Automatic Polyp Segmentation Using Modified Recurrent Residual Unet Network

SONG-TOAN TRAN¹, MINH-HOA NGUYEN¹, HUU-PHUC DANG¹ AND THANH-TAN NGUYEN¹

¹Department of Electrical and Electronic Engineering, Tra Vinh University, 87000 Tra Vinh, Viet Nam

Corresponding author: Song-Toan Tran (e-mail: tstoan1512@tvu.edu.vn).

This work was fully supported funded by Tra Vinh University under grant contract number 55/2022/HĐ.HDKH&D-T-DHTV.

ABSTRACT Colorectal cancer is a dangerous disease with a high mortality rate. To increase the likelihood of successful treatment, early detection of polyps is a useful solution. The Unet-architecture network model is showing success in medical image segmentation including analysis of polyps from colonoscopy images. Traditional Unet and Unet-based models are often huge, requiring training and deployment with a high-performance system. Designing models with compact size and high performance would be an important goal. In this study, we proposed to modify the Residual Recurrent Unet architecture to improve the size of the model while ensuring the model performance. The proposed model has flexibility in changing the number of filters in convolution units. By taking advantage of the strengths of residual and recurrent structures in terms of reuse of convolutional functions, the new model, therefore, was not only smaller in size but also has superior performance compared to the traditional Unet model and the others. The evaluations were performed on three public Colonoscopy image datasets: CVC-ClinicDB, ETIS-LaribPolypDB, and CVC-ColonDB. The Dice score on CVC-ClinicDB reached 94.59%, ETIS-LaribPolypDB reached 92.73% and 93.31% on CVC-ColonDB dataset. The experimental results obtained from the proposed network on datasets were better than those in recent related studies. The introduced model has a smaller size than the traditional model nevertheless has outstanding performance, therefore, it would be extremely productive for developing applications on low-performance devices.

INDEX TERMS Colonoscopy image, Medical image segmentation, Polyp segmentation, Recurrent Residual structure, Unet architecture

I. INTRODUCTION

COLORECTAL cancer has the third-highest rate of new cases of all cancers and the second-highest mortality rate in 2020 [1]. Polyps are the primary cause of colorectal cancer. Early detection of malignant polyps makes treatment more effective. Colonoscopy is one of the most effective solutions for the early detection of polyps. This makes treating and saving the patient’s life [2]. The polyps present in colonoscopy images are regularly diverse in size, shape, and color. It is, therefore, tremendously challenging to detect and distinguish the polyps that appear in the colonoscopy images [3]. This task requires plenty of time and effort from doctors. Therefore, automatic processing by the computer-aided would be the best solution to this challenge [4].

To improve the effectiveness of colorectal cancer screening and treatment, polyps inside the colon must be detected early and accurately. Accordingly, detecting and monitoring the size of polyps plays a significant role. With the developments in convolution neural networks (CNNs) and computer vision, the polyp segmentation challenge is becoming more efficient and attracting the attention of researchers [5]. Proposals for the detection and segmentation of polyps were introduced quite early using traditional image processing techniques. Karkanit et al. [6] suggested using color wavelet features to detect polyps in colonoscopy videos. Recently, Deeba et al. [7] proposed a three-part system that performs image enhancement, saliency detection, and feature extraction to classify polyps in colonoscopy images.

The rapid development of deep learning has achieved many outstanding results in the field of medical image processing as well as in the polyp segmentation problem [8]. Commonly used CNNs network architectures for medical im-
age segmentation, including Fully Convolutional Networks (FCN) [9], Unet [10], Deeplab [11]. . . have achieved superior results compared to traditional methods. Some recently proposed methods have combined multiple deep learning models to enhance the system’s accuracy [12]–[16]. Yamada et al. [12] introduced a system that used deep learning to diagnose colonoscopy in real-time. They built up the system that gets the video frames from the endoscopy system then applied the Faster R-CNN [17] method to detect and recognize the polyp in the images. The results will be displayed on the screen and warned by the sound. Guo et al. [13] proposed an ensemble model based on the pre-trained models of three famous models Unet, PSPNet [18], and SegNet [19]. They applied the transfer learning mechanism for Unet and SegNet; the VGG-16 [20] model was used as the backbone for these models. Nguyen et al. [14] proposed a multi-model deep encoder-decoder network that used multi-scale input images for multi-model. It collected more contextual information and captured the contour more precisely. Banik et al. [15] introduced a multi-model fusion for the polyp segmentation challenge. They enhanced the CNN model namely Dual-Tree Wavelet pooled CNN and the Local Gradient Weighting-embedded level set method. Two new models were fusion to achieve the results. Kang and Gwak [16] proposed an ensemble model that is based on Mask R-CNN [21] model. They used the transfer learning method with two different backbone structures on the Mask R-CNN. The results were the combination of two models. Fan et al. [22] proposed a Parallel Reverse Attention network for polyp segmentation. They used the parallel partial decoder to combine the high-level features. The global maps were generated and considered the interesting area for the next step with a reverse attention module that focuses on the boundary. This combination improved segmentation performance.

Unet is a CNN architecture that achieves outstanding results in medical image segmentation problems [23]. Many Unet-based deep learning models have been changed and upgraded to improve efficiency. Ibtehaz and Rahman [24] proposed the MultiRes-Unet that modified the Unet node structure by the residual structure and changed the skip connection by the forward structure based on the residual connection. Jha et al. [4] proposed ResUnet++ combined with Conditional Random Field and Test-time Augmentation. This network model was based on Unet combined with the residual structure [25]. Safarov and Whangbo [26] proposed A-DenseUnet that based on Unet++ structure [27]. The dilated convolution with different rates was applied. They also used an attention mechanism to enhance the contextual features. Recently, the polyp segmentation task was being interested and implemented with research on deep learning models. Mahmud et al. [28] proposed PolypSegNet by modifying the encoder-decoder architecture, the network reduced the semantic gap in skip connection of conventional Unet structure. Lin et al. [29] introduced the RefineUnet was improved from FCN architecture. The skip connection was generated with the intermediate layers. Yeung et al. [30] applied the Unet architecture combined with short-range skip connections, and deep supervision and introduced a dual attention gated. The performance of the deep learning model for polyp segmentation was compared by Trinh et al. [31], the best accuracy network required the high enough performance devices. An adaptive Markov random field was also proposed by Sasmal et al. [32] for the polyp segmentation task, this is an unsupervised learning method. The advantage of this approach is the smaller consumption time, however, the network efficiency is not better than the supervised method. In our previous research, we also proposed the TMD-Unet [33] that was based on a multiple-layer U^n-Net [34], multiple-input feature, and dense skip connection for medical image segmentation. Polyp segmentation was also used for the evaluation and testing of the TMD-Unet model. However, the evaluation is only on a dataset. In this study, we focus on analyzing and evaluating the model on many datasets with many evaluation criteria.

The combination and ensemble methods have enhanced the efficiency of the system. However, the disadvantage of these methods is that they consume a lot of computational time and memory. The Unet-based network models applied to the polyp segmentation problem have also significantly improved the performance. Nevertheless, the size of these models has not improved significantly compared to the previous models. Mingliang and Xiaolin Hu [35] introduced Recurrent Convolutional Neural Network (RCNN) that adopts the method of reusing the convolutional kernels. This improves the ability to extract context information. The main advantage of RCNN is that it can improve the size while increasing the depth of the network. It was inspired by the recurrence residual Unet network model [36] that combined the advantages of RCNN, residual structure, and Unet architecture for medical image segmentation. In this study, we proposed the mRR-Unet network model that changed connectivity in the nodes of the traditional Unet architecture. Moreover, to reduce the network size while keeping the amount of feature information. In the proposed model, we used 1x1 convolution at the output of the convolution node that doubles the number of feature maps and uses it for the subsequent layers and skip connection. In summary, our essential contributions in this study are:

- We proposed a new network model for polyp segmentation named mRR-Unet that combined the dominances of the recurrent convolution, residual structure, and Unet architecture. The proposed models show a performance improvement and a significant reduction in network model size compared to previous models.
- We implemented two versions of mRR-Unet that are mRR1-Unet and mRR2-Unet. The digits in the network model’s name represent the number of iterations of the convolutional unit in the node of the Unet architecture.
- To authenticate the effectiveness of the proposed models, we also re-built the traditional Unet and recurrence residual Unet models with different versions to make the comparison. The models have been evaluated on three popular data sets
of colonoscopy images, including CVC-ClinicDB, ETIS-LaribPolypDB, and CVC-ColonDB.

II. MODIFIED RESIDUAL RECURRENTUNET MODEL (MRR-UNET)
The proposed model is inspired by the Unet model with nodes in the encoder and decoder side built on the residual network and recurrent convolution architecture. In this study, we build up the model consisting of 5 levels that is followed the traditional Unet. In this section, the details of the proposed models are described.

A. MODEL ARCHITECTURE
FIGURE 1 shows the details of mRR-Unet. The model is based on the conventional Unet. The nodes on the Unet topology are designed to consist of a recurrent convolution layer (RCL) along with a double unit (D). In the proposed model, we included a double unit to improve the quality of the feature maps from the encoder nodes. The double unit allows increasing the number of feature maps to ensure the quality of the features. With this proposal, the network size will be significantly reduced while still ensuring the number of features for the model.

The structure of the RCLs block is depicted in FIGURE 2. Each RCLs block consists of two 3×3 convolution units $C_1$ and $C_2$. The $C_2$ is reused according to recurrent structure and combined with the residual network. The outputs of the RCLs are expressed as follow:

$$f_0 = ReLU (C_1 (I))$$

$$f_1 = ReLU (C_2 (f_0))$$

$$f_n = ReLU (C_2 (f_0 + f^{(n-1)})) \text{ with } n \geq 2$$

where $f_n$ indicates the output feature of the $n^{th}$ convolution units, $C_1(\bullet)$ and $C_2(\bullet)$ represent the convolution unit 1 and 2. Letter 'I' presents the input of the RCLs.

B. THE CONNECTIVITY
In this study, we deploy the proposed model with two versions: mRR1-Unet and mRR2-Unet, the digits '1' and '2' indicate the number of iterations of the convolution block $C_2$ in the RCLs. Let $F^i \in R^{n \times n \times k}$ is the output feature of the RCL unit, once $F^{i} = f^{n}$, $\chi^i \in R^{n \times n \times 2k}$ is the output of the node, $n \times n$ is the size of the feature map, $k$ is the filter number of the convolution in the node.

$$\chi^i = \left[C_{(1 \times 1)} (F^i) ; F^i\right]$$

The $[\bullet]$ denotes the concatenate function. The $C_{1 \times 1}(\bullet)$ is the 1x1 convolution. The number of filters in the convolution unit is the main factor to determine the network size. In this study, the 1x1 convolution was used to reduce the computation time by decreasing the parameter number while keeping the feature maps. The effectiveness of 1x1 convolution is proved in [37], [38]. In this study, by using 1x1 convolution, the output feature remains while the network size was dropped.
In the decoder nodes of mRR-Unet, the input of the nodes is defined as follow:

\[ I_D^i = B \left( \chi_E^i + U \left( \chi_D^{i-1} \right) \right) \]  

(5)

where B is the batch normalization function, \( \chi_E^i \) denotes the output feature from the \( i^{th} \) encoder node, the \( U(\bullet) \) denotes the up-convolution and \( \chi_D^{i-1} \) is the output feature from the decoder node. In this study, we built up the architecture with five layers including four encoder-decoder layers and one transition layer. The number of filters in the network nodes is \( 16 \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 128 \rightarrow 64 \rightarrow 32 \rightarrow 16 \). Details of the number of layers, convolution types, and functions used in the mRRn-Unet model are shown in TABLE 1.

### III. EXPERIMENTS

#### A. DATASETS

The datasets used in this study consist of three popular datasets on Polyp segmentation.

- **CVC-ClinicDB** [39] is a three-channel color image dataset, which includes 612 images in tiff format extracted from 25 different colonoscopy videos. The dataset was supplied by the 2015 MICCAI sub-challenge on automatic polyp detection. Most of the images are 384x288 in size. Each image contains polyps that are diverse in size and points of view. The annotation labels of the polyps are also included in the dataset.

- **CVC-ColonDB** [40] provides 300 frames from 15 short colonoscopy videos. The images are in RGB format with a size of 574x500. The annotation of the region of interest was supplied for all the images. The pixels belonging to the polyps are covered, and the ground truth images were presented in binary format. The images in the dataset cover polyps of various sizes and types.

- **ETIS-LaribPolypDB** [41] provides 196 frames with the size of 1225x966, which are extracted from 34 different videos. There are a total of 44 polyps included in the dataset. The polyps that appeared in the frames are diverse in shapes and sizes.

For training and inference, the original images were scaled to a size of 224x224. This strategy was applied to all the datasets used in this study. The downgraded size images would reduce the resolution of the image and may result in lost information. However, in this study, the networks need to have a fixed size input image and ensure synchronization. Therefore, the training, testing processes, and comparison of network performance were done in the same setting to ensure fairness when comparing and evaluating models. Because of the image number limitation, the data augmentation technique was applied in the training process. The image processing techniques used for creating the new data consist of rotation in the range of 50 degrees, height and width shift in the range of 0.2, horizontal flip, and shearing in the range of 0.5. Thereafter applying the augmentation technique, the network models were trained with a batch size of eight images and each epoch would be trained in one thousand batches. This strategy was applied to all datasets. FIGURE 3 presents some examples of the dataset used to evaluate the proposed network models.
B. EXPERIMENTAL SETTING

The early-stopping mechanism was applied to save the training time. The hyper-parameters of the training process are set as follows: 150 epochs, 0.2 of dropout rate, and $3 \times 10^{-4}$ of initial learning rate. A learning rate schedule was used to update the learning rate every ten epochs, determined by the following expression

$$\text{learning rate} = IRL \times (0.9)^{(\text{NoE}/10)}$$

where IRL is the initial learning rate, NoE shows the epochs number.

The loss function used in this study is a combination of dice-loss and weighted cross-entropy (WCE). The hybrid loss function aims to solve the imbalance between positive and negative classes. The expression of the loss function is as follow

$$L_{\text{hybrid}} = L_{DL} + L_{WCE}$$

$$L_{WCE} = -\frac{1}{P} \sum_{i=1}^{P} [(1 - \omega) y_{i\text{gt}} \log y_{i\text{pred}} + \omega (1 - y_{i\text{gt}}) \log (1 - y_{i\text{pred}})]$$

$$L_{DL} = 1 - \frac{2 \sum_{i=1}^{P} (y_{i\text{gt}} \times y_{i\text{pred}})}{\sum_{i=1}^{P} (y_{i\text{gt}} + y_{i\text{pred}}) + \varepsilon}$$

where $L_{\text{hybrid}}$ represents the total loss, $L_{DL}$ and $L_{WCE}$ show the dice loss and the weighted cross-entropy loss, respectively. The $y_{i\text{gt}}$ indicates the probability that pixel $i$ belongs to the background while the $y_{i\text{pred}}$ denotes the probability that pixel $i$ belongs to the positive class. The $P$ shows the pixel number, the $\omega$ denotes the weight of the foreground class that was 0.3 in this study, and the $\varepsilon$ is the small value that prevents the infinitive problem that was 10-8.

All experiments were performed on a personal computer with NVIDIA GeForce RTX 3070 Graphics Processing Unit, Intel Core i7 11370H CPU, and 24GB of RAM.

C. EVALUATION METRICS

The metrics used in this study include Dice Score (Dice), F1-score, mean of Intersection over Union (mIoU), Precision, and Recall. The metrics are expressed as follow

$$\text{Dice}(Y_{\text{true}}, Y_{\text{pred}}) = \frac{2 |Y_{\text{true}} \cap Y_{\text{pred}}|}{|Y_{\text{true}}| + |Y_{\text{pred}}|}$$

$$\text{mIoU}(Y_{\text{true}}, Y_{\text{pred}}) = \frac{|Y_{\text{true}} \cap Y_{\text{pred}}|}{|Y_{\text{true}} \cup Y_{\text{pred}}|}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where $Y_{\text{true}}$ indicates the case of truth values while $Y_{\text{pred}}$ denotes the predicted value number. The $TP, TN, FP, and FN$ denote true positives, true negatives, false positives, and false negatives case numbers.

IV. SEGMENTATION RESULTS

To prove the effectiveness of the proposed network model, we employed five different models that include traditional Unet, residual recurrent Unet with filter number of 16 and $t = 1$ (RR16-1), RR16-12 ($t = 2$), residual recurrent Unet with filter number of 32 (RR32-1), and RR32-2 ($t = 2$). All models are trained and evaluated with the same settings. We conducted the training and inference process fifteen times for every model on every dataset to ensure the precision.

A. THE SEGMENTATION RESULT ON THE CVC-CLINICDB DATASET

The segmentation results of the models on CVC-ClinicDB dataset are shown in TABLE 2. Observed in TABLE 2, the mRR2 model achieved the best value on all metrics (except Precision value, RR16-2 reached the best value of...
94.64%). Compared to Unet, the evaluated values increased by 2.68%, 2.64%, 4.6%, 3.18%, and 1.81% on Dice score, F1-score, mIoU, Recall, and Precision, respectively. The RR32-2 reached the second-best value in the models. The mRR1 has a smaller value than the RR32-2 but still outperforms traditional Unet and other models. The performance of mRR1 was not better than that of RR32, but the model size was significantly smaller.

B. THE SEGMENTATION PERFORMANCE ON THE ETIS-LARIBPOLYPDB DATASET

TABLE 3 shows the performance of the network models on the ETIS-LaribPolypDB dataset. Similar to the case on the CVC-ClinicDB dataset, on the ETIS-LaribPolypDB dataset, the mRR2 models achieved superior results than traditional Unet (11.11% increased on Dice score, 11.12% on F1-score, 16.7% on mIoU, 19.53% on Recall, and 8.1% on Precision). Both mRR1 and mRR2 have better results than the residual recurrence models on three crucial metrics Dice, F1-score, and mIoU. The residual recurrence models also outperform the conventional Unet. In summary, the mRR models achieved the best results on the ETIS-LaribPolypDB dataset.

C. THE RESULTS ON THE CVC-COLONDB DATASET

TABLE 4 shows the segmentation results on the CVC-ColonDB dataset. In TABLE 4, we can observe that the proposed model has the best performance on the metrics (except the Recall, the RR32-2 achieved the best of 94.12%). Comparing mRR2 and Unet, the results show the superiority of mRR2: improved of 3.22% on Dice, 3.09% on F1-score, 4.93% on mIoU, 4.5% on Recall, and 1.11% on Precision. The Residual Recurrence models also achieved better results than Unet. However, the mRR models still achieved better results.

TABLE 5 represents the calculated p-value with the 95% confidence interval between mRR with traditional Unet and RRUnet. mRR2 achieved a statistically significant value compared with the others. For mRR1, when compared with RR32-1 and RR32-2 is nonsignificant on the CVC-ClinicDB dataset. The p-value statistics table shows the reliability of the proposed model.

D. THE QUANTITATIVE RESULTS

FIGURE 4 illustrates the quantitative results of the models implemented in this study. For all cases, the Unet model performed the worst. The mRR2 model segmented the best results on both small and large polyps. In some cases, when the polyp size is small, traditional Unet cannot recognize the polyps, while RRUnet and mRR can both segment well on small polyps. As seen in Figure 4, mRR showed better segmentation results than other models. With extensive and differentiated polyps, the segmentation results of the
TABLE 5. The p-value was calculated between mRR-Unet on Dice score with the other.

|                  | $U_{\text{net}}$ | $R_{16} - 1$ | $R_{16} - 2$ | $R_{32} - 1$ | $R_{32} - 2$ | $m_{\text{RR}}$ |
|------------------|------------------|---------------|---------------|---------------|---------------|-----------------|
| CVC-ClinicDB     | $m_{\text{RR}}$1 | $1.50 \times 10^{-13}$ | $4.36 \times 10^{-2}$ | $8.46 \times 10^{-2}$ | $57.2 \times 10^{-2}$ | $36.8 \times 10^{-2}$ | -               |
|                  | $m_{\text{RR}}$2 | $1.51 \times 10^{-15}$ | $6.99 \times 10^{-9}$ | $3.37 \times 10^{-9}$ | $5.03 \times 10^{-4}$ | $3.10 \times 10^{-4}$ | $5.76 \times 10^{-6}$ |
| ETIS-LaribPolyDB | $m_{\text{RR}}$1 | $1.92 \times 10^{-20}$ | $4.70 \times 10^{-8}$ | $1.52 \times 10^{-3}$ | $1.89 \times 10^{-14}$ | $9.98 \times 10^{-6}$ | -               |
|                  | $m_{\text{RR}}$2 | $1.38 \times 10^{-20}$ | $2.09 \times 10^{-9}$ | $2.55 \times 10^{-5}$ | $3.66 \times 10^{-15}$ | $1.93 \times 10^{-7}$ | $7.14 \times 10^{-2}$ |
| CVC-ColonDB      | $m_{\text{RR}}$1 | $5.17 \times 10^{-15}$ | $7.84 \times 10^{-15}$ | $2.38 \times 10^{-3}$ | $2.98 \times 10^{-3}$ | $55.6 \times 10^{-3}$ | -               |
|                  | $m_{\text{RR}}$2 | $2.84 \times 10^{-17}$ | $6.95 \times 10^{-14}$ | $8.1 \times 10^{-10}$ | $1.29 \times 10^{-8}$ | $7.83 \times 10^{-5}$ | $3.03 \times 10^{-5}$ |

FIGURE 4. The illustration of some quantitative results on the datasets. The first and second rows show the results on the CVC-ClinicDB dataset, the next two rows indicate the CVC-ColonDB dataset results, and the last two rows show the ETIS-LaribPolyDB dataset results.

model are consistently above 95%. In the complex lighting conditions and noise environment, the analysis of polyps is complicated. However, the proposed model always shows reliable segmentation results in most cases.

V. DISCUSSION

A. TRAINING CURVES

FIGURE 5 depicts the training process of the models on the CVC-ClinicDB dataset. After the number of epochs, the loss and dice scores of the models have reached convergence, and the mRR model has a faster convergence speed than the traditional Unet. FIGURE 5(a) is the learning curve of
the models over the validation loss value. The mRR model shows stability during training compared to the other models. The residual recurrence models also show more stability than conventional Unet. The validation loss value also reached the smallest value. FIGURE 5(b) depicts the training curve on the validation dice value. The proposed model achieves a better validation value than Unet and residual recurrence. The residual recurrence models also outperform Unet, however, the proposed model has the best results.

![Validation Dice Curve](image1.png)

![Validation Loss Curve](image2.png)

**FIGURE 5.** The learning curve of the training process: (a) the validation dice curve and (b) the validation loss on the CVC-ClinicDB dataset.

### B. FEATURE MAP VISUALIZATION

Convolutional units with many filters will result in a significant increase in the network size. For the datasets with few samples, it can easily lead to overfitting problems. In this study, we have proposed to reduce the number of filters in the convolution layers in the nodes of the Unet architecture. This helps to reduce the network size while still ensuring the features for the model. FIGURE 6 illustrates the images of the feature map over the network nodes in the Unet structure. Observing FIGURE 6, we can see that the features passing through the network nodes of mRR, still ensure complete information, thereby leading to the quality assurance results compared to Unet.

TABLE 6 compares the inference time between the proposed models with the conventional Unet and RRUnet. The proposed model is smaller than traditional Unet (3.6 million of mRR1 and 4.6 million of mRR2 compared to 7.8 million of Unet). Compared to RR32-2, the mRR2 has a size reduced by nearly 3.2 times. Similarly, RR32-1 compared with mRR1 is three times larger in size. The reduced network size leads to a reduction in training and inference time. The mRR2 training time was reduced by 85 seconds per epoch for RR32-2. The mRR2 per-image inference time is also reduced by 7ms for the RR32-2 and 1.5 milliseconds for the traditional Unet.

**FIGURE 7** illustrates the comparison between the models’ size, performance, and inference time. As shown in **FIGURE 7**, although the size of the mRR model is smaller than that of the Unet and RR32 models, the performance is superior. The smaller the model size (smaller bubbles), the less inference time, but the performance of mRR1 and mRR2 is better than the rest of the models. The RR32 models are much larger than the rest, but the evaluation results are not superior. On all datasets, the proposed model has shown to be effective, which proves the effectiveness of the proposed model. With the modification in the node output, the model size is reduced, and the feature maps are used more efficiently.

### C. COMPARISON WITH CURRENT METHODS

TABLE 7 compares the performance of the proposed models with current models on the datasets used in the study. Many Unet-based models perform polyps’ segmentation. In this study, the proposed models are compared with DoubleUnet [42], ResUnet++ [4], TMD-Unet [33], MED-net [14], FCN-8s [43]. Observed in TABLE 7, the segmentation results of mRR2 are better than previous models on most of the evaluation metrics. With the single-model method, the proposed methods have smaller sizes on all datasets while still achieving better segmentation results.

### VI. CONCLUSION

This study proposed a new network model that modified the recurrence residual Unet. The proposed model is based on reusing convolutional units, which reduced the network size while still achieving better results. The number of filters in the convolution node is a problem for feature map optimization. In this paper, we have used double blocks to help reduce the network size while preserving the quality of
| Datasets          | Models                  | Metrics | Dice  | F1-score | mIoU  | Recall | Precision |
|-------------------|-------------------------|---------|-------|----------|-------|--------|-----------|
|                    |                         |         |       |          |       |        |           |
| CVC-ClinicDB      | Unet                    |         | 91.91 | 92.12    | 85.20 | 91.77  | 92.82     |
|                   | DoubleU-Net [42]        |         | 92.39 | 89.91    | 86.11 | 84.57  | 95.92     |
|                   | TMD-Unet [33]           |         | 92.65 | 92.94    | 84.41 | 92.25  | 93.76     |
|                   | MED-net [14]            |         | 91.30 | -        | 86.11 | -      | 95.30     |
|                   | ResUnet++ [4]           |         | 92.03 | -        | 88.98 | 93.93  | 84.59     |
|                   | mRR-2                   |         | 94.59 | 94.76    | 89.80 | 94.95  | 94.63     |
| ETIS-LaribPolypDB | Unet                    |         | 81.62 | 81.72    | 69.84 | 72.41  | 85.80     |
|                   | ResUnet++ [4]           |         | 63.64 | -        | 75.34 | 63.46  | 64.67     |
|                   | MED-net [14]            |         | -     | -        |     | -      | 90.93     |
|                   | Unet with Dilated Conv. [44] |     | 62.54 | 80.86    | -    | 81.25  | 80.48     |
|                   | mRR-2                   |         | 92.73 | 92.84    | 86.54 | 91.94  | 93.90     |
| CVC-ColorDB       | Unet                    |         | 90.09 | 90.32    | 82.61 | 89.16  | 92.11     |
|                   | FCN-8s [43]             |         | 70.14 | -        | -    | 75.66  | -         |
|                   | MED-net [14]            |         | 90.80 | -        | -    | 79.80  | -         |
|                   | ResUnet++ [4]           |         | 84.74 | -        | 84.66 | 84.34  | 81.18     |
|                   | mRR-2                   |         | 93.31 | 93.41    | 87.54 | 93.66  | 93.22     |

**FIGURE 6.** The illustration of the feature maps of the nodes of Unet, mRR1, and mRR2 models.

the feature map. Our proposed model obtains better results with a smaller size. This proves the effectiveness of the proposed model for segmenting polyps. Applying models with a smaller size but still ensuring performance will help research on devices with limited performance.

**REFERENCES**

[1] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA. Cancer J. Clin., vol. 71, no. 3, pp. 209–249, May 2021.
[2] L. F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F. M. Sánchez-Margallo, and J. B. Pagador, "Deep learning to find colorectal polyps in colonoscopy: A systematic literature review," Artif. Intell. Med., vol. 108, p. 101923, Aug. 2020.
[3] J. Bernal et al., "Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge," IEEE Trans. Med. Imaging, vol. 36, no. 6, pp. 1231–1249, Jun. 2017.
[4] D. Jha et al., "A Comprehensive Study on Colorectal Polyp Segmentation With ResUnet++, Conditional Random Field and Test-Time Augmentation," IEEE J. Biomed. Health Inform., vol. 25, no. 6, pp. 2029–2040, Jun. 2021.
FIGURE 7. The comparison of the models’ size, performance, and inference time. The bigger bubbles indicate the larger size models.

[5] D. Vázquez et al., “A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images,” J. Healthc. Eng., vol. 2017, pp. 1–9, 2017.

[6] S. A. Karkanis, D. K. Lakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras, “Computer-aided tumor detection in endoscopic video using color wavelet features,” IEEE Trans. Inf. Technol. Biomed., vol. 7, no. 3, pp. 141–152, Sep. 2003.

[7] F. Deeba, F. M. Bui, and K. A. Wahid, “Computer-aided polyp detection based on image enhancement and saliency-based selection,” Biomed. Signal Process. Control, vol. 55, p. 101530, Jan. 2020.

[8] J. Ker, L. Wang, J. Rao, and T. Lim, “Deep Learning Applications in Medical Image Analysis,” IEEE Access, vol. 6, pp. 9375–9389, 2018.

[9] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640–651, Apr. 2017.

[10] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015.

[11] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, Apr. 2018.

[12] M. Yamada et al., “Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy,” Sci. Rep., vol. 9, no. 1, p. 14465, Dec. 2019.

[13] X. Guo, N. Zhang, J. Guo, H. Zhang, Y. Hao, and J. Hang, “Automated polyp segmentation for colonoscopy images: A method based on convolutional neural networks and ensemble learning,” Med. Phys., vol. 46, no. 12, pp. 5666–5676, Dec. 2019.

[14] N. Q. Nguyen, D. M. Vo, and S. W. Lee, “Contour-Aware Polyp Segmentation in Colonoscopy Images Using Detailed Upsampling Encoder-Decoder Networks,” IEEE Access, vol. 8, pp. 99495–99508, 2020.

[15] D. Banik, K. Roy, D. Bhattacharjee, M. Nasipuri, and O. Krejcar, “PolypNet: A Multimodel Fusion Network for Polyp Segmentation,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–12, 2021.

[16] J. Kang and J. Gwak, “Ensemble of Instance Segmentation Models for Polyp Segmentation in Colonoscopy Images,” IEEE Access, vol. 7, pp. 26440–26447, 2019.

[17] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[18] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid Scene Parsing Network,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, Jul. 2017, pp. 6230–6239.

[19] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[20] S. Liu and W. Deng, “Very deep convolutional neural network based image classification using small training sample size,” in 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, Nov. 2015, pp. 730–734.

[21] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Oct. 2017, pp. 2980–2988.

[22] D.-P. Fan et al., “PraNet: Parallel Reverse Attention Network for Polyp Segmentation,” in Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, vol. 12266, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoeanne, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 263–273.

[23] A. S. Panayides et al., “AI in Medical Imaging Informatics: Current Challenges and Future Directions,” IEEE J. Biomed. Health Inform., vol. 24, no. 7, pp. 1837–1857, Jul. 2020.

[24] N. Itstehaz and M. S. Rahman, “MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation,” Neural Netw., vol. 121, pp. 74–87, Jan. 2020.

[25] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[26] S. Safarov and T. K. Whangbo, “A-DenseUNet: Adaptive Densely Connected U-Net for Polyp Segmentation in Colonoscopy Images with Atrous Convolution,” Sensors, vol. 21, no. 4, p. 1441, Feb. 2021.

[27] Z. Zhou, M. M. K. Siddiquie, N. Tjahjak, and J. Liang, “U-Net++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation,” IEEE Trans. Med. Imaging, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.
S.-T. Tran, C.-H. Cheng, and D.-G. Liu, “A Multiple Layer U-Net,” IEEE Access, vol. 13, no. 1, pp. 3752–3764, 2021.

P. Sasmal, M. K. Bhuyan, S. Dutta, and Y. Iwahori, “An unsupervised approach of colonic polyp segmentation using adaptive markov random fields,” Pattern Recognit., vol. 154, pp. 7–15, Feb. 2022.

S. T. Tran, C.-H. Cheng, T.-T. Nguyen, M.-H. Le, and D.-G. Liu, “TMD-Unet: Triple-Unet with Multi-Scale Input Features and Dense Skip Connection for Medical Image Segmentation,” Healthcare, vol. 9, no. 1, p. 54, Jan. 2021.

S.-T. Tran, C.-H. Cheng, and D.-G. Liu, “A Multiple Layer U-Net, U-Net for Liver and Liver Tumor Segmentation in CT,” IEEE Access, vol. 9, pp. 3752–3764, 2021.

Ming Liang and Xiaolin Hu, “Recurrent convolutional neural network for object recognition,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, Jun. 2015, pp. 3367–3375.

M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, “Recurrent residual U-Net for medical image segmentation,” J. Med. Imaging, vol. 6, no. 01, p. 1, Mar. 2019.

M. Lin, Q. Chen, and S. Yan, “Network In Network,” ArXiv13124400 Cs, Mar. 2014. [Online]. Available: http://arxiv.org/abs/1312.4400.

C. Szegedy et al., “Going deeper with convolutions,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, Jun. 2015, pp. 1–9.

J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, “WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physcians,” Comput. Med. Imaging Graph., vol. 43, pp. 99–111, Jul. 2015.

J. Bernal, J. Sánchez, and F. Vilariño, “Towards automatic polyp detection with a polyp appearance model,” Pattern Recognit., vol. 45, no. 9, pp. 3166–3182, Sep. 2012.

G. Fernández-Esparrach et al., “Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps,” Endoscopy, vol. 48, no. 09, pp. 837–842, Jun. 2016.

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), Rochester, MN, USA, Jul. 2020, pp. 558–564.

L. Zhang, S. Dolwani, and X. Ye, “Automated Polyp Segmentation in Colonoscopy Frames Using Fully Convolutional Neural Network and Textons,” in Medical Image Understanding and Analysis, vol. 723, M. Valdés Hernández and V. González-Castro, Eds. Cham: Springer International Publishing, 2017, pp. 707–717.

X. Sun, P. Zhang, D. Wang, Y. Cao, and B. Liu, “Colorectal Polyp Segmentation by U-Net with Dilatation Convolution,” in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, Dec. 2019, pp. 851–858.
THANH-TAN NGUYEN was born in Vinh Long province, Vietnam, in 1985. He graduated from Can Tho University in 2008, then received a Master’s degree in Electronic Engineering in 2015 from Ho Chi Minh City University of Technology and Education. Currently, he is a lecturer at Faculty of Engineering and Technology, Tra Vinh University, Vietnam. His research interests are design and control automation systems, embedded system, image processing, computer communication.