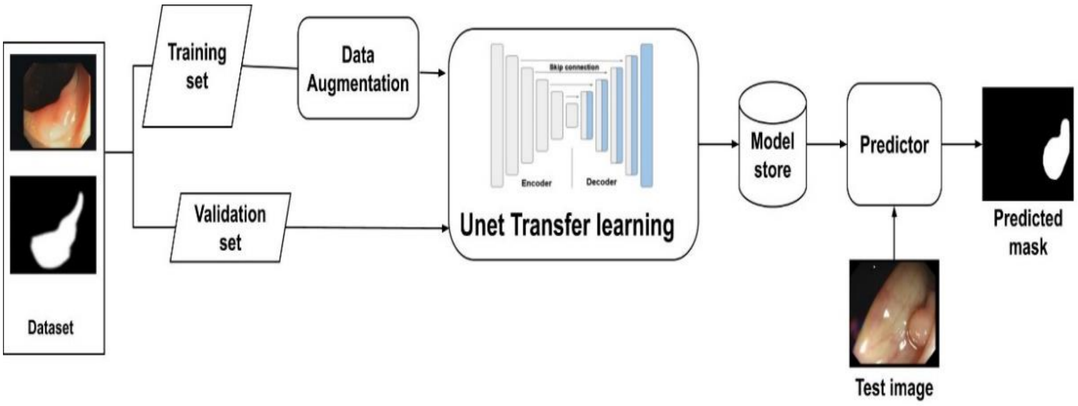


Figure 2. *Flowchart of the proposed polyp segmentation framework.*

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 B7 for our polyp segmentation framework because of the highest performance. We adopt a transfer learning approach with UNet architecture for polyp segmentation by using UNet with a CNN model pre-trained on the ImageNet dataset as the encoder.

To train the polyp segmentation network, we use a public polyp segmentation dataset consisting of colonoscopy images and their corresponding pixel-level annotated polyp masks that were annotated by colonoscopists. The asymmetric similarity loss function [19] is used for training networks to address the unbalanced data problem. The asymmetric similarity loss function is defined as:

$$\mathcal{L}_{AsymCE} = \alpha * \mathcal{L}_{CE} + \mathcal{L}_{Asym} \quad (1)$$

where $\mathcal{L}_{CE}$ is cross-entropy loss, $\mathcal{L}_{Asym} = 1 - F_\beta$ is asymmetric similarity loss which is based on F_β score and the hyperparameter α controls the amount of cross-entropy loss term contribution in the loss function. F_β score is defined as:

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

$\boldsymbol{F_\beta}$ score with the hyper-parameter $\boldsymbol{\beta}$ generalizes Dice similarity coefficient and Jaccard (IoU) index. When $\boldsymbol{\beta = 1}$, the $\boldsymbol{F_\beta}$ score is Dice score, $\boldsymbol{\beta = 2}$ generates F2 score, and $\boldsymbol{\beta = 0}$ transforms the score to precision.

3.2. Improved Unet for polyp segmentation

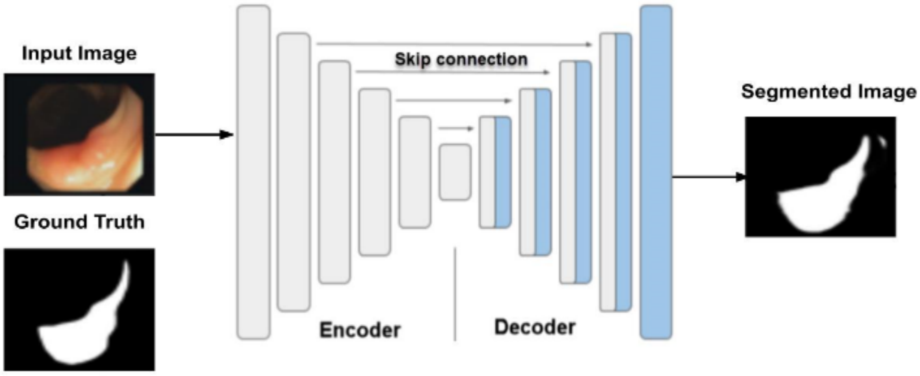


Figure 3. *Unet for polyp segmentation.*

The U-net was developed by Olaf Ronneberger et al. for BioMedical Image Segmentation [4], which 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 the outputs of the encoding layers are passed directly to the decoding layers so that all the important pieces of information can be preserved. Figure 3 depicts the general architecture of the Unet.

This work we improve Unet for polyp segmentation. We improved Unet with using pre-trained CNN as encoder. We select three pre-trained CNN as encoder to compare and evaluate their performance in polyp segmentation: MobileNet [9], ResNet [10], and EfficientNet [11]. MobileNet is a family of mobile-first computer vision models from Google. They are designed to maximize accuracy while being mindful of the restricted resources for an on-device or embedded application. ResNet is a residual learning framework that enables training deep networks easily. With ResNet, we can benefit from deeper CNN networks to obtain an even higher level of essential features for challenging tasks such as polyp segmentation. EfficientNets are the latest family of image classification models from Google, which achieves the state of the art accuracy on ImageNet. Mingxing Tan and Quoc V. Le [11] proposed the 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. The accuracies of networks are steadily increasing while maintaining a relatively small size from EfficientNet-B0 to EfficientNet-B7. The test results show that UNet with EfficientNet-B7 encoder gives the highest accuracy, so EfficientNet-B7 encoder is chosen. Figure 4 is architech of EfficientNet-B7.

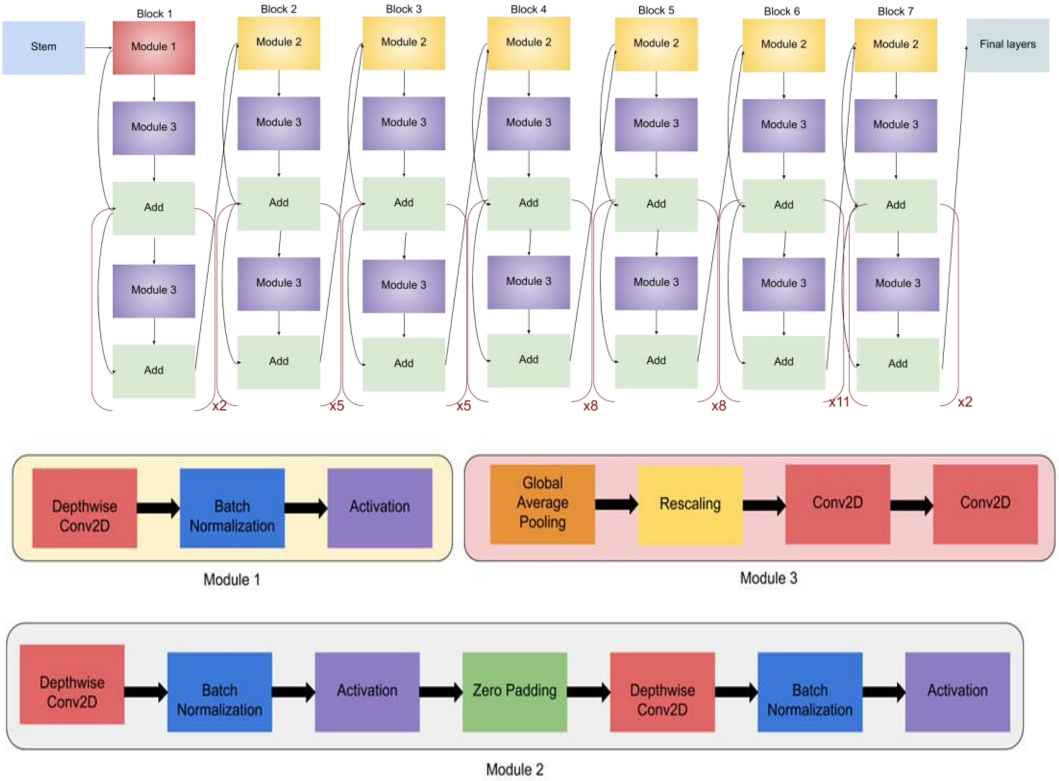


Figure 4. *EfficientB7 encoder architecture.*

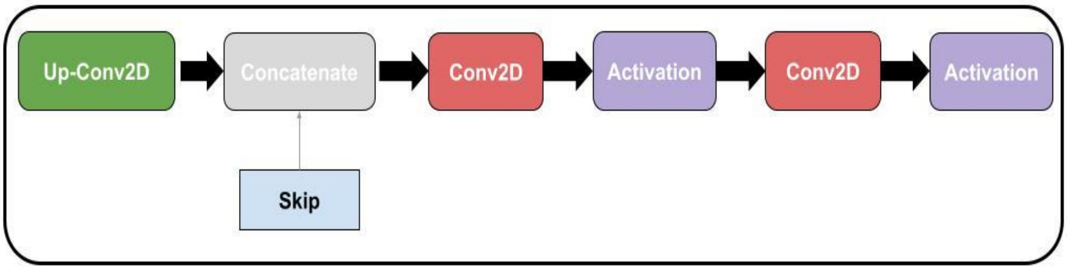


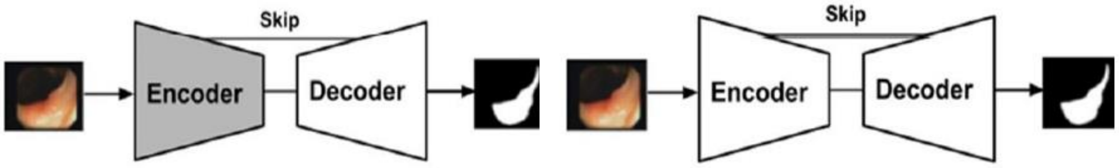
Figure 5. *Upsample2D block.*

The decoder of Unet for polyp segmentation is the encoder of the original Unet. The decoder consists of 4 Upsample2D blocks, each block consists of: Deconvolution layer with stride 2; Concatenation with the corresponding output of the skip layer from the encoder. To get better precise locations, at every step of the decoder we use skip connections by concatenating the output of the transposed convolution layers with the feature maps from the Encoder at the same level; and two 3 x 3 convolution layers +

ReLU activation function (with batch normalization). Figure 5 depicts the Upsample2D block. We implement four skip connections between the encoder and decoder of the proposed Unet network for polyp segmentation by Concatenation of Upsample2D block. Depending on different encoder architectures, skip layers are defined differently. With EfficientNet B7 encoders skip layers are: activation layers in Modul 2 of Block 6, Block 4, Block 3 and Block 2.

3.3. Transfer learning for polyp segmentation

The reuse of a pre-trained model on a new problem is known as transfer learning in machine learning. A machine uses the knowledge learned from a prior assignment to increase prediction about a new task in transfer learning. In computer vision, neural networks typically aim to detect edges in the first layer, forms in the middle layer, and task-specific features in the latter layers. The early and central layers are employed in transfer learning, and the latter layers are only retrained. It makes use of the labelled data from the task it was trained on. Transfer learning offers a number of advantages, the most important of which are reduced training time, improved neural network performance, and the absence of a large amount of data. To train a neural model from scratch, a lot of data is typically needed, but access to that data isn't always possible. Because the model has already been pre-trained, a good machine learning model can be generated with fairly little training data using transfer learning. This is especially useful in medical image analysis, where huge labelled datasets require a lot of expert knowledge. In this work, we use the transfer learning method on the improved Unet for polyp segmentation to achieve better performance. We use UNet with a CNN model pre-trained on the ImageNet dataset as the encoder. We investigate two different methods of transfer learning for polyp segmentation. The first method is to freeze the weights learned at the encoder and only finetune the decoder. The second way is to finetune all the weights, including the encoder, decoder. Figure 6 illustrates these methods, the gray area denotes the freezing of the weights learned from the pretext task.



(a) freezing the encoder and only finetunes the decoder

(b) finetuning all network

Figure 6. *Transfer learning for polyp segmentation.*

4. EXPERIMENTAL METHOD

4.1. Dataset

The proposed method was evaluated on three publicly available datasets: Kavasir-Seg [8], CVC-ColonDB [9], and ETIS-Larib Polyp DB [10]. Kavasir-SEG contains 1000 polyp images and their corresponding ground truth. The resolution of the images contained in Kvasir-SEG varies from 332x487 to 1920x1072 pixels. CVC-ColonDB contains 379 colonoscopy images with resolution 574 x 500, generated from 15 different video sequences, while ETIS-LaribPolypDB contains 196 colonoscopy images with a resolution of 1225 x 966, generated from 34 different video sequences. Each video sequence represented a subject (polyp) and background. These datasets are summarised in table 1. All of the colonoscopy images were associated with manually annotated polyp masks drawn by experts, and each image contained at least one polyp that was associated with its own individual polyp masks. In addition, we used an unlabeled dataset collected at 354 hospital to test and evaluate the accuracy of the proposed polyp segmentation model.

4.2. Implementation

The proposed models are implemented using Keras and Tensorflow backend. All algorithms have been programmed/trained on a PC with a GeForce GTX 1080 Ti GPU. The segmentation network is updated via Adam optimizer, the learning rate of Adam is set to 0.0001. All the training data is divided into mini-batches for network training, and the mini-batch size is set as four during the training stage. Data augmentation was performed on the fly, including vertical flipping, horizontal flipping, random rotation, random scaling, random shearing, random Gaussian blurring, random brightness, and random cropping and padding for training the proposed polyp segmentation network. The model is trained 200 epoch and the model generated at the epoch with a max dice value on the validation set is the final polyp segmentation model.

Table 1. *Details of datasets used for training and testing.*

Dataset	**Images**	**Training**	**Testing**	**Image Resolution**	**Label**
Kvasir-Seg [8]	1000	800	200	Varies from 332x487 to 1920x1072	Polyp mask
CVC-ColonDB [9]	300		300	574x500	Polyp mask
ETIS-Larib Polyp DB [10]	196		196	1225x996	Polyp mask
354 Hospital	*687*		*687*	*720x576*	*Unlabed*

4.3. Evaluation metric

For evaluation of polyp segmentation, we use common segmentation similarity score Dice coefficient as the main metric. Furthermore, in order 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), specificity (Spec), and accuracy 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 polyp 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:

$$Dice = \frac{2PR \cap GT}{|PR|+|GT|} \tag{3}$$

$$IoU = \frac{PR \cap GT}{PR \cup GT} \tag{4}$$

$$Re = \frac{TP}{TP+FN} \tag{5}$$

$$Prec = \frac{TP}{TP+FP} \tag{6}$$

5. RESULTS AND DISCUSSION

5.1. Performance evaluation on pretrained CNN as Encoders

In this section, we report the performance of U-net models for polyp segmentation with different pre-trained CNNs as encoders. We conduct experiments on Kvarsir-Seg dataset. The dataset is split 80/10/10 for training, validation, and testing. That is, there are 800 training images, 100 validating images and 100 test images. Several encoders are selected to evaluate their performance in polyp segmentation. The EfficientNet family from B0 to B7, MobileNetV2, ResNet variants, including ResNet18, ResNet34, ResNet101 have been used. Table 2 presents the overall results of the experiments and figure 6 illustrates the performance of UNet with different encoders. These show that EfficientNet family backbones significantly outperform ResNet and MobileNet in terms of Dice and IoU scores; EfficientNet backbones generally perform better as size increases. UNet-EfficientNetB7 gives the best segmentation performance with 94.79% Dice and 90.93% IoU.

Table 2. *The performance of UNet with different encoders.*

Network	**Dice(%)**	**IoU(%)**	**Re(%)**	**Pre(%)**
UNet_MobileNetV2	89.76	83.48	93.11	89.6
UNet_Resnet18	83.63	74.38	88.92	83.5
UNet_Resnet34	85.62	77.74	94.17	82.7
UNet_Resnet50	77.43	64.98	80.01	77.55
UNet_Resnet101	79.15	70.63	82.19	86.43
UNet_EfficientNetB0	91.99	86.21	97.67	88.21
UNet_EfficientNetB1	92.46	86.8	98.87	87.63
UNet_EfficientNetB2	92.54	87.28	95.52	90.93
UNet_EfficientNetB3	92.92	89.01	96.53	91.93
UNet_EfficientNetB4	93.35	89.67	95.27	93.9
UNet_EfficientNetB5	94.2	90.13	99.08	90.9
UNet_EfficientNetB6	94.42	90.08	96.98	92.5
UNet_EfficientNetB7	**94.79**	**90.93**	**98.68**	**92.07**

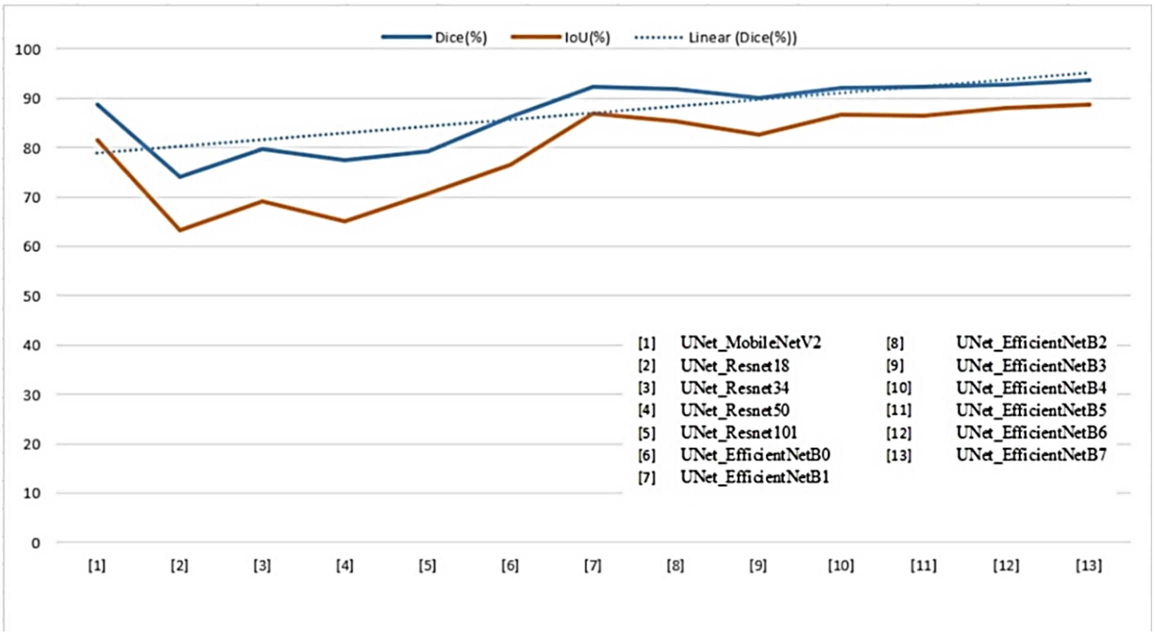


Figure 7. *The performance of UNet with different encoders.*

5.2. The effect of transfer learning

Table 3. *Comparison of Unet models trained from scratch and transfer learning.*

Network	**Trained from scratch**		**Transfer learning**			
			Fineturning Encoder		**Fineturning all network**	
	Dice (%)	**IoU (%)**	**Dice (%)**	**IoU (%)**	**Dice (%)**	**IoU (%)**
UNet_EfficientNetB0	86.03	77.85	87.09	80.81	**91.99**	**86.21**
UNet_EfficientNetB1	87.13	82.05	88.71	82.37	**92.46**	**86.8**
UNet_EfficientNetB2	87.97	81.14	88.17	81.52	**92.54**	**87.28**
UNet_EfficientNetB3	90.26	84.61	86.07	79.96	**92.92**	**89.01**
UNet_EfficientNetB4	89.35	82.88	88.93	82.49	**93.35**	**89.67**
UNet_EfficientNetB5	91.43	84.81	91.33	86.11	**94.2**	**90.13**
UNet_EfficientNetB6	91.77	85.94	91.28	87.78	**94.42**	**90.08**
UNet_EfficientNetB7	90.33	84.54	90.44	84.38	**94.79**	**90.93**

This study adopts a transfer learning approach with UNet architecture for polyp segmentation by using CNN models pre-trained on the ImageNet dataset as the encoder. To evaluate the effect of this transfer learning method, we train UNet from scratch and compare the received results with the result from the transfer learned UNet. Table 3 compares performance metrics for polyp segmentation between the UNet trained from scratch and transfer learning methods. As it shows, the performance of transfer learning method outperforms UNet trained from scratch in both IoU and Dice metrics with 4.46% in Dice and 6.39% in Dice for Unet_EfficientNet B7 the performance of models trained by the transfer learning method is significantly improved compared to those trained from scratched. In addition, when the models are deeper, the performance improvement is greater.

5.3. Comparison to exiting method

This section compares our proposed UNet_EfficientNetB7 to several recent SOTAs for polyp segmentation. We installed UNet_EfficientNetB7 and trained the model using the combined asymmetric loss function and the transfer learning method. We conduct experiments with different scenarios of training and testing data. We present and compare the results of the proposed method with existing methods in terms of learning ability, generalization capability on the same dataset, and cross-dataset.

- Results on the same datasets

We conduct two experiments to validate the model's learning ability when the training and test set are from the same dataset. The first experiment uses CVC-Clinic dataset consisting of 612 endoscopic images, and the second uses Kvasir-SEG dataset consisting of 1000 endoscopic images. These datasets are split 80/10/10 for training, validation, and testing. The results obtained will then be compared with recently published models, which have the same scenario using training and evaluation data. Table 4 and table 5 show the comparisons of the quantitative results on CVC-Clinic and Kvasir-SEG, respectively.

Table 4. *Comparison of quantitative results on CVC-ClinicDB dataset.*

Method	Dice(%)	IoU(%)	Re(%)	Pre(%)
UNet [4]	71.47	43.34	63.06	92.22
DoubleU-Net [20]	92.39	86.11	84.57	95.92
UNet++[15]	78.15	72.41	80.64	90.76
ResUNet++[16]	91.99	88.92	93.91	84.45
A-Dense UNet [22]	89.12	85.53	9448	92.66
PolypSegNet[23]	84.79	78.32	84.34	95.75
ResUNet++ CRF [21]	92.03	88.98	93.93	84.59
PraNet [24]	89.9	84.9	n/a	n/a
Unet_EfficientNetB7	**95.12**	**91.85**	**97.94**	**96.83**

Table 5. *Comparison of quantitative results on Kvasir-SEG dataset.*

Method	Dice(%)	IoU(%)	Re(%)	Pre(%)
UNet++[15]	80.21	72.15	79.14	93.21
ResUNet++[16]	81.19	80.68	85.78	7742

A-Dense UNet [22]	90.85	86.15	94.48	97.66
PolypSegNet[23]	88.72	82.86	91.68	92.54
SegNetImproved [17]	n/a	87.2	96.6	85.0
PraNet [24]	89.9	84.0	n/a	n/a
Unet_EfficientNetB7	**94.79**	**90.93**	**98.68**	**92.07**

As shown in these tables, our method outperforms all other methods in Dice and IoU metrics across both datasets. Specifically, Table 4 shows that our proposed methods achieve the best performance on CVC-Clinic dataset with Dice of 95.12% and IoU of 91.85%, outperforming the second-best ResUNet++ CRF by 3.09% in Dice and 2.87% in IoU. In table 5, on Kvasir-SEG dataset, our proposed method also gets the highest Dice of 94.79% and IoU of 90.93%. These results demonstrate that our model has a strong learning ability to segment polyps effectively.

- Result on cross-dataset

We carry out experiments with training and testing across different datasets to measure the generalization capability of the proposed method. Since different polyp datasets have different image properties and feature distributions, the models need to generalize well to have good performance. In this session, we train models on CVC-ClinicDB, Kvasir-SEG, respectively, and use the other independent datasets: ETIS-Larib, CVC-ColonDB for testing. Then, we compare the results with current works that have the same training and testing data scenarios.

Table 6 shows the results and comparison with other models using Kvasir-SEG as the training set. On ETIS-Larib test set, we obtain the best segmentation performance with 78.53% Dice, 66.95% IoU. On CVC-ColonDB test set, the proposed method gets the best results with 85.59% Dice, IoU 76.19%, recall of 88,07%, and precision of 86.78%.

Table 6. *Comparison results on cross-dataset using Kvasir-SEG as the training set.*

Method	**ETIS-Larib**		**CVC-ColonDB**	
	Dice(%)	**IoU(%)**	**Dice(%)**	**IoU(%)**
UNet [4]	60.25	n/a	66.12	n/a
UNet++[15]	58.43	n/a	65.21	n/a
ResUNet++ [16]	40.17	64.15	51.35	67.42
ResUNet++ TTA [21]	40.14	64.68	55.93	70.3
DoubleU-Net [20]	64.4	n\a	n\a	n\a
PolypSegNet[23]	71.8	n\a	n\a	n\a
Unet_EfficientNetB7	**78.53**	**66.95**	**85.56**	**76.19**

Table 7 presents the results and comparison with several SOTAs for polyp segmentation with CVC-ClinicDB as the training set. On ETIS-Larib test set, the proposed method gives the best segmentation performance with 79.37% Dice, 68.65% IoU, recall of 79.44%, and precision of 80.07%. The proposed method obtains the best results on CVC-ColonDB test set: 86.8% Dice, 77.43% IoU, recall of 86,4%, and precision of 85.52%. These results indicate that our method outperforms other SOTAs on both test sets. Especially with the CVC-ColonDB test set, our Dice score is 12.1% higher than PolypSegNet's [23], which is the second-highest method.

Table 7. *Comparison results on cross-dataset using CVC-ClinicDB as the training set.*

Method	**ETIS-Larib**		**CVC-ColonDB**	
	Dice(%)	**IoU(%)**	**Dice(%)**	**IoU(%)**
UNet [4]	57.25	n/a	65.32	n/a
U-Net++[15]	55.12	n/a	61.85	n/a
ResUNet++[16]	40.12	63.98	54.89	69.42
ResUNet++ TTA[21]	40.27	65.22	56.86	70.8
ResNet101-Mask-RCNN [11]	70.42	61.34	n\a	n\a
Ensemble Mask-RCNNs [12]	n\a	66.07	n\a	69.46
DoubleU-Net [20]	76.49	62.55	71.21	n/a
PolypSegNet[23]	68.6	n/a	74.7	n/a
Unet_EfficientNetB7	**79.37**	**68.85**	**86.8**	**77.43**

5.4. Result on 354 Hospital Dataset

We evaluate the accuracy of polyp segmentation on colonoscopy images on the 354_Hospital dataset. This is an unlabeled endoscopic image dataset collected at 354 Hospital. The dataset includes 4 colonoscopy videos with 867 colonoscopy images. The model will predict polyp segmentation on the endoscopic images of the test dataset. Predicted results are evaluated qualitatively by doctors. The evaluation metrics are defined in table 8.

Table 8. *Evalutated metrics for polyp segmentation on 354 Hospital Dataset.*

		Ground Truth	
		Polyp	None polyp
Prediction	Polyp	TP	FP
		(Right polyp segmentation)	(Wrong polyp segmentation)
	None polyp	FN	TN
		(None polyp segmentation)	(None polyp segmentation)
*Sensitive (%) =100*TP/(TP+FN)*			

Table 9 shows the results of the polyp segmentation assessment by doctors. Sensitivity (reflecting the probability that a case with polyps is correctly predicted) is quite high with averaging 86.8%, highest at 90.7%, and lowest at 85.7%.

Table 9. *Result of testing on 354 Hospital Dataset.*

	Video 1	**Video 2**	**Video 3**	**Video 4**	**Average**
Frames	94	287	213	87	681
TP	16	162	76	49	303
TN		9	32	5	46
FP	78	96	94	29	297
FN	2	27	12	5	46
Sensitive (%)	**88.9**	**85.7**	**86.3**	**90.7**	**86.8**

6. CONCLUSIONS

In this paper, we propose an improved UNet framework for polyp segmentation. We present a novel UNet-based architecture extended from UNet with the EfficientNet B7 encoder. Besides, we use the transfer learning method to train and validate the proposed method on various datasets, i.e., Kvasir-SEG, CVC-ClinicDB, CVC-ColonDB, EITS-Larib with different scenarios of using training and test data. Our experimental results show that the proposed method outperformed 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. Besides, we also continue to ensemble models to boost the performance of models.

Acknowledgments: *This research is funded by Academy of Military Science and Technology (AMST) under Logistics science research mission in 2022.*

REFERENCES

[1]. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, *“Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,”* CA, A Cancer J. Clinicians, vol. 71, no. 3, pp. 209-249, (2021).

[2]. J.-F. Rey and R. Lambert, *“ESGE recommendations for quality control in gastrointestinal endoscopy: Guidelines for image documentation in upper and lower GI endoscopy”*, Endoscopy, vol. 33, no. 10, pp. 901-903, (2001).

[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]. O. Ronneberger, P. Fischer, and T. Brox, *“U-Net: Convolutional networks for biomedical image segmentation”* in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, pp. 234-241, (2015).

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

[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]. D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, *"Kvasir-SEG: A segmented polyp dataset,"* in Proc. Int. Conf. Multimedia Modeling. Springer, pp. 451-462, (2020).

[9]. J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, *"Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer,"* Int. J. Comput. Assist. Radiol. Surg., vol. 9, no. 2, pp. 283-293, (2014).

[10]. J. Bernal, J. Sánchez, and F. Vilarino, *"Towards automatic polyp detection with a polyp appearance model,"* Pattern Recognit., vol. 45, no. 9, pp. 3166-3182, (2012).

[11]. H. A. Qadir, Y. Shin, J. Solhusvik, J. Bergsland, L. Aabakken, and I. Balasingham, *"Polyp detection and segmentation using mask R-CNN: Does a deeper feature extractorCNNalways perform better?"* in Proc. 13th Int. Symp. Med. Inf. Commun. Technol. (ISMICT), pp. 1-6, (2019).

[12]. J. Kang and J. Gwak, *"Ensemble of instance segmentation models for polyp segmentation in colonoscopy images"*, IEEE Access, vol. 7, pp. 26440-26447, (2019).

[13]. M. Akbari et al., *"Polyp segmentation in colonoscopy images using fully convolutional network,"* in EMBC. IEEE, pp. 69–72, (2018).

[14]. X. Sun, P. Zhang, D. Wang, Y. Cao, and B. Liu, *"Colorectal polyp segmentation by u-net with dilation convolution,"* in ICMLA. IEEE, pp. 851–858, (2019).

[15]. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, *"Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,"* IEEE Trans. Med. Imag., vol. 39, no. 6, p. 1856–1867, (2020).

[16]. D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen, *"ResUNet++: An advanced architecture for medical image segmentation"* in Proc. IEEE Int. Symp. Multimedia (ISM), pp. 225-2255, (2019).

[17]. P. Wang, X. Xiao, J. R. G. Brown, T. M. Berzin, M. Tu, F. Xiong, X. Hu, P. Liu, Y. Song, D. Zhang, and X. Yang, *"Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy"*, Nature Biomed. Eng., vol. 2, no. 10, pp. 741-748, (2018).

[18]. H. M. Afify, K. K. Mohammed, and A. E. Hassanien, *"An improved framework for polyp image segmentation based on SegNet architecture"*, Int. J. Imag. Syst. Technol., vol. 31, no. 3, pp. 1741-1751, (2021).

[19]. Le Thi Thu Hong, Nguyen Chi Thanh, and Tran Quoc Long, *"Polyp segmentation in colonoscopy images using ensembles of u-nets with efficientnet and asymmetric similarity loss function,"* in 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), IEEE, pp.1–6, (2020).

[20]. D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, *"Doubleu-net: A deep convolutional neural network for medical image segmentation," in 2020* IEEE 33rd International symposium on computer-based medical systems (CBMS), pp. 558–564, (2020).

[21]. D. Jha, P. H. Smedsrud, D. Johansen, T. de Lange, H. D. Johansen, P. Halvorsen, and M. A. Riegler, *"A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field andtest-time augmentation"*, IEEE J. Biomed. Health Informat., vol. 25, no. 6, pp. 2029-2040, (2021).

[22]. S. Safarov and T. K. Whangbo, *"A-DenseUNet: Adaptive densely connected UNet for polyp segmentation in colonoscopy images with atrous convolution,"* Sensors, vol. 21, no. 4, p. 1441, (2021).

[23]. T. Mahmud, B. Paul, and S. A. Fattah, *"PolypSegNet: A modified encoder-decoder architecture for automated polyp segmentation from colonoscopy images"* Comput. Biol. Med., vol. 128, Art. no. 104119, (2021).

[24]. D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, *"PraNet: Parallel reverse attention network for polyp segmentation"* in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, pp. 263-273, (2020).

TÓM TẮT

Phân vùng polyp trên ảnh nội soi đại tràng sử dụng mạng Unet cải tiến và phương pháp học chuyển giao

Polyp đại trực tràng là một trong những nguyên nhân gây ung thư đại tràng, một trong những dạng ung thư gây tỉ lệ tử vong cao. Để chẩn đoán được polyp, nội soi là phương pháp hàng đầu. Trí tuệ nhân tạo có thể sử dụng nâng cao chất lượng của phương pháp nội soi bằng cách tự động phân vùng các polyp trên ảnh nội soi hỗ trợ các bác sỹ trong quá trình chẩn đoán nội soi. Hơn nữa, nó có thể hỗ trợ các bác sỹ nâng cao chất lượng chẩn đoán với thời gian thực hiện chẩn đoán ngắn hơn. Trong bài báo này, chúng tôi đã đề xuất phương pháp tự động phân vùng polyp trên ảnh nội soi đại tràng theo hướng tiếp cận học sâu, cụ thể là sử dụng mạng nơ-ron tích chập. Mô hình đề xuất dựa trên kiến trúc mạng Unet cải tiến để phân vùng polyp trên ảnh nội soi đại tràng. Chúng tôi cũng đề xuất sử dụng phương pháp học chuyển giao để chuyển giao tri thức học được từ bộ ảnh ImageNet cho phân vùng polyp trên ảnh nội soi đại tràng. Mô hình phân vùng polyp trên ảnh nội soi đại tràng được huấn luyện sử dụng bộ dữ liệu Kvasir-SEG, bộ dữ liệu này chứa 1000 ảnh nội soi đại tràng có gán nhãn phân vùng polyp bởi các chuyên gia nội soi. Mô hình đạt được độ chính xác 94.79% dice, 90.08% IOU, 98.68% recall, and 92.07% precision. Kết quả đạt được đã khẳng định phương pháp đề xuất vượt trội so với các phương pháp phân vùng polyp trên ảnh nội soi đại tràng hiện đại gần đây.

Từ khóa: Trí tuệ nhân tạo; Nội soi đại tràng; Phân vùng polyp; Học chuyển giao; Mạng Unet.