



RCAR-UNet: Retinal vessel segmentation network algorithm via novel rough attention mechanism



Weiping Ding^{a,b,*}, Ying Sun^a, Jiashuang Huang^a, Hengrong Ju^a, Chongsheng Zhang^c, Guang Yang^{d,e}, Chin-Teng Lin^f

^a School of Information Science and Technology, Nantong University, Nantong 226019, China

^b Faculty of Data Science, City University of Macau, Macau 999078, China

^c School of Computer and Information Engineering, Henan University, Kaifeng 475001, China

^d Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK

^e National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK

^f Centre for Artificial Intelligence, FEIT, University of Technology Sydney, Ultimo, NSW 2007, Australia

ARTICLE INFO

Keywords:

Attention mechanism
Fundus retinal blood vessel image
Image segmentation
Rough set
Rough neuron
Residual connection

ABSTRACT

The health status of the retinal blood vessels is a significant reference for rapid and non-invasive diagnosis of various ophthalmological, diabetic, and cardio-cerebrovascular diseases. However, retinal vessels are characterized by ambiguous boundaries, with multiple thicknesses and obscured lesion areas. These phenomena cause deep neural networks to face the characteristic channel uncertainty when segmenting retinal blood vessels. The uncertainty in feature channels will affect the channel attention coefficient, making the deep neural network incapable of paying attention to the detailed features of retinal vessels. This study proposes a retinal vessel segmentation via a rough channel attention mechanism. First, the method integrates deep neural networks to learn complex features and rough sets to handle uncertainty for designing rough neurons. Second, a rough channel attention mechanism module is constructed based on rough neurons, and embedded in U-Net skip connection for the integration of high-level and low-level features. Then, the residual connections are added to transmit low-level features to high-level to enrich network feature extraction and help back-propagate the gradient when training the model. Finally, multiple comparison experiments were carried out on three public fundus retinal image datasets to verify the validity of Rough Channel Attention Residual U-Net (RCAR-UNet) model. The results show that the RCAR-UNet model offers high superiority in accuracy, sensitivity, F1, and Jaccard similarity, especially for the precise segmentation of fragile blood vessels, guaranteeing blood vessels' continuity.

1. Introduction

The retinal vessels can reflect the blood circulation dynamics and health status. Various ophthalmic and cardiovascular diseases can cause different degrees of deformation and hemorrhage in the retinal vessels [1,2]. Clinically, medical personnel can extract retinal vessels from the color fundus images to diagnose these diseases and analyze the morphological conditions of the blood vessels. However, the retinal blood vessels are distributed densely and irregularly. Many capillaries are indistinguishable from the background,

* Corresponding author at: School of Information Science and Technology, Nantong University, Nantong 226019, China.
E-mail address: dwp998@163.com (W. Ding).

and the contrast of them is poor. In addition, the blood vessel boundaries are blurred. Retinal vessel segmentation is more difficult because retinal vessels are susceptible to the influence of collection equipment, light and diseased tissue [3,4]. Relying on artificial segmentation of retinal vessels needs to a considerable workload, time, and cost. It is also severely affected by subjective factors [5]. Therefore, computer technology can identify an algorithm to quickly and accurately segment retinal vessels and extract blood vessel morphological structures from fundus images in real time. This computer technology can also perform an essential role in assisting medical personnel in diagnosing ophthalmic diseases and cardiovascular diseases.

The automatic and precise segmentation of retinal vessels using artificial intelligence to realize the auxiliary diagnosis of ophthalmic, cardiovascular, and cerebrovascular diseases has become a research hotspot. Compared with other methods for extracting morphological features of blood vessels, deep learning models can provide accurate results by automatically extracting features. The U-Net model has been extensively utilized in retinal vessel segmentation algorithms based on its unique network structure [6–13]. Alomd et al. [11] introduced the residual and recurrent networks into the U-Net model. They designed recurrent U-Net and recurrent residual U-Net in which the residual unit helps train deep architecture and accumulate the extracted features using recursive residual convolution layers. This approach facilitates a better representation of the segmentation task. Yue et al. [12] applied a multi-scale input and dense blocks into traditional U-Net, allowing the network to utilize abundant spatial context information and improve segmentation results. Their methods improve the results of the problematic fine blood vessels that are hard to distinguish due to low contrast between blood vessels and background. Zhang et al. [13] incorporated RNN into CNN to transmit the context and generate the probability map of retinal blood vessels. Despite excellent results of retinal vessel segmentation, it still has some shortcomings. U-Net uses skip connections to combine the down-sampled spatial information with up-sampled spatial details and compensate for inaccurate reconstructed spatial information during up-sampling. However, this approach brings a great deal of redundant low-level feature extraction, resulting in many invalid calculations and wasting many computing resources.

The attention mechanism [14–16] permits the model concentrate on meaningful information. Two forms of attention mechanisms are hard attention and soft attention. Hard attention works by cropping the image to highlight the relevant regions. The hard attention mechanism is non-differentiable and cannot be used for standard backpropagation. Moreover, it requires reinforcement learning for training. Soft attention is implemented by weighting different parts of the image and determining the correlation by assigning the weights to different image regions. Soft attention is differentiable and can be trained through backpropagation. Ozan designed the Attention U-Net model and applied soft attention to skip connections [17] to avoid the excessiveness of low features in U-Net caused by skip connections. This model can effectively suppress irrelevant regions' activation and decrease excessive features without increasing huge parameters and calculations, extensively implemented in medical image segmentation [18,19].

Cui et al. [20] combined the attention mechanism and U-Net model to carry out cardiac segmentation in short-axis magnetic resonance imaging. Guo et al. [21] proposed a lightweight network of spatial Attention U-Net (SA-UNet) to segment retinal vessels precisely. Li et al. [22] proposed Attention UNet++ to introduce an attention mechanism between nested convolution blocks. Punn et al. [23] imported an effective residual cross-spatial attention-guided inception U-Net model which is used for tumor segmentation. Yang et al [24] came up with a residual Attention U-Net (RAU-Net), and constructed a residual module to combine with attention mechanism which is capable of enriching the character representation.

However, the models mentioned above have some shortcomings because they only integrate attention gate into the U-Net network. The Attention U-Net model only considers the vital distinction of the image's spatial location and ignores the critical distinction between different channels in the feature map [25]. Moreover, the characteristics of retinal blood vessels, such as subtlety, irregular distribution and blurred edges, lead to channel uncertainty when using convolution operation for feature extraction of retinal blood vessel images. The feature channel uncertainty will affect the channel attention coefficient, failing the depth network to pay attention to the detailed features of retinal blood vessels. Moreover, the segmentation effect of thin blood vessels is not ideal.

Granular computing theory is a novel tool to deal with uncertain information, including rough set theory, fuzzy theory, etc. In the existing research, many researchers use the combination of rough set theory and fuzzy theory and deep neural network model to deal with uncertain information. Ding et al. [26] combined fuzzy theory with deep learning techniques, and redesign the Project & Excite

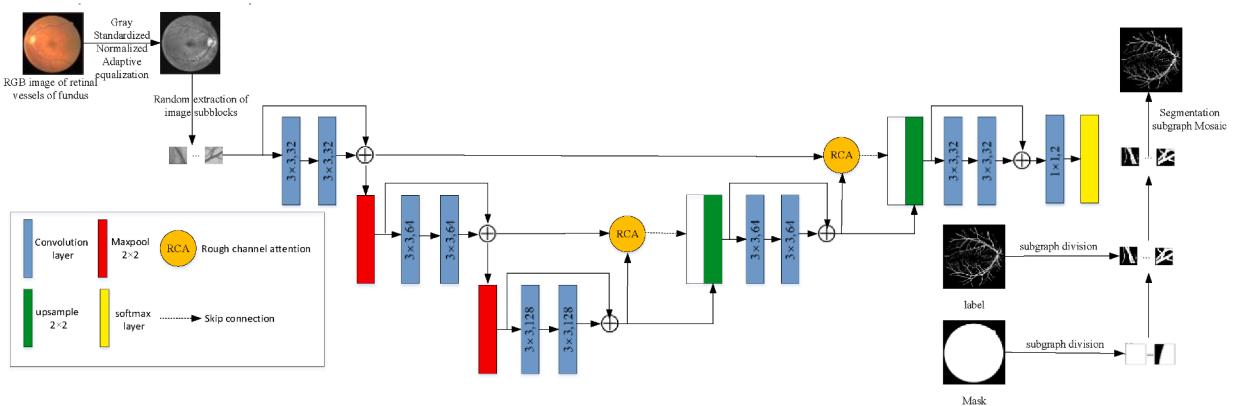


Fig. 1. Framework of the U-Net via the rough channel attention (RCAR-UNet).

module using the VPF layer to enable modeling uncertainty during feature recalibration to improve the efficiency and accuracy of multimodal infant brain segmentation problems. Fang et al.[27] presented Fuzzy MA2C which integrates 2 s-FDNN into multiagent deep RL to deal with uncertain communication information for improving the robustness and generalization under partially observed environments. Zheng et al.[28] designed a novel generation of the fuzzy random forest to get the essential features of leaves. And a large number of fuzzy random forests are integrated and cascaded for deep fuzzy representation learning to get higher classification accuracy. These researches fully demonstrate the effectiveness of combining rough sets or fuzzy theory with deep learning.

We propose a network which combines rough neurons and a channel attention mechanism for the uncertainty of retinal vessel feature channels. Fig. 1 illustrates its structure taking the U-Net as the backbone. First, the method integrates the deep neural network's ability to learn complex features and rough sets' ability to deal with uncertainty and design rough neurons. Second, we build a rough channel attention mechanism module based on rough neurons and integrate it in the skip connection of the U-Net model for the fusion of high and low features. Third, residual connections are added to pass the low-level features to high layers directly, enriching network feature extraction and helping back-propagate the gradient when training the model. Finally, the Rough Channel Attention Residual U-Net's (RCAR-UNet) feasibility and validity are verified on three retinal vascular datasets. The comparative experimental results indicate that the RCAR-UNet model significantly improves accuracy and sensitivity indices.

The main contributions of RCAR-UNet are following:

- The capacity of deep neural networks to learn complex features and rough sets to deal with uncertainty is integrated into the design of rough neurons. The output of the upper and lower neurons is multiplied by a certain weight to distinguish the contribution of different neurons to the final output, so as to obtain a better semantic output value.
- The rough channel attention mechanism module is constructed based on rough neurons, and the upper and lower approximate neurons are constructed using global max and average pooling respectively, and the weighted summation is performed. The dependency relationship contains both global information and local features, thus realizing feature re-calibration effectively.
- The residual connection is introduced to transfer the low-level features directly to the high-level, capable of improving network degradation to a certain extent.
- The application of RCAR-UNet in the segmentation of retinal blood vessels offers high superiority in accuracy, sensitivity, F1, and Jaccard similarity, especially for the segmentation of fragile blood vessels, guaranteeing the continuity of blood vessels.

The remaining chapters are arranged as follows: Section 2 introduces a rough neural network overview. Section 3 presents the rough neuron structure designed using the upper and lower approximate set. This section also presents the U-Net model structural design based on the rough channel attention mechanism that combines the rough channel attention mechanism with the residual connection. Section 4 compares the outcomes of the experiments with multiple models of three fundus retinal image datasets to illustrate the feasibility and validity of RCAR-UNet. Section 5 gives a discussion, and Section 6 sums up the paper.

2. Background theories

2.1. Rough set theory

The rough set [29] is a computational paradigm proposed by Pawlak in 1982 to effectively handle inaccurate, inconsonant, and fragmental information. Rough set partitions data by indiscernibility relation. It describes the target using the upper and lower approximate sets, forming three non-overlapping regions of positive, negative, and boundary domains.

The upper and lower approximation sets can describe the target concept $X \subseteq U$ in the rough set. The equivalence class of x under the conditional attribute R is expressed as $[x]_R$. The rough set divides the upper and lower approximation sets of objects in any subset $X \subseteq U$ of the universe U concerning the attribute subset $R \subseteq A$. Then, the upper approximation set $\bar{R}(X)$ and the lower approximation set $R(X)$, based on the attribute set, are given by [30–33]. It can be seen from the definition that the objects in the upper approximate set may belong to the target concept according to the existing knowledge. The objects in the lower approximate set must belong to the object concept according to the judgment of the existing knowledge.

2.2. Attention u-net

The Attention U-Net model takes U-Net encoding-decoding U-shaped network structure as its backbone. U-Net network structure comprises a shrinking path, an extended path, and a skip connection. The principle of it is to capture global features on the shrinking path, achieve a precise position on the extended path, and connect the upper and lower network information through the skip connection directly, to solve the segmentation of complex neuron structure. However, the U-Net network has the following shortcomings:

The skip connection directly splices the feature maps at the symmetrical positions in the U-shaped structure. Although skip connection facilitates information fusion between high-level and low-level features, it leads to many redundant low-level features, resulting in massive invalid calculations and wasting additional computing resources.

However, many feature maps are obtained by several convolutions and pooling operations. For example, we are interested in the blood vessel regions relevant to the image segmentation task of fundus retinal blood vessels. Therefore, different regions of the same image have different degrees of attention.

In this regard, soft attention is introduced into U-Net to propose the Attention U-Net model realized by weighting different image regions. Moreover, the correlation of different regions in images determines the weights. This model effectively suppresses the activation in unrelated areas, removes excessive features, and does not introduce extra parameters and calculations. Fig. 2 shows the construction of the spatial attention gate [17].

The feature graph $g(H_g \times W_g \times C)$ is obtained through the contraction path, and the feature graph $x(H_x \times W_x \times C)$ is obtained through the expansion path with the same number of channels. After a series of linear transformations, we obtain an attention coefficient $\alpha(H_x \times W_x)$. By multiplying the coefficient of attention with the feature graph x , the attention degree of different regions of the feature graph x is recalibrated to obtain a new feature graph x' .

3. Case study and exploratory analysis

3.1. Rough neurons

The image world is colorful, containing uncertain information such as randomness and ambiguity. For example, acquisition equipment and environment easily affect the fundus image during acquisition process. This issue causes retinal blood vessel distribution disorder, irregular shape, and blurred edges. The uncertain information in these fundus retinal blood vessel images makes the deep network ineffective. Lingras presented rough neural network based on the general concept of upper and lower bounds [34]. The integration of rough neural networks with deep architectures is promoted owing to the advantage of deep neural networks in learning complicated features and the capability of rough neural networks to deal with inconclusive information.

The concept of upper and lower limits has been used in various artificial intelligence applications. Especially, the upper and lower approximate sets prove effective for upper and lower limits in rule generation. The rough set theory approximates the target concept using upper and lower approximation sets. We introduce the upper and lower approximation sets and construct upper and lower approximation neurons to roughen attention coefficient obtained by attention module reasonably [35–40]. Fig. 3 indicates the structure of the rough neuron.

The parameters of the upper approximation neuron $\theta_U = \{W_U, b_U, \alpha\}$, and the lower approximation neuron parameters $\theta_L = \{W_L, b_L, \beta\}$, where W_U and b_U describe the weight and bias of upper approximation neuron severally. W_L and b_L express the weight and bias of the lower approximate neuron, respectively. Unlike the individual output values of the conventional neurons, the outputs of the rough neurons are a pair of upper and lower limits, calculated as follows:

$$O_U = f(W_U \cdot X + b_U), \quad (1)$$

$$O_L = f(W_L \cdot X + b_L). \quad (2)$$

Finally, given certain weights α and β of the upper and lower boundary neurons, the output of the final rough neuron O is obtained using weighted summation, expressed as follows:

$$O = \alpha \cdot O_U + \beta \cdot O_L. \quad (3)$$

The rough neuron maps the input value into a rough range, and the final output depends on the upper and lower neurons. Certain weight value multiplies different outputs to distinguish the contribution of different neurons to the final output. The purpose is to coarsen the deterministic values of the input and recalibrate the deterministic values to obtain a more reasonable and accurate output value.

3.2. Channel attention mechanism

Images have two dimensions: spatial and channel. Spatial attention mainly considers the spatial scale of the image, whereas the channel dimension considers the channel attention mechanism. We obtain a feature map with n channels if n convolution kernels are used to capture features. The benefits of the information extracted from these n channels are the same, and there is no need to

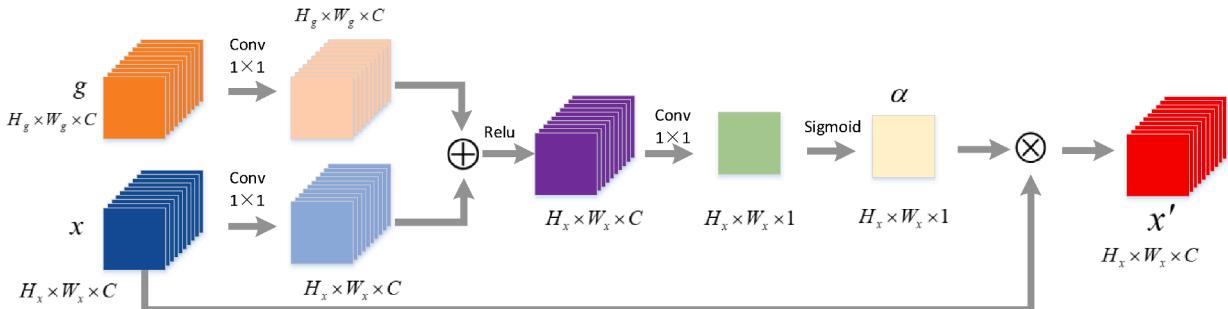


Fig. 2. Structure of the spatial attention gate.

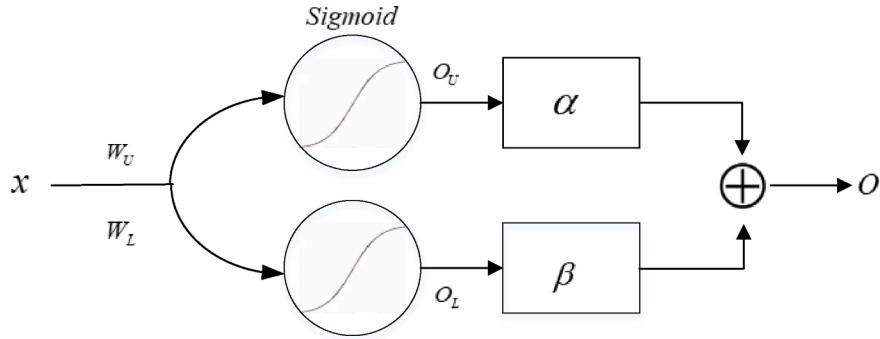


Fig. 3. Structure of rough neurons.

discriminate between these n channels. However, different regions of the same image have different attention, and a practical difference is remarked about the information in different feature channels of the feature map. If the blood vessels are the target segmentation for a fundus image, we will focus on the feature channels of blood vessels and ignore those irrelevant to the task. In this regard, we must construct an essential representation of the information in every feature channel and correct of the characteristic values in the characteristic mapping. The results are obtained by multiplying weight coefficients and characterizing the significance of every channel with the corresponding channel in the original feature map.

Fig. 4 provides channel attention's structure. First, the compression operation merged the feature map of which makes the two-dimensional data of each channel with $H_x \times W_x \times C$, which is converted into a number Z using global pooling to characterize the significance of every channel. Second, the excitation operation consists of two full connection layers. The first full connection layer reduces the model's parameters by scaling r to the feature channels and compressing channels into C/r to decrease parameters and calculation time. The second full connection layer recovers channels' original number making it more non-linear and better integrated into the sophisticated relationship within different channels. Thus, the two fully connected layers are activated by the Relu and Sigmoid functions. The Relu function reduces the probability of gradient disappearance. In the Sigmoid function, the feature weights of each channel are between 0 and 1. Finally, the rescaling process is achieved by correcting feature values in the feature map and multiplying the weight coefficients characterizing every channel's significance with the corresponding channel in original feature map.

The feature graph $g(H_g \times W_g \times C)$ is obtained through the contraction path, and the feature graph $x(H_x \times W_x \times C)$ is obtained through the expansion path with the same number of channels. Then, we obtain the channel attention coefficient ($1 \times 1 \times C$) through global max pooling and two fully connection layers. Through multiplying the coefficient of attention with the feature graph x , the attention degree of different channels of the feature graph x is recalibrated to obtain a new feature graph x' .

3.3. U-net model based on rough channel attention mechanism

Based on the rough neuron and channel attention mechanism, rough channel attention (RCA) mechanism is proposed and embedded into the U-Net to complete the segmentation task of retinal vessels. Fig. 1 shows the U-Net structure on the basis of rough channel attention mechanism (RCAR-UNet), and Table 1 shows its specific information.

In the feature encoder of the U-Net model, we construct simple feature extraction module includes two 3×3 convolutional layers connected by residuals and a 2×2 max pooling layer. The residual connection is added because it can directly propagate the shallow characteristics to deep layer and alleviate the network degradation phenomenon.

Our team propose rough channel attention module by combining the rough neurons with learning intricate characteristics and the channel attention advantage to distinguish the significance of the information obtained by the image, whose detail is given in Fig. 5.

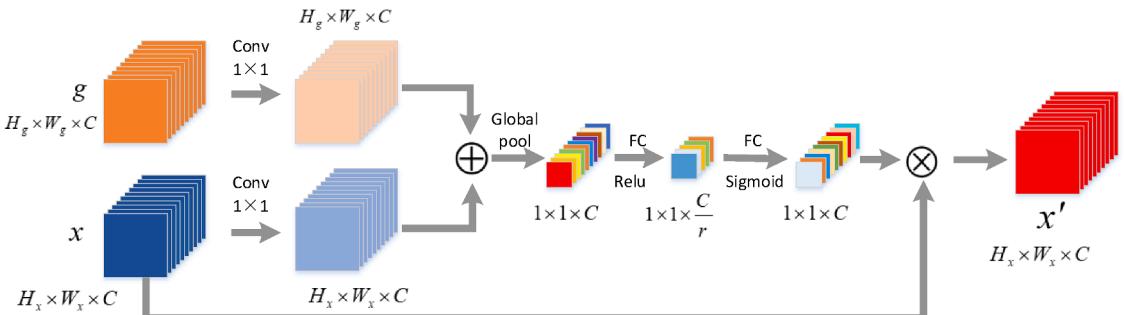


Fig. 4. Structure of channel attention.

Table 1

U-NET model based on rough channel attention architecture.

No	Block name	Layer name	Image size
1	Input $48 \times 48 \times 1$		
2	Encoder Block 1	Conv_1(kernel size = 3×3),activation = Relu Conv_2(kernel size = 3×3),activation = Relu Residual	$48 \times 48 \times 32$ $48 \times 48 \times 32$ $48 \times 48 \times 32$
3	Encoder Block 2	2 × 2 max pooling, stride = 2 Conv_3(kernel size = 3×3),activation = Relu Conv_4(kernel size = 3×3),activation = Relu Residual	$24 \times 24 \times 64$ $24 \times 24 \times 64$ $24 \times 24 \times 64$
4	Encoder Block 3	2 × 2 max pooling, stride = 2 Conv_5(kernel size = 3×3),activation = Relu Conv_6(kernel size = 3×3),activation = Relu Residual	$12 \times 12 \times 128$ $12 \times 12 \times 128$ $12 \times 12 \times 128$
7	Up-sampling, kernel size = 2×2 Decoder Block1	Concat1 Up_1($24 \times 24 \times 128$)Rough Attention_1($24 \times 24 \times 64$)	$24 \times 24 \times 192$
8		Conv_7(kernel size = 3×3),activation = Relu Conv_8(kernel size = 3×3),activation = Relu Residual	$24 \times 24 \times 64$ $24 \times 24 \times 64$ $24 \times 24 \times 64$
9	Up-sampling, kernel size = 2×2 Decoder Block1	Concat2 Up_2($48 \times 48 \times 64$)Rough Attention_1 ($48 \times 48 \times 32$)	$48 \times 48 \times 96$
10		Conv_9(kernel size = 3×3),activation = Relu Conv_10(kernel size = 3×3),activation = Relu Residual	$48 \times 48 \times 32$ $48 \times 48 \times 32$ $48 \times 48 \times 32$
11		Conv_11(kernel size = 1×1)	$48 \times 48 \times 2$
12		Softmax	$48 \times 48 \times 2$
13	Output $48 \times 48 \times 2$		

The channel attention distinguishes the significance of every channel in characteristics mappings. However, the feature map must be compressed when constructing a feature map that can characterize the channel's importance. Moreover, the global average pooling is chosen to compress the two-dimensional matrix into a real number to measure the channel importance. However, the retinal blood vessels vary in thickness, especially with the fine vessels ending. Thus, we must focus on the local image information and compress the feature maps using global max pooling, whose values contain local information. When constructing rough neurons, The global max pool is selected as the upper neuron and the global average pool is selected as the lower neuron. The upper and lower limits of the importance of feature channels are described respectively. This approach makes the Attention coefficient values have global and local information. We implement feature rescaling and use the learned attention coefficients to correct the values and achieve a new retinal vessel feature map. The specific steps are as follows:

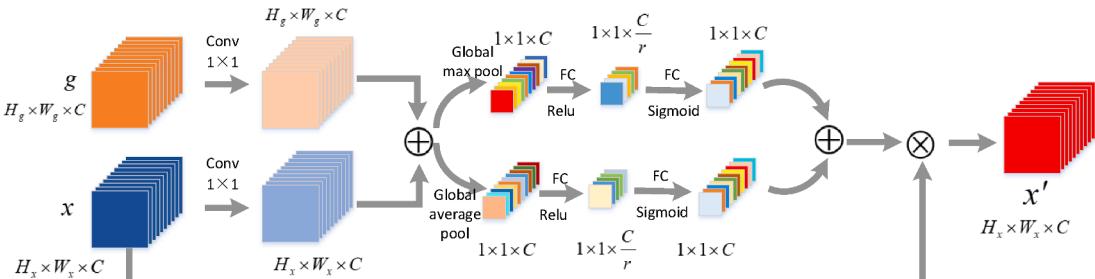
First, we define a low-level feature map $L_{feature} = [L_{feature_1}, \dots, L_{feature_C}]$ with C channels, where $L_{feature_k} \in R^{H \times W}$ is the k -th feature channel of the low-level feature map H and W indicate the width and height of the low-level feature map, respectively. We define a high-level feature map $H_{feature} = [H_{feature_1}, \dots, H_{feature_C}]$ with C channels, where $H_{feature_k} \in R^{H' \times W'}$ is the k -th feature channel of the deep layer characteristics mappings, H' and W' express the width and height of the high-level characteristics mapping.

We add high and low feature maps to obtain the fusion feature map $New_{feature} = [New_{feature_1}, \dots, New_{feature_C}]$, expressed as follows:

$$New_{feature_c} = L_{feature_c} + H_{feature_c}. \quad (4)$$

Then, we used the global max-pooling layer to establish the dependency between channels and preserve the local information, expressed as follows:

$$F_{feature_c}^H = \text{Max}(New_{feature_c}(i, j)), \quad (5)$$

**Fig. 5.** Structure of the rough channel attention mechanism.

where $0 < i \leq H, 0 < j \leq W, 0 < c \leq C$, the global max pooling is used to calculate the max value of pixels in every feature channel, and a $1 \times 1 \times C$ tensor is obtained after the global average pooling layer.

The global average pooling layer is applied to establish dependency between channels and preserve global information, expressed as follows:

$$F_{feature_c}^L = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^{W'} New_{feature_c}(i, j), \quad (6)$$

where $0 < i \leq H, 0 < j \leq W', 0 < c \leq C$. The global average pooling is utilized to compute the average value of the pixel sum in each feature channel. After the global average pooling layer, a $1 \times 1 \times C$ tensor is obtained.

Excitation operations are performed on two $1 \times 1 \times C$ tensors of global average pooling and global max pooling, containing two full connection layers. One of them implements compression of channels to reduce model parameters and compute time. The other has the function of restoring channels to increase the nonlinearity of the model so that it can better acclimate to the complicated relationships within channels; Relu and Sigmoid activation functions are used to activate these two full connection layers. The upper and lower limits of the attention coefficients are obtained, expressed as follows:

$$F_{feature_c}^H = \mu(W_2 \cdot \sigma(W_1 \cdot F_{feature_c}^L)), \quad (7)$$

$$F_{feature_c}^L = \mu(W_2 \cdot \sigma(W_1 \cdot F_{feature_c}^H)), \quad (8)$$

where σ represents the Relu function; μ represents the Sigmoid function. $W_1 \in R_r^{C \times C}$ and $W_2 \in R^{C \times r}$ represent the weight matrix of the

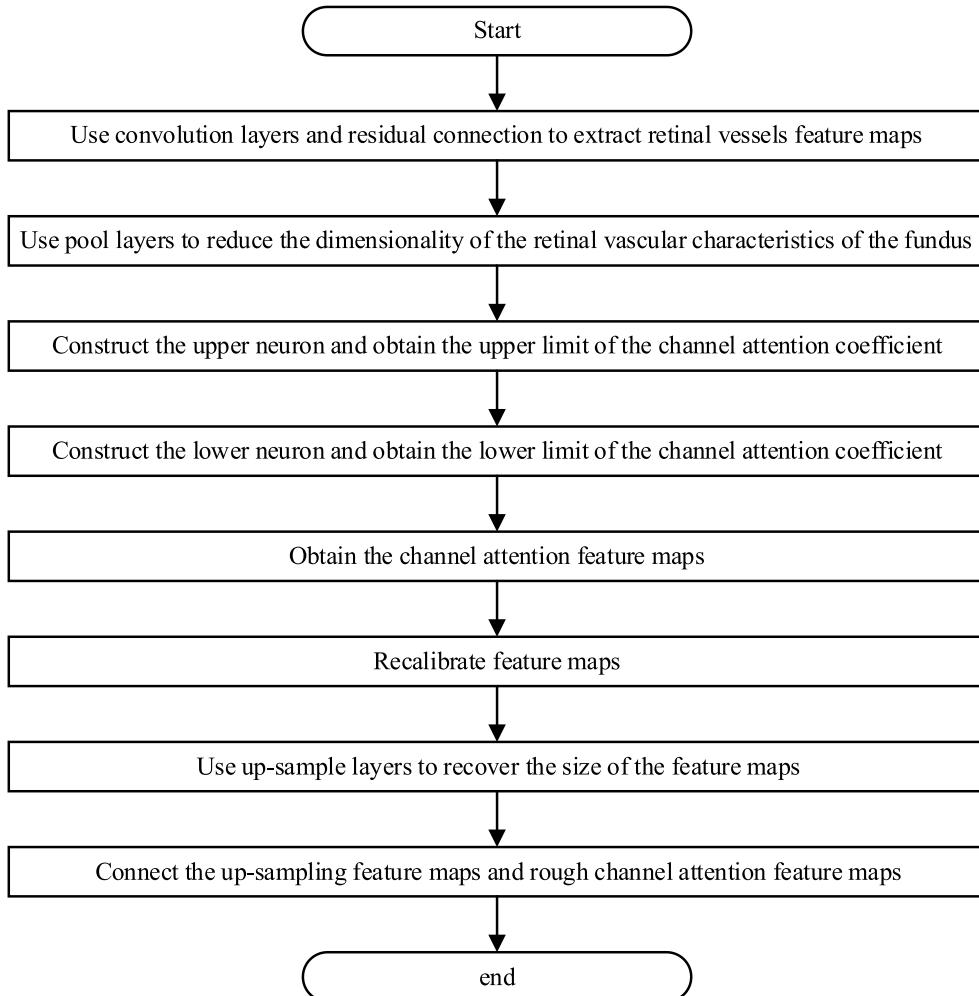


Fig. 6. Flow chart of the Algorithm: RCA.

two fully connected layers.

A weighted summation operation is performed on the upper and lower limits of the obtained channel importance values, so that the dependencies of different channels contain global and local details, and a new dependency relationship between channels is obtained:

$$F_{feature_c} = \alpha \cdot F_{feature_c}^H + \beta \cdot F_{feature_c}^L. \quad (9)$$

in which α and β indicate the weight information of the upper and lower limits.

The dependency relationship is used between the output channels to recalibrate the feature map representation as follows:

$$\vec{F}_{feature_c} = F_{feature_c} \cdot L_{feature_c}, \quad (10)$$

Then, a new characteristic mapping is stitched with the up-sampled characteristic mappings $F_{up-feature}$ of the same layer with the same dimensionality and used as the input of the next network layer.

In the feature decoding section of the U-Net model, a simple recovery feature module includes an up-sampling layer and two convolutional layers. These channels are in two categories: blood vessels and non-vascular. After restoring the feature size, two 1×1 convolution kernels are used for convolution operation to obtain a feature map of two channels: one channel represents the blood vessel category, and the second channel represents the non-vascular category. Finally, the softmax layer is developed to obtain the possibility of output belonging to every category. The algorithm flow chart is shown in Fig. 6.

Algorithm: RCA

Input: Fundus retinal blood vessel image *input*

Output: Retinal blood vessel feature maps *F*

Parameter: α, β

Step 1: Use convolution layers and residual connection to extract retinal vessels feature maps;

Step 2: Use pool layers to reduce the dimensionality of the retinal vascular characteristics of the fundus;

Step 3: Construct the upper neuron and obtain the upper limit of the channel attention coefficient by (7);

Step 4: Construct the lower neuron and obtain the lower limit of the channel attention coefficient by (8);

Step 5: Obtain the channel attention feature maps by (9);

Step 6: Recalibrate feature maps by (10);

Step 7: Use up-sample layers to recover the size of the feature maps;

Step 8: Connect the up-sampling feature maps and rough channel attention feature maps.

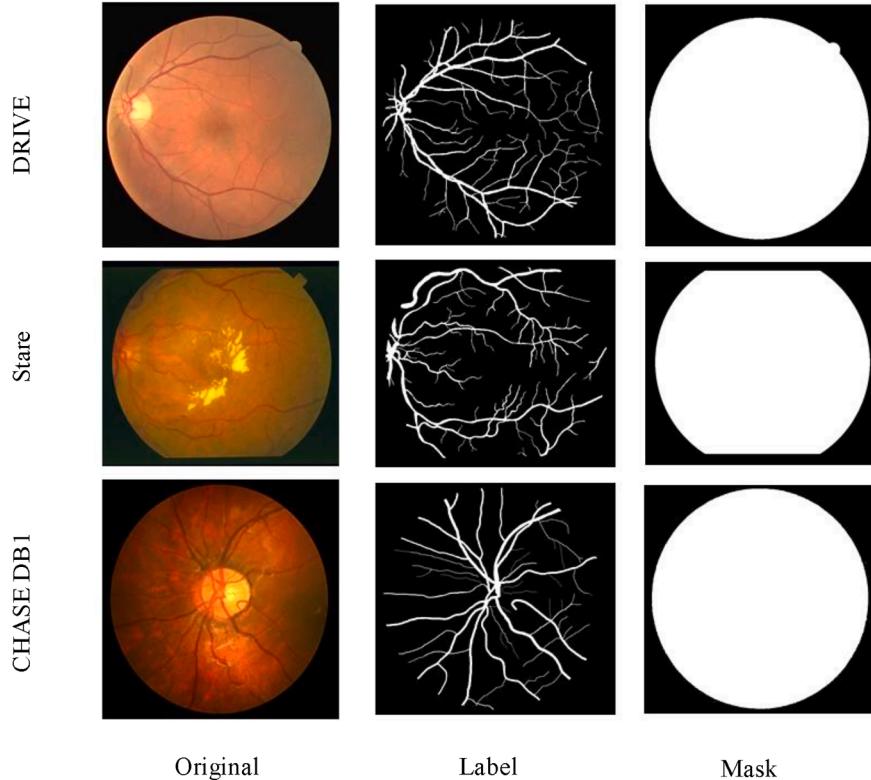


Fig. 7. Dataset information.

4. Experiment

4.1. Dataset

The DRIVE dataset [41–44] comprises 40 color fundus images and corresponding labeled images. This dataset was released in 2004 and established by Niemeijer et al. in the Dutch diabetic retinopathy screening study. The size of each color fundus image is 565×584 .

Michael Goldbaum built the Stare dataset [41–44] in 1975, comprising a color fundus map database for retinal blood vessel segmentation. This dataset consists of 20 pictures with an image size of 605×700 . For every image in it, two experts are assigned to manually segment the standard map. Thus, we must manually set the mask because no mask is available in this database.

The CHASE DB1 dataset [41–44], acquired from the eyes of 14 school children, composed of 28 retinal images in jpg format with size of 999×960 . Two experts conducted manual segmentation labels of each image, and the corresponding mask must be coded. Fig. 7 shows the information on the three-fundus retinal vessel image datasets.

4.2. Pre-processing

The acquisition environment and equipment generally affect fundus images during acquisition, resulting in uneven brightness and the contrast between the retinal vessels and the background is low. Fundus images input to the network model are preprocessed to make the model capture finer vessel features. Colorful eye fundus images are expanded by channel, G channel was moderately bright and could distinguish the retinal vessels from the background. The contrast-constrained adaptive histogram equalization is used to strengthen the distinction between the two without amplifying the noise of the fundus images. This technique makes the structure and features of the retinal vessels more easily noticed. Gamma transform is used for image enhancement to correct overly bright or dark image areas. Fig. 8 represent the original and pre-processed images.

4.3. Dataset expansion

The network model will not generalize the features learned with few images in the training set. The model does excellently in training dataset but poorly in test dataset. For example, approximately 10 image samples in the three datasets are used. Thus, the training images, the corresponding expert manual labels, and mask maps are divided into image sub-blocks of size 48×48 , as shown in Fig. 9. A total of 190,000 image blocks are randomly extracted in the DRIVE and Stare dataset, and 140,000 image patches are randomly acquired from CHASE DB1 dataset.

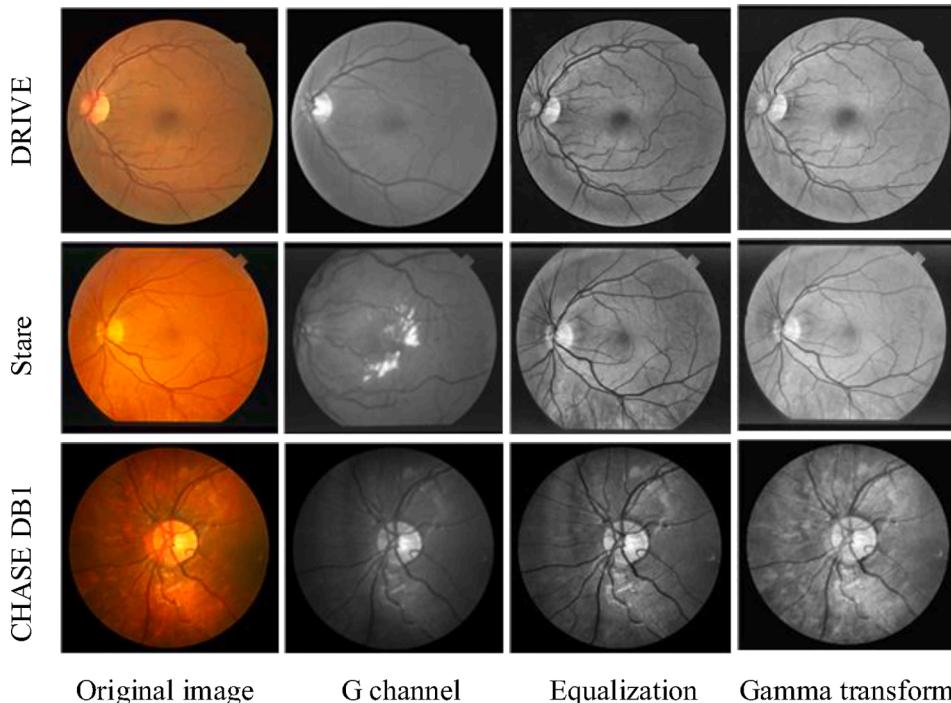


Fig. 8. Original image and preprocessed image.

4.4. Experimental parameter settings

This study uses a workstation based on Windows 10, running on Intel i7-10750H CPU, with 16 GB of memory and 2060 6.0 GB GPU. TensorFlow and Keras frameworks are adopted to construct the network. In the training phase, the cross-entropy loss function is adopted as the training loss function. The batch size is 32. The model iterations' number is 20. The original learning rate of the model is 0.01, and SGD is chosen as the optimizer.

4.5. Evaluation criteria

Fundus images contain two classes: blood vessels and background. The blood vessel class is the target that must be detected and segmented, which is also called the positive class. The non-vascular class is also known as the negative class. The results of the segmentation algorithm are compared with the real value to obtain N_{TP} , N_{FP} , N_{FN} and N_{TN} , where N_{TP} indicates the number of pixels of the correctly classified blood vessel class; N_{FP} indicates the number of misclassified background class; N_{FN} indicates the number of pixels of the misclassified blood vessel class; N_{TN} indicates the number of pixels of correctly classified background class. The accuracy Acc , sensitivity Sen , specificity Spe , and precision Pre evaluation indices are selected to estimate the retinal vessel segmentation algorithm's performances, where Acc denotes the probability of correctly classifying blood vessels and background classes. Accuracy is a statistic for all classification classes, without differentiation between categories. However, the cost of error may vary between classes. Sen represents the probability of correctly classifying the blood vessels class. Spe represents the probability of correctly classifying the background class. Pre represents the proportion of samples labeled as blood vessels class among those predicted as blood vessels class. We expect that all pixels predicted to be retinal vessels are correctly classified, or all pixels labeled as retinal vessels are predicted to be retinal vessels, i.e., high accuracy and high sensitivity. However, the pursuit of precision or sensitivity is one-sided and meaningless, and the model is evaluated using the summed average of precision and sensitivity. In addition, RCAR-UNet is a supervised learning capable of obtaining a segmentation map consistent with the labeled map. Jaccard used similar supervised learning to calculate the similarity between the gold standard map *truth* and the segmentation map *result*. The Jaccard value measures similarity between sets; the formula is as follows:

$$Acc = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \quad (11)$$

$$Sen = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (12)$$

$$Spe = \frac{N_{TN}}{N_{TN} + N_{FP}} \quad (13)$$

$$Pre = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (14)$$

$$J(truth, result) = \frac{|truth \cap result|}{|truth \cup result|} \quad (15)$$

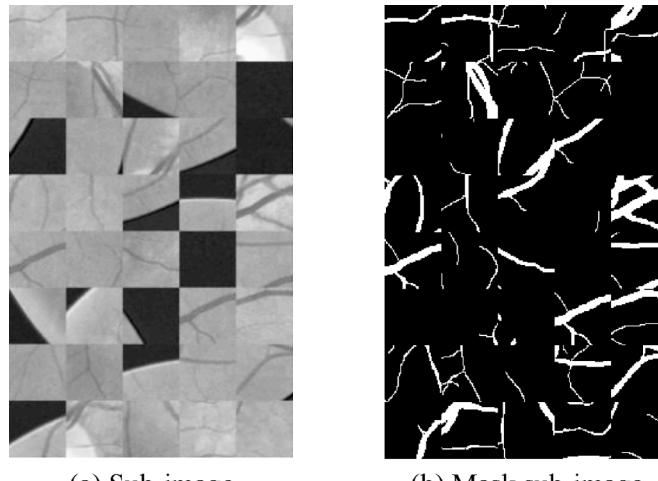


Fig. 9. Sub-image and corresponding mask sub-image.

$$F_1 = 2 \times \frac{Pre \times Sen}{Pre + Sen} \quad (16)$$

4.6. Analysis of experimental results

The proposed model takes the U-Net network as basic network and integrates it with rough channel attention mechanism. First, the segmentation networks FCN and Seg-Net [45] were selected to confirm the effectiveness of the basic U-Net network and compare experiments on DRIVE, Stare, and CHASE DB1 datasets of retinal vascular images. Experimental results were analyzed to evaluate the accuracy, sensitivity, specificity, and Jaccard similarity.

Table 2 displays the comparison results of FCN, Seg-Net, and U-Net models on the three datasets of retinal blood vessels. The table shows that U-Net model outperforms the other two models. Specifically, U-Net has higher segmentation accuracy and better recognition ability in the blood vessel pixels. A higher similarity is obtained between the segmentation map, gold standard map, and competitive recognition ability of background pixels. Considering the segmentation accuracy and sensitivity, a higher value is obtained in U-Net.

The above results show that networks based on codec structures like Seg-Net and U-Net have superior competitiveness in image segmentation. The U-Net model with skip connection structure enable the network concatenate and fuse the characteristic mappings of the encoder symmetry layer in the up-sampling process of every layer. Through the fusion of shallow and deep layer characteristics, the network is able to reserve details in high-level feature maps, thereby enhancing image segmentation accuracy. Thus, choosing U-Net as the basic network has certain effectiveness.

We divide experiments into four sections for verification and illustration. First, the efficiency of the rough channel attention mechanism is fully illustrated by comparing U-Net, Attention U-Net, and Rough channel attention U-Net (RCA-UNet). Second, the comparative experiments of Channel attention U-Net (CA-UNet) and Rough channel attention U-Net (RCA-UNet) fully demonstrate the effectiveness of rough neurons. Third, the effectiveness of the residual connection is illustrated through comparative RCA-UNet and RCAR-UNet experiments. Based on the numerical analysis, a visual comparison was performed to demonstrate more intuitively the superiority of RCAR-UNet in segmenting fine vascular structures.

We compare the RCA-UNet and CA-UNet model on three datasets to illustrate the effectiveness of the rough neurons. **Tables 3-5** show that the RCA-UNet model has better segmentation performance than others in terms of *Acc*, *Sen*, *F₁*, and Jaccard similarity. Compared with the CA-UNet model, the RCA-UNet improves 0.02 %~0.07 % in *Acc*, 0.01 %~2.83 % in *Sen*, 0.19 %~1.33 % in Jaccard similarity, and 0.14 %~1.05 % in *F₁*.

We selected U-Net, Attention U-Net, CA-UNet, and RCA-UNet to compare three fundus retinal vascular datasets and illustrate the validity of the rough channel attention, which combines rough neurons and the channel attention mechanism. Tables 3-5 show that the RCA-UNet model performs relatively better in the three datasets. Specifically, **Fig. 10** shows that the RCA-UNet model has high sensitivity, which is in three datasets, 2.97 %, 2.9 %, and 10.33 % higher than U-Net model; 0.31 %, 1.23 %, and 0.54 % higher than Attention U-Net model; and 0.01 %, 0.02 %, and 2.83 % higher than that of the CA-UNet. RCA-UNet model also offers better prediction accuracy for retinal vessels.

Pursuing the improvement of precision and sensitivity does not have much effect when considering the model's performance. Combining the proportion of positive and negative samples is essential for comprehensively evaluating the actual segmentation task. Moreover, the property of the network is appraised by harmonic mean *F₁* index, simultaneously affected by precision and sensitivity. As shown in **Fig. 11**, the RCA-UNet model performs excellently and improves by nearly 0.19 %~1.43 %, 0.88 %~2.03 %, and 1.24 %~4.71 %. As shown in **Fig. 12**, the Jaccard similarity between segmentation map and corresponding gold standard is higher, which is increased by 0.26 %-1.96 %, 1.12 %-2.57 %, and 1.57 %-5.78 % compared with other models.

Such results indicate the effectiveness of global max pooling and global average pooling for constructing rough channel attention mechanism with upper and lower approximate neurons. Thus, we consider the dependence between channels and introduce the rough set's upper and lower approximation principle, applying global max pooling and global average pooling to construct upper and lower approximation neurons, respectively. Moreover, we assign certain adaptive weights to obtain more reasonable attention coefficients, and perform corresponding re-calibration operations on the feature maps.

RCA-UNet and RCAR-UNet are compared on three datasets to show the advantage of rough channel attention module and verify residual connection effectiveness. **Tables 3-5** give comparative results intuitively. In tables, the best results are annotated using

Table 2

Comparison results of three public blood vessel datasets.

Dataset	Method	Acc	Sen	Spe	Pre	J	<i>F₁</i>
DRIVE	FCN	0.9001	0.2589	0.9946	0.8754	0.2497	0.3996
	Seg-Net	0.9440	0.6950	0.9774	0.8222	0.6193	0.7649
	U-Net	0.9509	0.7031	0.9871	0.8882	0.6459	0.7849
Stare	FCN	0.9204	0.2681	0.9978	0.9379	0.2635	0.4171
	Seg-Net	0.9289	0.3608	0.9963	0.9208	0.3499	0.5185
	U-Net	0.9526	0.6124	0.9929	0.9118	0.5782	0.7327
CHASE DB1	FCN	0.9340	0.4756	0.9784	0.6809	0.3889	0.5600
	Seg-Net	0.9377	0.5439	0.9758	0.6856	0.4353	0.6066
	U-Net	0.9517	0.6365	0.9822	0.7667	0.5379	0.6995

Table 3

Comparison results of drive retinal blood vessel datasets.

Method	Acc	Sen	Spe	Pre	J	F_1
U-Net	0.9509	0.7031	0.9871	0.8882	0.6459	0.7849
Attention U-Net	0.9527	0.7297	0.9853	0.8786	0.6629	0.7973
CA-UNet	0.9526	0.7327	0.9846	0.8742	0.6636	0.7978
RCA- UNet	0.9531	0.7328	0.9852	0.8788	0.6655	0.7992
RCAR-UNet	<u>0.9537</u>	<u>0.7487</u>	0.9836	0.8696	<u>0.6732</u>	<u>0.8047</u>

Table 4

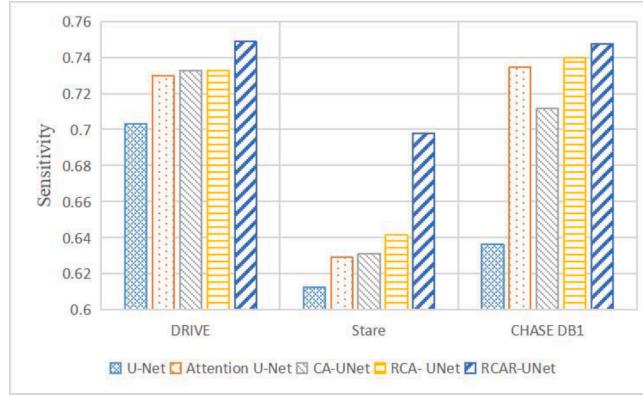
Comparison results of stare retinal blood vessel datasets.

Method	Acc	Sen	Spe	Pre	J	F_1
U-Net	0.9526	0.6124	0.9929	0.9118	0.5782	0.7327
Attention U-Net	0.9541	0.6291	0.9927	0.9109	0.5927	0.7442
CA-UNet	0.9551	0.6312	0.9936	0.9212	0.5989	0.7491
RCA- UNet	0.9553	0.6414	0.9926	0.9119	0.6039	0.7530
RCAR-UNet	<u>0.9594</u>	<u>0.6979</u>	0.9905	0.9069	<u>0.6461</u>	<u>0.7850</u>

Table 5

Comparison results of chase db1retinal blood vessel datasets.

Method	Acc	Sen	Spe	Pre	J	F_1
U-Net	0.9517	0.6365	0.9822	0.7667	0.5379	0.6995
Attention U-Net	0.9530	0.7344	0.9742	0.7340	0.5800	0.7342
CA-UNet	0.9549	0.7115	0.9785	0.7625	0.5824	0.7361
RCA- UNet	0.9556	0.7398	0.9765	0.7536	0.5957	0.7466
RCAR-UNet	<u>0.9566</u>	<u>0.7475</u>	0.9798	<u>0.7747</u>	<u>0.5983</u>	<u>0.7470</u>

**Fig. 10.** Sensitivity of different datasets.

underline and the second-best results are in bold.

For the RCA-UNet model, Acc , Sen , Jaccard similarity, and F_1 were improved by 0.06 %, 1.59 %, 0.77 %, and 0.55 %, respectively, on DRIVE dataset, as illustrated in [Table 3](#). As shown in [Table 4](#), RCAR-UNet showed significant improvement in Acc , Sen , J , and F_1 , with an increase of 0.41 %, 5.65 %, 4.22 %, and 3.2 %, respectively. As shown in [Table 5](#), compared with RCA-UNet model, RCAR-UNet improvement in the CHASE DB1 dataset is insignificant. Still, but they are all improved; Acc increases by 0.1 %, Sen increases by 0.33 %, Pre increases by 2.11 %, Jaccard similarity increases by 0.26 %, and F_1 increases by 0.04 %. The above experimental results fully prove the effectiveness of adding residual connections to the model for achieving feature maps. Transferring shallow characteristics to deep layer directly enriches the features extracted through the network. It helps the backpropagation of gradients when training the network, which improves the phenomenon of network degradation.

We employ ROC curve and PR curve to assess model performance. [Fig. 13](#) and [Fig. 14](#) describe the experimental results of ROC curves and PR curves of the four network models on different datasets. As shown in [Fig. 13](#), the area under curve (AUC) values of ROC curves of RCAR-UNet model on the three datasets are 0.9686, 0.9708, and 0.9540. RCAR-UNet is 0.73 % higher than classic U-Net model. The ordinate of the ROC curve is true positive rate (TPR), representing the proportionality of the predicted positive samples in the actual positive samples, namely prediction accuracy of actual positive samples, so the larger, the better. Moreover, the abscissa,

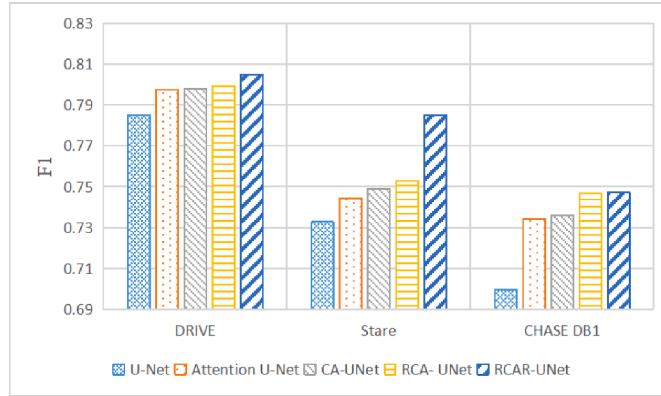


Fig. 11. Of different datasets.

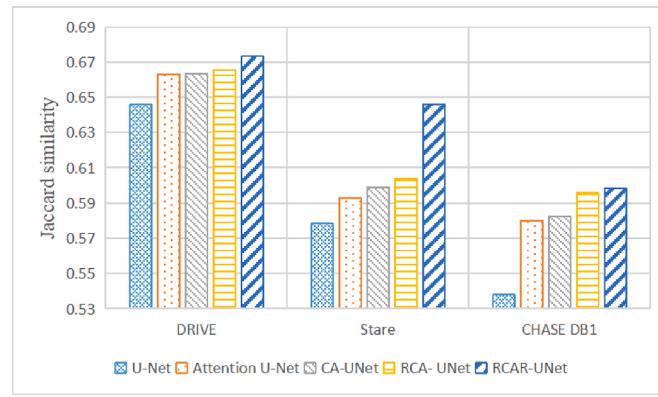


Fig. 12. Jaccard similarity of different datasets.

which is false positive rate (FPR), means the overall shall of the actual negative sample predicted to be positive. This FPR is the prediction error rate of the actual positive sample, so the smaller, the better.

In summary, the higher AUC value on ROC curve indicates better performance. Thus, RCAR-UNet performs excellently on the three fundus retinal vessels datasets. PR curve describes the relationship between precision rate and recall rate.

The segmentation process can have a precision rate and a high recall rate value, and the two can achieve a better balance point. The AUC area enclosed by the curve represents the AUC area, and the coordinate system is used to evaluate the model. Fig. 13 shows that RCAR-UNet has better performance, and its AUC values on three datasets are 0.9003, 0.8980, and 0.8285, improved compared with other models.

In summary, RCAR-UNet can obtain better segmentation accuracy of retinal blood vessels matched up against classic U-Net, Attention U-Net and other models. The analysis is conducted because the RCAR-UNet model integrates the deep neural network's capacity to learn complex features and the rough set's capacity to deal with uncertainty. A rough channel attention mechanism module is built and embedded in the U-Net model's skip connection to design rough neurons and achieve fusion between high and low features. Afterward, residual connections are added to transfer shallow features to higher levels to enrich network feature extraction and help the backpropagation of gradients when training the model.

We compared the segmentation effect diagrams of U-Net, Attention U-Net, and other models to intuitively illustrate RCAR-UNet's good segmentation effect on retinal blood vessels. It can be intuitively seen from Fig. 15 that the segmentation effect of each model for the main retinal blood vessel is basically consistent with the expert segmentation gold standard, but some blood vessel endings have segmentation loss and some blood vessels have segmentation errors. The segmentation results of RCAR-UNet model are basically consistent with (Fig. 16) the expert segmentation standard diagram (Fig. 17).

In order to support the quantitative evaluation and comparison against other approaches, we added actual image data to enable a clear comparison of the claimed improved performance for at least one each of cases highlighting (a) improved detection of thin vessels, (b) continuity of blood vessels. The segmentation effect of fine blood vessels in the region promising, effectively alleviating the problem of blood vessel segmentation and fracture.

Finally, the proposed RCAR-UNet model was compared with recent advanced models to confirm its effectiveness. We selected the model proposed by previous studies [13,46] for comparative experimental verification. Table 6 shows the results, analyzed by accuracy, sensitivity, and specificity.

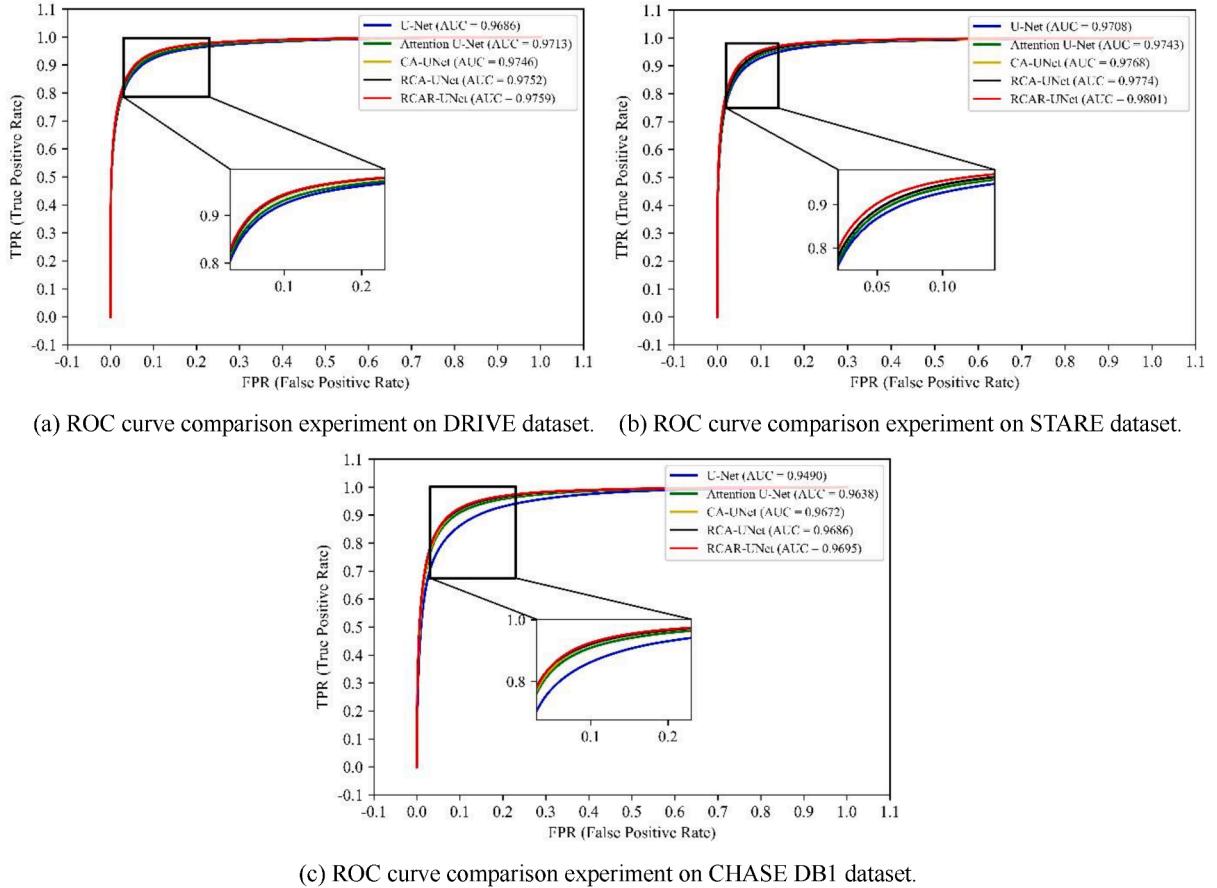


Fig. 13. Comparison of ROC curves of different models on different datasets.

As shown in Table 6, the experimental data of the RCAR-UNet model is not much different from the other two models in the three datasets. However, the results show a slight advantage in accuracy and sensitivity with certain competitiveness in specificity. Thus, the RCAR-UNet model has more robust blood vessel detection ability and better segmentation accuracy. Our proposed rough channel attention module makes the model pay more attention to details and features. Moreover, it has a more vital detection ability of fine blood vessels and achieves higher segmentation accuracy.

However, the complexity analysis of the model involves a significant amount of calculation (time/computational complexity) and a significant number of model parameters (spatial complexity). The model time complexity is described below:

$$Time \sim o\left(\sum_{i=1}^N F_i^2 \cdot K_i^2 \cdot C_{i-1} \cdot C_i\right) \quad (17)$$

The time complexity is related to the output feature map's size for every layer F_i , the size dimension of each convolution kernel K_i , neural network layers' number N , channel's number in output characteristic mappings of previous layer C_{i-1} , and channel's number in output feature graphs at current layer C_i . Meanwhile, the output feature graph's size at each layer F is related to the input feature map's size X , step size of every convolution kernel $stride$, fillings' number $padding$ and convolution kernel's size K , defined as:

$$F = (X - K + 2 \times padding) / stride + 1 \quad (18)$$

Spatial complexity has two components: total weight parameters in all parametric layers of the model and the calculated output feature graph's size per layer. Thus, total weight parameters in all the model parametric layers are related to each convolution kernel's size K_i , the network's layer number N , channels' number in the output characteristic mapping of previous layer C_{i-1} , and channels' number in output characteristic mapping of current layer C_i , but not input data's size. The output characteristic mapping's space depends on the multiplication of the output characteristic mapping's size F_i and channel's number per layer C_i . Therefore, model spatial complexity is expressed as follows:

$$Space \sim o\left(\sum_{i=1}^N K_i^2 \cdot C_{i-1} \cdot C_i + \sum_{i=1}^N M^2 \cdot C_i\right) \quad (19)$$

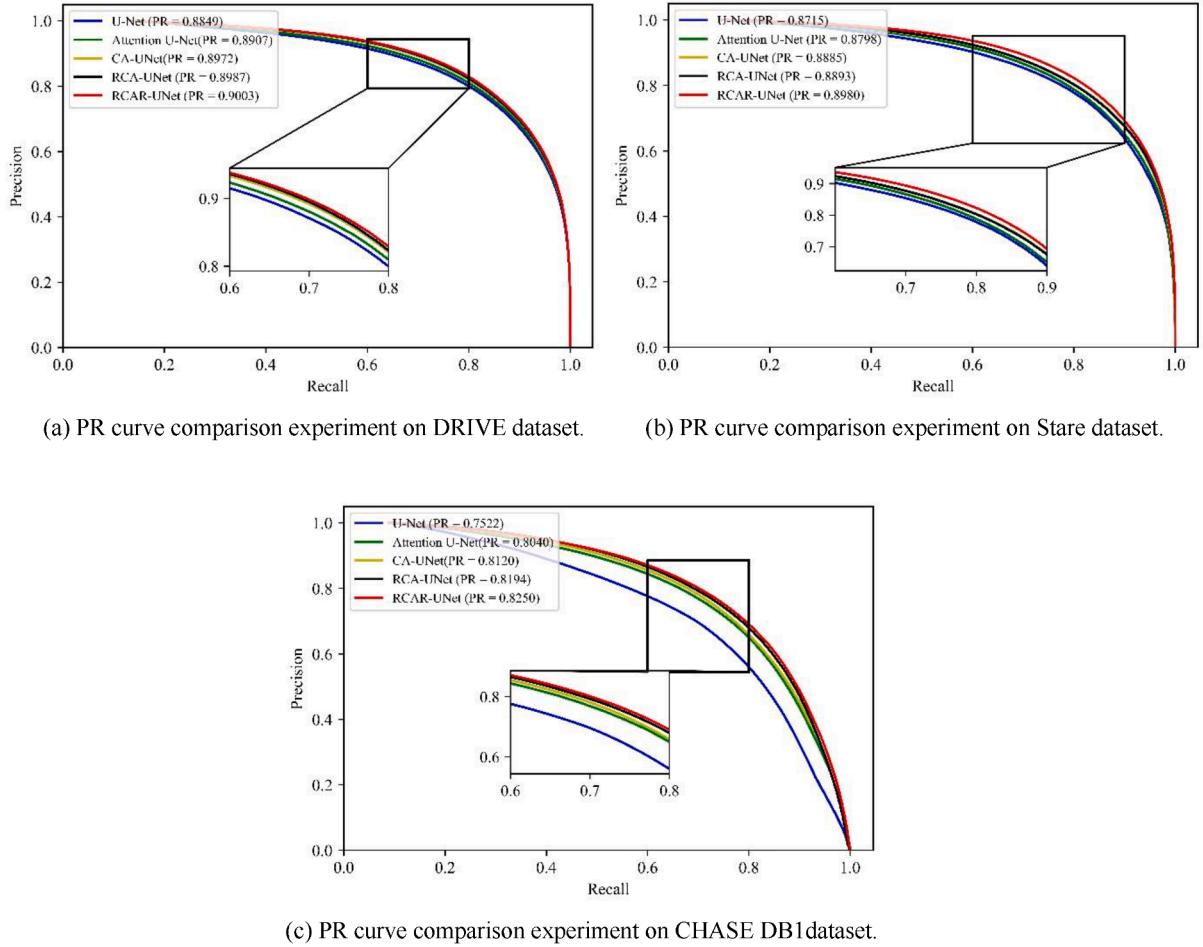


Fig. 14. Comparison of PR curves of different models on different datasets.

Although adding a rough channel attention module based on rough neurons to the skip connection to measure uncertainty does increase the time and space complexity, RCAR-UNet will achieve higher segmentation accuracy and better recognition ability of retinal blood vessels, based on the comparative experimental data. The resulting segmentation graph is closer to the gold standard graph. A higher F_1 value is obtained by considering the accuracy and sensitivity. The segmentation effect drawing shows directly that the retina's blood vessel shape is basically distinguished, especially for thin blood vessels segmentation effect is better, and guaranteed its continuity. Considering model's spatiotemporal complexity and its effect, we believe it is worthwhile to increase the spatiotemporal complexity to obtain better segmentation results.

5. Discussion

The width and thickness of retinal blood vessels in fundus are different, and the boundary is blurred. The existing retinal blood vessel segmentation methods are not effective in the segmentation of fine blood vessels, and the retinal blood vessels are easily broken. In this paper, several sets of comparative experiments were carried out, and the research results fully showed that the RCAR-UNet model was superior to several representative models on the three fundus retinal blood vessel data sets. RCAR-UNet model has better ability to identify blood vessels, and can obtain higher segmentation accuracy, and the segmentation result map is more similar to the gold standard map. At the same time, it can be seen intuitively from the schematic diagram of the segmentation results that the RCAR-UNet model has a better effect on the segmentation of extremely fine blood vessels at the vascular end, and the continuity of blood vessels is also guaranteed, which effectively improves the shortcomings of the previous model. The reasons for the good segmentation effect of this model are as follows:

- (1) We demonstrate that the structure and image acquisition environment affect the retinal blood vessel image. Moreover, uncertain information is identified in the retinal vessel image. The ability of deep neural networks to learn complex features and rough sets to deal with uncertainty is integrated into the design of rough neurons. Rough neurons map the input value into a rough range. The final output depends on the output of the upper and lower neurons. We multiply a certain weight value to

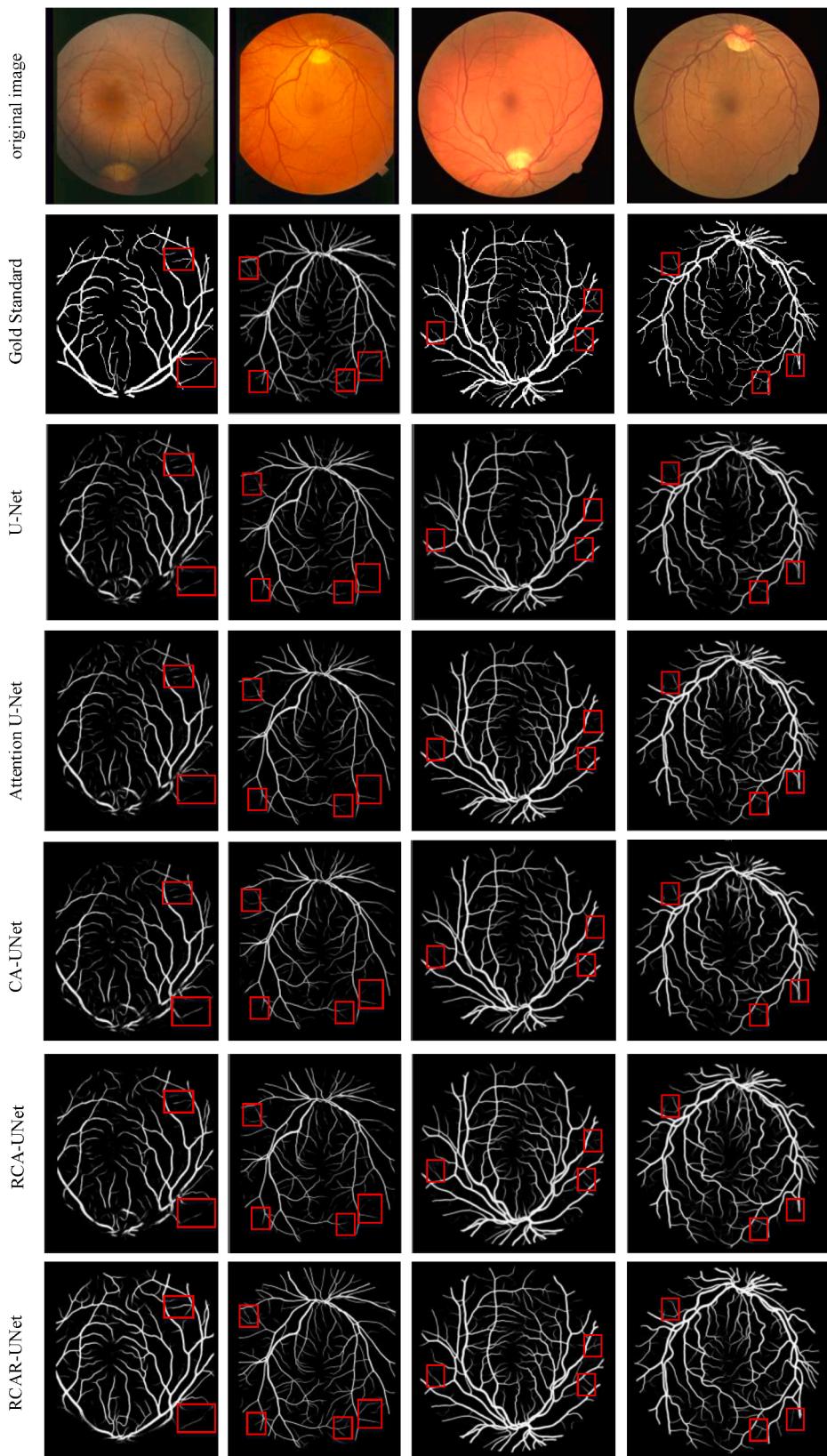


Fig. 15. Segmentation results of each model for retinal vessels.

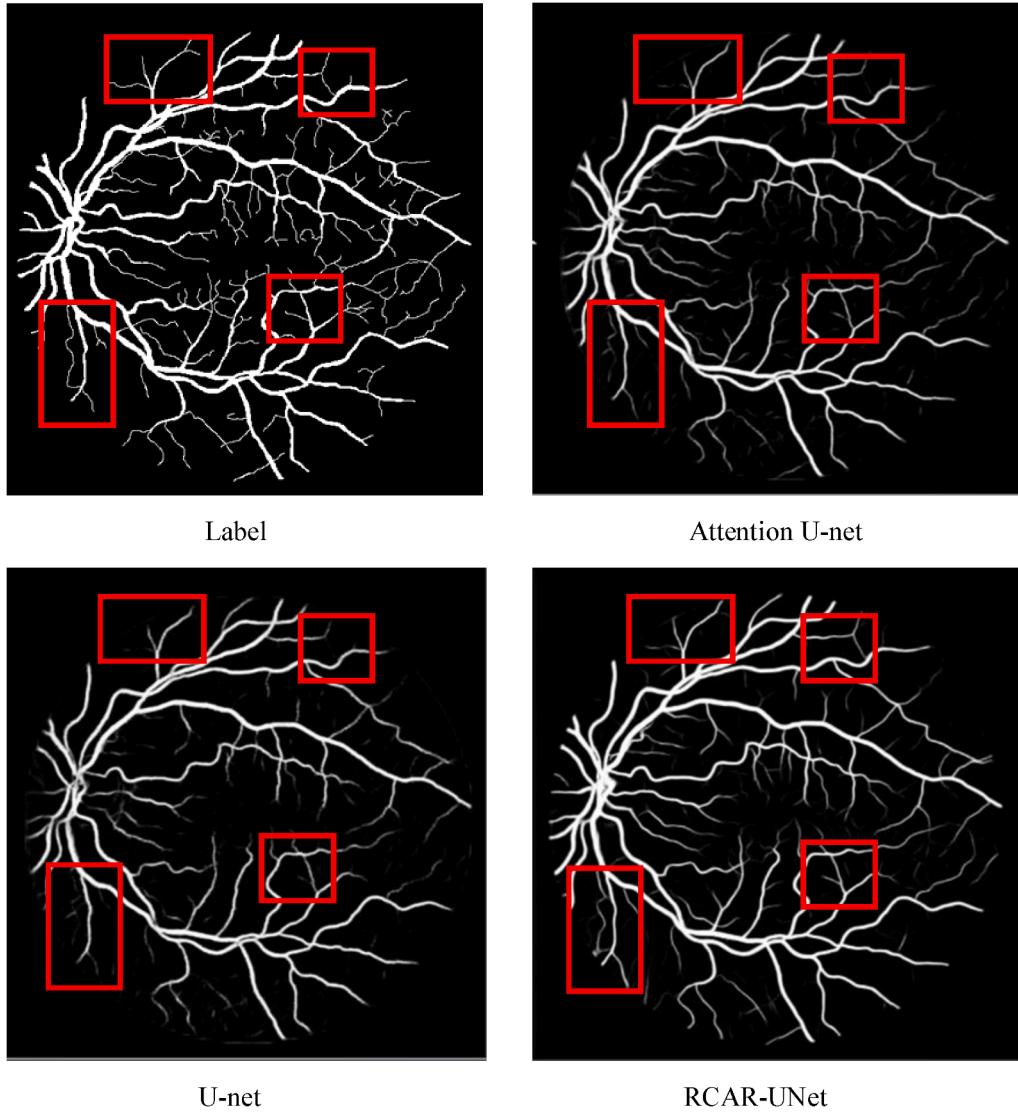


Fig. 16. Comparison of the effect of fine blood vessel segmentation.

distinguish the contribution of different neurons to the final output. The purpose of the design is to rough the deterministic values of the input and recalibrate the deterministic values to obtain more semantic output values.

- (2) We argue that the feature extraction of retinal blood vessel images using convolution operation will face channel uncertainty, which will affect the channel attention coefficient and make the deep network incapable of paying attention to the detailed features of retinal blood vessels. Based on rough neurons, a rough channel attention mechanism module is constructed and embedded in the jump connection of the U-Net model for the fusion of high and low features. The upper and lower approximate neurons were constructed using global maximum pooling and global average pooling. Moreover, the weighted summation between neurons was conducted. The established channel dependence relationship was roughed. The dependency relationship contained global information and had local characteristics, effectively realizing the accurate re-calibration of the extracted retinal vascular features.
- (3) We introduced residual connection for the direct transmission from low-level features to high-level features, capable of improving network degradation to a certain extent.

In this paper, the rough channel Attention mechanism, which combines the rough set and the attention mechanism, is applied to the retinal blood vessel segmentation. The experimental results fully verify that the model has a good segmentation effect on the subtle retinal blood vessels. In addition, experiments show that compared with the general attention mechanism, the rough channel attention mechanism proposed in this paper has a better ability to extract fine-grained features. This feature is not only suitable for retinal vessel segmentation tasks, but also can be extended to more general applications, such as natural scene image classification, face recognition,

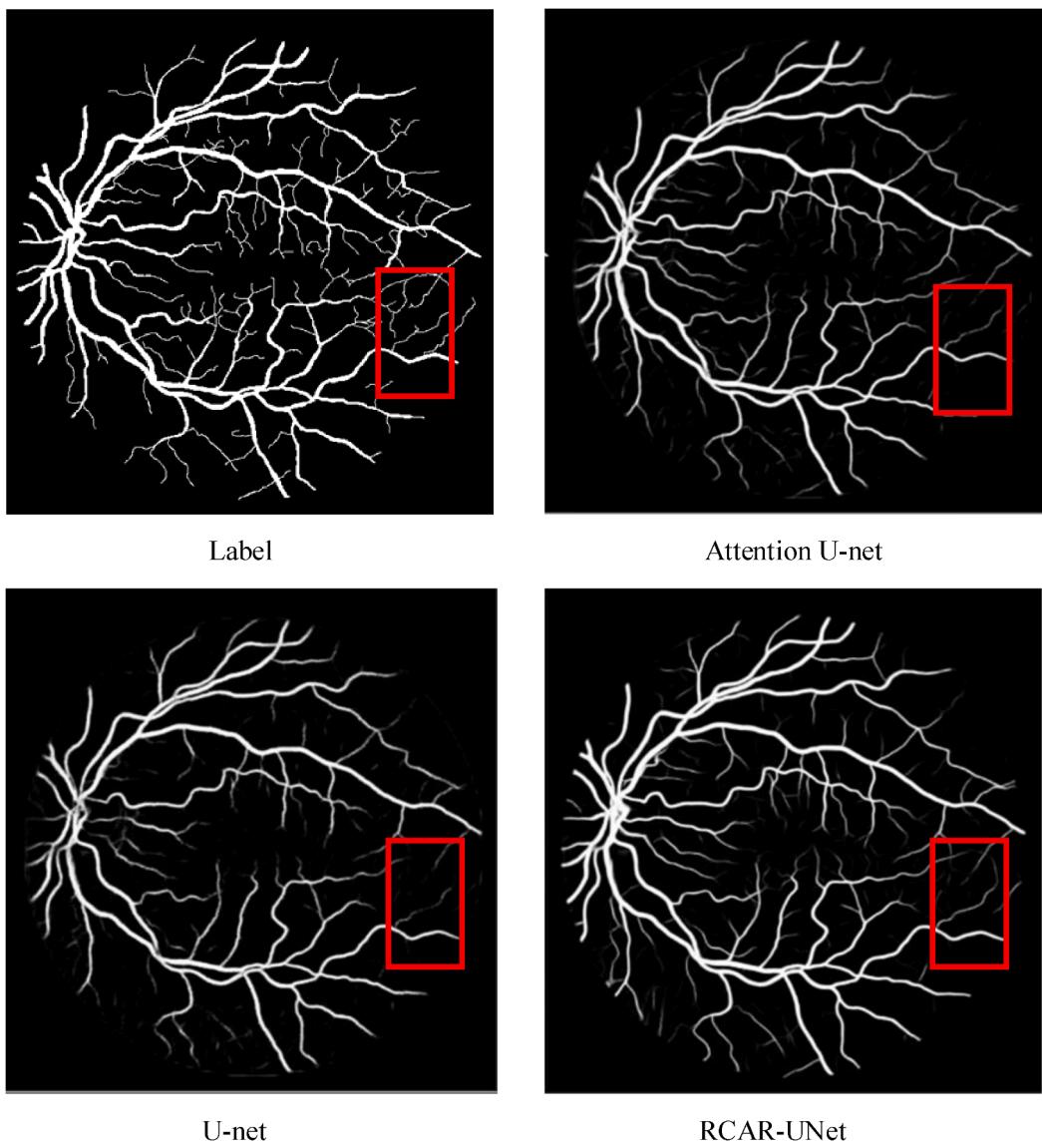


Fig. 17. Continuous comparison of vascular segmentation effect.

Table 6

Comparison results of three public blood vessel datasets.

Dataset	Method	Acc	Sen	Spe
DRIVE	Reference [43]	0.9514	0.7477	0.9702
	Reference [13]	0.9527	0.7469	0.9846
	RCAR-UNet	0.9537	0.7487	0.9836
Stare	Reference [43]	0.9587	0.6870	0.9974
	Reference [13]	0.9591	0.6973	0.9924
	RCAR-UNet	0.9594	0.6979	0.9905
CHASE DB1	Reference [43]	0.9551	0.7134	0.9789
	Reference [13]	0.9562	0.7442	0.9792
	RCAR-UNet	0.9566	0.7475	0.9798

behavior detection and other applications. However, this model also has some shortcomings, which need to be followed up:

- (1) Fig. 6 shows that there are two fully connected layers in the rough channel attention mechanism. Compared with the convolutional layer, the fully connected layer generates more parameters that need to be adjusted adaptively. Although the model performance is improved, the complexity and computation of the network are also increased.
- (2) The deep convolutional neural network model is dominated by convolution and pooling operations. The uncertain information in the image will be propagated in various stages of the network. Combining rough set theory into convolution and pooling operations can deal with uncertain data in images.

6. Conclusions and future directions

We considered the relationship between characteristic channels and introduced a new rough channel attention, which can increase the identification ability of network, given the complexity of retinal vascular structure, delicate blood vessels, and susceptibility to light. Specifically, the rough attention module is constructed using the upper and lower approximation concepts to set limits for attention coefficient. The upper and lower neurons are constructed using global max pooling as the upper limit of the attention coefficient. The lower limit of the attention coefficient is used as the lower limit of the global average pooling structure. The upper and lower limits are given a certain weight coefficient and summed to obtain a more semantic attention coefficient. The attention coefficient contains global information and has certain partial information. Then, the residual connection is introduced to map the shallow features directly to the high layer to improve the network degradation. The characteristic of proposed model is proved by comparing with the existing models on three fundus retinal blood vessel datasets. The RCAR-UNet model can segment the small branches of the retinal vessels and obtain better segmentation accuracy. In future studies, our team will continue to work on investigate the fusion of rough set into convolution and pooling layers of deep neural network models to handle the uncertainty and imprecision information in image features.

CRediT authorship contribution statement

Weiping Ding: Conceptualization, Methodology, Investigation, Formal analysis, Supervision, Writing – review & editing, Funding acquisition. **Ying Sun:** Data curation, Visualization, Writing – original draft, Writing – review & editing. **Jiashuang Huang:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Hengrong Ju:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Chongsheng Zhang:** Formal analysis, Visualization, Writing – review & editing. **Guang Yang:** Review & editing, Supervision. **Chin-Teng Lin:** Formal analysis, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 61976120, Grant 62006128 and Grant 62102199, the Natural Science Foundation of Jiangsu Province under Grant BK20231337, and the Natural Science Key Foundation of Jiangsu Education Department under Grant 21KJA510004.

References

- [1] Y. Ma, H. Hao, J. Xie, et al., ROSE: a retinal OCT-angiography vessel segmentation dataset and new model, *IEEE Trans. Med. Imaging* 40 (3) (Mar 2020) 928–939.
- [2] Y. Tan, K.F. Yang, S.X. Zhao, et al., Retinal Vessel Segmentation with Skeletal Prior and Contrastive Loss, *IEEE Trans. Med. Imaging* 41 (9) (Sept. 2022) 2238–2251.
- [3] X. Li, Y. Jiang, M. Li, et al., Lightweight Attention Convolutional Neural Network for Retinal Vessel Image Segmentation, *IEEE Trans. Ind. Inf.* 17 (3) (May 2021) 1958–1967.
- [4] S. Feng, Z. Zhuo, D. Pan, et al., CcNet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features, *Neurocomputing* 392 (Jun 2020) 268–276.
- [5] J. Son, S.J. Park, K.H. Jung, Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks, *J. Digit. Imaging* 32 (3) (Oct 2019) 499–512.
- [6] Q. Jin, Z. Meng, T.D. Pham, et al., DUNet: A deformable network for retinal vessel segmentation, *Knowl.-Based Syst.* 178 (15) (Aug 2019) 149–162.
- [7] Y. Li, et al., Global Transformer and Dual Local Attention Network via Deep-Shallow Hierarchical Feature Fusion for Retinal Vessel Segmentation, *IEEE Trans. Cybern.* 53 (9) (Sept. 2023) 5826–5839.
- [8] N. Ibtehaz, M.S. Rahman, MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation, *Neural Netw.* 121 (Jan 2020) 74–87.

- [9] Y. Ye, C. Pan, Y. Wu, et al., MFI-Net: Multiscale Feature Interaction Network for Retinal Vessel Segmentation, *IEEE J. Biomed. Health Inform.* 26 (9) (Sept. 2022) 4551–4562.
- [10] W.P. Ding, Mohamed Abdel-Basset, Hossam Hawash, et al., “Multimodal Infant Brain Segmentation by Fuzzy-informed Deep Learning”, *IEEE Trans. Fuzzy Syst.* 30 (4) (Apr 2022) 1088–1101.
- [11] M.Z. Alom, C. Yakopcic, M. Hasan, et al., Recurrent residual U-Net for medical image segmentation, *J. Med. Imaging* 6 (1) (Mar 2019).
- [12] K. Yue, B. Zou, Z. Chen, et al., Retinal vessel segmentation using dense U-net with multiscale inputs, *J. Med. Imaging* 6 (3) (Sep 2019).
- [13] Zhang, Y, Miao H, Chen Z N, et al. “Bridge-Net: Context-involved U-net with patch-based loss weight mapping for retinal blood vessel segmentation.” *Expert Systems with Application* Jun, 2022.
- [14] Vaswani A, Shazeer N, Parmar N, et al. “Attention is all you need,” in 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017, pp. 5998–6008.
- [15] P. Xuan, L. Zhan, H. Cui, et al., Graph Triple-Attention Network for Disease-Related LncRNA Prediction, *IEEE J. Biomed. Health Inform.* 26 (6) (June 2022) 2839–2849.
- [16] P. Fang, J. Zhou, S.K. Roy, et al., Attention in Attention Networks for Person Retrieval, *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (9) (Sept. 2022) 4626–4641.
- [17] Oktay, Ozan, et al. “Attention unet: Learning where to look for the pancreas,” arXiv preprint arXiv:1804.03999, 2018.
- [18] R. Gu, G. Wang, T. Song, et al., CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation, *IEEE Trans. Med. Imaging* 40 (2) (Feb 2020) 699–711.
- [19] Z. Chen, M. Wu, R. Zhao, et al., Machine remaining useful life prediction via an attention-based deep learning approach, *IEEE Trans. Ind. Electron.* 68 (3) (Mar 2020) 2521–2531.
- [20] H. Cui, C. Yuwen, L. Jiang, et al., Multiscale attention guided U-Net architecture for cardiac segmentation in short-axis MRI images, *Comput. Methods Programs Biomed.* 206 (July 2021), 106142.
- [21] Guo C, Szemenyei M, Yi Y, et al. “Sa-unet: Spatial attention u-net for retinal vessel segmentation,” in 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 1236–1242.
- [22] Li C, Tan Y, Chen W, et al. “Attention Unet++: A Nested Attention-Aware U-Net for Liver CT Image Segmentation,” in 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, pp. 345–349, 2020.
- [23] N.S. Punn, S. Agarwal, RCA-JUNet: a residual cross-spatial attention-guided inception U-Net model for tumor segmentation in breast ultrasound imaging, *Mach. Vis. Appl.* (Feb 2022).
- [24] Y. Yang, W. Wan, S. Huang, et al., RADCU-Net: residual attention and dual-supervision cascaded U-Net for retinal blood vessel segmentation, *Int. J. Mach. Learn. Cybern.* (Nov, 2022.) 1–16.
- [25] X. Tang, B. Zhong, J. Peng, et al., Multi-scale channel importance sorting and spatial attention mechanism for retinal vessels segmentation, *Appl. Soft Comput.* 93 (Aug 2020), 106353.
- [26] W. Ding, M. Abdel-Basset, W. Pedrycz, et al., Multimodal Infant Brain Segmentation by Fuzzy-Informed Deep Learning, *IEEE Trans. Fuzzy Syst.* 30 (4) (April 2022) 1088–1101.
- [27] B. Fang, C. Zheng, H. Wang, et al., Two-Stream Fused Fuzzy Deep Neural Network for Multiagent Learning, *IEEE Trans. Fuzzy Syst.*, Feb. 31 (2) (2023) 511–520.
- [28] W. Zheng, L. Yan, C. Gou, et al., Fuzzy Deep Forest With Deep Contours Feature for Leaf Cultivar Classification, *IEEE Trans. Fuzzy Syst.* 30 (12) (Dec. 2022) 5431–5444.
- [29] Z. Pawlak, Rough sets, *Int. J. Comput. Inform. Sci.* 11 (5) (Oct 1982) 341–356.
- [30] W.P. Ding, C.T. Lin, Z.H. Cao, Deep neuro-cognitive co-evolution for fuzzy attribute reduction by quantum leaping PSO with nearest-neighbor memplexes, *IEEE Trans. Cybern.* 49 (7) (Jul 2019) 2744–2757.
- [31] W. Wei, J. Liang, Information fusion in rough set theory: An overview, *Information Fusion* 48 (Aug 2019) 107–118.
- [32] Q. Wang, Y. Qian, X. Liang, et al., Local neighborhood rough set, *Knowl.-Based Syst.* 153 (Aug 2018) 53–64.
- [33] M. Hu, Y. Yao, Structured approximations as a basis for three-way decisions in rough set theory, *Knowl.-Based Syst.* 165 (Feb 2019) 92–109.
- [34] Lingras P. “Rough neural networks” in Proc. of the 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge based Systems, 1996, pp. 1445–1450.
- [35] J. Jelonek, K. Krawiec, R. Slowinski, Rough set reduction of attributes and their domains for neural networks, *Comput. Intell.* 11 (2) (May 1995) 339–347.
- [36] M. Khodayar, O. Kaynak, M.E. Khodayar, Rough deep neural architecture for short-term wind speed forecasting, *IEEE Trans. Ind. Inf.* 13 (6) (Dec 2017) 2770–2779.
- [37] B. Cao, J. Zhao, Z. Lv, et al., Multi objective evolution of fuzzy rough neural network via distributed parallelism for stock prediction, *IEEE Trans. Fuzzy Syst.* 28 (5) (May 2020) 939–952.
- [38] H. Liao, S. Ding, M. Wang, et al., An overview on rough neural networks, *Neural Comput. & Applic.* 27 (7) (Aug 2016) 1805–1816.
- [39] H. Jahangir, H. Tayyari, S. Baghali, et al., A novel electricity price forecasting approach based on dimension reduction strategy and rough artificial neural networks, *IEEE Trans. Ind. Inf.* 16 (4) (Apr 2019) 2369–2381.
- [40] Sabzalian M H, Mohammadzadeh A, Lin S, et al. “A robust control of a class of induction motors using rough type-2 fuzzy neural networks,” *Soft Computing*, pp. 1–11, Nov, 2019.
- [41] S. Li, Q. Yan, P. Liu, An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism, *IEEE Trans. Image Process.* 29 (Aug 2020) 8467–8475.
- [42] P. Liskowski, K. Krawiec, Segmenting retinal blood vessels with deep neural networks, *IEEE Trans. Med. Imaging* 35 (11) (Nov 2016) 2369–2380.
- [43] S.A.K. Usmani, S. Gulnaz, M. Sultana, et al., Segmenting Retinal Blood Vessels with Deep Neural Networks, *J. Informat. Storag. Process. Syst.* 19 (1) (Apr 2020).
- [44] J. Almotiri, K. Elleithy, A. Elleithy, Retinal vessels segmentation techniques and algorithms: a survey, *Appl. Sci.* 8 (2) (Jan 2018) 155.
- [45] V. Badrinarayanan, A. Kendall, R. Cipolla, Seg-net: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12) (2017) 2481–2495.
- [46] Francia, Gendry Alfonso, et al. “Chaining a U-net with a residual U-net for retinal blood vessels segmentation.” *IEEE Access*, vol. 8, pp. 38493–38500, Feb, 2020.