Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

doi:10.3233/FAIA230872

Research on Segmentation of Intestinal Polyps Based on Improved Transformer and FCN

Xinjie CHEN[1] and Chen WANG
Wuhan Textile University, WuHan 430200, China

Abstract. The rising prevalence of colon cancer necessitates early detection through colonoscopy. Deep learning segmentation technology has emerged as a reliable tool for detecting lesions, with convolutional neural networks (CNNs) leading advancements in medical image processing. However, manual polyp segmentation during colonoscopic examinations is time-consuming, highlighting the need for automated approaches. This study introduces a novel parallel branching structure to address limitations and improve contextual information and low-level detail handling. The structure extracts complementary features, enriching image analysis. Additionally, a quadratic complete volume integral branch enhances segmentation performance. We present our innovative model and methodology for polyp segmentation, advancing automated detection in colonoscopy. Experimental results demonstrate the effectiveness and robustness of our approach. By integrating deep learning techniques and leveraging our parallel branching structure, our method achieves superior segmentation accuracy for efficient and accurate polyp detection.

Keywords. colon cancer, CNN, Transformer, deep learning

1. Introduction

Colon fistula is the main cause of colon cancer, with the incidence and mortality of cancer ranked third among all malignant tumors [1], and it is widely believed that most colon cancers evolve from the colon fistula .In the past, colonoscopy was manually examined and marked by doctors, but this was very inefficient, and with the rise of machine learning, deep learning had excellent results in the division of medical images. Based on the 2022 FCN Transformer [2] algorithm, we made some improvements and introduced a new type of network MFT. The detection and classification of pulses are the two tasks that we need to optimize, and we want to improve the accuracy of colonoscopy screening, and our intelligent separation tools are crucial in solving both tasks.

We need to optimize, and we want to improve the accuracy of colonoscopy screening, and our intelligent separation tools are crucial in solving both tasks.

Medical image segmentation has evolved significantly due to recent advances in Convolutional Neural Networks (CNNs), on which excellent automatic segmentation algorithms such as Unet and ResUnet have been developed [3]. However,

[1] Corresponding Author. Xinjie CHEN, Wuhan Textile University, China;
Email: 2115363047@mail.wtu.edu.cn.

convolutional neural networks exhibit common limitations in modeling the interaction of remote semantic information. For the problems that arise, we need a better solution that not only improves the efficiency of model training and prediction, but also preserves local details. In this paper, we propose a new segmentation architecture called MFT, which consists of a fully convolutional FCB and an attention mechanism module for better segmentation of colon polyps. They are a parallel branching structure, both start from the input image, a Fully Convolutional branch which returns full size feature mapping and a Transformer branch which returns reduced size feature mapping, the reduced feature mapping will be returned to normal size size in the up-sampled module. Finally after fusion the prediction is done and a final segmentation result is obtained.

In designing the TB module, we borrowed the structure of SSFormer [4], which predicts the h/4 × w/4 spatial dimension of the segmentation map, and the module performs well in the reduced size segmentation. For the decoder in the TB module, we did not refer to the structure in SSFormer, but made some improvements, the structure of this decoder (PLD+) is progressive local decoding, and the residual module is added in the up-sampling process, which can get a better grasp of the local details. The FCB is in the form of a fully convolutional structure, which consists of residual blocks (RBs), and the residual module includes the group normalization layer [5], SiLU activation function [6] and a convolutional layer with residual connectivity [7]. The encoder and decoder use dense skip connections [8]. The PH then consists of two RB blocks and a final pixel-level prediction layer that uses a convolution with 1 × 1 kernels. In a general test, we achieved state-of-the-art performance on the Kvasir-SEG[9] and CVC-ClinicDB [10] datasets for the mDice, mIoU, mprecision, and m-recall metrics, and we trained the models on the Kvasir-SEG[9] and on the CVC-ClinicDB[10] and vice versa.

Thus, the main new contribution of this work is:

- By incorporating the new Transformer architecture to extract the most important features in the image, we demonstrated through experimental comparisons the superior performance of our segmentation module.
- To decode the features extracted by the transformer encoder, we adopted the improvement made by SSFormer [4] to the progressive local area decoder (PLD), using residual blocks (RBs) consisting of group normalization [5], SiLU activation function [6], convolutional layer, and residual connections [7].

2. Method

2.1 TB Module

The TB module takes inspiration from the current characteristics of SSFormer [4], Its structure is as shown in Figure 1 b. The encoder of this architecture uses the pyramid module pre-trained on ImageNet[11] . In the input image, the TB encoder module downsamples the image four times, obtaining progressively deeper features during the downsampling process that will contain more pixel information. The encoder ultimately returns a four-level feature pyramid that serves as input to the progressive local decoder (PLD+). In the decoder, the features obtained by the encoder's four feature maps are processed by the local emphasis module (LE) to address our network's

shortcomings in local feature representation. We used the architecture shown in Fig. 1b (PLD+) to improve the originally introduced progressive local decoder (PLD). During feature fusion, we used residual blocks for connection, which can better capture subtle changes in features while greatly reducing the number of parameters required for network training.

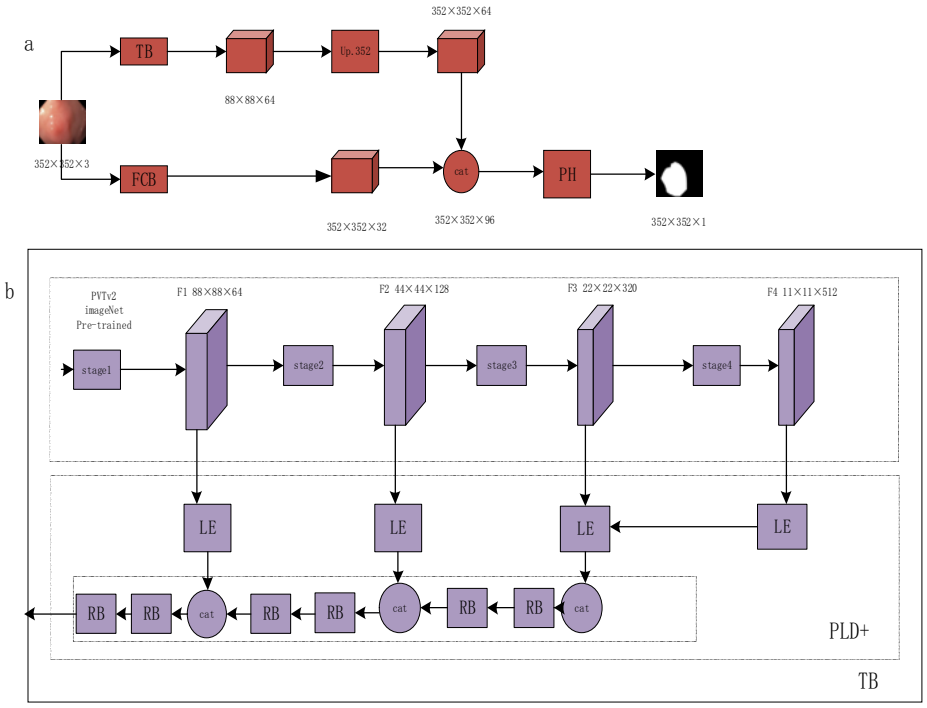

Figure 1. MFT network structure and TB branch structure

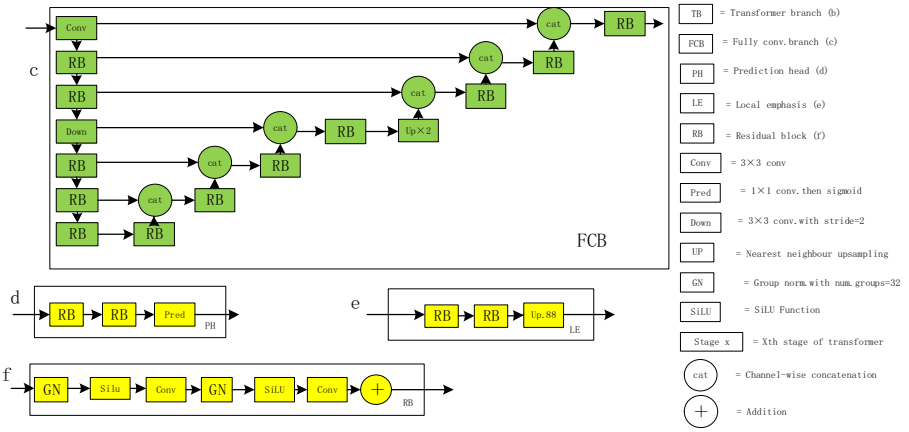

Figure 2. FCB Network Structure and Other Module Structures

2.2 FCB module

The structure of the full-volume branches as shown in Figure 2 c. Most of the residual blocks (RBs) in the fully convolutional branch play an important role in both upsampling and downsampling, as well as in skip connections. The FCB (Fully Connected Block) encoder and decoder are commonly used in semantic segmentation models for image segmentation tasks, and their main purpose is to improve the model's receptive field and reduce the number of parameters. Meanwhile, in the decoder, since the original image's resolution needs to be restored, a skip connection is also necessary. This involves adding the feature map of each downsampling layer in the encoder with the feature map of the corresponding upsampling layer in the decoder to retain more image detail information.

2.3 Prediction Module

The prediction module employs a full-sized tensor, which connects the outputs of the TB branch and the FCB branch. By combining the coarse features extracted by TB with the fine-grained features extracted by FCB, a predictive segmentation map is generated. This novel approach, which had not been used in previous experiments, has proven to be highly effective in polyp segmentation. Our experiments show that FCN and Transformer working in parallel, prior to the fusion of features and pixel-wise prediction on the fused features, provide a powerful foundation for dense prediction.

3. Experiments

To evaluate the performance of our network, we conducted experiments on two publicly available datasets: Kvasir SEG [8] and CVC ClinicDB [9], which provided both training images and labels. The Kvasir SEG dataset contains polyp images of varying sizes, while all samples in the CVC ClinicDB dataset have a spatial dimension of 288 × 384. These datasets include polyps of different shapes, which provide a strong benchmark for the development of polyp segmentation models. To determine the robustness and generalization of our network, we conducted four comparative experiments on these datasets and recorded the experimental results.

Table 1. the experimental results when both the training and testing sets are Kvasir-SEG.

Training data		Kvasir-SEG		
Test data		Kvasir-SEG		
Metric	mDice	mIoU	mPrec	mRec
ResUnet++	0.8133	0.7927	0.7064	0.8774
FCBFormer	0.9011	0.8423	0.9347	0.9029
MFT	0.9069	0.8482	0.9219	0.9195

Table 2.the experimental results when both the training and testing sets are CVC-ClinicDB.

Training data		CVC-ClinicDB		
Test data		CVC-ClinicDB		
Metric	mDice	mIoU	mPrec	mRec
ResUnet++	0.7955	0.7962	0.8785	0.7022
FCBFormer	0.9145	0.8492	0.9124	0.9225
MFT	0.9214	0.8680	0.9323	0.9162

Table 3. the experimental results when the training set is Kvasir-SEG and the testing set is CVC-ClinicDB.

Training data	Kvasir-SEG			
Test data	CVC-ClinicDB			
Metric	mDice	mIoU	mPrec	mRec
ResUnet++	0.5560	0.4542	0.6775	0.5795
FCBFormer	0.8650	0.7875	0.8546	0.9070
MFT	0.8764	0.8008	0.8705	0.9165

Table 4. the experimental results when the training set is CVC-ClinicDB and the testing set is Kvasir-SEG.

Training data	CVC-ClinicDB			
Test data	Kvasir-SEG			
Metric	mDice	mIoU	mPrec	mRec
ResUnet++	0.5147	0.4082	0.7181	0.4860
FCBFormer	0.7626	0.6741	0.9163	0.7189
MFT	0.7744	0.6798	0.8958	0.7563

3.1 Evaluation

Our segmentation model demonstrates excellent performance even on polyp images with unclear boundaries, which can be attributed to the successful combination of attention mechanisms and full convolution. The TB modules process the main structure of the polyps, while the FCB branch provides reliable full-size boundaries. We evaluate the performance of our model on each dataset using several metrics, including MDICE, MIOU, Mprecision, and MRECALL, where "M" represents the average of the values measured on the test set. These evaluation results, as shown in Table 1, Table 2, Table 3 and Table 4, indicate that MFT is superior to existing models in all indicators.

4. Conclusion

This doctoral dissertation introduces a groundbreaking network architecture named MFT, specifically designed for the task of colon polyp segmentation. The research comprises a comprehensive investigation, including four comparative experiments conducted on two distinct datasets. The experimental findings reveal substantial enhancements across four crucial metrics: mDice, mIoU, mPrec, and mRec. These notable improvements serve to emphasize the exceptional performance achieved by our proposed network structure when compared to the existing FCBFormer and ResUnet++ model.

References

[1] Salmo, E., Haboubi, N.: Adenoma and malignant colorectal polyp: pathological considerations and clinical applications. Gastroenterology 7(1), 92–102 (2018)

[2] Sanderson E, Matuszewski B J. FCN-transformer feature fusion for polyp segmentation[C]//Medical Image Understanding and Analysis: 26th Annual Conference, MIUA 2022, Cambridge, UK, July 27–29, 2022, Proceedings. Cham: Springer International Publishing, 2022: 892-907.

[3] Isensee,F.,J ¨ ager,P.F.,et al.:Automated design of deep learning methods for biomedical image segmentation.arXiv preprint arXiv:1904.08128 (2019).

[4] Wang, J., Huang, Q., Tang, F., Meng, J., Su, J., Song, S.: Stepwise feature fusion: Local guides global. arXiv preprint arXiv:2203.03635 (2022)
[5] Wu, Y., He, K.: Group normalization. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
[6] Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
[7] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016).
[8] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
[9] Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., Lange, T.d., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: International Conference on Multimedia Modeling. pp. 451–462. Springer (2020)
[10] Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics 43, 99–111 (2015)
[11] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A largescale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)