

Image Segmentation of Intestinal Polyps using Attention Mechanism based on Convolutional Neural Network

Xinyi Zheng¹, Wanru Gong², Ruijia Yang³, Guoyu Zuo^{4*}

Faculty of Information Technology, Beijing University of Technology, Beijing 100083, China^{1,2,4}
School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China³

Abstract—The intestinal polyp is one of the common intestinal diseases, which is characterized by protruding lining tissue of the colon or rectum. Considering that they may become cancerous, they should be removed by surgery as soon as possible. In the past, it took a lot of manpower and time to identify and diagnose intestinal polyps, which greatly affected the treatment efficiency of medical staff. Because the polyp part looks similar to the normal structure of the human body, the probability of human eye misjudgment is high. Therefore, it is necessary to use advanced computer technology to segment the intestinal polyp image. In the model established in this paper, an image segmentation method based on convolution neural network is proposed. The Har-DNet backbone network is used as the encoder in the model, and its feature processing results are converted into three feature images of different sizes, which are input to the decoding module. In the decoding process, each output first expands the receptive field module and then fuses the feature image processed by the attention mechanism. The fusion results are input to the density aggregation module for processing to improve the operation efficiency and accuracy of the model. The experimental results show that compared with the previous Pra-Net model and Har-DNet MSEG model, the accuracy and precision of this method are greatly improved, and can be applied to the actual medical image recognition process, thus improving the treatment efficiency of patients.

Keywords—Image segmentation; intestinal polyps; block convolutional attention mechanism; Har DNet

I INTRODUCTION

Rectal polyps generally refer to the tissues protruding from the surface of the rectal mucosa to the intestinal cavity, which is also a high-risk factor for rectal cancer. Studies have pointed out that 95% of rectal cancer is caused by colorectal adenocarcinoma. To effectively avoid the incidence of rectal cancer, regular colonoscopy or resection of rectal polyps are effective means. During these years, with medical knowledge popularization and publicity about the causes of rectal cancer and rectal polyps, more and more people are aware of the importance of regular colonoscopy. This medical examination also depends on the doctor's own experience and competence, so it is a very arduous task for mankind. As an item of physical examination, a manual colonoscopy is a great test for doctors' clinical experience and ability. Most doctors need a lot of time to identify intestinal polyps, and also need a lot of experience to ensure the correctness of the results. Therefore, manual examination of intestinal polyp images by medical personnel is

very time-consuming work and depends on some experienced doctors. Therefore, people begin to use computers to process images instead of human to identify and segment intestinal polyps by combining image processing technology in the computer field to develop its application in the medical field.

In the medical science field, image segmentation technology is the basis of image processing in the relevant area of computers. Medical imaging (radiology) is a field where medical staff recreates images for diagnosis and treatment purposes. In order to improve the efficiency and accuracy of judging the pathological position, some valuable methods for information processing in the computer science area are adopted. After the computer processes the image and then crops the images with the elements such as brightness and texture in the image as the standard for segmenting, the model of image segmentation through a neural network has been more clearly established. Moreover, as the model is constantly improved with the algorithm, together with the adopting of a convolution neural network to process the image, the quality of the outcome has been greatly improved.

This model also applies the method of image segmentation in computer science. The core is the joint action of the convolution Neural Network (CNN) and Convolutional Block Attention Module (CBAM). The selected intestinal polyp data set is the actual medical image of the human intestinal cavity, and the polyp part has similar surface characteristics to the normal structure of the human body, which is not easy to distinguish. Therefore, segmentation of intestinal polyp images demands both the accuracy and precision of computer image processing. This model chooses to enhance the feature extraction ability to strengthen the feature of the intestinal polyp feature different from the normal intestinal cavity tissue of the human body. This model reduces the possibility of misdiagnosis in the process of computer processing by enhancing the feature extraction ability, and reduces the time of feature processing by feature enhancement, so as to jointly improve the performance of the model in terms of time cost and accuracy. The choice of CBAM mainly depends on the difficulty of intestinal polyp image processing. Through the joint processing of the attention mechanism and receptive field expansion module, the intestinal polyp features of the image are effectively enhanced, which provides excellent pre-processing for feature extraction and image segmentation.

*Corresponding Author.

In the part of image feature extraction, this model selects the Har DNet convolution network with perfect function as the backbone network in the encoder stage. However, due to the full integration of upper and lower image features by Har DNet, the amount of calculation required to output the feature image is large. Although the accuracy and effectiveness of segmented image features are retained to a great extent, there is still room for improvement in processing speed. This model improves the processing speed based on Har DNet. Through the expansion of the receptive field and the addition of an attention mechanism, it not only improves the accuracy of feature processing but also accelerates the processing, so as to effectively improve the performance of overall image processing.

A more rapid and precise processing of the images of intestine polyps has a huge and profound impact on the medical world. In an age when computers were not used to help with medical images processing, homeopathy had to invest in a number of people in the determination of the images of intestine polyps. This does not only require the abilities of doctors, but also the time and effort of the medical personnel. This paper proposes and uses an image-processing model that can free therapeutic staff in a maximum degree. With the aid of computer technology, medical personnel do not have to spend much time identifying the pathological tissue that is very close to normal tissue. The doctor can devote more energy to the treatment of the follow up. In addition, considering the high accuracy of the model, when small colonic polyps are encountered, the model can still accurately distinguish to achieve the effect of early warning.

This paper will start with the evolution of various neural network models, emphasize the performance improvement of numerous models over time, and lead to the optimization of subsequent models in this paper. After that, the main structure of the model is the core content of the article, which is described in combination with the structure diagram. Finally, this paper will list the results of relevant experiments for readers to intuitively experience the performance improvement of this model.

II RELATED WORK

CNN has achieved great success in the field of medical image segmentation because of its excellent feature extraction ability and good feature expression ability [1,7], and it does not need manual extraction or much pre-processing work. The U-Net neural network [2-5] is one of the earliest adopted models in the semantic segmentation network [26]. Due to the data augmentation by elastic distortions, it only needs very few annotated images. However, since each pixel needs to take an image block [17] centered on itself, two adjacent pixels are highly similar in their block information, which causes much redundancy and slow network training. Therefore, on this basis, the U-Net++ network [14-15] is developed. Through the effective integration of U-Net in different depths, that is, these U-Net parts share an encoder, and effectively recover the fine-grained details of the target object under complex background utilizing in-depth supervision and joint learning. In short, U-Net++ is equivalent to splicing four U-Net networks with different depths through multiple skip connections. The first

advantage is the improvement of accuracy, which should be brought by integrating the characteristics of different levels. The second is the flexible network structure combined with deep supervision so that the deep network with huge parameters can greatly reduce the parameters within the acceptable accuracy range, which has been developed to more advanced models [27].

The ResUNet model [18-19] performs better in-depth and avoids the degradation of the network through residual learning. The main idea of this model is to add a direct channel in the network and control the neural network of the receiving layer by allowing the original input information to be directly transmitted to the later layer. Instead of learning all the outputs after complete calculation, it can learn the residual of the output of the previous network [19]. This kind of information processing protects the integrity of information. The whole network only needs to learn the part of the difference between input and output, simplify the learning objectives and difficulties, and significantly improve the output results.

The PraNet model [9], which draws reverse attention, can effectively improve the style progress and accurate characterization of the results by mining boundary clues through the reverse attention module and establishing the connections between regions and boundaries. This model also gives inspiration for our model: the establishment of the attention module helps to build the relationship between the actual image pixels. The attention mechanism can be used as a weighting module to basically process the features in the image before processing the image so that the features in the image have a judgment similar to "important" or "unimportant". This judgment has high accuracy because of the strong correlation of the information around the attention mechanism's domain. As a kind of attention mechanism, the block attention mechanism proposed in this model can help the model give different weights to each part of the input and extract the more critical information so that the model can make a more accurate judgment. It is a lightweight attention mechanism module, so it will not bring more overhead to the calculation and storage of the model and improves the accuracy and calculation speed of the model from multiple layers.

HarDNet, as an improved model of L2-Net, measures the amount of data accessing memory through convolution input-output (CIO) [10-11]. The amount of data in storage that can be accessed through HarDNet can improve the density of calculation. Compared with the previous convolutional neural network [6,8], it has a very significant improvement in the accuracy and speed of calculation.

In recent years, the HSNet [28] model has emerged as a new neural network model for image segmentation. The model uses a convolutional neural network (CNN) to predict the 3D model of the human body, thus avoiding the limitations of handmade features and postures. At the same time, it is applied in virtual fitting and human body size measurement. In the paper on the HSNet model, four possible situations are analyzed as the input of the network: (a) The single binary contour of a human is scaled to a fixed size to prevent the loss of camera calibration information; (b) The human shadow image is scaled to a fixed size because the shadow will retain

the information complementary to the contour; (c) Assume the front profile of known camera parameters; (d) Assume that the front and side profiles of the camera parameters are known. HSNNet [28] uses CNN to accurately predict 3D human models from contour or shadow images and tries to find the global mapping of deformation parameters from 2D images to 3D models. The model has been widely evaluated and tested on thousands of people and real people. In addition, this model has been proved by comprehensive experiments that better prediction results can be obtained if there is shadow information. The model further assumes that humans wear tight clothes. Applying the proposed method to people wearing other clothes will increase the error. The limitation of the method of this model is that in the current training, it cannot deal with the posture that is obviously different from the neutral posture and contains self-occlusion [20-23]. This can be solved by generating a larger training set (including more prominent poses).

The encoder-decoder mechanism is often used in the CNN model. CNN encoder can basically be regarded as the feature extraction network, while the decoder interprets the received information and enlarges the received image features to the size of the original image through the decoded part [16,24], which greatly facilitates the prediction of the part of the image segmentation area and simplifies many subsequent calculation processes.

Considering the calculation speed and accuracy of the above models, based on the Har DNet convolution neural network, a segmentation model of intestinal polyp image with a block convolution attention mechanism is proposed in this paper. The model uses Har DNet as the basic network structure, and the image processing results obtained after passing through the backbone network are augmented to varying degrees and input into the subsequent decoder. Firstly, the decoder uses the Convolutional Block Attention Module and RFB module [25] to process the image features, so that the model can improve the feature extraction ability of the network by simulating the human receptive field. At the same time, the input feature information is weighted through the block attention mechanism, combined with the low-level and high-level feature input combination representation, so that the overall accuracy and speed of feature extraction of the model are greatly improved.

In the paper "An Imperial Study on Ensemble of Segmentation Approvals" [29], it is proposed that many machine learning models cannot be continuously learned especially neural networks. When they receive new training tasks, they will forget what they have learned before. This phenomenon is often called catastrophic amnesia. One of the reasons for the assumption of catastrophic forgetting in neural networks is the change of input distribution in different tasks [12], for example, the lack of common factors or structures may cause the optimization method to converge to completely different results each time. This paper studies the catastrophic forgetting phenomenon that occurs when the model is learning the same task. The author tries to understand the optimization

process in depth by analyzing the interaction between samples during learning. The first goal is whether to continue compressing data sets based on some forgotten data, so as to improve the efficiency of data training without affecting the generalization accuracy. The second goal is to identify "important" samples, outliers and other special data by forgetting statistical data. The model used in this paper also has the defect of unsustainable learning. In the future improvement, we will consider the method of using statistical forgetting data to identify "important samples", combined with the attention mechanism module to extract special data for image processing.[13]

III METHODS

A. HarDNet

Compared with the traditional CNN, Har DNet optimizes its sampling strategy and loss function based on L2-Net, reduces the inference time, and improves the accuracy of the results by using 1 convolution. Taking sampling in a batch as an example, Har DNet's sampling strategy selects the nearest negative sample by comparing the distance between the descriptors of the corresponding patch in the batch, so as to improve the inference speed of the network by reducing shortcuts. At the same time, Har DNet also optimizes its loss function based on this idea: the distance between matching pairs is less than the distance of unmatched patch time, which reduces unnecessary calculation and improves the inference speed of the network. Compared with the speed of the optimized network, the accuracy of the optimized network is better than that of the optimized network. Therefore, Har DNet is selected as the basic network structure of this model. Fig. 1 shows the network structure of Har DNet.

B. Decoder

The decoder generates more accurate mapping by combining low-resolution, high-level features, and high-resolution, low-level features, and refines deep features by cascade connection. However, this method has the following two disadvantages. First, the low-level features contribute less to the network performance than the high-level features. Secondly, the use of high-resolution low-level features makes the computation more complex.

According to the above two shortcomings, it is proposed to refine the deep-seated features to improve the representation ability and abandon the shallow features to reduce the computational complexity, that is, the cascaded partial decoder is used. By cascading, the original modules are connected across to improve the expression ability and reduce the computational complexity.

C. Convolution Block Attention Mechanism

The decoder, CBAM is also adopted to improve the computing power of the network. The visual attention mechanism of the human visual system can assist the brain in signal processing. The CBAM works similarly: the ultimate goal is to select the information that is more critical to the current task goal from a large amount of information.

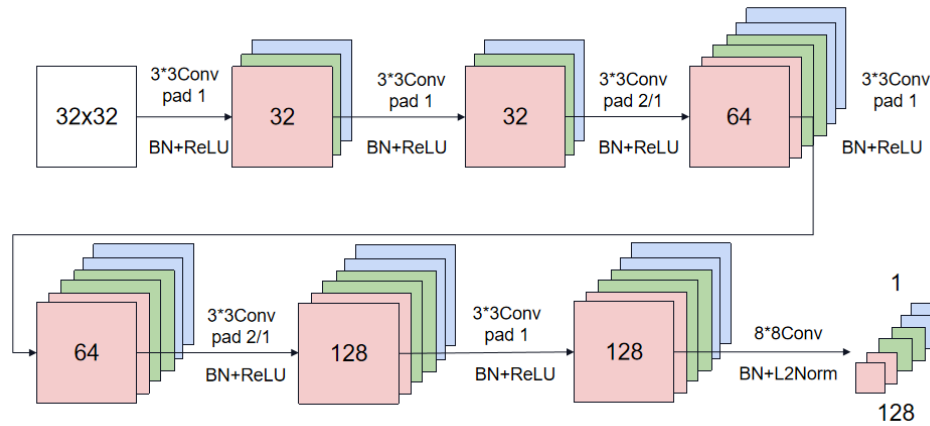


Fig. 1. HarDNet.

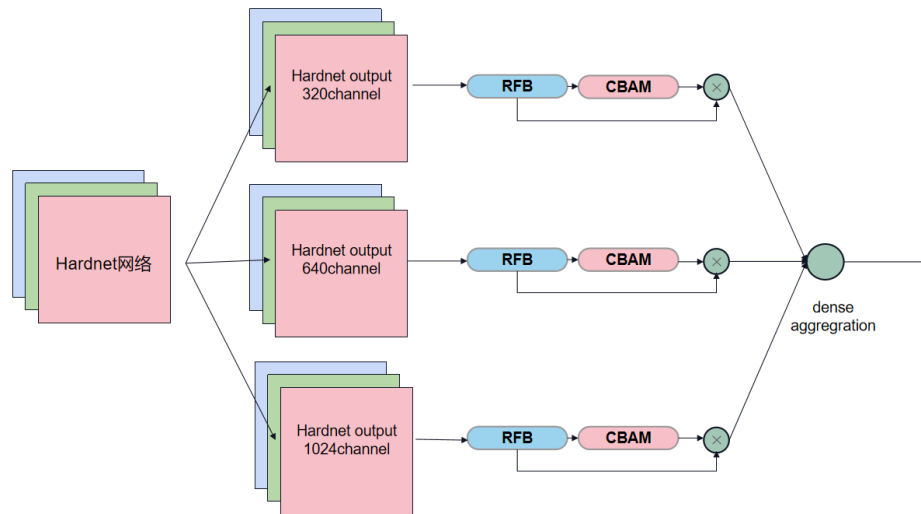


Fig. 2. Elaborate of convolutional block attention module.

The human visual attention mechanism receives the location of focal regions by quickly scanning all image information that needs to be processed by the brain and then pays more attention to these focal regions to obtain the detailed information needed by the brain. Through the judgment and selection of the focus of attention, the human eye pays more attention to the image information worth attention to rather than the non-focal, unimportant information. This attention mechanism of human vision greatly improves the information-processing efficiency of the human eye. At the same time, by investing more attention resources in the focus, the visual information obtained also has higher accuracy. Fig. 2 shows the action mechanism of the attention mechanism CBAM in the model.

The process of using the attention mechanism to improve the computing power of the network is similar to the human eye attention mechanism. Its core idea is to introduce specific weights into the input information in order to give priority to

the location of relevant information. This part of processing refers to the content of the Convolutional Block Attention Module (CBAM), and the input information is weighted through channel attention convolution and spatial attention convolution to enhance the useful features and suppress the useless features.

The channel attention model mainly focuses on the meaningful part of the input characteristic diagram, calculates the internal relationship between each channel, and finally uses the channel attention diagram to represent the internal relationship between each channel in the input characteristic diagram, and confirms the corresponding weighting of different channels to determine which channels should be ignored. The spatial attention model mainly focuses on the internal relationship of the feature map at the spatial level, that is, the weighting corresponding to different regions, which is complementary to the processing results of the channel attention model at the spatial level.

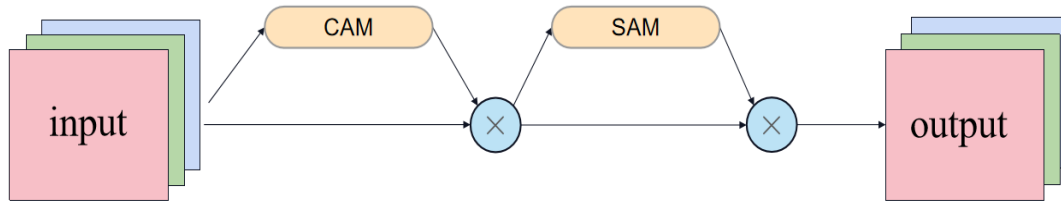


Fig. 3. Convolutional block attention module.

Fig. 3 shows the implementation mechanism of the channel attention mechanism in the model. It mainly processes the results obtained by the RFB model through a channel attention mechanism to generate the characteristic results of the channel attention mechanism and then takes it as the input characteristic diagram of the spatial attention mechanism to obtain the final results of the attention mechanism processing. The input features of RFB pass through the two pooling layers of Max pooling and average pooling respectively and then pass through MLP respectively. The output features are added based on element-wise and sigmoid activation to generate the final channel attention feature map. Then the feature map is multiplied by the input feature map, and the output feature map is used as the input feature of the spatial attention mechanism.

Similar operations are carried out in the part of the spatial attention mechanism. The feature map obtained through the channel attention mechanism passes through the two pooling layers of max pooling and average pooling, respectively. The two results are combined based on the channel, and then the feature map of the spatial attention mechanism is generated through the sigmoid operation. Finally, the feature map is multiplied with the input feature map to obtain the final

feature, which can be input into the subsequent calculation of the density aggregation part.

As Fig. 4 implies, the receptive field module network is composed of the convolutional layer with different expansion rates and convolution kernels of different sizes. By integrating the deep and shallow feature maps, the feature recognition effect can be equivalent to that of human vision. Through this method, the features are processed to achieve the effect of expanding the receptive field.

The decoder of this model enhances or suppresses the target features by adding a convolution block attention mechanism, which effectively improves the operation efficiency and feature expression ability of the model, and generally improves the performance of the model.

When processing the input image, rich feature representations is required. Combining these feature representations can improve the inference effect of category and location. Therefore, deep aggregation is used to enhance the network structure in order to better integrate information. The bottom stage is connected to the high level by jumping to fuse the size and resolution, to achieve the effect of combined expression of feature representation.

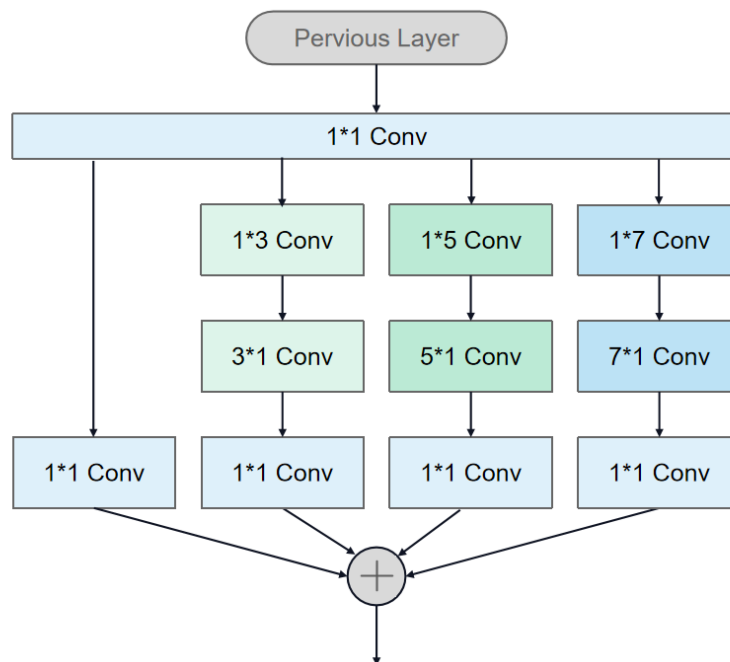


Fig. 4. Receptive field module network.

D. Algorithm Flow

Based on the HarDNet network, this model improved the processing speed and accuracy of the model by adding a

receptive field and block convolution attention mechanism. The overall algorithm flow is shown in Fig. 5.

Algorithm 1: Algorithm flow

```
initialization of the HarDNet network;  
while different size of the channel(320/640/1024) do  
    Processing through HarDNet module;  
    Processing through RFB module;  
    while input data processed by RFB do  
        Processing through CBAM module;  
        Multiplication CBAM and RFB modules' resylt;  
    end  
    Combine results through the dense aggregation module;  
end
```

Fig. 5. Algorithm

IV EXPERIMENT

Based on the results of previous data processing in related polyp experiments, five polyp segmentation data sets were taken as the experimental data sets. At the same time, the network structure used in this experiment was compared with the previous network structure, and the reasoning speed and the accuracy of the results are compared.

A. Datasets

In this experiment, five polyp segmentation data sets mentioned earlier were used: Kvasir-SEG, CVC-ColonDB, EndoScene, ETIS-Larib Polyp DB, and CVC-ClinicDB. This section also compared the results obtained this time with the results obtained from several previous models: U-Net, U-Net++, ResUNet, SFA, and PraNet. Table I shows the experimental results of these models on CVC-T Data Sets.

B. Segmentation Results of Intestinal Polyp Image

The CVC-T data set mentioned earlier in the experiment is the largest publicly released gastrointestinal image data set, which includes images of pathological tissues and images of normal phenomena. According to the needs of this experiment, 1,000 pictures of intestinal polyps were selected for the experiment (polyps of different sizes, pathological positions, and different qualities), which can reflect the more

comprehensive learning ability of this model to a certain extent. At the same time, this experimental result was the calculation result of the above-mentioned evaluation index. Each column in the table below corresponds to an evaluation index respectively. Each behavior is different and used to compare the model. The model result of the last behavior experimental design, and n/a means that the result cannot be obtained.

In the above experiment, 512×512 training input size and 1e-2 learning rate were adopted. There were 100 epochs used to test the model. It can be seen that the reasoning speed with the attention mechanism is much faster than that of other models.

In addition, according to the above calculation results, the two evaluation indexes of mDice and mIoU are improved, in which the average intersection and union ratio is increased from 0.834 to 0.851, up to 0.17, and the accuracy and processing speed are improved. It shows that the experimental model can effectively segment the meat in the actual pathological picture. However, the SM index is 0.02 lower than that of the HarDNet model, because in the process of adding attention mechanism, the image is weighted and the processing weight of some pixels is reduced, which makes this experiment lack the similarity with the actual image.

TABLE 1. COMPARISON OF EXPERIMENTAL RESULTS OF DIFFERENT MODELS ON CVC-T DATA SETS

Models	mDice	mIoU	F_{β}^{ω}	S_m	MAE	E_m^{\max}
U-Net	0.818	0.746	0.794	0.858	0.055	0.893
U-Net++	0.821	0.743	0.808	0.862	0.048	0.91
ResUNet-mod	0.791	n/a	n/a	n/a	n/a	n/a
ResUNet++	0.813	0.793	n/a	n/a	n/a	n/a
SFA	0.723	0.611	0.67	0.782	0.075	0.849
PraNet	0.898	0.84	0.885	0.915	0.03	0.948
HarDNet-MSEG	0.901	0.834	0.871	0.935	0.008	0.974
HarDNet-CBAM	0.902	0.836	0.881	0.933	0.03	0.978

TABLE 2. EXPERIMENTAL RESULTS ON DIFFERENT DATA SETS

Models	Clinicdb		ColonDB		ETIS		Kvasir	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
U-Net	0.823	0.755	0.512	0.444	0.398	0.335	0.71	0.627
U-Net++	0.794	0.729	0.483	0.41	0.401	0.344	0.707	0.624
Res UNet-mod	0.779	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Res UNet++	0.796	0.796	n/a	n/a	n/a	n/a	n/a	n/a
SFA	0.700	0.607	0.469	0.347	0.297	0.217	0.467	0.329
Pra Net	0.899	0.849	0.709	0.640	0.628	0.567	0.871	0.797
Har DNet-MSEG	0.909	0.864	0.721	0.641	0.678	0.605	0.891	0.824
Har DNet-CBAM	0.926	0.879	0.756	0.677	0.735	0.661	0.897	0.841

On the basis of the above experiments, four other data sets are used to test the results of the model, namely, ETIS, CVC-ClinicDB, Kvasir-SEG, and CVC-ColonDB. Table II shows the quantitative results of these data sets.

The four data sets used in the experiment have different characteristics. CVC-ClinicDB and CVC-ColonDB are both small-scale data sets, so they are all used in the test experiments; In addition, ETIS is a data set for early diagnosis of colorectal cancer, containing 196 images that are also used in test experiments. In this experiment, the Endosece dataset is divided into three subsets (training, verification, and test) as the previous ones did. This experiment mainly adopted the kvasir subset as the test set.

It can be seen from Table II that the results of the model with attention mechanism under the two indicators of mDice and mIoU show higher accuracy and better performance.

C. Evaluation Index

In the training, this experiment employed 900 images in kvasir SEG and 550 images in CVC-ClinicDB, a total of 1450 images as training images. The 80% of the pictures in the five datasets mentioned above (Kvasir-SEG, CVC-ColonDB, EndoScene, ETIS-Larib Polyp DB, and CVC-ClinicDB) are randomly selected as the training set, and 10% for validation, 10% for the testing set. In addition, meanDice and meanIoU were used to evaluate the accuracy and correlation of image-cutting results. The calculation method of the index is as follows:

$$\text{MeanDice} = \frac{2 \bullet \text{tp}}{2 \bullet \text{tp} + \text{fp} + \text{fn}} \quad (1)$$

$$\text{mIoU} = \frac{\text{tp}}{\text{tp} + \text{fp} + \text{fn}} \quad (2)$$

In addition, in order to further evaluate the network model, the experiment also took into account other metrics for evaluation: F_{β}^{ω} , S_m , MAE, E_m^{\max} and FPS. Among them, F_{β}^{ω} is used to modify the Dice process to avoid the deviation of the results when the samples are relatively balanced. MAE is used to evaluate the accuracy of the result, and the size is at the pixel level. Because the above two metrics are pixel-level evaluation indicators, they lack the relationship between macro and global structure and ignore the evaluation of the similarity between them. Therefore, another evaluation index S_m is used as the evaluation standard of the similarity between the prediction result and the actual image, so as to make up for the defect of only pixel-level evaluation. The alignment measure for enhancement is also used E_m^{\max} to evaluate the similarity between the image processing at the pixel level and the global image.

Fig. 6 is an exemplificative picture of the segmentation in the processing test data set. The masked part is the display part of the intestinal polyps after image segmentation, and the white part is the judgment position of the trained intestinal polyps. After comparison, we can roughly feel the approximate accuracy of the experimental results.

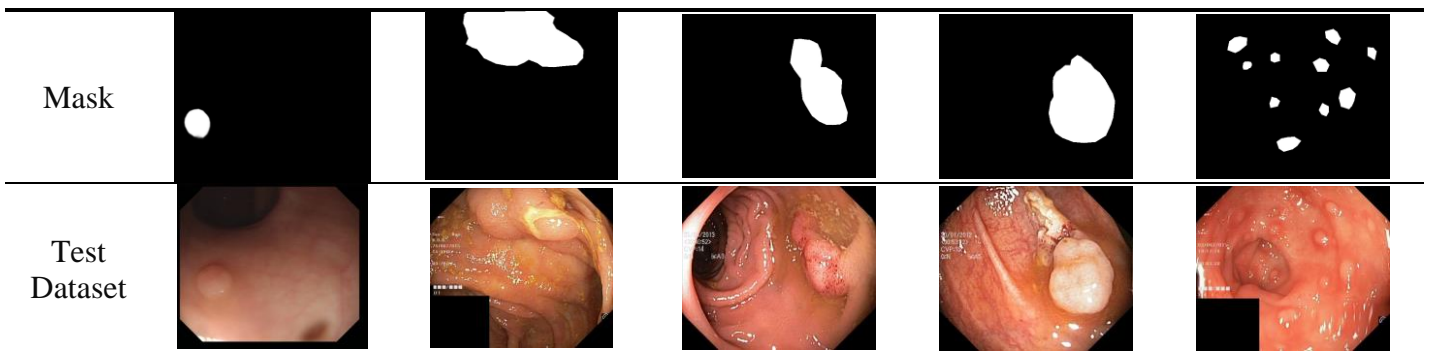


Fig. 6. Sample of model image segmentation results.

V SUMMARY

In this paper, a segmentation model structure of intestinal polyp image based on attention mechanism was proposed. This model added Convolutional Block Attention Module (CBAM) and RFB module based on the HarDNet backbone to optimize the feature extraction ability. By enhancing or suppressing the target features in the computer operation process, the operation efficiency of the model and the expression ability of features were effectively improved. The fusion of the CBAM module and RFB module in the calculation details not only increases the receptive field of the model through RFB model processing but also makes the result of the final feature map more prominent through the attention mechanism, which greatly improves the ability of the model to extract image features and improves the processing speed and accuracy of the model.

The HarDNet-CBAM model has also obtained very good experimental results through experiments. The average precision and processing speed of four test data sets have been greatly improved, indicating that this model has been greatly improved in feature extraction. At the same time, the HarDNet-CBAM model has been tested in different data sets. Compared with other models, this model has higher accuracy and greater advantages in a variety of test sets. In a word, the HarDNet-CBAM model has excellent learning ability and universality in polyp image segmentation. At the same time, it also reflects its advantages in feature extraction performance and accuracy of results.

VI FUNDING STATEMENT

The research is funded by the National Key R&D Program (2018YFB1307004); Supported by the National Natural Science Foundation of China (61873008).

REFERENCES

- [1] Y. Cheng, et al. "Text Sentiment Orientation Analysis of Multi-Channels CNN and BiGRU Based on Attention Mechanism". *Journal of Computer Research and Development* vol. 12, pp. 2583-2595, 2020.
- [2] D. Xu, H. Li, L. X. Zhou, "Model of Automatic Identification of Diabetic Macular Edema via Convolutional Neural Networks UNet". *Recent Advances in Ophthalmology* vol. 4, pp. 357-361, 2020.
- [3] T. Zhou, et al. "U-Net and its applications in medical image segmentation: a review". *Journal of Image and Graphics*, vol. 9, pp. 2058-2077, 2021.
- [4] H. Zhang, et al. "Improved U-Net Models and Its Applications in Medical Image Segmentation: A Review". *Laser & Optoelectronics Progress* vol. 2, pp. 55-71, 2022.
- [5] Y. W. Zhou, et al. "Breast Mass Image Segmentation Algorithm Based on Improved Residual U-Net". *Journal of Southwest University of Science and Technology (Natural Science Edition)* vol. 2, pp. 68-74, 2021.
- [6] Y. Luo, "A Method for Segmentation Brain Tumor in MRI Based on ResUnet Embedded in GA", *Microcomputer Applications*, vol. 7, pp. 13-15, 2020.
- [7] Akbari, M., et al. (2018). "Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network." 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 69-72, 2018.
- [8] D. P. Fan, et al, "Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images". *IEEE Transactions on Medical Imaging*, vol. 39, pp. 2626-2637, 2020.
- [9] D. P. Fan, et al. "PraNet: Parallel Reverse Attention Network for Polyp Segmentation.". *MICCAI 2020: Medical Image Computing and Computer Assisted Intervention–MICCAI*, voi. *Medical Image Computing and Computer Assisted Intervention–MICCAI*, pp. 263-273, 2020.
- [10] P. Chao, et al. "HarDNet: A Low Memory Traffic Network". *CoRR*, pp. 00948, 2019.
- [11] C. H. Huang, et al., "HarDNet-MSEG: A Simple Encoder-Decoder Polyp Segmentation Neural Network that Achieves over 0.9 Mean Dice and 86 FPS." *ArXiv* pp. abs/2101.07172, 2021.
- [12] F. Zhang, et al., "Persistent Fault Attack in Practice." *IACR Trans. Cryptogr. Hardw. Embed. Syst.* pp. 172-195, 2020.
- [13] S. Engels, et al. "SPFA: SFA on Multiple Persistent Faults." 2020 Workshop on Fault Detection and Tolerance in Cryptography (FDTC), pp. 49-56, 2020.
- [14] L. D. Huynh, & N. Boutry, "A U-Net++ With Pre-Trained EfficientNet Backbone for Segmentation of Diseases and Artifacts in Endoscopy Images and Videos". *EndoCV@ISBI*, 2020.
- [15] D. Jha, et al. "DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation". 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS). Pp. 558-564, 2020.
- [16] Chen, L.C et al., "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation". *ArXiv* abs/1802.02611, 2018.
- [17] X. F. Yang, et al., "Road Detection and Centerline Extraction Via Deep Recurrent Convolutional Neural Network U-Net". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 7209-7220, 2019.
- [18] Diakogiannis, F. I., (2019). "ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data". *ArXiv* abs/1904.00592.
- [19] D. Jha, , et al. "ResUNet++: An Advanced Architecture for Medical Image Segmentation". *IEEE International Symposium on Multimedia (ISM)*, 225-2255, 2019.
- [20] W. T. Zhu, Y. F. Huang, L. Zeng, "AnatomyNet: Deep Learning for Fast and Fully Automated Whole-volume Segmentation of Head and Neck Anatomy". *Medical Physics* 2, 2019.
- [21] C. E. He, H. J. Xu, & Wang, Z. (2020). "Automatic Segmentation Algorithm for Multimodal Magnetic Resonance-Based Brain Tumor Images". *Acta Optica Sinica* 6, 2020.
- [22] W. J. Wang, J. X. Chen, J. Zhao, et al. (2019). "Automated Segmentation of Pulmonary Lobes Using Coordination-Guided Deep Neural Networks[C]". 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), April 8-11, Venice, Italy. New York: IEEE Press, pp. 1353-1357, 2019.
- [23] Y. Li, et al., "Asymmetric GAN for Unpaired Image-to-Image Translation". *IEEE Transactions on Image Processing*, vol. 12, pp. 5881-5896, 2019.
- [24] L. C. Chen, et al. (2018). "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. " *ArXiv* abs/1802.02611 (2018): n. pag, 2018.
- [25] S. T. Liu, et al. (2018). "Receptive Field Block Net for Accurate and Fast Object Detection. " *ArXiv* abs/1711.07767: n. pag, 2018.
- [26] Oktay, Ozan et al. (2018). "Attention U-Net: Learning Where to Look for the Pancreas. " *ArXiv* abs/1804.03999: n. pag, 2018.
- [27] C. L. Guo, et al., "SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation. " 2020 25th International Conference on Pattern Recognition (ICPR), pp. 1236-1242, 2021.
- [28] W. C. Zhang, C. Fu, Y. Zheng, F. Y. Zhang, Y. L. Zhao, "HSNet: A hybrid semantic network for polyp segmentation". *Computers in Biology and Medicine*. Volume 150. 2022.
- [29] Nanni, Loris, Alessandra Lumini, Andrea Loreggia, Alberto Formaggio, and Daniela Cuza "An Empirical Study on Ensemble of Segmentation Approaches" *Signals* 3, vol. 2, pp. 341-358, 2022.