RESEARCH ARTICLE

Ophthalmic OCT Segmentation Method Based on RCNN-Attention

BINGBING LI[1,2], YAO CONG[1], AND HONGWEI MO [ID][1]

[1]College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China
[2]Technology School, Jilin Business and Technology College, Changchun 130102, China

Corresponding author: Hongwei Mo (mohongwei2023@163.com)

ABSTRACT As the growth of science and technology, ophthalmic optical coherence tomography (OCT) image segmentation plays a key role in ophthalmic diagnosis. To improve the accuracy of segmentation, the experiment proposed an ophthalmic OCT segmentation method that combines the recurrent residual network (ResNet) and the attention mechanism (AM). In the process, the graph search (GS) algorithm is first used to perform fine segmentation operations on the OCT image, then a recurrent residual convolution network is introduced to correct the phenomenon of image drill and rapid degradation, and finally the AM is integrated to improve the utilization of the global information of the image and achieve accurate segmentation of the image. The results show that the research methodis tested on different data sets, and the research method shows the largest fitness value. There are four models, research methods, medical image segmentation technologies based on full-size connection Unet 3+ (UNet 3plus), recurrent residual convolutional neural network (ResNet-RCNN) and deep learning and graphicssearch (DL-GS). The areas under the ROC curve of the four models of the patient retinal layer boundary segmentation method searching for automatic segmentation technology are 0.956, 0.911, 0.897 and 0.856 respectively. When the precision rate is 0.900, the recall rates of the four models corresponding to the research method, Unet 3+, ResNet-RCNN and DL-GS are 0.801, 0.663, 0.574 and 0.438 respectively. At the same time, the system can reach a stable state within 0.501s of the running time of the research method. When applying the research method to the segmentation of practical visual ophthalmic OCT images, the outcomes are closer to the results of artificial expert annotation. Based on the above, the research method's segmentation accuracy and system stability are better, providing a certain reference for the optimization of image segmentation technology in the medical field.

INDEX TERMS Residual neural network, attention, OCT images, ophthalmology, image segmentation.

I. INTRODUCTION

Optical Coherence Tomography (OCT) in ophthalmology is a non-invasive imaging technique that offers ophthalmologists with high-resolution cross-sectional images similar to tissue microscopes [1], [2]. Since its introduction in 1991, OCT had produced great impact on the field of ophthalmic diagnosis and treatment, especially in the early detection and treatment planning of common diseases such as macular diseases and

glaucoma [3]. However, due to the complexity, noise, and differences between different devices of OCT images, their segmentation remains a challenging issue. To accurately extract various levels of the eye from OCT images, researchers and engineers have tried various algorithms and technologies. In recent years, deep learning, especially Convolutional Neural Networks (CNN), has made breakthrough progress in medical image analysis, including segmentation of OCT images [4], [5]. Although traditional CNN has realized quite good achievements in OCT image segmentation, how to further promote the accuracy, stability, and robustness of

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

segmentation is still the focus of research. Recurrent residual convolutional neural network (RCNN)for medical, as an efficient network structure, has shown excellent performance in multiple fields. It captures long-distance information by introducing circular connections and alleviates the problem of gradient disappearance through residual connections, thus enabling better processing of complex structural information. Meanwhile, attention mechanism (AM) has been widely utilized in various machine learning tasks. Its core idea is to assign different weights to different parts of input data, enabling the network to focus on more key information. This study aims to explore a novel OCT image segmentation method that integrates RCNN and AM. The experiment suggests that this approach can better utilize the contextual information in OCT images, thereby achieving more accurate and stable eye structure segmentation. In addition, the method that integrates RCNN and AM can also provide better interpretability for the model. In the field of medicine, it is crucial to understand how models make decisions, especially when it comes to critical medical diagnoses and decisions. By introducing these two mechanisms, the research aims to provide an OCT image segmentation method that has both high segmentation performance and good interpretability.

The main contributions of this article can be divided into three points. Firstly, an image segmentation method based on recurrent ResNet and improved graph search(GS) algorithm is proposed in the experiment. It effectively combines the advantages of ResNets and recurrent networks. Specifically, RR Net uses ResNets as the backbone network to protect the integrity of underlying information transmission and further optimizes the retinal layer segmentation results using GS algorithm. Secondly, an ophthalmic OCT boundary segmentation algorithm based on RCNNis proposed in the experiment to focus more on finding significant useful information related to segmentation results in retinal QCT images. Thirdly, the proposed method is evaluated from different perspectives to demonstrate that the research method has the best segmentation accuracy.

The innovation of this article is as follows: the experiment first uses the AM. This method can automatically locate and enlarge the most critical parts of the image for segmentation, significantly improving the model's ability to recognize complex structures and edges, thereby improving the accuracy of segmentation. Then the recurrent ResNet is fused to enable the model to use the same features multiple times for learning and optimization, which strengthens the recognition of subtle structures that are difficult to distinguish in OCT images.

The research content can be mainly divided into four parts. The first part mainly summarizes the current research status at home and abroad; The second part is an introduction to the method. Firstly, the GSalgorithm is applied to raise the image segmentation's accuracy. Secondly, a RCNN is introduced to avoid image degradation. Finally, Attention is fused to search for global information in the image, and the system is successfully constructed. The third part is the performance testing and application effect analysis of the model constructed by the research institute; The fourth part is a summary statement of the research content.

II. RELATED WORKS

In recent years, many different theories have been widely applied to the study of retinal layer boundary segmentation, such as threshold method, active contour model, level set model, Markov random field model, graph theory, etc. Some scholars use a combination of structural tensor and kernel regression models to segment the retinal layer boundaries to preserve the complex structural information in OCT images. Generally speaking, these methods mostly rely on prior knowledge or complex processing to obtain handmade features. However, due to the presence of noise and low contrast in OCT images, the segmentation of retinal layer boundaries remains a challenge. In view of this, many scholars have conducted detailed analysis on the application of different image segmentation methods. S Minaee et al. summarized image segmentation techniques and found that a series of methods such as scene understanding, medical image analysis, and robot perception can be applied to image segmentation, achieving good segmentation results [6]. Liu X et al. proposed an image segmentation method based on adaptive region growth and Otsu algorithm to segment insulator images generated during the power detection process. During the experiment, 8 neighboring pixels were applied to region growth to process images of different shapes. The results showed that this method could completely extract rough image contour information and had better performance [7]. To effectively improve the detection efficiency of the COVID-19 diagnosis system, researchers such as Pushokhin DA proposed a bidirectional LSTM like attention layer based on residual network(RCAL-BiLSTM). The system first preprocessed the data during operation, and then used the SM layer to categorize the feature vectors. The outcomes denoted that the sensitivity of the constructed model was 93.28%, the accuracy was 94.90%, and all performance indicators were significantly better [8]. Zhang's team proposed an image segmentation method based on edge information and Deeplabv3+ network for effective segmentation of tongue images. During the process, the network structure was optimized and a new fitness function was proposed, which utilizes prior knowledge to post-process the results. The results showed that this method had a high accuracy in image segmentation and could effectively improve the ambiguity of image segmentation [9]. Xu and other scholars proposed a method based on phase attention ResNet to effectively model multiphase features to achieve segmentation of liver tumors in patients. The fusion of multi-scale architecture and 3D boundary enhancement fitness during the process trained the system model, making the network more aware of tumor boundaries. The findings indicated that the raised method could accurately detect multiphase CT datasets and was superior to other state-of-the-art methods [10].

Meanwhile, scholars have also conducted numerous discussions on image segmentation techniques in the medical field. Srivastava A's team proposed a new method based on multi-scale residual fusion network (MSRF-Net) for medical image segmentation. During the process, the model could use dual scale dense fusion blocks to interchange features of different receptive fields. Through a large number of experiments, it has been proven that the effectiveness of the research method was significantly better than other excellent methods on four different public datasets, and it had a relatively universal applicability [11]. Gao et al. put forward an innovative MSRF-Net algorithm to raise the capture of spatial features in remote sensing image processing, and applied it to hyperspectral image classification. The proposed method could not only extract features of different sizes from different feature maps, but also effectively reduce the learnable parameters in the network. Through data experimental verification, it was found that the proposed method had the most effective resolution ability and accuracy for remote sensing images [12]. Fu and other researchers proposed a RCNN algorithm based on frugal attention for effective recognition of remote sensing scene images, and applied it to spatiotemporal long-range multi-hop communication of remote sensing images. The effectiveness of the algorithm was tested through different image datasets. The findings denoted that the model parameters proposed by the research institute were very accurate and could accurately classify and detect target detection objects [13]. Scholars such as Yu proposed an improved ResNet-based apple leaf disease recognition model to ensure the healthy growth of apples. During the experiment, the model effectively reduced the number of model parameters and improved the recognition accuracy and efficiency by decomposing the convolutional kernel, updating the mapping method, and reducing residual modules. The outcomes expressed that the average accuracy and recall rate of the research model were superior, and the overall effectiveness of the model was significantly better than other models, which can provide effective technical support for apple leaf disease [14]. Li and his team members proposed a method based on progressive recursive image enhancement networks to enhance the visual effectiveness of low light images. During the process, local feature extraction was executed on the input system image through a recursive unit composed of a recursive layer and residual blocks. Then, the global features of dual attention were extracted in a gradual manner. The outcomes expressed that the constructed method could effectively reduce the number of model parameters and maintain stable system performance [15].

From historical literature, it can be seen that the early OCT image segmentation technology was mainly based on traditional image processing methods, such as edge detection, threshold segmentation, etc. However, with the rapid development of deep learning and neural network technology, cyclic residual networks (CRNs) have been widely used in OCT image segmentation, as they can effectively capture detailed information in images and handle complex background noise. However, there is currently limited research on the accurate segmentation of ophthalmic images using cyclic residual networks and AM. In view of this, the study innovatively introduces graph search algorithms and combines RCNN and AM to achieve precise segmentation of ophthalmic images. We look forward to providing new technical support for ophthalmologists and effectively capturing detailed information in images.

III. OPHTHALMIC OCT IMAGE SEGMENTATION METHOD BASED ON FUSION OF RCNN-AM

To provide clinicians with more accurate ophthalmic OCT image segmentation tools, an innovative method that integrates RCNN and AM has been proposed. It is expected that the research method will open up a new path for ophthalmic OCT image segmentation research and bring important value for future applications.

A. IMAGE SEGMENTATION METHOD BASED ON IMPROVED GS ALGORITHM

OCT images can clearly reveal each layer of the retina's structure, timely detect early small lesions, and provide a standard for the etinal diseases diagnosis [16], [17]. The current retinal layer segmentation technology based on image block classification may bring redundant information when selecting image blocks, and the segmentation process may lead to the fitness of global information. For full image segmentation methods, the contextual and significant information related to segmentation are often overlooked. To effectively utilize the correlation between image pixels and learn the contextual information of the image, a retinal boundary segmentation method based on recurrent RCNN is proposed. To further improve segmentation accuracy, a new cyclic residual network method was designed in the experiment on the segmentation of retinal layer boundaries, named RR-Net. Four GRUs were used to form a cyclic block similar to ReNet structure, in which the encoder and decoder were included. Applying RR Net to feature learning of ophthalmic OCT images is expected to achieve coarse segmentation results.

The operation of the RR-Net network structure can be broken into four stages. In the first stage, a total of 3 residual blocks are used, with a convolutional layer filter of 16 and a step size of 1 for all residual blocks. Additionally, 2 recurrent blocks are used, with the corresponding number of GRU units set to 8. The second stage applies 2 recurrent blocks and 4 residual blocks. The convolutional layers corresponding to residual and recurrent blocks are multiplied by the number of GRUs. In the third stage, 6 residual and 2 recurrent blocks are used; The fourth stage utilizes3 residual and 2recurrent blocks. After all stages are set, upsampling deconvolution and downsampling convolution are added before and after the four stages, respectively, to enlarge and shrink the image. Simultaneously, deconvolution is used to achieve restoration of image size. Before using the RR-Net network
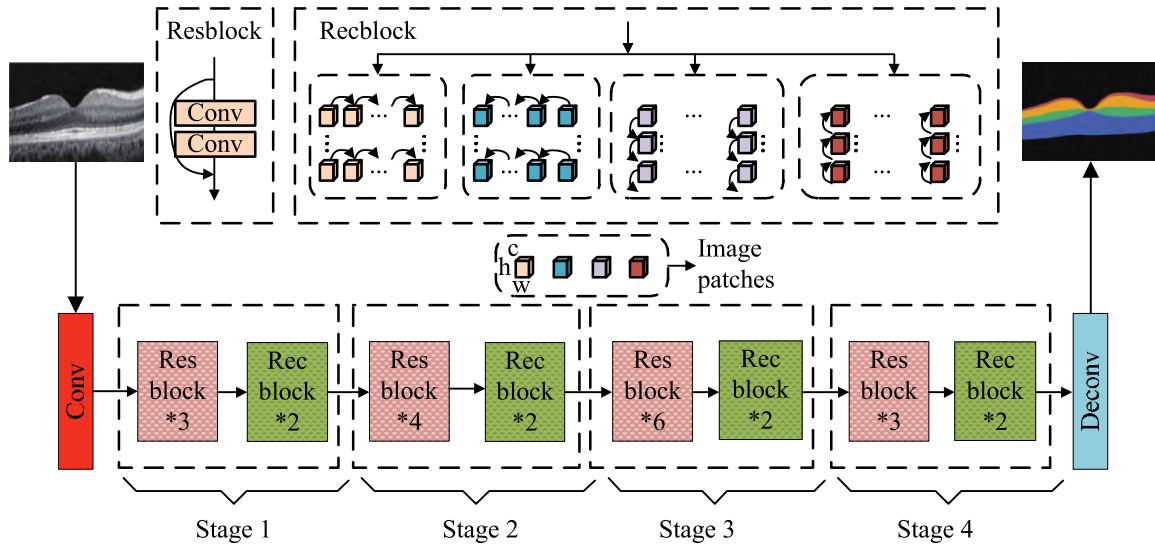
FIGURE 1. Overall network structure of RR-Net.

structure to process images, it is necessary to preprocess the data information in the images to remove redundant information. Then, residual blocks are used to form a backbone network. Usually, there are certain constraints and continuity constraints on the boundaries of the retinal layer. However, in the data feature learning on the retinal layer using network structures, this constraint is not taken into account in the system, resulting in the structure of layer segmentation being prone to small holes or isolated regions. To eliminate this phenomenon, the experiment introduces a GS algorithm for fine segmentation of OCT images. In the OCT image composition GS algorithm, each graph contains different nodes, and the connection weights of the nodes are calculated, as shown in equation (1) [18], [19].

$$\begin{cases} w_{ij} = 2 - (g_i + g_j) + w_{\min} \\ w_{ij,t} = 2 - (P_{i,t} + P_{j,t}) + w_{\min} \end{cases} \quad (1)$$

In equation (1), $w_{\min}$ means a small positive number used to stop the value of the connection weight from being 0; g_i, g_j mean the nodes i and j's weight. $P_{i,t}$, $P_{j,t}$ denote the pixels i, j's probability values in the probability map labeled t. The value of its weight is inversely proportional to the amount of candidate pixels corresponding to the two connected nodes. Generally speaking, traditional GS algorithms must find candidate regions for the boundary when using networks for classification, and cannot decide the final retinal layer boundary. Thus, the experiment improves the GS algorithm and utilizes the Dijkstra shortest path algorithm to detect layer boundaries, and constructs a sparse adjacency matrix to effectively improve the algorithm's efficiency [20], [21].

B. EYE OCT BOUNDARY SEGMENTATION ALGORITHM BASED ON RCNN

In computer vision, the processing of visual signals perceived by humans is modulated by context. In image segmentation,
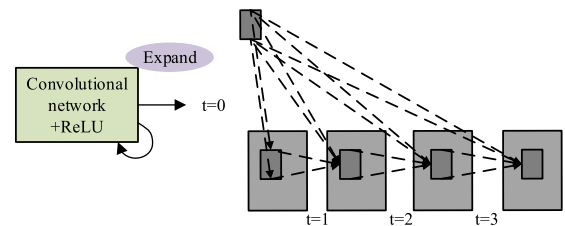

FIGURE 2. Structure of the ResNet-RCNN module.

contextual information can be used to extract more comprehensive feature maps and more accurately classify different pixels. To fully apply the obtained information to practical applications, the experiment uses a recurrent convolutional neural network (ResNet-RCNN) to connect in the same layer of the deep learning algorithm [22]. The structure of the ResNet-RCNN module is shown in Figure 2.

Large squares represent feature maps in ResNet-RCNN. The smallest square represents the convolution kernel. In the output of the module, both recurrent connections and feedforward connections have shared weights between different positions and local connections. When the module processes the image, the output calculation of the unit located at the k feature map (i, j) at time step t is indicated in equation (2).

$$O_{ijk}^l(t) = (W_k^f)^T * x_l^{f(i,j)}(t) + (W_k^r)^T * x_l^{r(i,j)}(t-1) + b_k \quad (2)$$

In equation (2), $x_l^{f(i,j)}(t)$ represents the standard convolutional layer input; $x_l^{r(i,j)}(t-1)$ means the input of the recurrent convolutional layer in the lth layer; W_k^f and W_k^r denote the weight of the standard feedforward and current convolutional layers of the kth feature map; r indicates a recurrent convolutional layer; f expresses the standard feedforward convolutional layer; b_k is deviation. With the
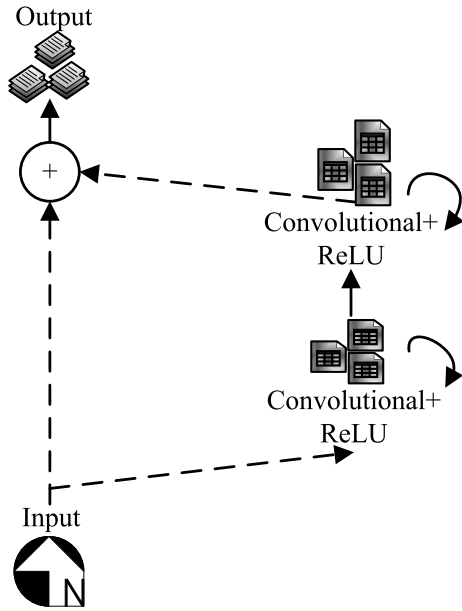
learning and in-depth training of ResNet-RCNN, the depth of the network will affect the final results obtained. When the network reaches a certain depth, rapid degradation occurs, ultimately leading to higher training errors [23]. In view of this, some scholars have proposed a ResNet network, which effectively solves the problem of gradient disappearance and network digestion. Therefore, the experiment combines ResNet-RCNN with residual learning to form recurrent residual convolutional units for better feature enhancement. The specific structure is shown in Figure 3.

In the recurrent residual convolution unit, the ReLU activation function is used to alleviate the problem of over fitting in network learning, as defined in equation (3).

$$f(x) = \max(0, x) \qquad (3)$$

In equation (3), $f(x)$ indicates the ReLU activation function. In practical applications, the ReLU function will directly take the region of the network output that is smaller than zero as 0 to enhance the sparsity of the network. After the ReLU activation function operation, the output representation of the lth layer of the recurrent convolution module is shown in equation (4).

$$F(x_l, w_l) = f\left(O_{ijk}^l(t)\right) = \max\left(0, O_{ijk}^l(t)\right) \qquad (4)$$

In equation (4), F means the module output. This output can be used for the upper and lower sampling layers. Next, the recurrent residual and convolutional layer learning are combined, and the final recurrent convolution module is output through the residual unit. The calculation is shown in equation (5).

$$x_{l+1} = x_l + F(x_l, w_l) \qquad (5)$$

In equation (5), x_{l+1} denotes the residual module. The output of this module is utilized as the input for the upper and lower sampling layers of the experimental method. The network of encoder and decoder structures is widely used in medical image segmentation. To meet the high-precision requirements, the U-Net network is combined with full scale jump connections in the experiment [24]. The structure of full scale jump connections is shown in Figure 4.

Whether to use full scale jump connections does not affect the parameter quantity. During the decoder operation phase, the amount of channels in the ith decoding stage of the U-Net network decoding section is displayed as 32×2^i, and the corresponding parameter quantity calculation is shown in equation (6).

$$P_{SC}^i = D_C \times D_C \times \left[d\left(X_{De}^{i+1}\right) \times d\left(X_{De}^i\right) + d\left(X_{De}^i\right)^2 \right.$$
$$\left. + d\left(X_{En}^i + X_{De}^i\right) \times d\left(X_{De}^i\right) \right] \qquad (6)$$

In equation (6), D_C refers to the size of the convolutional kernel; $d(\cdot)$ expresses the depth of the node; X_{En} represents the node of the encoder; X_{De} severs as the node of the decoder. Then, full scale jump connections are used, and different decoders are connected to N at different scales, resulting in a channel count of $64 \times N$. The calculation of some parameters of the decoder is indicated in equation (7).

$$P_{FSC}^i = D_C \times D_C \times \left[d\left(X_{De}^{i+1}\right) \times d\left(X_{De}^i\right) + d\left(X_{De}^i\right)^2 \right.$$
$$\left. + d\left(X_{De}^i + X_{En}^i\right) \times d\left(X_{De}^i\right) \right] \qquad (7)$$

According to equation (7), a decrease in the number of decoder channels will result in a decrease in the final parameter count. To seamlessly integrate shallow and deep feature information, the experiment further utilizes feature aggregation mechanisms to aggregate the concatenated feature maps. Formally, this mechanism can effectively represent jump connections as follows: node i samples down the encoder index layer, and N represents the total amount of encoders. The feature map stack is figured out, as expressed in equation (8).

$$X_{De}^i = \begin{cases} X_{En}^i, & i = N \\ H\left(\left[C\left(D\left(X_{En}^k\right)\right)_{k=1}^{i-1}, \right. \right. \\ \left. \left. C\left(X_{En}^i\right), C\left(U\left(X_{De}^k\right)\right)_{k=i+1}^N \right] \right) \\ i = 1, 2, \cdots, N-1 \end{cases} \qquad (8)$$

In equation (8), X_{De}^i refers to the feature map stack; X_{De}^i indicates convolutional operation; $H(\cdot)$ can implement feature aggregation mechanism; $D(\cdot), U(\cdot)$ denote up and down sampling operations respectively; $[\cdot]$ indicates series connection. The fitness function, also known as the cost function, is located in the last layer of ResNet-RCNN training, and is usually utilized to measure the closeness
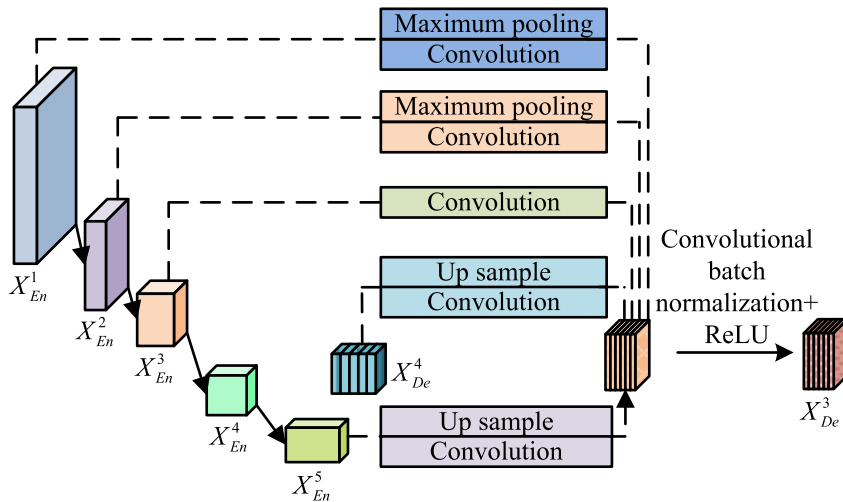
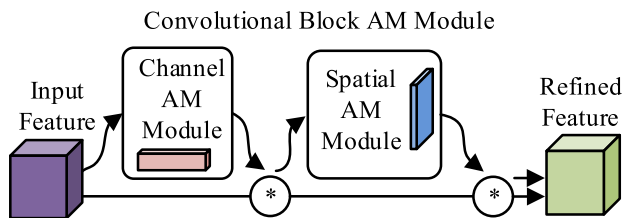**FIGURE 4.** Structure of full scale jump connections.

FIGURE 5. AMstructure.

between the predicted and the true values of the model. The specific calculation is shown in equation (9).

$$l_n = -w_n \left[y_n \cdot \ln x_n + (1 - y_n) \cdot \ln(1 - x_n) \right] \qquad (9)$$

In equation (9), l_n represents the actual output segmented image of the nth sample image in each batch of deep CNN samples; y_n stands for the label image; w_n severs as weight.

C. OPHTHALMIC RETINAL LAYER BOUNDARY SEGMENTATION ALGORITHM WITH RCNN-AM GLOBAL RESIDUAL NETWORK (AGR-NET) FUSION

The correlation between image pixels using the proposed ResNet-RCNN is analyzed, and then the contextual information of the image is learned. At the same time, to more effectively utilize the global information of the image and learn useful information related to segmentation results, AGR-Netmodel is introduced in this experiment to finely segment the retinal layer boundaries of OCT images. The AM structure is shown in Figure 5.

The AGR-Net embeds all global feature modules into the ResNet to capture the global information features of the image more comprehensively. At the same time, to make the network to pay more AMon finding useful information significantly related to segmentation results in patients' rigorous retinal OCT images, it effectively embeds the channel AM module into the network to improve the

performance of retinal layer boundary segmentation. Finally, a GS algorithm is used to perform more precise segmentation operations on the image, allowing the segmentation results closer to the truth map. The experiment utilizes AGR-Net for training, which can effectively address the network degradation. At this point, the definition of network residual is shown in equation (10).

$$y_l = F(x_l, w_l) + x_l \qquad (10)$$

In equation (10), x_l, y_l mean the input and output mapping, respectively; $F(x_l, w_l)$ indicates the residual mapping that needs to be studied. In the U-Net network, not all the features are conductive to segmentation. For this, the experiment utilizesAGs to replace the original jump connections. The main function of AGs here is to suppress irrelevant feature responses, highlight favorable feature information, and ultimately improve network performance. The structure of the AG is shown in Figure 6.

In Figure 6, x_l refers to the encoding path feature; g_l denotes decoding path features; σ_1, σ_2 denote the ReLU and Sigmoid functions. Image features are extracted using the encoding path, thenthe extracted features are into the AG, the decoding path's feature g_l is utilized as a gating signal to choose the focus area of the image. During the process, the x_l and g_l features are first added, and then the activation value of the repeating region is increased. Then, the AM coefficient α_l is calculated through the AM gate, with a value range of 0 to 1. The feature response related to the target task is exported to suppress useless feature information. The output feature of the final AG is obtained by multiplying the encoding path features x_l and α_l by elements. Finally, it concatenates the decoding path features with the obtained features. The calculation of AM gate output characteristics is shown in equation (11).
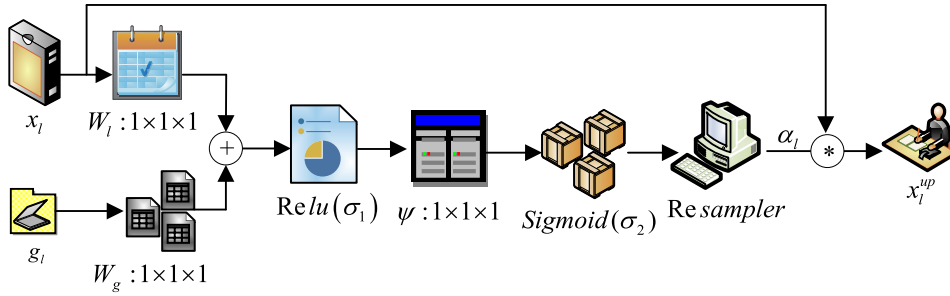
$$x_i^{up} = x_l + \alpha_l \qquad (11)$$
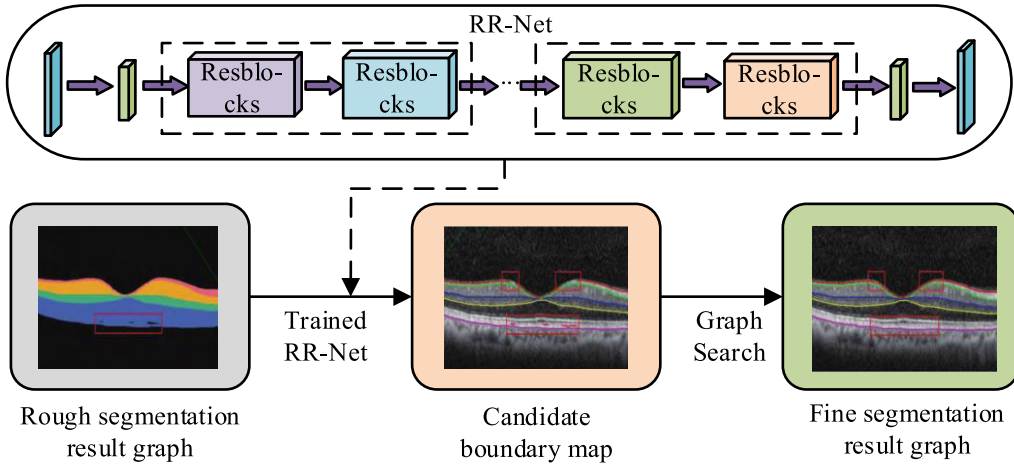
FIGURE 6. AGnetwork structure.

FIGURE 7. Overall process of OCT boundary image segmentation algorithm in ophthalmology.

In equation (11), x_i^{up} refers to the obtained feature, which is the output feature of the AM gate; α_l severs as the AM coefficient, which is obtained through additive AM and tends to obtain larger values in the ROI region and smaller values in the background region, ultimately achieving the purpose of feature filtering. The expression of the additive function is shown in equation (12).

$$\begin{cases} Q_l = \psi\left(\sigma_1\left(W_x x_l + W_g g_l + b_g\right)\right) + b_\psi \\ \alpha_l = \sigma_2\left(Q_l\right) \end{cases} \tag{12}$$

In equation (12), W_x, W_g stand for the weight; b_g, b_ψ denote the bias term; σ_1, σ_2 mean the ReLU and Sigmoid activation functions. The experiment utilizes AM based on vector concatenation, and linear transformation uses tensor channel direction 1 * 1 * 1 convolution to calculate. In addition, the update of GA parameters is trained using standard back propagation methods. The overall process of the obtained ophthalmic OCT boundary image segmentation algorithm is shown in Figure 7.

In Figure 7, first, the original OCT image is downsampled to half the size of the original image through one convolution layer, and then passes through the first part of the network: 3 residual layers (composed of two convolutions plus short connections: One residual layer), 2 global feature modules, and 1 channel AM module. Then it goes through the third

part of the network: four residual layers, 2 global feature modules, and 1 channel AM module. The fourth part of the network consists of six residual layers, 2 global feature modules, and 1 channel AM module. Finally, the fourth part of the network is passed: three residual layers, 2 global feature modules, and 1 channel AM module. To ensure that image information is not lost as much as possible, the experiment only performed image downsampling on the first convolutional layer. Therefore, in the decoding stage, only a simple transposed convolution is used for upsampling to restore the image size.

IV. PERFORMANCE TESTING AND APPLICATION EFFECTIVENESS OF OPHTHALMIC OCT IMAGE SEGMENTATION MODEL

To gain a more comprehensive understanding of the specific performance of the constructed model, the experiment verified the effectiveness of the improved CRNN-AM method in ophthalmic image segmentation systems. The experiment first tested the effectiveness of the constructed system. All experiments were conducted under the PyTorch framework, and the simulation environment settings are shown in Table 1.

The study first selected the same type of medical image segmentation method based on ResNet-RCNN, patient retinal layer boundary segmentation method based on deep learning

TABLE 1. The experimental basic environmental parameters.

Parameter variables	Parameter selection
The overall implementation platform of the system	Simulink
Operating system	Windows10
Operating environment	MATLAB
System PC side memory	24G
CPU dominant frequency	2.62Hz
GPU	RTX-2070
Central Processing Unit	i7-8700
Data storage	MySQL
Data regression analysis platform	SPSS26.0

and graphics search automatic segmentation technology (DL-GS), and the Unet 3+, the performance of the medical image segmentation technology based on full-size connection Unet 3+ (Unet 3+) was compared with the method constructed in the experiment [29], [30]. To ensure the fairness and impartiality of the experiment throughout the entire process, all algorithms were conducted under unified experimental conditions. The number of iterations set for all algorithms was the same. Then two publicly available datasets were selected to assess the actual effectiveness of different algorithms. Next, add U The of Miami OCT dataset, as the validation experimental task dataset, was collected by the Palmer Ophthalmology Institute in Bath, Miami after obtaining informed consent from patients. Select multiple experimental datasets that meet the standards, and after data preprocessing to remove redundant data, obtain 5000 valid data. Firstly, 70% of the data will be selected as the training dataset, which means that the experiment will have 3500 pieces of data (i.e. 5000 pieces) × 0.7) Used for training research models. This part of the data is used to train the model, where the model will learn the features and patterns of the data. Next, the experiment selects 30% of the data as the validation dataset, which is 1500 pieces of data (i.e. 5000 pieces) × 0.30). The validation dataset is mainly used for model tuning, such as adjusting hyperparameters, and to prevent overfitting of the model. All experimental images in the dataset were three-dimensional images. The experiment selected convergence degree, mean square error(MSE), ROC curve, PR curve, time to reach a stable state and IoU evaluation results as model performance evaluation indicators. The superior performance of the model constructed by the research was verified by comparing the performance of different algorithms. Firstly, it compared the convergence rates of different algorithms, as shown in Figure 8.

Figure 8 (a) shows the convergence comparison of different models running on dataset 1. As different models operated, the numerical changes of all algorithms showed a steep upward trend. The fitness values of the three models, Unet 3+, ResNet-RCNN and DL-GS, have been constantly changing [25]. When the number of iterations of the system was 109, the fitness value of the research method was 98.12. This also indicated that the convergence of the research method was more accurate and efficient. Figure 8(b) indicates the change test of convergence on dataset 2. The fitness of
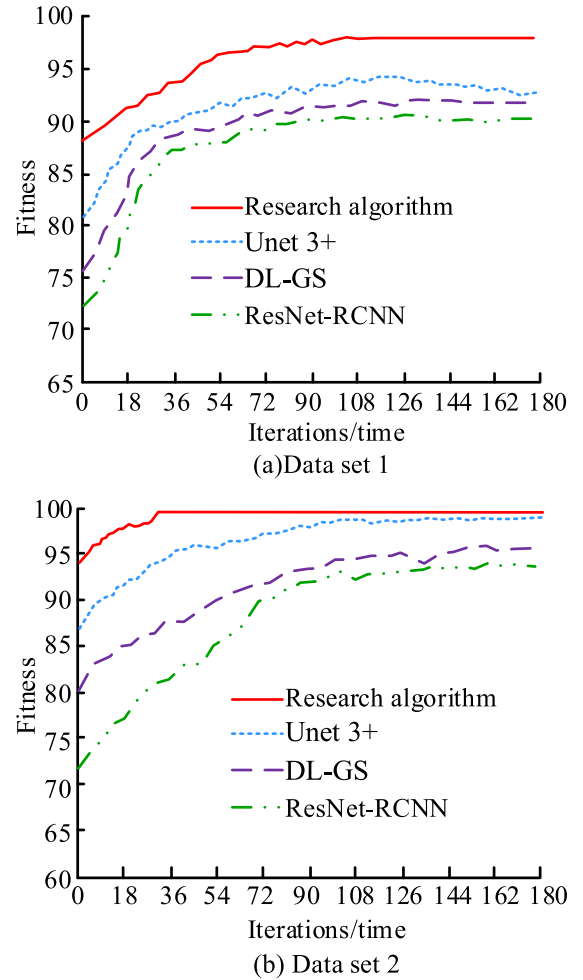

FIGURE 8. Convergence comparison of different model runs.

all algorithms was in an extreme state of change, with the research method showing a maximum fitness value that was infinitely close to 100 when the system was iterated 41 times. At the same time, the fitness values of other algorithms were still changing and have not shown stable values. The above results indicated that the research method had the highest fitness value and could quickly reach convergence speed within the same time. The ROC curves obtained by running four algorithms on dataset 2 are shown in Figure 9.

In Figure 9, the ROC curve areas of the research methods, Unet 3+, ResNet-RCNN, and DL-GS models were 0.956, 0.911, 0.897, and 0.856, respectively. The AUC value of the proposed method wasobviously higher than the other three algorithms. This indicated that the ophthalmic OCT image segmentation results obtained using the research method had the highest authenticity, and the detection accuracy of the research method during application was high, which was beneficial for ophthalmologists to better diagnose the patient's condition and achieve good disease intervention. On this basis, the Mean Squared Error(MSE) of image segmentation detection using four algorithms on two datasets was compared, and the results are shown in Figure 10.
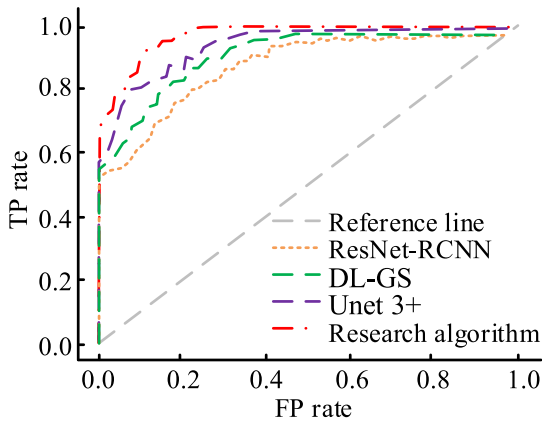
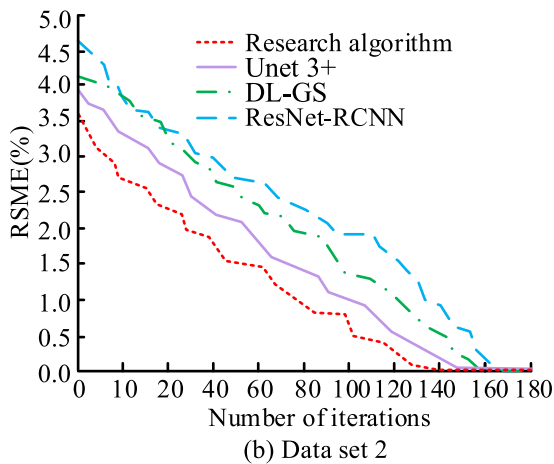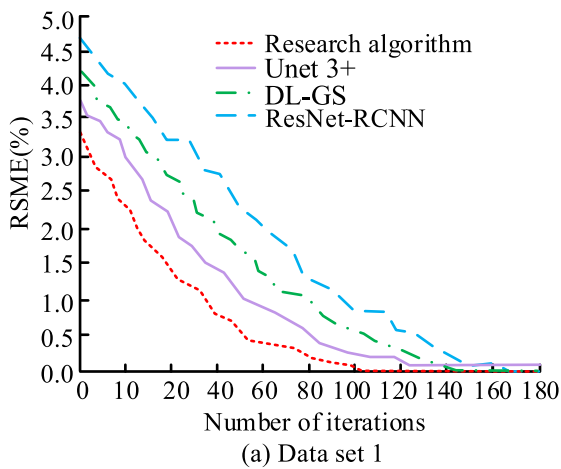**FIGURE 9.** Comparison of ROC curve changes for different models running.

(a) Data set 1

(b) Data set 2

FIGURE 10. Comparison of MSE values of different models.
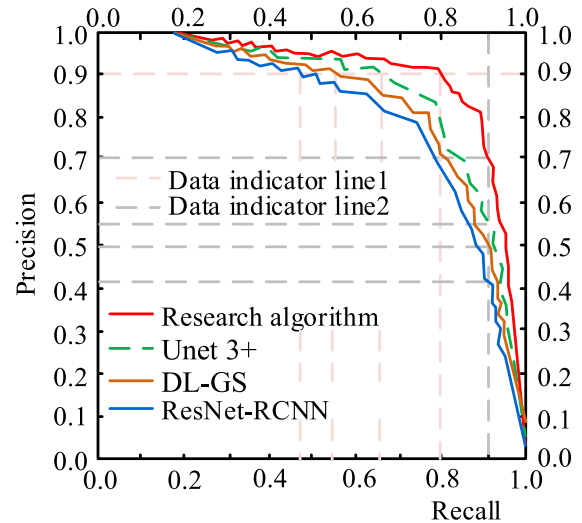
FIGURE 11. Comparison of accuracy and recall rates of different methods.

5.04%, and 7.89%, respectively, which were higher than the MSE values of the research method. Figure 10 (b) shows the MSE changes of four models for segmentation detection on dataset 2. When the MSE value of the research method started to approach 0.00, the corresponding number of iterations for the system was 138. At this time, the MSE values of Unet 3+, ResNet-RCNN, and DL-GS models were significantly greater than 0.00, with values of 1.97%, 5.12%, and 9.88%, respectively. Based on the analysis of significance results, from the comparison, there was a significant difference in MSE values among the research method and the other three methods. This indicated that the detection and measurement errors of the research method in ophthalmic OCT image segmentation were relatively small, resulting in higher clarity and accuracy of the segmented images. Next, the PR curves obtained from dataset 1 of the four models were compared, and the results are shown in Figure 11.

In Figure 11, when the accuracy was 0.900, the recall rates of the corresponding research methods, Unet 3+, ResNet-RCNN, and DL-GS models were 0.801, 0.663, 0.574, and 0.438, respectively. When the recall rate was 0.900, the accuracy of the corresponding research methods, Unet 3+, ResNet-RCNN, and DL-GS models were 0.699, 0.554, 0.498, and 0.421, respectively. By comparison, under the same experimental conditions, the accuracy and recall of the research method had always been high. This also indicated that the method constructed by the research institute had high adaptability in ophthalmic OCT image segmentation and could provide more assistance to ophthalmologists. Next, a comparison was made of the time required for the system to achieve stable operation when running four methods on two datasets, as shown in Figure 12.

From Figure 12 (a), when the system gradually ran162 iterations, the running time of the research method, Unet 3+, ResNet-RCNN, and DL-GS models was 0.501s, 0.687s, 0.774s, and 0.683s, respectively. At this point, the research

Figure 10 (a) shows different models'MSE changes on dataset 1. As the system iteration changed, the MSE values of all models showed a downward trend. When the iteration ran102 times, the research method could obtain the minimum MSE value and began to infinitely approach 0.0. At this time, the MSE values of Unet 3+, ResNet-RCNN, and DL-GS models did not reach stability, with values of 3.25%,
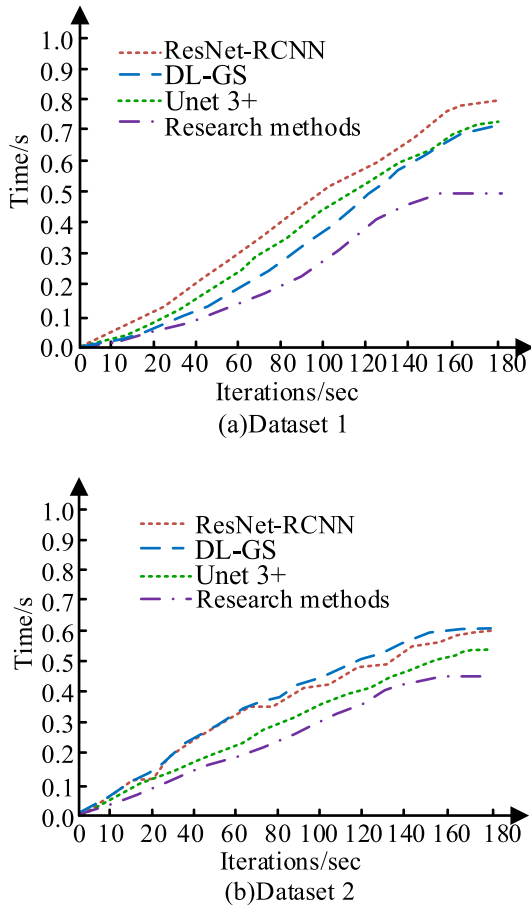
(a)Dataset 1

(b)Dataset 2

FIGURE 12. Time taken for different models to reach a stable state.

TABLE 2. IoU evaluation results of five models in pascal and grabcut datasets.

Model Type	Dataset 1 (IoU/%)	Dataset 2 (IoU/%)
DL-GS	55.53	54.38
ResNet-RCNN	57.48	74.24
Unet 3+	79.82	80.31
Research methods proposed	88.79	90.36

method has begun to approach a stationary state. In Figure 12 (b), as the system gradually iterated to 160 times, the system constructed by the research institute began to approach stability, while all other models were in a state of change. The running time of the research methods, Unet 3+, ResNet-RCNN, and DL-GS models was 0.451, 0.503, 0.562, and 0.600 seconds, respectively. By comparison, the time required to achieve stable operation of the research model was significantly lower than other algorithms, which can to some extent accelerate the efficiency of successful model image segmentation. The Intersection over Union (IoU) evaluation results of the four models in Datasets 1 and 2 are illustrated in Table 2.

IoU is anindicator for assessing the accuracy of detecting corresponding objects in a specific dataset. Table 2 showed that in dataset 1, the IoU values of the research methods, Unet 3+, ResNet-RCNN, and DL-GS models were 88.79%,
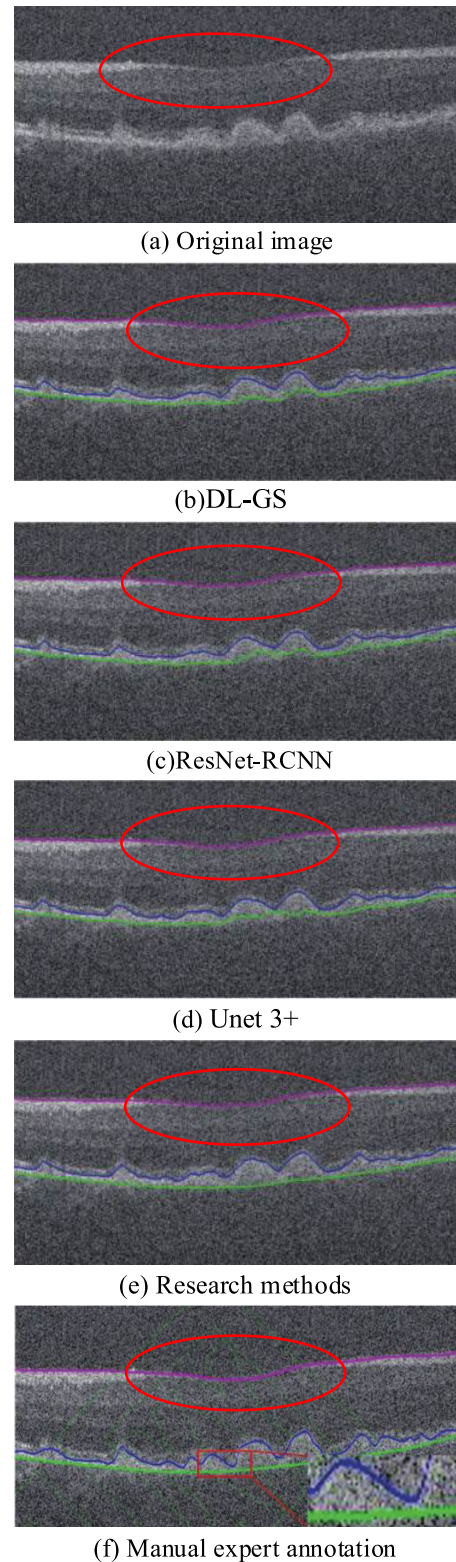

(a) Original image

(b)DL-GS

(c)ResNet-RCNN

(d) Unet 3+

(e) Research methods

(f) Manual expert annotation

FIGURE 13. Comparison of segmentation performance with existing methods on dataset 2.

79.82%, 57.48%, and 55.53%, respectively. In dataset 2, the IoU values of the research methods, Unet 3+, ResNet-RCNN, and DL-GS models were 90.36%, 80.31%, 74.24%,

and 54.38%, respectively. Overall, the performance of the research method was superior, and the accuracy of edge image segmentation was higher.

Finally, the actual segmentation effects of different algorithms on Data Set 2 were analyzed. For the convenience of observation, the experiment compared the two-dimensional images of the segmentation results. The specific results are shown in Figure 13.

Figure 13 shows the comparison of visual segmentation effects between the research method and other methods. The segmentation results of research methods were closer to the results of manual expert annotation. The research method yielded smoother results when segmenting ophthalmic OCT images, without obvious "fault" phenomena, and more accurate results when processing image boundary deformation.

V. CONCLUSION

After in-depth research and experiments, a deep learning model integrating RCNN and AM was successfully proposed and validated for the segmentation of ophthalmic OCT images. During the process, the GS algorithm was first used to achieve fine segmentation of the image, and then a recurrent residual structure was introduced to strengthen the internal information flow of the network. Finally, the AM was fused to focus on key areas in the image, successfully achieving accurate segmentation of the image. The data showed that when the system had 109 iterations, the fitness value of the research method was 98.12. The fitness values of the other three models have been constantly changing. Among the changes in ROC curves, the areas under the ROC curves of the research method were 0.956, with the highest values. In the PR curve variation, when the recall rates of all models were 0.900, the accuracy rates of the research methods, Unet 3+, ResNet-RCNN, and DL-GS models were 0.699, 0.554, 0.498, and 0.421, respectively. When the iteration ran 102 times, the research method could obtain the minimum MSE value and began to infinitely approach 0.0. At this time, the MSE values of Unet 3+, ResNet-RCNN, and DL-GS models did not reach stability, with values of 3.25%, 5.04%, and 7.89%, respectively, which were higher than the MSE values of the research method. It visualized the segmentation of images on dataset 2, and the research method's segmentation outcomes were closer to manual expert annotations. The above results indicated that the research method had high feasibility and effectiveness in segmenting ophthalmic OCT images, effectively reducing the error rate of image segmentation, and had good universality. This provided an efficient and accurate auxiliary diagnostic tool for ophthalmologists, and also opened up new directions for the application of deep learning in medical image processing. However, the experiment failed to verify the segmentation effectiveness of the algorithm in other types of medical images. At the same time, the fine segmentation processing used by the two methods proposed in this article was aGS algorithm. Although it played a certain role in segmenting retinal layer boundaries in OCT images, But it

also brought about an increase in computational complexity. In the future, it will continue to consider other subdivision processing or post-processing methods to explore how to achieve a relative balance between segmentation accuracy and computational complexity. Therefore, follow-up work will continue to focus on these issues and continue to improve and optimize.

EXPLANATION

This article proposes a novel method that integrates recurrentResNet and AM for automatic segmentation of ophthalmic OCT images. Although the automatic segmentation of OCT images has received widespreadattention in recent years, many traditional methods still suffer from insufficient accuracy and low computational efficiency due to their complex texture characteristics and various noise interferences. The main originality of this study is reflected in the following points: firstly, the experiment introduces a recurrentResNet into OCT image segmentation for the first time, enabling the network to capture and utilize image details at multiple scales, thereby more accurately identifying the structure of each layer. Secondly, traditional ophthalmic OCT segmentation often overlooks key regions in the image. By incorporating the AM, research methods can automatically focus on the crucial parts of the image that are crucial for segmentation tasks, improving segmentation accuracy. Thirdly, experiments on different publicly available OCT datasets have shown that the fusion method of recurrent ResNet-AM outperforms existing segmentation techniques while ensuring computational efficiency, exhibiting higher accuracy and robustness.

REFERENCES

[1] E. O. Rodrigues, A. Conci, and P. Liatsis, "ELEMENT: Multi-modal retinal vessel segmentation based on a coupled region growing and machine learning approach," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3507–3519, Dec. 2020, doi: 10.1109/JBHI.2020.2999257.

[2] K. Li, X. Qi, Y. Luo, Z. Yao, X. Zhou, and M. Sun, "Accurate retinal vessel segmentation in color fundus images using fully attention-based networks," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 2071–2081, Jun. 2021, doi: 10.1109/JBHI.2020.3028180.

[3] H. Wu, J. Pan, Z. Li, Z. Wen, and J. Qin, "Automated skin lesion segmentation via an adaptive dual attention module," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 357–370, Jan. 2021, doi: 10.1109/TMI.2020.3027341.

[4] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022, doi: 10.1109/TPAMI.2021.3059968.

[5] X. Liu, H. Tian, Y. Wang, F. Jiang, and C. Zhang, "Research on image segmentation algorithm and performance of power insulator based on adaptive region growing," *J. Electr. Eng. Technol.*, vol. 17, no. 6, pp. 3601–3612, Jun. 2022, doi: 10.1007/s42835-022-01118-y.

[6] D. A. Pustokhin, I. V. Pustokhina, P. N. Dinh, S. V. Phan, G. N. Nguyen, and G. P. Joshi, "An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19," *J. Appl. Statist.*, vol. 50, no. 3, pp. 477–494, Feb. 2023, doi: 10.1080/02664763.2020.1849057.

[7] X. Zhang, H. Bian, Y. Cai, K. Zhang, and H. Li, "An improved tongue image segmentation algorithm based on Deeplabv3+ framework," *IET Image Process.*, vol. 16, no. 5, pp. 1473–1485, Apr. 2022, doi: 10.1049/ipr2.12425.

[8] Y. Xu, M. Cai, L. Lin, Y. Zhang, H. Hu, Z. Peng, Q. Zhang, Q. Chen, X. Mao, Y. Iwamoto, X. Han, Y. Chen, and R. Tong, "PA-ResSeg: A phase attention residual network for liver tumor segmentation from multiphase CT images," *Med. Phys.*, vol. 48, no. 7, pp. 3752–3766, May 2021, doi: 10.1002/mp.14922.

[9] A. Srivastava, D. Jha, S. Chanda, U. Pal, H. D. Johansen, D. Johansen, M. A. Riegler, S. Ali, and P. Halvorsen, "MSRF-Net: A multi-scale residual fusion network for biomedical image segmentation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 2252–2263, May 2022, doi: 10.1109/JBHI.2021.3138024.

[10] H. Gao, Y. Yang, C. Li, L. Gao, and B. Zhang, "Multiscale residual network with mixed depthwise convolution for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3396–3408, Apr. 2021, doi: 10.1109/TGRS.2020.3008286.

[11] L. Fu, D. Zhang, and Q. Ye, "Recurrent thrifty attention network for remote sensing scene recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8257–8268, Oct. 2021, doi: 10.1109/TGRS.2020.3042507.

[12] H. Yu, X. Cheng, C. Chen, A. A. Heidari, J. Liu, Z. Cai, and H. Chen, "Apple leaf disease recognition method with improved residual network," *Multimedia Tools Appl.*, vol. 81, no. 6, pp. 7759–7782, Jan. 2022, doi: 10.1007/s11042-022-11915-2.

[13] J. Li, X. Feng, and Z. Hua, "Low-light image enhancement via progressive-recursive network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 11, pp. 4227–4240, Nov. 2021, doi: 10.1109/TCSVT.2021.3049940.

[14] S. Huang, J. Li, Y. Xiao, N. Shen, and T. Xu, "RTNet: Relation transformer network for diabetic retinopathy multi-lesion segmentation," *IEEE Trans. Med. Imag.*, vol. 41, no. 6, pp. 1596–1607, Jun. 2022, doi: 10.1109/TMI.2022.3143833.

[15] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "CPFNet: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020, doi: 10.1109/TMI.2020.2983721.

[16] L. Zhu, Z. Deng, X. Hu, H. Xie, X. Xu, J. Qin, and P.-A. Heng, "Learning gated non-local residual for single-image rain streak removal," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 6, pp. 2147–2159, Jun. 2021, doi: 10.1109/TCSVT.2020.3022707.

[17] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive spectral–spatial kernel ResNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7831–7843, Sep. 2021, doi: 10.1109/TGRS.2020.3043267.

[18] A. Khandouzi, A. Ariafar, Z. Mashayekhpour, M. Pazira, and Y. Baleghi, "Retinal vessel segmentation, a review of classic and deep methods," *Ann. Biomed. Eng.*, vol. 50, no. 10, pp. 1292–1314, Aug. 2022, doi: 10.1007/s10439-022-03058-0.

[19] Y. Xu and Y. Fan, "Dual-channel asymmetric convolutional neural network for an efficient retinal blood vessel segmentation in eye fundus images," *Biocybern. Biomed. Eng.*, vol. 42, no. 2, pp. 695–706, Apr. 2022, doi: 10.1016/j.bbe.2022.05.003.

[20] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Opt. Exp.*, vol. 18, no. 18, p. 19413, 2010, doi: 10.1364/oe.18.019413.

[21] W. Liu, G. Wu, F. Ren, and X. Kang, "DFF-ResNet: An insect pest recognition model based on residual networks," *Big Data Mining Anal.*, vol. 3, no. 4, pp. 300–310, Dec. 2020, doi: 10.26599/BDMA.2020.9020021.

[22] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNet: A fast and lightweight network for single-image super resolution," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1443–1453, Mar. 2021, doi: 10.1109/TCYB.2020.2970104.

[23] L. Yang, J. Fan, B. Huo, E. Li, and Y. Liu, "Image denoising of seam images with deep learning for laser vision seam tracking," *IEEE Sensors J.*, vol. 22, no. 6, pp. 6098–6107, Mar. 2022, doi: 10.1109/JSEN.2022.3147489.

[24] G. Song, K. Song, and Y. Yan, "EDRNet: Encoder–decoder residual network for salient object detection of strip steel surface defects," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9709–9719, Dec. 2020, doi: 10.1109/TIM.2020.3002277.

[25] W. Cai, Y. Song, H. Duan, Z. Xia, and Z. Wei, "Multi-feature fusion-guided multiscale bidirectional attention networks for logistics pallet segmentation," *Comput. Model. Eng. Sci.*, vol. 131, no. 3, pp. 1539–1555, Dec. 2022, doi: 10.32604/cmes.2022.019785.

[26] T. Anand, S. Sinha, M. Mandal, V. Chamola, and F. R. Yu, "AgriSegNet: Deep aerial semantic segmentation framework for IoT-assisted precision agriculture," *IEEE Sensors J.*, vol. 21, no. 16, pp. 17581–17590, Aug. 2021, doi: 10.1109/JSEN.2021.3071290.

[27] N. Shakeel, P. Teradata, and S. Shakeel, "Context-free word importance scores for attacking neural networks," *J. Comput. Cognit. Eng.*, vol. 1, no. 4, pp. 187–192, Sep. 2022, doi: 10.47852/bonviewjcce2202406.

[28] M. A. A. Al-qaness, A. Dahou, M. A. Elaziz, and A. M. Helmi, "Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 144–152, Jan. 2023, doi: 10.1109/TII.2022.3165875.

[29] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Exp.*, vol. 8, no. 5, p. 2732, 2017, doi: 10.1364/boe.8.002732.

[30] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 1055–1059, doi: 10.1109/ICASSP40776.2020.9053405.

BINGBING LI received the B.S. degree in communication engineering from the Department of Communication Engineering, Jilin University, and the M.S. degree in measurement technology and instruments from Jilin University, in 2015. She is currently pursuing the Ph.D. degree in control science and engineering with Harbin Engineering University. Her research interests include computer vision, medical image processing, deep learning, and self-supervised learning.

YAO CONG received the M.Eng. degree in instrument science and technology from Northeast Petroleum University, China, in 2020. She is currently pursuing the Ph.D. degree with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, China. Her research interests include scene understanding, visual question answering, and embodied intelligence.

HONGWEI MO received the Ph.D. degree in engineering, in March 2005. He was a Visiting Scholar with the University of California, Davis. He is currently with the School of Automation, Harbin Engineering University. He has undertaken and completed 17 projects, such as NSFC and national defense pre-research. He has published more than 70 papers. He was the Winner of the Heilongjiang Outstanding Youth Fund, in 2012.

• • •