IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 December 2022, accepted 26 December 2022, date of publication 29 December 2022, date of current version 5 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3233078

RESEARCH ARTICLE

Segmentation of White Blood Cells Based on CBAM-DC-UNet

DONGMING LI[1], SHIYU YIN[1], YU LEI[1], JINGNING QIAN[2], CHUNXI ZHAO[3], AND LIJUAN ZHANG[4]

[1]School of Information Technology, Jilin Agricultural University, Changchun, Jilin 130118, China
[2]Faculty of Engineering and Information Technology (FEIT), The University of Melbourne, Carlton, VIC 3053, Australia
[3]Information Center, Jilin Agricultural University, Changchun, Jilin 130118, China
[4]College of Computer Science and Engineering, Changchun University of Technology, Changchun, Jilin 130012, China

Corresponding authors: Chunxi Zhao (zcx@jlau.edu.cn) and Lijuan Zhang (zhanglijuan@ccut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China Youth Science Foundation Project under Grant 61801439, in part by the Scientific Research Project of Jilin Provincial Department of Education under Grant JJKH20210747JK, in part by the Project of Jilin Provincial Environmental Protection Department under Grant 202107, and in part by the Scientific Research Project of Jilin Provincial Department of Ecology and Environment.

ABSTRACT Monitoring the morphology of blood leukocytes, plays an important role in medical research, especially in the treatment of diseases such as immunodeficiency. Traditional manual detection methods are susceptible to numerous interference factors. Therefore, blood cells are often segmented using deep-learning algorithms. This study proposes a U-Net model based on a combination of an attention mechanism and dilated convolutions. First, the traditional convolution in a double convolutional module in a network is replaced by dilated convolution, and multi-scale features are obtained by expanding the receptive field. Second, after each convolution layer in the upsampling layer, an attention mechanism module is combined to refine the adaptive features and improve the segmentation performance of the model. Finally, the RAdam optimizer was used to enhance the robustness of the learning rate variations. Through the ablation experiment of the three improvement directions, it was concluded that all three improvement directions had a positive effect on the segmentation result, and the improvement was the most effective when the three improvements were combined. The experimental results show that compared with the original U-Net model, the segmentation indicators of blood leukocytes, intersection over union (IOU), recall and accuracy were increased by 5.1%, 5.7% and 1.2%, respectively, which more accurately segmented blood leukocytes, which may be used for a greater degree of auxiliary leukocyte detection in the application of immunodeficiency and other diseases.

INDEX TERMS Attention, blood leukocytes segmentation, dilated convolution, deep learning, image segmentation, U-Net.

I. INTRODUCTION

The detection of leukocyte morphology in the blood is useful in medical research for the monitoring of leukemia, immunodeficiency and other diseases. Therefore, the accurate segmentation of leukocyte morphology plays an important role in medical research. The variety of leukocyte types and morphologies makes segmentation difficult, and the presence of adhesions between leukocytes and a large number of blood cells in the blood makes it a challenging to achieve accurate segmentation of leukocytes with clear boundaries after segmentation.

Currently, research on blood leukocyte segmentation algorithms has focused on traditional methods of image segmentation. These include threshold segmentation [1]. This determines the threshold value based on the difference in grayscale between the image target and background regions, and uses the threshold value to separate the target from the background. However this method is affected by contrast, and the effect is not ideal when the contrast between the target and the background is low. Another approach is edge detection segmentation [2]. Edge detection segmentation

The associate editor coordinating the review of this manuscript and approving it for publication was Amin Zehtabian.

performs feature extraction at the target boundary to separate the target from the background. This method is also affected by contrast, and the segmentation effect is poor at boundaries with low contrast. Biswas et al. [3] proposed a blood cell detection method based on frequency domain sobel filter threshold estimation watershed transform (SFD-TEW). The detection of cell contours in blood cell images using the SFD-TEW method can show excellent detection results, but this method can be disturbed by noise in complex backgrounds, producing some false contour information, and can produce false contours for the concave points in the middle of red blood cells close to the background color. Nayak et al. [4] proposed a blood cell detection method based on morphological transformation and improved fuzzy scatter, which can largely solve the interference of background noise and obtain more accurate cell contours, but has poor ability to deal with the complex texture of the cells themselves, which makes the cells with complex texture incomplete The method is able to resolve the background noise to a large extent and obtain a more accurate cell outline, but it is not good at handling the complex texture of the cells themselves, so that the cells with complex texture are not completely represented. Such algorithms suffer from a lack of global applicability, which limits their applications. In contrast, deep learning methods can process data in their raw form, eliminating the need for hand-crafted features. This approach has been widely used for the segmentation of images with good results.

With the continuous development of information technology, various deep learning models have been proposed and widely used for image segmentation [5]. The AlexNet model was proposed by Krizhevsky et al. [6]. Uses GPU parallel acceleration for training, whereas dropout is proposed to prevent overfitting. However the model is relatively simple and not very accurate in comparison. Simonyan et al. [7] proposed the VGG model, which uses consecutive 3×3 convolutions instead of the larger convolutional kernels in AlexNet. Deepening the network depth under the condition of ensuring the same perceptual field, which improves the effectiveness of neural networks to some extent, but will consume more computational resources and consume more memory. He et al. [8] proposed the deep residual network ResNet, which is less complex than the traditional VGG, and also improves the phenomenon of gradient disappearance caused by deepening the network depth. However the training period is longer when the network model is deeper the fully convolutional network (FCN) proposed by Long et al. [9], which removes the fully connected layer from the traditional network model and replaces it with a convolutional network, and uses deconvolution to roll up a large feature map from a small one, but the network segmentation results are not fine enough. Later, scholars proposed SegNet [10], DeepLabv1 [11], DeepLabv2 [12], DeepLabv3 [13], DeepLabv3+ [14], RefineNet [15], PSPNet [16], U-Net [17], etc. for semantic segmentation. All of these methods are widely used in various image segmentation tasks. SegNet is based on the full convolutional network, which is a semantic segmentation network obtained by modifying the structure of the VGG16 network, and achieves target segmentation through end-to-end, pixel-to-pixel training. The DeepLab series network has been continuously improved from v1 to v3+, and has become one of the most mainstream algorithms in the field of semantic segmentation. It introduced the concept of null convolution in the network, and also introduced deep separable convolution in DeepLabv3+ to optimize the speed of the segmentation network, so that it achieves better performance while reducing the computation and computing time of the model, and the effect of the model is nearly perfect. U-Net is improved based on the FCN, which is a symmetric coding-decoding structure. The network applies the image enhancement method, which can obtain good accuracy even with limited data sets, and is widely used in medical images, achieving good segmentation results. U-Net++ [18] is an improvement on U-Net, which connects all layers 1 to 4 of U-Net together, can capture features of different depths, share a feature extractor, and integration through a feature overlay. This allows the model to achieve better segmentation accuracy. Stringer [19] et al. propose a general, deep learning-based cell segmentation algorithm that can accurately segment cells from a variety of image types and does not require model retraining or parameter tuning. Schmidt et al. [20] propose a deep learning method for locating cell nuclei by star-convex polygons, which, compared to bounding boxes, have a much better shape representation is much better and thus does not require shape refinement. Teng et al. [21] also achieved good segmentation results by using a simplified MobileNetv2 [22] modified with DeepLabv3+ to build a lightweight leukocyte segmentation network, a semi-supervised leukocyte segmentation method in an adversarial learning framework.

Compared to traditional unsupervised segmentation methods. Deep learning methods have better applicability and superior performance. Due to the blurred boundary and complex gradient of medical images, more high-resolution information is needed, and high-resolution is used for accurate segmentation. And the internal structure of human body is relatively fixed, the distribution of segmentation target in human body image is very regular and the semantics is simple and clear, and the low-resolution information can provide this information for target object recognition. However, U-Net combines low-resolution information and high-resolution information, which is perfectly suitable for medical image segmentation. Although the U-Net model has a good performance in medical image segmentation [23], [24], the model was proposed relatively early, and some newly proposed modules with improvements to the deep learning network do not exist in the network, and the segmentation of blood leukocytes is not effective for the segmentation of adherent cells. Therefore, this study made improvements to the U-Net model. First, in order to increase the receptive field of feature extraction, the conventional convolution in the network encoder is replaced by dilated convolution [25], which expands the perceptual field without losing resolution,

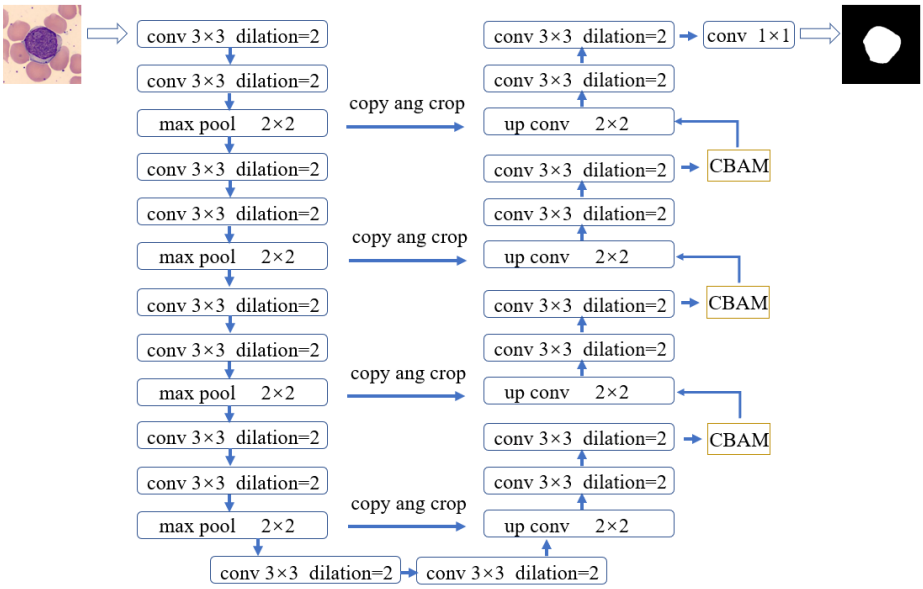


FIGURE 1. Architecture of the proposed method.

while adjusting the null rate to obtain multi-scale information. Second, the original model lack of attention mechanism, a convolutional block attention module (CBAM) [26] at the decoder side. By multiplying the obtained attention map with the input feature map to perform adaptive feature refinement. Finally, in order to improve the deep defects of the original optimizer, replacing the RMSProp optimizer [27] with the RAdam optimizer [28] to make the learning rate change more robust. The segmentation accuracy of the model for blood leukocytes is improved. The specific work is organized as follows: the ablation experiments of improvement points are conducted to address the shortcomings of U-Net, and the ablation experiments are done separately for the null convolution, attention mechanism and replacement optimizer, and one improvement point is replaced alone, while all other network parameters are unchanged, to observe the effect of improvement points on the network model. Then the improvement points are fused one by one to observe the effect of fusing the improvement points together on the model segmentation effect. Finally, the improved network model is compared with the mainstream segmentation model. The evaluation indexes IOU, Recall, Acc are used to determine the effectiveness of the model segmentation, and it is confirmed that the improved model has better segmentation effect on blood leukocytes.

II. CBAM-DC-UNet

A. MODEL INTRODUCTION

The structure of U-Net consists of encoder, decoder structure, and skip connections. The encoder performs down-sampling feature extraction through convolutional and pooling layers. The down-sampling module of each layer is first passed through a max pooling (size 2×2). Then go through two convolution kernels (size 3×3). In the down-sampling process, the number of channels is doubled. There are four layers of down-sampling modules in the structure. The decoder recovers the feature map size through upsampling. The up-sampling module of each layer first passes through a convolution kernel (size 2×2, stride 2), after passing through two convolution kernels (size 3×3) to complete the upsampling operation. The number of channels of the picture is halved during the up-sampling process. Skip connections are used to fuse the corresponding features of the encoder and decoder. The convolution part in the original model is a 3×3 traditional convolution. The effect of the feature extraction is general. To increase the receptive field in feature extraction and obtain multi-scale information. This study uses a 3×3 dilated convolution with a dilation rate of 2 to replace the 3×3 traditional convolution in the original model. Simultaneously, the original model lacked an attention module. Therefore, an attention module was introduced in the up-sampling module to enhance the segmentation effect of the model. Finally, to improve the deep defects of the original optimizer, we replaced it with a new optimizer to adjust the loss function better and increase the robustness of the learning rate. Through the above optimization to improve the shortcomings of the original model, a better segmentation effect is obtained. The improved model structure is shown in Figure 1.

B. DILATED CONVOLUTION

In deep learning, the role of convolution is to extract image features from the shallow edge structure information to a deep texture semantic structure. Compared to traditional convolution, dilated convolution adds voids to traditional convolutions and increases the receptive field. Thus, each convolution output contained a larger range of information. Different dilation rates control the module size of dilated convolutions. As shown in Figure 2 (b), the original 3×3 convolution is

changed to a dilated convolution when the dilation rate is 2. The dilation rate affects the size of the convolution kernel after dilation as follows:

$$k' = k + (k-1)(d-1) \tag{1}$$

where k is the original convolution kernel size, d is the dilation rate, and k' is the dilated convolution kernel size.

The main purpose of the dilated convolution is to increase the receptive field. As shown in Figure 2 (b), the dilated convolutional kernel is 3×3, but it has a void rate of 2, therefore, the field of perception becomes 7×7. In this study, the original convolutional module with a convolutional kernel size of 3×3 and padding of 1 was replaced by a convolutional module with a void rate of 2 and padding of 2. It can be calculated that the output of the replacement The size of the feature map remains the same, but the null convolution increases the perceptual field so that each convolutional output contains more feature information. Therefore, in this study, we replaced the traditional convolution with null convolution to improve of the target segmentation effect.

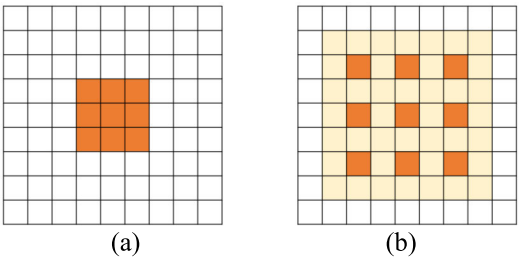


FIGURE 2. Dilated convolutions with different dilated rates a. rate = 1; b. rate = 2.

C. CONVOLUTIONAL BLOCK ATTENTION MODULE

The attention module is a method of processing data in machine learning, which can be used to enable the neural network to focus more on a feature and focus on local information. It is mainly divided into a channel attention module and spatial attention module. The convolutional block attention module is a module combines channel attention module and spatial attention module. This model is illustrated in Figure 3.

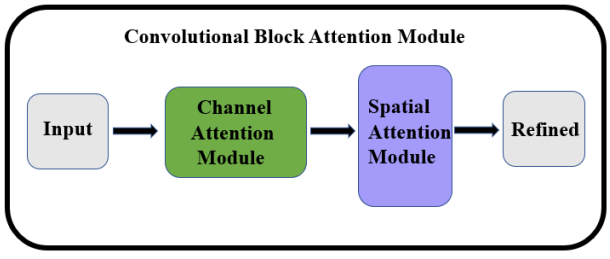


FIGURE 3. Convolutional block attention module.

The model of the channel attention module in the convolutional block attention module is shown in Figure 4. An intermediate feature map $F \in C \times H \times W$ is used as the input. Its main objective is to obtain two different spatial context descriptors (AvgPool (F) and MaxPool (F)) from the input feature map through max pooling and average pooling respectively. The two descriptors are then passed through a shared network to generate a channel attention map. A shared network is a multilayer perceptron (MLP) [29] with one hidden layer. The number of neurons in the first layer is C/r (r is the reduction rate). After applying the shared network to each descriptor, the output feature vectors are merged using an element-wise summation. For a graph, channel attention focuses on the content of this image. Average pooling provides feedback for each pixel in a feature map. When the maximum pooling is back-propagating the gradient. The only largest response in the feature map is gradient feedback. This process can be expressed as:

$$M\_c(F) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F) + \mathrm{MLP}\,(\mathrm{MaxPool}\,(F)) \tag{2}$$

where σ is the sigmoid operation, F is the input of channel attention, and $M\_c(F)$ is the output of the channel attention module.

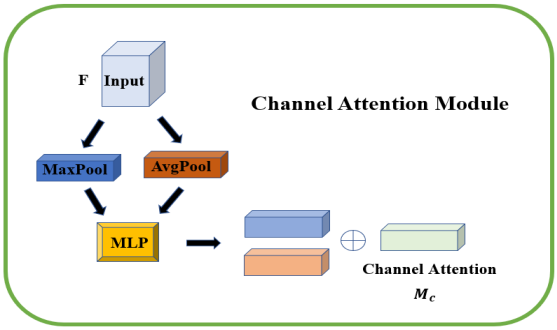


FIGURE 4. Channel attention module.

The spatial attention module of the structure is illustrated in Figure 5. Its main objective is to use the feature map output from the channel attention module as the feature map input for this module. First, two feature maps of $H \times W \times 1$ are obtained through max pooling and average pooling. They are then concatenated and convolved by standard convolution to generate a spatial attention module map. For a graph, the spatial attention focuses on "where" is an informative part, and applying pooling operations along the channel axis can effectively highlight information area, which is a complement

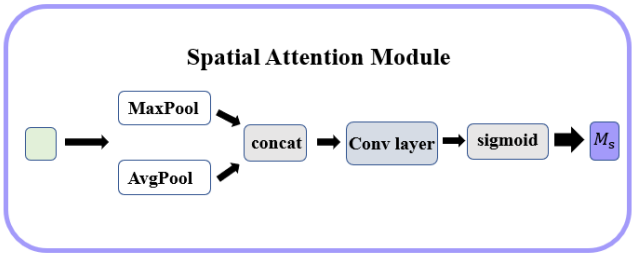


FIGURE 5. Spatial attention module.

to the channel attention. The process can be expressed as:

$$M_s(F) = \sigma\left(f^{7\times7}\left(\left[\begin{array}{c}\text{AvgPool}(F),\\ \text{MaxPool}(F)\end{array}\right]\right)\right) \tag{3}$$

where $f^{7\times7}$ is a convolution operation with a convolution kernel of 7×7 and $M_s(F)$ is the output of the spatial attention module.

It can be seen that CBAM can solve the problem of what and where the picture features are by combining the channel attention module and the spatial attention module. Therefore, in this study, CBAM is combined with U-Net network, as shown in Figure. 6, the input is first convolved through the original upsampling module, namely W1, W2 and W3 three-layer convolutions, and then convolved through the CBAM attention module to form a new upsampling module. The intermediate feature map $F(R^{C\times H\times W})$ after the three-layer convolution, first obtains a 1D channel attention map M_c ($R^{C\times1\times1}$) through the channel attention mechanism, and then obtains a 2D spatial attention map M_s ($R^{1\times H\times W}$) through the spatial attention mechanism. This process increases the ability of the network to extract local features and enables the network model to obtain a better segmentation effect. The process can be expressed as:

$$F' = M_c(F)\cdot F \tag{4}$$

$$F'' = M_s(F')\cdot F' \tag{5}$$

where F' is the result of the dot product of the feature map output after the channel attention module and the intermediate feature map F, F'' is the dot product of the feature map output after the spatial attention module and the intermediate feature map F′ is the result of.

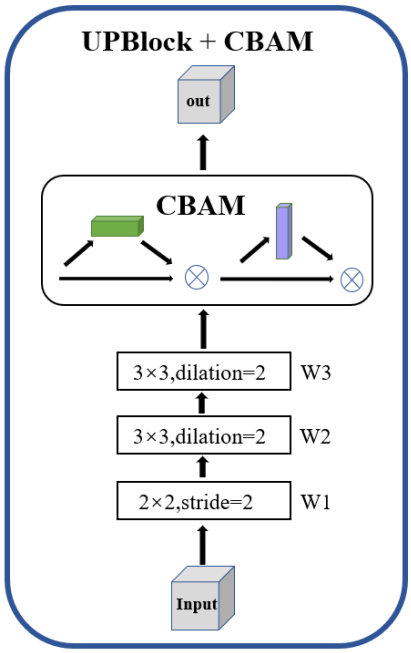


FIGURE 6. Combination of convolutional attention module and U-Net.

D. RADAM OPTIMIZER

The role of the optimizer in deep learning is to adjust the loss function by training the optimization parameters. The loss function is used to reflect the deviation of the actual value of the target value in the test set from the predicted value. Therefore, the optimization algorithm in the optimizer affects the parameters of model training and output. Commonly used optimizers are SGD [30], SGDM, RMSprop, and Adam [31]. The SGD optimizer uses stochastic gradient descent to update model parameters, which solves the problem of random small batch samples, but has problems such as adaptive learning rate; The SGDM optimizer adds a first-order momentum mechanism to the SGD optimizer to reduce the shock effect; The RMSProp optimizer is an improvement on the SGD optimizer by adding second-order momentum to better improve the problem of excessive amplitude of the function appearing during the training process, while speeding up the convergence; The Adam optimizer adds a correction term based on the SGDM optimizer and the RMSProp optimizer. By correcting the first-order and second-order momentum deviations, the algorithm can assign different adaptive learning rates to different weight parameters. The RAdam optimizer is an improvement on the basis of the Adam optimizer, and a new correction factor is introduced. The formula of the correction factor r_t is given by Equation (6).

$$r_t = \sqrt{\frac{(\rho_t-4)(\rho_t-2)\rho_\infty}{(\rho_\infty-4)(\rho_\infty-2)\rho_t}} \tag{6}$$

where ρ_t is the degree of freedom at moment of t, the calculation formula is shown in Equation (7), and ρ_∞ is the specified value, as shown in Equation (8).

$$\rho_t = f(t,\beta_2) = \frac{2}{1-\beta_2} - 1 - \frac{2t\beta_2^t}{1-\beta_2^t} \tag{7}$$

$$\rho_\infty = \lim_{t\to\infty} f(t,\beta_2) = \frac{2}{1-\beta_2} - 1 \tag{8}$$

where β_2 is the hyperparameter of the exponential moving average.

By introducing this correction coefficient, the variance term of the adaptive learning rate is corrected to alleviate the convergence problem and make the learning rate change more robust. After replacing this optimizer, the accuracy of the model is further improved, and a better segmentation effect is obtained.

III. EXPERIMENT AND ANALYSIS

In this study, the experimental software and hardware configuration are shown in Table 1.

A. DATASET AND PREPROCESSING

The blood image data used in this study were obtained from the public blood cell dataset, which contained 100 images with a resolution of 300 × 300 pixels(Graviti Open Datasets/WBC Image Dataset 2 | Graviti). Manual annotation was performed using LabelMe software. The dataset

TABLE 1. Software and hardware configuration of the test.

Name	Model and parameters
CPU	Intel® Core™ i9-11900k
GPU	GTX 1080Ti
OS	Ubuntu
Development language	Python
Environment	PyTorch1.10

was expanded to 500 blood cell images using augmentation processes such as rotation, zooming, and random cropping to enhance the data sample. Eighty percent of the entire dataset was used as the training set, and 20% was used as the test set.

The experimental comparison network model U-Net is referred to as the prototype framework in this study. The number of training rounds was set to 80, the initial learning rate was set to 0.01, the optimizer RMSProp was used for gradient descent, and the batch size was set to 4.

B. SEGMENTATION EVALUATION CRITERIA

To quantitatively evaluate the segmentation results of the algorithm, the following evaluation indicators were mainly used: intersection over union (IOU), recall, and accuracy (ACC), The formula is shown in Equations (9)-(11).

$$\mathrm{IOU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{9}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{10}$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{11}$$

where TP is the correct number of pixels for white blood cell segmentation in pixel-level segmentation, that is, the pixel point that is predicted to be a white blood cell and is correctly predicted. TN is the number of pixels for background segmentation in pixel-level segmentation, that is, the pixel points that are predicted to be background and are correctly predicted. FN is the number of pixels for background segmentation in pixel-level segmentation, that is, the pixel points that are predicted to be background but are predicted wrong. FP is the number of pixel points with incorrect leukocyte segmentation in pixel-level segmentation, that is, pixels predicted to be leukocytes but predicted incorrectly.

C. ABLATION EXPERIMENT

First, ablation experiments were conducted on the attention module CBAM improvement points by adjusting their positions durning the process of model upsampling. First they are added behind the first layer upsampling, second layer upsampling and third layer upsampling, and their segmentation results are observed and analyzed, as shown in Table 2. It can be seen that the effect of adding the attention mechanism to the model is positive, and the results are all better than the segmentation effect of the original model. Then combined addition experiments were performed, and the results after adding to the first and second layer, the second and third layers, the first and third layer upsampling, and after adding to the first three layers upsampling at the same time are shown in Table 2. By observing and analyzing the above segmentation results and evaluation metrics, it can be concluded that the optimal segmentation effect can be obtained by adding the attention module CBAM after upsampling the first three layers at the same time. Therefore, the attention mechanism module of this model was added to increase the segmentation accuracy of the model after the first three layers of upsampling modules.

Second, in terms of the choice of optimizers, Adam and RAdam were selected for comparison with the original optimizer RMSProp, that, the experiments were conducted by replacing the optimizer only based on the original model and analyzing the effect of their evaluation metrics as shown in Table 3. It can be observed that the RAdam optimizer yields better results in this model. Therefore, the optimizer of this model is chosen as the RAdam optimizer to improve the robustness of the learning rate.

Finally, ablation experiments are performed for three improvement directions of the model. The segmentation results for the test set are shown in Figure 7. Figure 7 (a) shows the image information of blood cells, (b) shows the labels of manually segmented blood leukocytes, (c) shows the segmentation result of the original U-Net model, and (d) shows the model segmentation after changing only the hole convolution result. (e) shows the model segmentation result after only adding the attention mechanism CBAM, (f) shows the model segmentation result after only replacing the optimizer Radam, (g) shows the segmentation results of the model with a mixture of replacing the dilated convolution and adding the attention mechanism, (h) shows the segmentation result of the model that not only replaces the atrous convolution, but also adds the attention mechanism and replaces the optimizer. It can be seen that the improved model exhibits a better performance for the boundary segmentation of leukocytes.

The evaluation indices of each model were calculated as shown in Table 4. It can be seen that by changing the dilated convolution, adding the attention module CBAM and replacing the optimizer RAdam, the segmentation effect of the original model has been enhanced to varying degrees. When combined, the model obtained the best segmentation result. The IOU of the improved model exceeded the benchmark U-Net by 5.1%, the recall exceeded 5.7%, and the Acc exceeded 1.2%. Compared with the benchmark U-Net model, the model proposed in this study has better stability and segmentation effect.

D. COMPARATIVE EXPERIMENT

To verify the segmentation performance of the improved model, two mainstream medical image segmentation models were selected for comparison: ResUNet and DeepLabv3+. At the same time, it is compared with the model MIF-Net

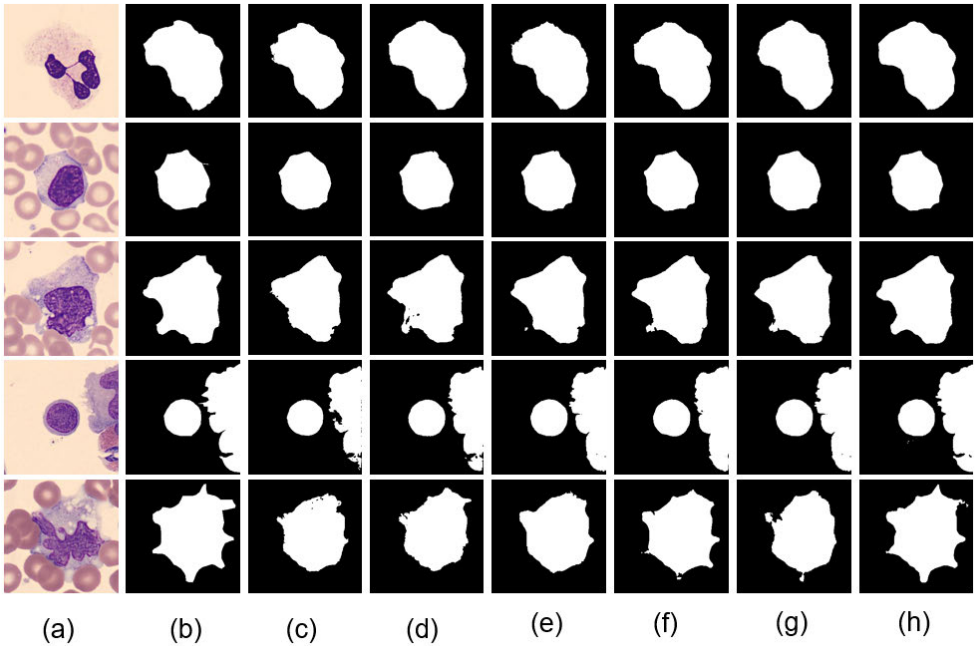


FIGURE 7. Schematic diagram of blood leukocyte segmentation.

TABLE 2. Experimental results of attention mechanisms.

CBAM add location	IOU (%)	Recall (%)	Acc (%)
/	91.1	91.7	97.9
After layer 1 upsampling	93.6	94.4	98.5
After layer 2 upsampling	93.0	95.7	98.3
After layer 3 upsampling	93.3	94.6	98.5
After upsampling of layers 1 and 2	94.1	94.9	98.6
After upsampling of layers 2 and 3	94.4	96.2	98.7
After upsampling of layers 1 and 3	93.8	95.8	98.4
After upsampling on layers 1, 2, and 3	94.8	95.8	98.8

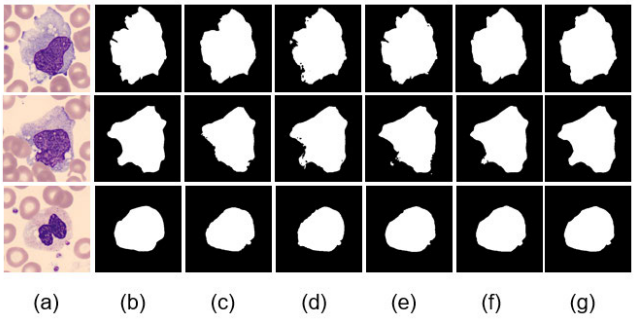


FIGURE 8. Comparative test segmentation results.

TABLE 3. Optimizer experimental results.

Optimizer	IOU (%)	Recall (%)	Acc (%)
RMSprop	91.1	91.7	97.9
Adam	94.9	96.1	98.8
RAdam	95.1	96.7	98.8

proposed in 2022 for blood cell segmentation. The segmentation results of the different models are shown in Figure 8, where (a) is the image information of blood cells, (b) is the labels of manually segmented blood leukocytes, (c) is the segmentation result of the original U-Net model, (d) is the segmentation result of the ResUNet model, (e) is the segmentation result of the DeepLabv3+ model, (f) is the segmentation result of the MIF-Net model, and (g) is the segmentation result of the model in this study. The quantitative evaluation indicators for the segmentation results are presented in Table 5.

It can be seen from the figure 8 that the model in this study can segment the morphology of the leukocytes more accurately. Compared to the comparison model, the boundary

TABLE 4. Statistical analysis of segmentation results of different models.

Basic Network	Dilated convolution	CBAM	RAdam optimizer	IOU (%)	Recall (%)	Acc (%)
U-Net				91.1	91.7	97.9
	√			93.8	94.8	98.5
		√		94.8	95.8	98.8
			√	95.1	96.7	98.8
	√	√		95.3	96.5	98.9
	√		√	96.0	97.0	99.0
		√	√	95.9	97.1	99.0
	√	√	√	96.2	97.4	99.1

TABLE 5. Comparative model segmentation results.

Model	IOU (%)	Recall (%)	Acc (%)
U-Net[17]	91.1	91.7	97.9
ResUNet[32]	92.0	92.5	98.1
DeepLabV3+[14]	93.9	94.5	98.5
MIF-Net[33]	95.3	96.2	98.4
CBAM-DC-UNet	96.2	97.4	99.1

segmentation is clearer, which can reduce the influence of the adhesion of surrounding erythrocytes. It can be seen from the quantitative evaluation indicators that the CBAM-DC-UNet model in this study showed better results for all indicators. As shown in Table 5, CBAM-DC-UNet can achieve 96.2% and 99.1% in IOU and Acc evaluation indicators,repectively, which are 4.2% and 1% higher than the results of the ResUNet network, respectively. Compared with DeepLabv3+, these two indicators are 2.3% and 0.6%, respectively. Compared to MIF-Net, these two indicators improved by 0.9% and 0.7%, respectively. It can be seen that the network in this paper has better performance in the segmentation of blood leukocytes.

IV. CONCLUSION AND OUTLOOK

In view of the important role of the detection of white blood cell morphology in medical research on diseases such as immunodeficiency, and the lack of clear boundaries in the segmentation process. This study proposes a CBAM-DC-Unet based on U-Net. (i) Using dilated convolution to replace the traditional convolution, which increases the receptive field of the model and improves the model's ability to extract the characteristic information of white blood cells, (ii) in the upsampling process, combined with the attention mechanism CBAM, the problem of what the image features and where the features can be better solved, and the model's ability to extract white blood cell features is strengthened, (iii) using the optimizer RAdam to adjust the loss function better and make the learning rate changes more robust, which improves the segmentation accuracy of the model. On the blood cell dataset, the IOU of the proposed model reached 0.962, the recall rate reached 0.974, and Acc reached 0.991. Compared with the U-Net model, the improvements were 0.051, 0.057, and 0.012, respectively, which verifies the effectiveness of the model. The next step will be to study the multi-class segmentation of blood cell images, that is, to segment white blood cells, white blood cell nuclei and red blood cells in blood cell images at the same time, and to increase the usefulness of the segmentation results for medical research.

REFERENCES

[1] C. Wang, H. Zhang, Z. Li, X. Zhou, Y. Cheng, and R. Chen, "White blood cell image segmentation based on color component combination and contour fitting," *Current Bioinf.*, vol. 15, no. 5, pp. 463–471, Oct. 2020.

[2] Z. Zhong, T. Wang, K. Zeng, X. Zhou, and Z. Li, "White blood cell segmentation via sparsity and geometry constraints," *IEEE Access*, vol. 7, pp. 167593–167604, 2019.

[3] S. Biswas and D. Ghoshal, "Blood cell detection using thresholding estimation based watershed transformation with Sobel filter in frequency domain," *Proc. Comput. Sci.*, vol. 89, pp. 651–657, Jan. 2016.

[4] D. R. Nayak, N. Padhy, and B. K. Swain, "Blood cell image segmentation using modified fuzzy divergence with morphological transforms," *Mater. Today, Proc.*, vol. 37, pp. 2708–2718, Feb. 2021.

[5] T. A. Soomro, L. Zheng, A. J. Afifi, A. Ali, S. Soomro, M. Yin, and J. Gao, "Image segmentation for MR brain tumor detection using machine learning: A review," *IEEE Rev. Biomed. Eng.*, early access, Jun. 23, 2022, doi: 10.1109/RBME.2022.3185292.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

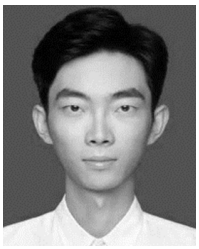
[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

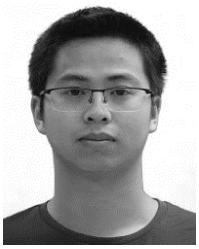
[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[10] 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.
[11] K. Chen, K. Fu, M. Yan, X. Gao, X. Sun, and X. Wei, "Semantic segmentation of aerial images with shuffling convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 173–177, Feb. 2018.
[12] 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.
[13] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
[14] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 801–818.
[15] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1925–1934.
[16] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.
[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image SEG-mentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Cham, Switzerland: Springer, Nov. 2015, pp. 234–241.
[18] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, Cham, Switzerland: Springer, Sep. 2018, pp. 3–11.
[19] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: A generalist algorithm for cellular segmentation," *Nature Methods*, vol. 18, no. 1, pp. 100–106, Jan. 2021.
[20] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, "Cell detection with star-convex polygons," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Jun. 2018, pp. 265–273.
[21] S. Teng, J. Wu, Y. Chen, H. Fan, X. Cao, and Z. Li, "Semi-supervised leukocyte segmentation based on adversarial learning with reconstruction enhancement," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.
[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[23] S. Yin, H. Deng, Z. Xu, Q. Zhu, and J. Cheng, "SD-UNet: A novel segmentation framework for CT images of lung infections," *Electronics*, vol. 11, no. 1, p. 130, Jan. 2022.
[24] B. Li, S. Liu, F. Wu, G. Li, M. Zhong, and X. Guan, "RT-UNet: An advanced network based on residual network and transformer for medical image segmentation," *Int. J. Intell. Syst.*, vol. 37, no. 11, pp. 8565–8582, Aug. 2022.
[25] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.
[26] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.
[27] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," 2019, *arXiv:1908.03265*.
[28] T. Tieleman and G. Hinton, "RMSProp: Divide the gradient by a running average of its recent magnitude," *COURSERA Neural Networks Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
[29] I. Tolstikhin, "MLP-mixer: An all-MLP architecture for vision," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 24261–24272.
[30] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[32] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet–A: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020.
[33] N. Akram, S. Adnan, M. Asif, S. M. A. Imran, M. N. Yasir, R. A. Naqvi, and D. Hussain, "Exploiting the multiscale information fusion capabilities for aiding the leukemia diagnosis through white blood cells segmentation," *IEEE Access*, vol. 10, pp. 48747–48760, 2022.

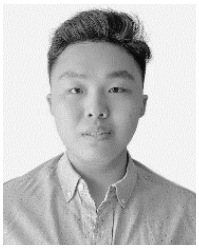
DONGMING LI received the Ph.D. degree from the Changchun University of Science and Technology, Changchun, China, in 2021. In 2001, he joined Jilin Agricultural University, Changchun, where he is currently a Professor carrying out teaching and research on projects. He was a Visiting Scholar at CSIRO, Australia, from 2016 to 2017. His current research interests include computer vision, image analysis, and pattern recognition.



SHIYU YIN was born in Jilin, China. He received the B.E. degree from the Changchun University of Science and Technology, in 2018. He is currently pursuing the master's degree with Jilin Agricultural University. His research interests include deep learning and medical image processing.



YU LEI was born in Jilin, China. He received the B.S. degree from Jilin Agricultural University, in 2020, where he is currently pursuing the master's degree. His research interests include deep learning and medical image processing.



JINGNING QIAN was born in Anhui, China. He received the degree in computer science from the Changchun University of Science and Technology, in 2022. He is currently pursuing the master's degree in software engineering with The University of Melbourne, Australia. His research interests include deep learning and software project.

CHUNXI ZHAO received the master's degree in education from Jilin Agricultural University, in 2008. He is currently working with the Center for Informatization, Jilin Agricultural University. His main research interests include computer networks and artificial intelligence applications.

LIJUAN ZHANG received the M.E. and Ph.D. degrees from the Changchun University of Science and Technology, Changchun, China, in 2004 and 2015, respectively. She was a Visiting Scholar at CSIRO, Australia, from 2016 to 2017. She is currently a Professor with the College of Computer Science and Engineering, Changchun University of Technology. Her research interests include image restoration, computer vision, and image analysis.