



# OPEN An effective vessel segmentation method using SLOA-HGC

Zerui Liu<sup>1</sup>, Junliang Du<sup>3</sup>, Weisi Dai<sup>1</sup>, Wenke Zhu<sup>2</sup>, Ziqing Ye<sup>1✉</sup>, Lin Li<sup>1✉</sup>, Zewei Liu<sup>1</sup>, Linan Hu<sup>4</sup>, Lin Chen<sup>5</sup> & Lixiang Sun<sup>1</sup>

Accurate segmentation of retinal blood vessels from retinal images is crucial for detecting and diagnosing a wide range of ophthalmic diseases. Our retinal blood vessel segmentation algorithm enhances microfine vessel extraction, improves edge texture clarity, and normalizes vessel distribution. It stabilizes neural network training for complex retinal vascular features. Channel-aware self-attention (CAS) improves microfine vessel segmentation sensitivity. Heterogeneous adaptive pooling (HAP) facilitates accurate vessel edge segmentation through multi-scale feature extraction. The ghost fully convolutional Rectified Linear Unit (GFCReLU) module in the output convolutional layer captures deep semantic information for better vessel localization. Optimization training with Sparrow-Integrated Lion Optimization Algorithm (SLOA) employs sparrow stochastic updating and annealing to fine-tune parameters. The results of the experiments on our homemade dataset and three public datasets are as follows: Mean Intersection over Union (MIoU) of 80.61%, 76.14%, 76.90%, 74.11%; Dice coefficients of 78.97%, 72.51%, 72.84%, 68.93%; and accuracies of 94.83%, 95.74%, 96.67%, 95.81% respectively. The model effectively segments retinal blood vessels, offering potential for diagnosing ophthalmic diseases. Our dataset is available at <https://github.com/ZhouGuoXiong/Retinal-I-blood-vessels-for-segmentation>.

**Keywords** Vessel segmentation, Sparrow updating mechanism, Artificial intelligence applied medical image processing

Retinal blood vessel morphology, such as shape and branching, can help diagnose diseases like diabetes mellitus<sup>1</sup>, hypertension<sup>2</sup>, and atherosclerosis<sup>3</sup>. However, this information is mainly obtained by manual visual inspection, which is not only inefficient but also subjective. Moreover, retinal vessel segmentation is challenging due to the presence of microfine blood vessels<sup>4</sup>, blurred edge texture<sup>5</sup>, and irregular blood vessel distribution<sup>6</sup> in fundus images. Therefore, utilizing topological features of blood vessels without requiring expert-labeled vessel information is essential. These features include Gabor wavelets<sup>7</sup>, combinatorial filters<sup>8</sup>, and vector field divergence<sup>9</sup>. The unsupervised learning method does not utilize labeled data and lacks a learning process. This makes it unable to adapt to differences in data distribution between different domains and capture microfine blood vessels. Thus, it fails to meet the needs of practical clinical assistance. Supervised learning segmentation methods mainly extract the structural features of retinal blood vessels through algorithms and train segmentation models based on expert labels, such as support vector machine (SVM)<sup>10</sup>. The above algorithms still have the problems of being affected by subjective factors when manually extracting features, and not being able to adapt to the morphological-geometrical transformations of blood vessels. Supervised deep learning segmentation algorithms have gained popularity in relevant domains in recent years. These algorithms can learn semantic representations more effectively and detect the vessel pixels more accurately<sup>11,12</sup>. Moreover, they can be self-improved or optimized, which enhances their stability and precision for fundus vessel segmentation. UNet<sup>13</sup> is a U-shaped fully convolutional network<sup>14</sup> that can obtain features from various scales of an image and maintain the fine details by using skip connections. This structure allows accurate segmentation of targets in different domains, so UNet and its variants<sup>15–17</sup> have been widely used in various image segmentation tasks with excellent results. Guo et al.<sup>18</sup> proposed SA-UNet with the introduction of a spatial attention mechanism. They used spatial attention modules to highlight important features and suppress unnecessary features to improve the expressive power of the network. But there are still models with insufficient receptive field and the problem of missing local features of microfine blood vessels. The appearance of Transformer model<sup>19</sup> brought the vitality to the vision

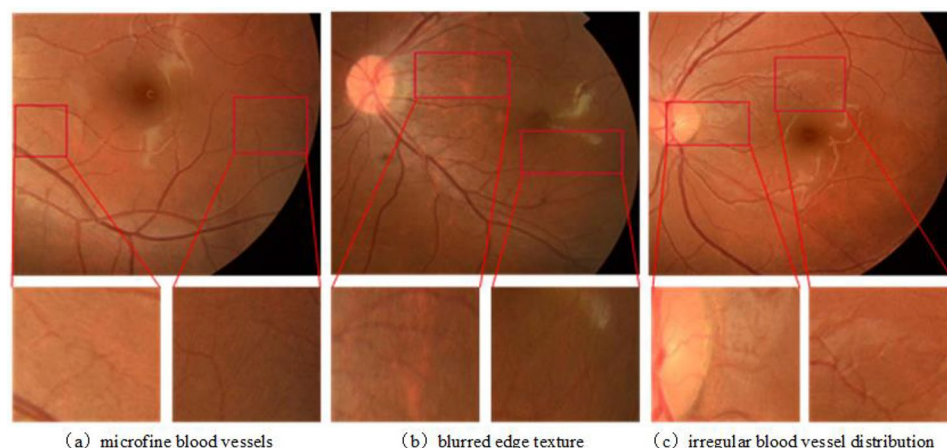
<sup>1</sup>Faculty of Electronic Information and Physics, Central South University of Forestry and Technology, Changsha 410004, Hunan, China. <sup>2</sup>College of Bangor, Central South University of Forestry and Technology, Changsha 410004, Hunan, China. <sup>3</sup>Artificial Intelligence Institute, Shanghai Jiao Tong University, Shanghai 200240, China. <sup>4</sup>Zhuzhou Central Hospital, Zhuzhou 412000, Hunan, China. <sup>5</sup>Zhuzhou Sansanyi Eye Hospital, Zhuzhou 412000, Hunan, China. ✉email: 20011081@qq.com; edalin@163.com

field, which was the first to be used in the natural language processing (NLP) area, and then the proposed Vision Transformer model<sup>20</sup> solved the problem of difficult high-level semantics extraction to a certain extent. However, Transformer has obvious drawbacks, such as large number of parameters, high arithmetic requirements, and difficulty in model training. Sultana et al.<sup>21</sup> proposed RIMNet, an image enhancement network and residual block model that uses the traditional encoder-decoder structure. It introduces the image enhancement network and residual block for accurate retinal blood vessel segmentation, but this also increases the model capacity and degrees of freedom. With insufficient data volume, it is more likely to overfit and miss microvascular features in segmentation. Guo et al.<sup>22</sup> proposed multi-scale short-connected divine supervisory network BTS-DSN, which uses short-connections to transfer semantic information between side output layers, and top-down short-connections enhance the high-level side outputs with low-level semantic information. The algorithm achieves more ideal segmentation results, but still has problems such as blurred boundary texture.

In conclusion, the existing segmentation algorithms, despite their good performance, still struggle with extracting microfine blood vessels, handling retinal vessel edges robustly, and locating vessel positions accurately. The field of retinal vessel segmentation needs more research to reveal deeper features of diseases and improve the specialization of the network. To this end, we apply UNet as a contrast network, perform related work analyses, and demonstrate the practicality of enhancing retinal vessel segmentation based on deep learning methods.

The current process of segmenting retinal vessels faces four urgent issues: (1) As shown in Fig. 1a, microvessels form a part of the retinal vasculature. They reflect the microcirculatory status of the retina, which is important for assessing the severity of ophthalmic diseases<sup>23</sup>. However, the microvessels themselves are difficult to distinguish and extract in images, which creates difficulties and challenges for accurate segmentation of retinal vessels. It is important to note that microvessels, defined as those with small diameters and belonging to tertiary branches, only account for approximately 10% of the total surface area of the retinal vascular network in the entire retina<sup>24,25</sup>, and thus the ratio shown in Fig. 1a mainly reflects the distribution of vessels in a specific region, rather than the entire retina as a whole. (2) As shown in Fig. 1b, some of the blood vessels have blurred edges due to the small gray value difference between them and the background. This makes it challenging to determine the width of the vessels, which affects the precision and consistency of retinal vessel segmentation, and consequently, the subsequent retrieval and evaluation of vessel features. (3) As shown in Fig. 1c, there are large variations and differences in the distribution characteristics of blood vessels, and the irregular distribution of blood vessels affects the robustness and generalization of retinal blood vessel segmentation, thus affecting the subsequent feature extraction and analysis. Therefore, the segmentation algorithm should be adaptable to these small differences for effective task completion. (4) The complexity and variability of retinal vascular structures, including diabetic retinopathy (DR), can lead to severe vascular changes that challenge machine learning models for detection and classification. Advanced stages of DR often result in new vessel formation, as well as changes in existing vessels, such as intraretinal microvascular anomalies (IRMA) and vein beading<sup>26,27</sup>. Since these vascular changes are similar in appearance and irregular in morphology to other retinal features, they complicate the detection and segmentation tasks of machine learning algorithms. Therefore, new optimization algorithms are needed that can adapt to these different data features and effectively address the complexity associated with vascular pathology changes. This complexity tends to affect the robustness of the neural network training process, and traditional optimizers may have difficulty in maintaining stable learning of the model in the presence of these highly variable features, thus new optimization algorithms need to be developed to accommodate diverse data features.

Several studies have also suggested some solutions to these problems. To address the difficult task of extracting the microfine blood vessels of the retina, Guo et al.<sup>28</sup> proposed a modified efficient channel attention (MECA) and proposed channel attention double residual fast (CADRB), which formed CAR-UNet to segment retinal vascular and non-vascular pixels accurately. Cao et al.<sup>29</sup> proposed Swin-UNet, a medical segmentation model with a pure UNet-like transformer that feeds labeled image patches into a transformer-based U-shaped encoder-decoder architecture with skip connections to segment microfine blood vessels accurately. These methods show



**Fig. 1.** Retinal vascular characterization.

how channel attention and self-attention can improve the network's feature representation and segmentation ability, which motivates our research. By combining multiple attentions, the network's receptive field is expanded to capture retinal vessel features from different directions, including microfine blood vessels.

To address the challenges of blurred retinal vessel edges, Laibacher et al.<sup>30</sup> added a pre-trained MobileNetV2 component and a shrinking bottleneck module for encoding and decoding, respectively, to estimate the vessel width. Liang et al.<sup>31</sup> proposed FMVG-Net, an algorithm that incorporates Mobile Vit and inverted gating codecs for segmenting retinal vessels. It can dynamically extract and accurately segment the pixels of retinal trunk and branch vessels. By using multi-scale feature extraction, these two methods improve the network's ability to analyze features and edges, which motivates us to solve the problem of blurred retinal vessel edges.

To localize irregular retinal blood vessels accurately, Liu et al.<sup>32</sup> proposed an Attention Augmented Generative Adversarial Network (AA-WGAN) that captures the pixel dependencies in the whole image and segments the complex vascular structures. Samuel et al.<sup>33</sup> used a vessel-specific convolutional block (VSC) that skips the chain convolution (SC) and feature map summation, which define the vessel features learned from the global features, to extract irregular vessel distribution. These methods show how to handle redundant vessel features and prevent overfitting by stacking different features, as well as improve the vessel localization accuracy.

Building upon these insights and inspired by the aforementioned advancements, our work aims to address the persistent challenges in retinal vessel segmentation. By leveraging novel techniques and incorporating domain-specific knowledge, we propose a more advanced solution that not only improves segmentation accuracy but also tackles issues like microfine vessel extraction, blurred edge resolution, and irregular vessel localization. The key contributions of this paper are:

- (1) The retinal vessel segmentation dataset RVSD was constructed for model training. RVSD contains 420 retinal vessel images provided by Zhuzhou Sansanyi Aier Eye Hospital. With the help of doctors, we annotate the vessel part of the images with high precision to help the model get better training results.
- (2) We propose a SLOA-HetConv Adaptive Pooling and GFCReLU UNet with Channel-Aware Self-Attention (SLOA-HGC) algorithm to tackle the difficulties of retinal vessel segmentation. The design is as follows:
  - a. We design channel-aware self-attention (CAS) in the decoder, which adapts the channel dimension of the feature map by inserting a channel-aware self-attention module before each bilinear interpolation to enable the dynamic extraction of microfine blood vessels.
  - b. We propose heterogeneous adaptive pooling (HAP) placed after the decoder to improve the fusion and representation of multi-scale features, capturing contextual information at different scales while maintaining spatial resolution, thus achieving segmentation and localization capabilities for blurred edge texture.
  - c. We optimize the output convolutional layer and design ghost fully convolutional Rectified Linear Unit (GFCReLU) to process the output feature maps by extracting rich and redundant features respectively, output two parts of the feature maps using  $1 \times 1$  convolution and pointwise convolution. And finally, by splicing the two portions of the feature map that result from distinct convolutional processes, the network can accurately localize and extract irregular blood vessel.
  - d. Sparrow-Integrated Lion Optimization Algorithm (SLOA) is proposed to replace the RMSprop optimizer to optimize the parameters of network training, and by combining the sparrow random update to increase the exploration range of the optimizer and the annealing mechanism to stabilize the exploration and utilization, to achieve finer parameter tuning and to improve the stability and segmentation accuracy of the model.
3. The SLOA-HGC proposed in this paper achieved 79.64% MIoU and 94.73% ACC on the self-constructed dataset RVSD, and the method can effectively extract microfine blood vessels and blur the edge texture, and can effectively localize and segment blood vessels with irregular distribution. And the training and inference durations are 4341 s and 3.6 s, respectively. In short, the method can segment retinal blood vessels quickly and accurately, which provides a reference for the application of deep learning methods in retinal segmentation.

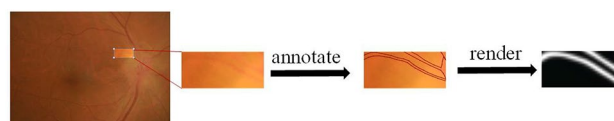
## Materials and methods

### Data acquisition

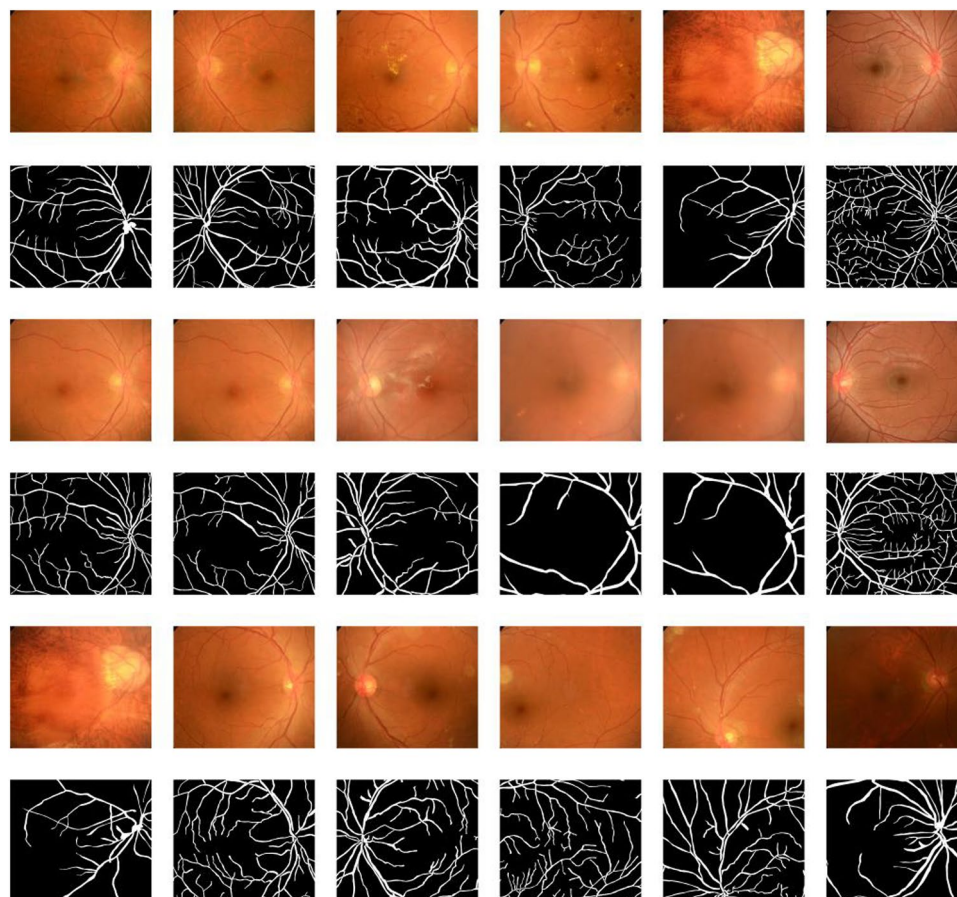
Our research work is based on the retinal vessel dataset, which was provided by Zhuzhou Sansanyi Aier Eye Hospital consisting of 420 retinal vessel images, of which 378 were used as the training set and 42 as the test set (Ethical approval number: Zhuzhou Sansanyi Eye Hospital, 2022-06-01, 202206). The resolution of each image was  $1620 \times 1400$ , and we zoomed in on each image using Labelme software with the help of a physician, manually labeled the vascular sites in the images, and then assessed the accuracy of the labeling based on the criteria used to manually label the images. We stored each labeled image in JSON format and converted it to PNG format. As shown in Fig. 2, during the labeling process, we encountered some cases of blurred edges of retinal vessels, which were often caused by uneven lighting or too thin vessels, etc. To solve this problem, we invited a professional ophthalmologist to assist us in judging the edge position and correctly labeled according to the doctor's instruction. This ensured the accuracy and consistency of our dataset. The final partial labeling map generated is shown in Fig. 3.

### Data augmentation

To extract effective features and avoid overfitting, deep learning neural network models require many images for training. Therefore, the original dataset needs to be augmented. The data enhancement in this paper includes



**Fig. 2.** Example of manual labeling process.



**Fig. 3.** Partial labeling legend.

(1) perspective, including performing 70% and 130% resizing; (2) flipping, including up and down flipping and left and right flipping; and (3) contrast adjustment, including adjusting to 70% and 130% of the original contrast. Taking an image in RVSD as an example, the enhanced image is shown in Fig. 4. By using the above data augmentation, different image qualities are simulated, so that it can handle the low contrast, motion blur, lighting changes and other disturbances existing in the actual shooting, to improve the segmentation robustness.

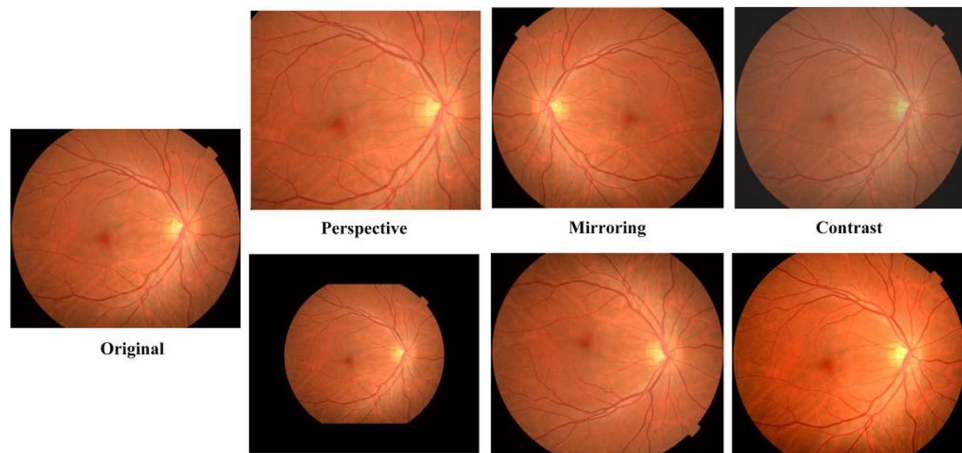
### SLOA-HGC

Here we introduce SLOA-HGC, a new UNet variant to segment retinal vessels, as shown in Fig. 5. The network's encoder uses downsampling to capture the global vessel information. The decoder alternates CAS with bilinear interpolation to adaptively weight and restructure the dimensions of the channels in the feature maps. Then, we feed the images of different scales into HAP to fuse the multi-scale feature information. Subsequently, we use GFCReLU to map the last feature map of the expansion path to the target categories and obtain the pixel-wise predictions. Finally, SLOA is used to optimize the network training to enhance the robustness and segmentation accuracy of the network. Each module is described in detail below.

#### *Channel-aware self-attention (CAS)*

We can enhance the network's performances by using the attention mechanism<sup>34–36</sup>. As well, it able the network to focus adaptively on the vessel information from various directions. Early SE<sup>37</sup> implemented attention in the channel dimension, adaptively learned the weights of each channel, and weighted each feature according to its importance. This not only improves the performance of the network, but also shows good robustness. The subsequent CCNET<sup>38</sup> used a criss-cross attention module on top of the SE adaptive learning of each channel





**Fig. 4.** Three types of data enhancement for retinal vascular images.

weight, and demonstrated that the combination of the two can show even better effectiveness. Instead of channel attention that emphasizes the features, criss-cross attention captures the contextual information of all pixels along the cross-paths, which ultimately captures the dependencies of the whole map through further loop operations.

For microfine blood vessels, the vessel color is usually similar to the retinal background. This often causes the segmentation network to mistake the retinal background for microfine blood vessels. At this point, if only channel attention (like SE) or spatial attention is used, it often leads to unsatisfactory segmentation because of the insufficient contextual information. Criss-cross attention in CCNET solves this problem and has good results. As mentioned in the introduction, we mentioned channel-aware self-attention, while the Parnet block<sup>39</sup> has a similar structure that includes both channel and spatial information. This lets the Parnet block learn the relation among different channels and spatial locations. Such relation learning is crucial for microfine blood vessels segmentation.

To address the challenge of segmenting microfine blood vessels, we have adopted an approach that combines cross attention, channel attention, and spatial attention. By cascading criss-cross attention and Parnet block, we can extract and fuse blood vessel features from different directions comprehensively. This enables a more complete retinal vessel segmentation, and we name the improved attentional mechanism CAS, as shown in Fig. 5c. The CAS module consists of two main parts:

(1) Feature extraction.

Suppose the input retinal vessel feature map, where  $C$  is the number of input channels,  $H$  and  $W$  are input feature map heights and widths, after three  $1 \times 1$  convolutional layers, it is filtered to generate three feature maps  $Q$ ,  $K$  and  $V$ .  $Q$ ,  $K$  undergoes the Affinity operation together with  $V$  undergoes the Aggregation operation as follows:

① Affinity works as follows:

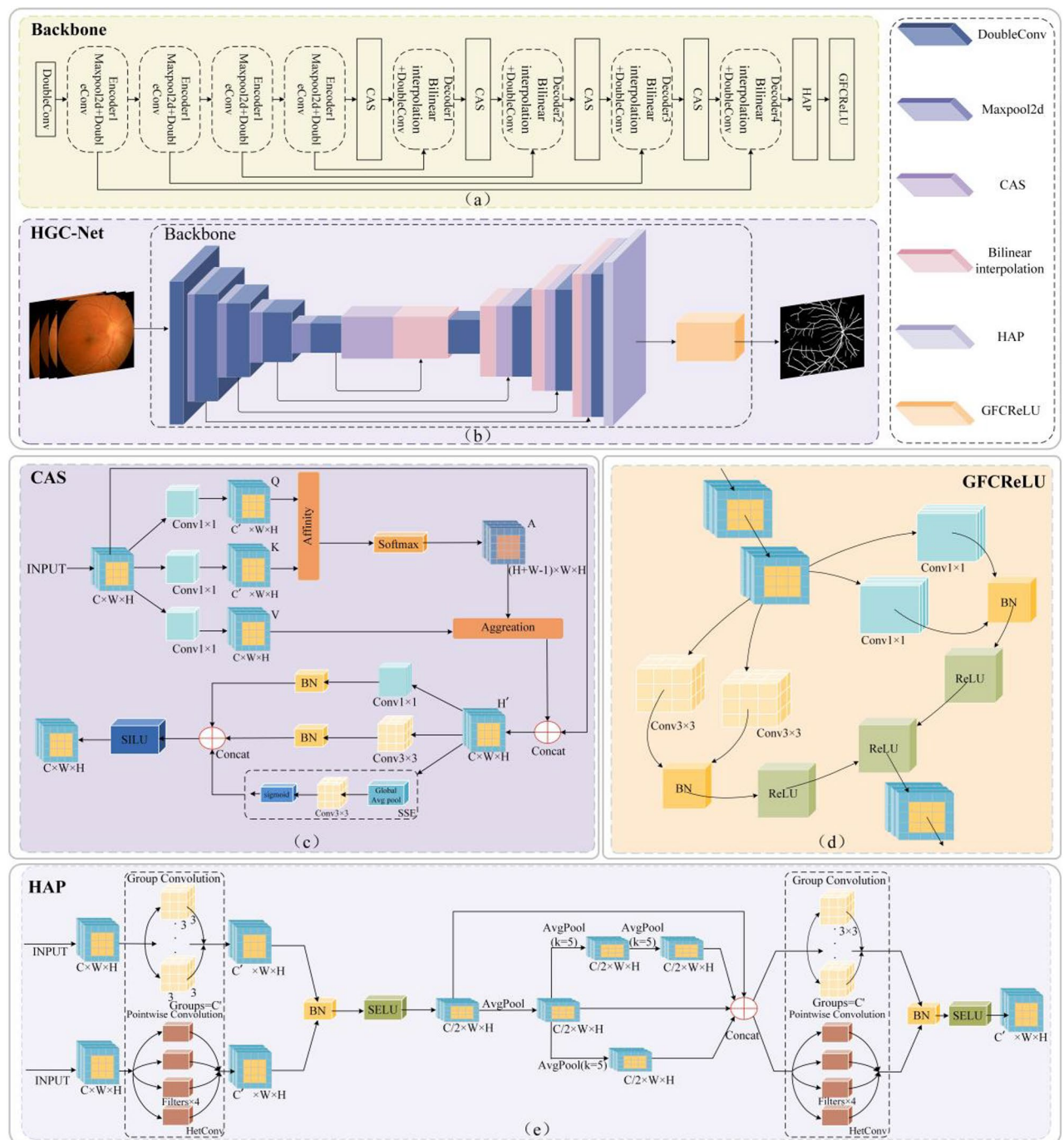
First, the feature map generation  $Q$  and  $K$  are reshaped into feature matrices, then a matrix multiplication will be performed on  $Q$  and  $K$  to acquire a matrix  $A'$ , where each element  $A'[i,j]$  denotes the similarity between the  $i$ th pixel and the  $j$ th pixel. Next, a softmax operation is performed on  $A'$  such that the total of the elements in each row is 1. This results in a normalized matrix  $A$ . Finally, a reshape operation is performed on  $A$  to transform it into a feature map  $X$ .

② Aggregation works as follows:

A linear variation of  $X$  is performed to obtain  $g(X) \in R^{C' \times N \times W}$ , given the input feature map  $X \in R^{C \times H \times W}$ , where  $C'$  is the number of channels after dimensionality reduction, a reshaping operation is performed on  $g(X)$  to obtain  $g(X) \in R^{C' \times N}$ , where  $N' = W \times N$  is the number of pixels, and then output the matrix product of the attention weight matrix and the feature matrix.

(2) Feature weight update.

The input to the Parnet block is the feature matrix output from CCNET, which consists of three parallel substructures, SSE (Skip-Squeeze-and-Excitation),  $1 \times 1$ Conv + BN, and  $3 \times 3$ Conv + BN. After inputting the feature matrix, the Parnet block lets each partial feature be weighted according to the relationship of the other partial features, and the features are transformed with attention in terms of channel and space, and finally the outputs of the three parallel substructures are spliced, and the SILU activation function is traversed to obtain a new weighted feature vector.



**Fig. 5.** SLOA-HGC structure.

Through the above two operations, the three attention mechanism transformations, namely channel attention, cross-attention, and spatial attention were employed to fuse different levels of feature information, and enhance the network segmentation performance for microfine blood vessels.

#### *Heterogeneous adaptive pooling (HAP)*

In retinal vessel segmentation, vessel edge blurring can affect the segmentation results. Vessel edge blur refers to the presence of a certain degree of transition region between the vessel and the surrounding tissue, making the vessel boundary blurred. Although the common Maxpool operation has been able to capture regular-scale vessel features, the operation only retains the maximum value in each sliding window, resulting in the loss of important feature information, which affects the localization and accuracy of segmentation. In addition, the common  $3 \times 3$  convolutional structure is unable to dynamically extract retinal vessel details and edge information, which may

limit the model's ability to segment retinal vessel edges. Therefore, we try to use more convolution and pooling operations to obtain retinal vessel edge features.

Pyramid pooling enriches the vessel information by performing multi-scale pooling operations on the input feature maps. This idea is inspired by the spatial pyramid pooling (SPP) proposed by He et al.<sup>40</sup>. They demonstrate the theoretical feasibility of spatial pyramid pooling for improving the CNN architecture. Spatial pyramid pooling enhances the segmentation quality by performing pooling operations at different scales on the feature maps, capturing multi-scale features, while ordinary pooling only extracts single-scale features with poor results. The grouped convolution operation adjusts the size and shape of the convolution kernel according to the different channels and spatial locations of the input feature maps, thus improving the discriminative and expressive abilities of the features. HetConv proposed by Singh et al.<sup>41</sup> applies grouped convolution to deep learning architectures by dynamically adjusting the size and shape of the convolution kernel and achieves good results. A HAP module is proposed to solve the problem of irregular blood vessel distribution. The HAP structure is shown in Fig. 5d. At first, the module applies a HetConv + BN + SELU layer for initial feature extraction and enhancement of the input feature map. Then, it downsamples the feature maps multiple times with three parallel Avgpool operations, the generated feature maps at different scales are spliced and fused. Finally, it processes the fused feature maps with a HetConv + BN + SELU. The retinal vessel edge features are effectively captured by pointwise and group convolution in HetConv and multi-scale Avgpool operations, which reduce the likelihood of the network misclassifying retinal vessel widths, and improve the segmentation accuracy of the network at the same time.

The ratio of  $3 \times 3$  convolutions to  $1 \times 1$  convolutions in the HetConv structure can be controlled by hyperparameterizing  $G$  (number of groups). Let  $h_i \times w_i \times c_i$  be the input size of the eye vessel feature map in the standard convolutional layer, where  $h_i$  is the height of the input feature map,  $w_i$  is the width of the input feature map, and  $c_i$  is the number of input channels. Consider  $X$  is the output feature map,  $h_0$  be its height,  $w_0$  be the input feature map's width, and  $c_0$  be the input channels' number. Apply a filter of size  $K \times K \times C$ , where  $x$  is the kernel size, so the total computational cost of this layer is:

$$FLOPs = h_0 \times w_0 \times c_0 \times K \times K \quad (1)$$

From Eq. (1), assuming the input image size is fixed, the kernel size  $K$  and the feature map are two important factors that affect the size of the computational cost, and HetConv reduces the computational cost by designing the convolutional operation. It defines the  $G$  part,  $1/G$  part of the whole kernel is  $K \times K$  and the rest  $(1 - 1/G)$  is  $1 \times 1$  size. In the filter of HetConv with  $G$ -part, the computational cost of the  $K \times K$  kernel is as follows:

$$FLOPs = \frac{h_0 \times w_0 \times c_0 \times c_i \times K \times K}{G} \quad (2)$$

The cost is reduced by a factor of  $G$ , since there are only  $c_0/G$  kernels of size  $K \times K$ . In addition, HetConv can adapt to different sizes of input feature maps, with pointwise and grouped convolutions increasing the diversity and complexity of the feature maps, which ensures that the model correctly segments the edges of fuzzy blood vessels.

#### *Ghost fully convolutional Rectified Linear Unit (GFCReLU)*

The distribution of retinal blood vessels is extremely irregular. This makes it difficult for traditional models to extract them accurately, and leads to misdetection and omission of pixels representing the vessels in the results. If the fusion of these features is insufficient, the subtle information of the blood vessels will be lost; this loss of information affects both the performance of the segmentation process and the accuracy of network segmentation.

The traditional approach to feature extraction is to perform convolutional operations on every channel of the input feature map with multiple convolutional kernels. However, in deep networks, this leads to a large number of parameters and computations, as well as many rich or redundant feature maps. Therefore, several studies have proposed model compression methods: pruning, quantization, knowledge distillation, etc. However, these methods have drawbacks such as intricate model design and training challenges, despite their ability to lower the parameter count. Han et al.<sup>42</sup> proposed a lightweight network, GhostNet, which employs a number of cheap linear transformations to produce various feature maps that can fully expose the inherent feature information for precise localization and segmentation of blood vessels. Although GhostNet uses fewer convolutional kernels and cheaper linear transformations, it can lead to excessive correlation between features, which results in less robustness of the features.

Therefore, we adjusted the structure of the output convolutional layer. Replacing the original  $1 \times 1$  convolution with GhostConv, we create the input retinal vessel feature map  $\times$  first and obtain the initial feature map  $\times 1$  by  $1 \times 1$  convolution,

$$(batch\_size, in, height, width) \xrightarrow{Conv1 \times 1 \text{ BNReLU}} x_1: (batch\_size, in\_channels, height, width) \quad (3)$$

$$inchannels = match.ceil(oup/ratio), ratio = 2$$

The remaining feature maps are then obtained using the cheaper PointwiseConv  $\times 2$ ,

$$(batch\_size, inp, height, width) \xrightarrow{PWConv\ BNReLU} x_2: (batch\_size, new\_channels, height, width) \quad (4)$$

$$new\_channels = inchannels \times (ratio - 1), ratio = 2$$

The final step was to merge the two parts of the feature maps together, and then based on the diversity and selectivity of the GhostConv convolution, we decided to add the ReLU function on top of GhostConv to obtain the final retinal vessel output feature map Out (Eq. (5)).

$$x_1, x_2 \xrightarrow{Conv1 \times 1, slice} Out = ReLU[x_1, x_2] [:, :oup, :, :] \quad (5)$$

The output convolutional layer operation, as described above, enhances its nonlinearity and sparsity, and also the robustness of the features. It is very effective for the network to accurately localize and segment irregular blood vessels, which we named GFCReLU, and the structure is shown in Fig. 5e.

The ability of the network to localize retinal vessels is enhanced by further processing the retinal vessel feature map through the two operations described above.

#### *Sparrow-Integrated Lion Optimization Algorithm (SLOA)*

In order to increase the stability of the model during training and improve the segmentation accuracy, we propose Sparrow-Integrated Lion Optimization Algorithm (SLOA) to optimize the parameters of the SLOA-HGCNet segmentation model during training. The dynamic changes of parameters (e.g., learning rate, momentum coefficients, etc.) during training affect the robustness and accuracy of the model; however, traditional parameter tuning strategies may not be flexible enough in some cases, especially when dealing with complex features such as retinal blood vessels. Therefore, we choose the more robust Lion algorithm<sup>43</sup> as the basis and propose SLOA for parameter optimization. At the beginning of training, SLOA starts with a higher initial learning rate to accelerate the initial global search process, and gradually reduces the learning rate through the annealing mechanism as the training advances to achieve stable convergence. This adaptation not only accelerates the model to approach the global optimal solution, but also helps to refine the parameter tuning in the later stages of training, thus avoiding oscillations or deviations that may be caused by too large learning step sizes. In addition, in order to increase the parameter exploration space of the optimizer, SLOA introduces a unique sparrow updating mechanism<sup>44</sup>, which introduces random perturbations at each update interval to enhance the exploration capability of the model and prevent it from falling into local minima. In this way, the SLOA optimizer provides a more robust training method for deep learning models.



---

**Input:**

- params: Iterator for the parameter to be optimized
- lr: learning\_rate
- weight\_decay: Coefficients used calculate the exponential moving average of the gradient and its squared value
- sparrow\_factor: Coefficients for the sparrow update term
- anneal\_rate: Learning rate annealing coefficients for annealing mechanisms
- update\_interval: Frequency of implementation of the sparrow update

**Output:**

- Loss: Loss value
- Other parameters:
  - global\_step: global step counter

**Pseudocode:**

```

def step (self, closure=None)
    loss=None
    if closure is not None:
        with torch. enable. grad():
            self.global_step+=1
            self. anneal_sparrow_factor()

    for group in self.param_groups:
        for p in group ['params']
            if p.grad is None
                Continue

    #Use the original lion formula:
    Weight decay:
        p.data.mul_(1-group['lr']*group['weight_decay'])
    Gradient update:
    Initialization status:
    if len(state)=0:
        state['exp_avg']=torch.zeros_like(p.data)
    Update Momentum:
        exp_avg=state['exp_avg']
        exp_avg.mul_(beta 1).add_(grad, alpha=1-beta 1)
    Application updates:
        p.data.add_(exp_avg.sign(), alpha=-group['lr'])
    Momentum decay:
        exp_avg.mul_(beta 2).add_(grad, alpha=1-beta 2)
    Sparrow update (every update_interval step):
        if self.global_step % self.update_interval == 0:
            for p in group['params']:
                sparrow_update=group['sparrow_factor']*np.random.randn(*p.data.size())
                sparrow_update_tensor=torch.from_numpy(sparrow_update).to(p.data.device)

                p.data.add_(sparrow_update_tensor, alpha=group['lr'])

    Annealing mechanism:
    For group in self.param_groups:
        group['sparrow_factor']=group['sparrow_factor']*(1-self. anneal_rate
    self.global_step)
  
```

**#Usage:**

Create a SLOA object with specified parameters

SLOA=SLOA (lr, beats, weight\_decay, sparrow\_factor, anneal\_rate, update\_interval)

---

#### Algorithm: Sparrow-Integrated Lion Optimization Algorithm (SLOA).

After training begins, SLOA first initializes the parameters, validating the learning rate and the coefficients used to compute the running average of the gradient and gradient squared. To prevent training overfitting, a weight decay operation is first performed, penalizing large weight values and encouraging the model to learn smaller weights. Then a momentum update operation is performed, which is applied to the real-time parameter updates by calculating new exponential moving averages, with the goal of accelerating learning and reducing parameter oscillations. Next, by combining current and past gradient information, momentum decay update is performed to smooth the gradient, accelerate convergence, and adjust the current gradient update, the algorithm

is able to cope with various challenges in the training process more effectively. The above process is represented as follows:

$$p = p \times (1 - lr \times weight\_decay) \quad (6)$$

$$\exp\_avg = \exp\_avg \times beta1 + grad \times (1 - beta1) \quad (7)$$

$$p = p - lr \times sign(\exp\_avg) \quad (8)$$

$$\exp\_avg = \exp\_avg \times beta2 + grad \times (1 - beta2) \quad (9)$$

In order to increase the explorability of the network, we spaced out the sparrow updates in the optimizer, after every 10 epochs, the optimizer simulates the natural behavior of the sparrow and performs a random perturbation on the parameters, and this periodic random perturbation is achieved by multiplying the set `sparrow_factor_t` with a Gaussian random number, which has the following mathematical expression:

$$p = p + lr \times sparrow\_factor\_t \cdot N(1 - t) \quad (10)$$

Finally, considering that the increase in computation brought about by the continuous application of the sparrow algorithm throughout the training process tends to affect the efficiency of model training, we integrated the annealing mechanism<sup>45</sup>, whose core idea is borrowed from the simulated annealing process in physics, where annealing is a heat treatment process that reduces the internal defects of a material by gradually lowering the temperature of the material so as to achieve a more stable state. In SLOA, we apply the annealing mechanism to gradually adjust the optimization of `sparrow_factor_t` to improve the efficiency and quality of finding the optimal solution. Our annealing mechanism achieves this by gradually decreasing the value of `sparrow_factor_t`. In the early stage of training, its higher value of `sparrow_factor_t` allows for larger random perturbations, promotes extensive parameter space exploration, and helps the optimizer to jump out of the local optimal solution, while as the training progresses, the value of `sparrow_factor_t` gradually decreases, reducing the magnitude of the random perturbations, thus reducing the exploration intensity and allowing the optimizer to tune the parameters more carefully to find a more accurate optimal solution. The mathematical expression is as follows:

$$sparrow\_factor\_t = sparrow\_factor_{\{t-1\}} \times (1 - anneal\_rate \times global\_step) \quad (11)$$

The adaptive gradient adjustment capability of SLOA allows the algorithm to be optimized for different parameters and different training phases, which effectively solves the potential drawbacks of parameter optimization that may lead to low accuracy and slow convergence of the model in the segmentation phase. SLOA improves the stability of the training of the SLOA-HGCNet model, and enhances the accuracy and robustness of segmenting retinal blood vessel images.

## Results and analysis

This section experimentally verifies that SLOA-HGC solves the challenges posed by microfine blood vessels, blurred vessel edges, and irregular vessel distribution in the retinal vessel segmentation task, which demonstrates the superiority of the model. The other subsections are divided into (1) describing the experimental environment and setup, including the hardware and software environments, as well as the setup of the training parameters; (2) evaluating the experimental metrics of the SLOA-HGC; (3) analyzing the performance of the SLOA-HGC and verifying the superiority of the model in this paper; (4) evaluating the validity of the individual modules of the SLOA-HGC, and determining the roles produced by the modules; (5) performing ablation experiments on the SLOA-HGC ablation experiments to validate the effectiveness of the method proposed in this paper; (6) Compare SLOA-HGC with other deep learning network models and visualize the detection results of SLOA-HGC to analyze the advancement of SLOA-HGC intuitively and to prove that SLOA-HGC outperforms other methods in retinal vessel segmentation tasks; (7) Compare the performance of SLOA-HGC on the public datasets DRIVE, CHASEDB1 and HRF for generalization experiments to demonstrate the generalization ability of the SLOA-HGC model.

## Experimental environment and training details

In order to ensure that the results of SLOA-HGC are not disturbed by different experimental environments, we conducted all the experiments in this paper in a unified hardware and software environment. The RTX 3080 GPU and 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU provided by AutoDL were borrowed as the main hardware devices for this experiment. The versions of Python, CUDA, and CUDNN had no effect on the experimental results as long as they matched the software and hardware. We have implemented SLOA-HGC on Pytorch 1.11.0. The specific hardware table is shown in Table 1.

The input image size is  $1620 \times 1400$  pixels, and we conducted experiments using a tenfold cross-validation method to properly assess the performance of the model. The 420 retinal vessel images were divided into training and test sets at a ratio of 9:1, and the dataset was augmented with perspective transformations, mirror flips, and contrast adjustments to avoid model overfitting before training. The experiment contains 200 training cycles, and the final training images obtained are 75,600, with a learning rate of  $1e-4$ . The optimizer uses SLOA, with `sparrow_factor_t` set to 0.01, annealing rate set to  $1 \times 10^{-5}$ , weight decay factor set to  $1e-8$ , and momentum 0.9.

Hardware environment	CPU	12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU
	GPU	RTX 3080
	RAM	375GB
	Video Memory	10GB
Software environment	OS	Ubuntu 20.04.4
	CUDA Toolkit	V11.3
	CUDNN	V8.2.4
	Python	3.8.10
	torch	1.11.0
	torchvision	0.14.1

**Table 1.** Experimental hardware and software parameters.

Considering the performance of the hardware device and the training effect, batch\_size is set to 1. Train the network from scratch without using pre-training weights.

### Analysis metrics

The retinal vessel segmentation task boils down to classifying each pixel as vascular or non-vascular. The vascular pixels, which are the target for segmentation, are the positive class, and the non-vascular pixels are the negative class. The comparison of the results of the segmentation algorithm with the true values can be obtained as True Positive TP, False Positive FP, False Negative FN, and True Negative TN in the confusion matrix. In this paper, we use evaluation metrics such as Mean Intersection and MIoU, Dice coefficients, Accuracy (ACC), Specificity, Sensitivity, FPS, and so on. Among them, MIoU reflects the overlap between the predicted segmentation region and the real segmentation region. The calculation formula is as follows:

$$MIoU = \frac{1}{k+1} \sum \frac{TP}{FN + FP + TN} \quad (12)$$

The ACC reflects the proportion of correctly predicted pixel points to all pixel points. The formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)$$

The Dice coefficient indicates the similarity between the predicted blood vessel pixels and the real blood vessel pixels. It is calculated as follows:

$$Dice = \frac{TP \cap Label}{TP \cup Label} \quad (14)$$

FPS indicates the average number of frames captured per second. The formula is as follows:

$$FPS = \frac{1}{T} \quad (15)$$

where T is the time required to process one frame of the image. To assess the effectiveness of the segmentation model, Hausdorff distance<sup>46</sup> (HD) was used to describe the boundary difference between the segmentation result and the real label, and the smaller Hausdorff distance indicates the higher ability of the segmentation model in judging the edge of the blood vessel. Its calculation formula is expressed as:

$$\begin{aligned} h(A, B) &= \max_{a \in A} \left\{ \min_{b \in B} \|a - b\| \right\} \\ h(B, A) &= \max_{b \in B} \left\{ \min_{a \in A} \|b - a\| \right\} \\ H(A, B) &= \max(h(A, B), h(B, A)) \end{aligned} \quad (16)$$

$\|a - b\|$  is the range of distances between point set A and point set B.

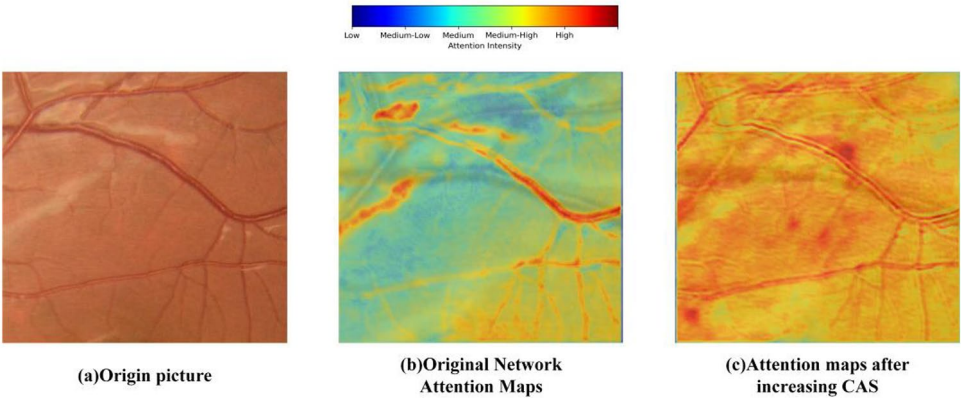
### Results and analysis

#### Validity of CAS

In this paper, we use CAS in a decoder alternately connected with bilinear interpolation, and to verify its validity, we choose to compare it with CBAM<sup>47</sup>, SE<sup>37</sup>, Coordinate Attention (CA)<sup>48</sup>, Criss-cross Attention (CCA)<sup>38</sup>,

Method	MIoU	ACC
Without attention	0.7506	0.9394
SE	0.7311	0.9276
CA	0.7540	0.9348
PB	0.7635	0.9405
CCA	0.7624	0.9417
CA + CCA	0.7601	0.9413
CA + PB	0.7603	0.9401
CAS	<b>0.7678</b>	<b>0.9436</b>

**Table 2.** Explore the best mechanisms for attention. Significant values are in bold.



**Fig. 6.** Comparison of attention graphs before and after adding CAS.

Parnet Block<sup>39</sup>. The experimental results are presented in Table 2 (All experimental results below are obtained on the test set).

CA, CCA and PB obtained better results in it after joining the UNet network alone, as shown by the experimental results. However, the SE attention mechanism cannot accurately segment retinal vessels, only considering attention in the channel dimension and ignoring spatial and positional information of the target in the image. Therefore, we performed a pairwise fusion of CA and CCA with PB to test its effect. The experimental results showed that since the fusion of CCA with PB (CAS) took into account the interaction of information on cross-attention, channel attention, and spatial attention at the same time, while the fusion of CA with CCA and PB suffered from the problem that it could only capture long-distance relationships in a single direction, resulting in insufficient ability to segment the microfine blood vessels, the CAS performed the best in blood vessel segmentation, achieving the highest MIoU and ACC (0.7678 and 0.9436), so we selected CAS as the attention mechanism for SLOA-HGC.

In addition, we also compared the attention maps of the networks before and after adding CAS, as shown in Fig. 6. It can be seen that the attention of the original network is focused on the larger and more obvious blood vessel regions, and the attention to the fine blood vessels is weaker. After adding CAS, the attention weights around the fine blood vessels are significantly higher, especially in the smaller branching blood vessel regions in the image, where the pseudo-colors change from blue (low attention) to yellow or red (high attention). This suggests that the CAS module is better able to capture these difficult-to-attend detailed features.

*Effectiveness of the HAP*

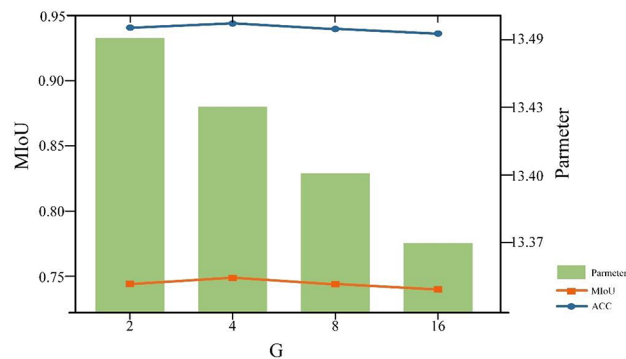
In this paper, we mentioned that the HAP is placed after the decoder to perform multi-scale feature extraction on the feature map. The effect of the hyperparameterization G in HetConv on the convolutional architecture is mentioned in section “Heterogeneous adaptive pooling (HAP)”. In order to select the most appropriate hyperparameterization G, we evaluated the experimental results for four different G values. The experimental results of adjusting the parameter G are shown in Fig. 7.

Through the experimental results, we found that the number of parameters of HetConv shows a decreasing trend as G increases, but when G exceeds 4, the MIoU and ACC of the segmentation results also seem to decrease, which indicates that the segmentation performance of HetConv is affected, so we adopt HetConv with G = 4 as the base structure of the HAP module.

To evaluate the overall performance of HAP, we inserted other pooling operations such as Maxpool, Avgpool, SPP<sup>40</sup>, SPPF at the same locations in the model, and the experimental results are shown in Table 3.

The experimental results show that HAP has the highest ACC (0.9469) and also outperforms Maxpool, Avgpool and SPP in terms of FPS, thanks to the structure of HetConv and the operation of multiple concurrent





**Fig. 7.** Experiments with tuning parameter G in HAP.

Method	ACC	FPS
MaxPool	0.9341	12.843
AvgPool	0.9353	12.914
SPP	0.9414	13.141
SPPF	0.9375	<b>14.255</b>
HAP	<b>0.9469</b>	14.237

**Table 3.** Comparison of different pooling operations. Significant values are in bold.

Method	MIoU	ACC
$1 \times 1$ Conv	0.7506	0.9394
GhostConv	0.7715	0.9403
GhostConv + SELU	0.7511	0.9375
GhostConv + SiLU	0.7641	<b>0.9427</b>
GFCReLU (GhostConv + RELU)	<b>0.7733</b>	0.9421

**Table 4.** Convolutional comparison experiment. Significant values are in bold.

average pooling, although SPPF has a slightly better FPS (14.255), its point convolution fails to extract and fuse features effectively, resulting in poor segmentation accuracy. Therefore, we choose HAP as the model's pooling layer.

#### Effectiveness of GFCReLU

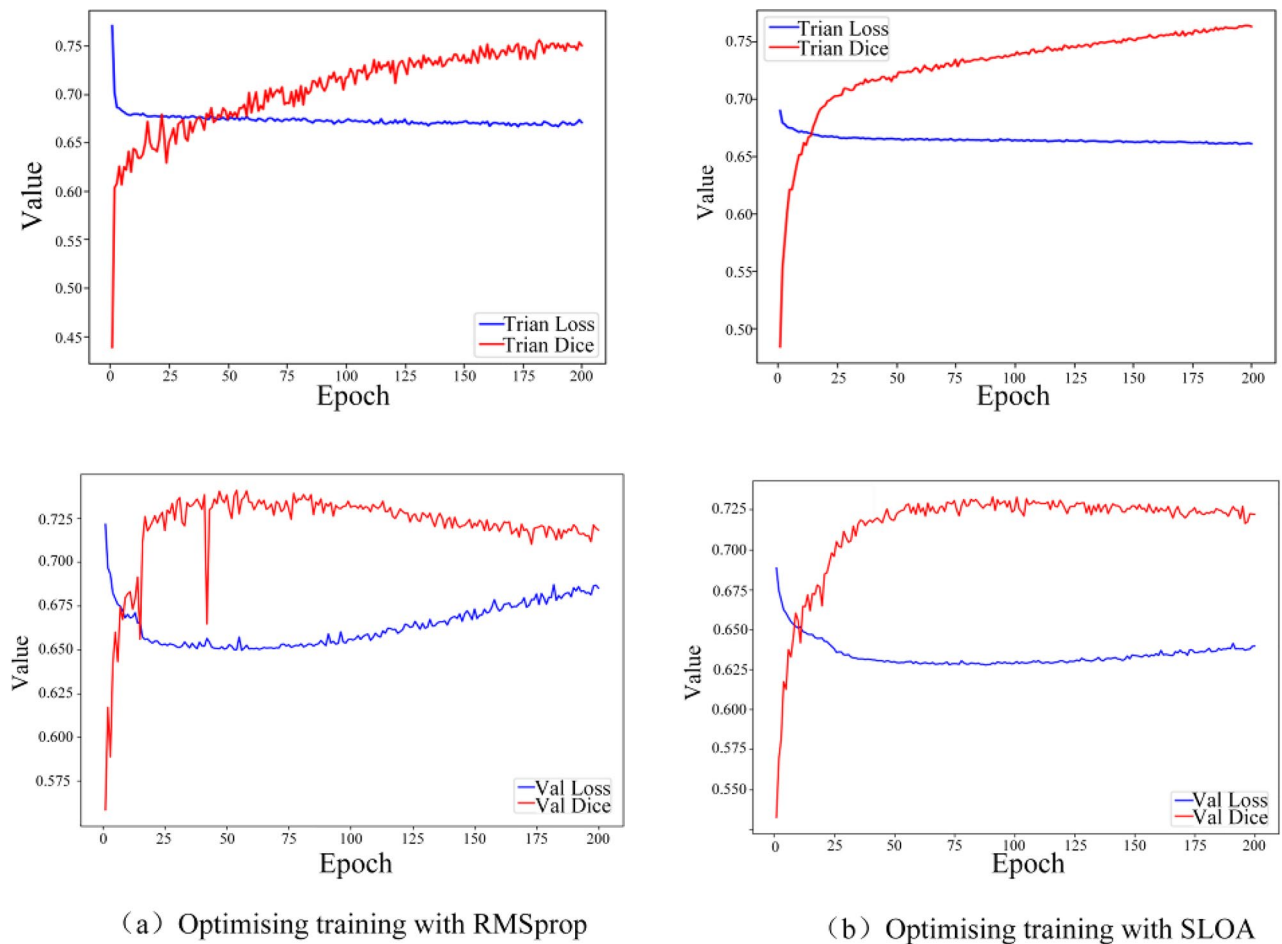
We replace the original  $1 \times 1$  convolution of the output convolution layer with GFCReLU to process the output feature maps. To determine the shape of the output convolution layer, we evaluate the performance of  $1 \times 1$  convolution, GhostConv, GhostConv + SELU, GhostConv + SiLU, and GFCReLU. Table 4 shows the results.

The experimental results show that the model performs well when we use GhostConv instead of the original convolution to process the output. Applying some activation functions after GhostConv can enhance the nonlinear capacity of the network and obtain better outcomes in segmentation performance. GhostConv + SiLU has the highest value of ACC, but it has a lower MIoU than GhostConv and GFCReLU ( $-0.74\%$  and  $-0.92\%$ ). GFCReLU achieves the best MIoU and has a slightly lower value of ACC than GhostConv + SiLU ( $-0.06\%$ ). Based on the MIoU and ACC results, we choose GFCReLU (GhostConv + RELU) to enhance the model segmentation ability.

#### Effectiveness of SLOA

In this paper, SLOA is used instead of RMSprop optimizer to optimize the network training in order to adjust the parameters more finely so as to make the network training more stable and to improve the network segmentation accuracy, and a comparison of the Loss and Dice change curves during its training process is shown in Fig. 8.

Furthermore, from the validation results, SLOA demonstrates a significant advantage over RMSprop in terms of both validation loss and validation Dice coefficient. The validation loss under SLOA exhibits a more consistent and steady decline compared to RMSprop, indicating a stronger resistance to overfitting during the training process. Simultaneously, the training Dice coefficient under SLOA achieves a higher peak (approximately 0.78) and maintains a relatively stable trajectory throughout the training epochs, underscoring its superior ability to generalize on data and deliver improved segmentation performance in practical applications.



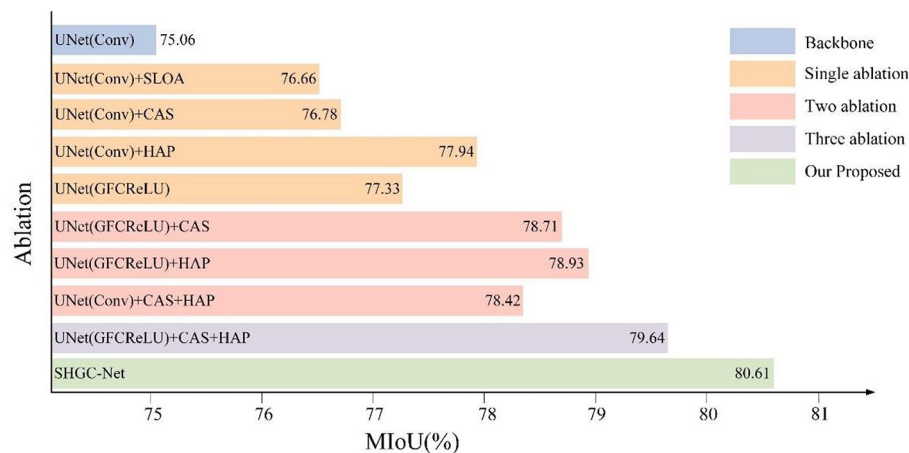
**Fig. 8.** Before and after using SLOA.

Method	MIoU	ACC
RMSprop	0.7506	0.9394
Adam	0.7432	0.9389
AdamW	0.7606	0.9433
PSO	0.7448	0.9384
Lion	0.7576	0.9419
SLOA without AM	0.7631	0.9420
SLOA	<b>0.7666</b>	<b>0.9440</b>

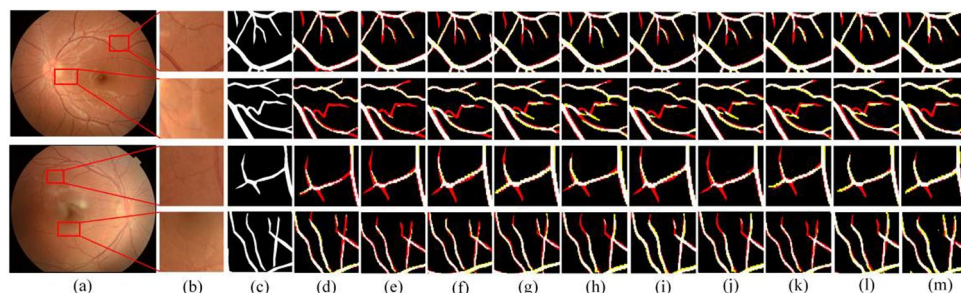
**Table 5.** Comparison of different optimizers. Significant values are in bold.

In this paper, we use SLOA instead of RMSprop optimizer in order to optimize the training of the network, in order to evaluate the overall performance of SLOA, we compare the performance of RMSprop<sup>49</sup>, Adam<sup>50</sup>, AdamW<sup>51</sup>, Lion<sup>43</sup>, PSO<sup>52</sup>, and SLOA without Annealing Mechanism (SLOA without AM), the experimental results are shown in Table 5.

The experimental results show that combining the two metrics, SLOA performs optimally on MIoU, reaching 0.7666, which indicates that it has the highest consistency between its predictions and the true labeling in the segmentation task. Meanwhile, SLOA is also the highest in accuracy at 0.9440, which means that it has the strongest ability to correctly classify pixels on the whole. Compared with the Lion optimizer, we introduce the sparrow update mechanism, an improvement that significantly increases the optimizer's exploration range in the parameter space, thus improving the accuracy of retinal vessel segmentation. In addition, by incorporating the annealing mechanism, we further enhance the robustness of the optimizer in parameter search, which in turn achieves improved segmentation accuracy. Based on these advantages, we choose SLOA as the optimizer for network training.



**Fig. 9.** Ablation experiments with SLOA-HGC.



**Fig. 10.** Visualization of ablation experiments on RVSD. (a) Original image; (b) Detailed view; (c) Ground truth; (d) UNet ( $1 \times 1$ Conv); (e) UNet ( $1 \times 1$ Conv) + SLOA; (f) UNet ( $1 \times 1$ Conv) + CAS; (g) UNet ( $1 \times 1$ Conv) + HAP; (h) UNet (GFCReLU); (i) UNet (GFCReLU) + CAS; (j) UNet (GFCReLU) + HAP; (k) UNet ( $1 \times 1$ Conv) + CAS + HAP; (l) UNet (GFCReLU) + CAS + HAP; (m) Ours (SLOA-HGC).

### Ablation experiments

To validate the effectiveness of SLOA-HGC, we performed ablation experiments on the proposed SLOA-HGC network on the RVSD dataset. Taking the UNet network as the baseline network, we used the control variable approach to gradually add CAS, HAP, and replace  $1 \times 1$ Conv with GFCReLU in it, and optimally trained the UNet and the UNet with CAS, HAP, and GFCReLU added by the SLOA optimizer. By comparing the differences in MIoU, we have analyzed the performance of each module and their importance to the network. In the visualization, yellow (False Positive) and red (False Negative) are used to represent over-segmentation and missed segmentation respectively. The corresponding experimental results are shown in Figs. 9 and 10.

Our experimental results show that our proposed HAP can significantly improve the segmentation performance of the network compared to traditional pooling methods, and HAP can improve MIoU by 1.76% on average under different conditions, which proves that heterogeneous adaptive pooling can be adapted according to the features of different scales for better extraction of vascular features. In addition, we also used GFCReLU and CAS to optimize the network structure, although their improvement in network performance is relatively small, 1.6% and 1.07%, respectively, they are still effective, and the optimization training by SLOA on the initial and the network with the addition of HAP, CAS, and GFCReLU also improves the network, with an improvement of 1.6% and 0.97%, respectively. In summary, each module of SLOA-HGC has a positive effect on the MIoU of the model segmentation, which verifies that our proposed replacement of CAS, HAP, GFCReLU and SLOA is beneficial.

### Comparing different models

In order to analyze the performance of SLOA-HGC in more depth, we compared the performance of old and new retinal vessel segmentation methods in the same testing setting. The experimental results are shown in Table 6 and Fig. 11.

To further analyze the performance advantages of our proposed SLOA-HGC, we compare it with the classical encoder-decoder networks and some advanced networks proposed in recent years. We found that in the classical encoder-decoder network, SA-UNet performs feature extraction on retinal blood vessels by replacing the initial convolution with the integrated DropBlock and batch structured dropout convolution blocks, and introducing a spatial attention module. However, SA-UNet only considers spatial attention but not channel

Model	MIoU	Dice	ACC	HD	FPS
UNet <sup>13</sup>	0.751 ± 0.023	0.724 ± 0.017	0.940 ± 0.007	339.765 ± 3.210	13.848 ± 0.106
SA-UNet <sup>17</sup>	0.752 ± 0.013	0.726 ± 0.024	0.939 ± 0.004	360.389 ± 5.116	13.924 ± 0.217
UNET++ <sup>53</sup>	0.767 ± 0.019	0.748 ± 0.022	0.943 ± 0.010	329.644 ± 3.107	13.975 ± 0.901
RIMNet <sup>21</sup>	0.776 ± 0.011	0.761 ± 0.017	0.944 ± 0.011	339.801 ± 7.230	<b>15.461 ± 0.723</b>
SegR-Net <sup>54</sup>	0.778 ± 0.003	0.763 ± 0.009	0.945 ± 0.010	343.850 ± 2.210	14.237 ± 0.509
BTS-DSN <sup>22</sup>	0.779 ± 0.009	0.764 ± 0.048	0.945 ± 0.003	342.891 ± 5.310	12.843 ± 0.145
U <sup>2</sup> Net <sup>55</sup>	0.784 ± 0.007	0.771 ± 0.031	0.946 ± 0.005	353.866 ± 6.380	10.325 ± 0.137
SLOA-HGC	<b>0.806 ± 0.011</b>	<b>0.790 ± 0.030</b>	<b>0.948 ± 0.007</b>	<b>319.985 ± 3.491</b>	12.541 ± 0.225

**Table 6.** Comparing the performance of SLOA-HGC with other models. Significant values are in bold.

attention, which again leads to lost feature information of some microfine blood vessels. UNet++ introduces a convolutional layer over the jump connections to increase the dense connections. Although that reduces the semantic gap between the encoder and decoder feature maps, the model tends to overlook multiscale features during the training process. U<sup>2</sup>Net, based on UNet, uses seven nested UNet for feature fusion over jump connections, but its drawbacks are also obvious, the computational volume of the network increases significantly and the computational speed decreases significantly. SegR-Net employs dense multi-scale fusion to produce segmentation masks, but neglects the weights and importance of different features, leading to low segmentation accuracy. In contrast, our proposed HGC network has obvious advantages in terms of accuracy and only slightly lags behind in terms of computational speed. Among other networks, the speed advantage of RIMNet and BTS-DSN suggests that they are more suitable for practical application environments. However, they are not accurate enough. SLOA-HGC outperforms the networks of recent years in terms of overall performance. Among them, channel-aware self-attention and heterogeneous adaptive pooling can better extract the features of microvessels and edge parts in retinal blood vessels, and utilize GFCReLU for effective feature fusion, and finally optimize the training by SLOA to enhance the network robustness. Finally, the experimental accuracy reaches 94.83% and the FPS stays at 12.541. Although the speed is slower than UNet, it meets the real-time criterion for retinal vessel segmentation. This demonstrates that our proposed SLOA-HGC is able to improve the accuracy of retinal vessel segmentation while maintaining the speed.

A comparison of segmentation experiments between our method and other models on five typical images from the RVSD dataset is shown in Fig. 10. The figure shows that our network model segmented retinal blood vessels better than the other networks. For example, UNet, SA-UNet, and BTS-DSN or inaccurately localized the microfine blood vessels. In conclusion, by comparing the segmentation index parameters and the segmentation results, we can see that SLOA-HGC performs well in RVSD.

Compared to other models, our SLOA-HGC network has the following advantages: (a) using CAS with bilinear upsampling (Upsample) as the decoder part of the network, the features are extracted emphatically in cross-attention, channel-attention, and spatial-attention to improve the accuracy of the segmentation of the microfine blood vessels; (b) adding HAP in the tail part of the decoder, the features are subjected to multiscale fusion to filter out some interfering information and make the blood vessel edge expression clearer; (c) using GFCReLU to increase the feature fusion capability of the network and localize the blood vessel part more accurately. (d) SLOA is used instead of RMSprop for optimization training to increase the exploration of parameter space and improve network stability and segmentation accuracy. (e) The self-constructed dataset and data enhancement approach used eliminates some blurred and low-quality images, which facilitates model training.

Generalization experiment

The generalization experiment is a crucial metric in retinal vessel segmentation experiments, which can evaluate the performance of different segmentation methods on different data sets, as well as the sufficiency of their generalization ability and robustness. The generalization experiments can help us to deeply analyze the advantages and disadvantages of the segmentation methods in dealing with different image qualities, resolutions, illumination conditions, and other factors, as well as whether they can adapt to different variations in retinal structure and vessel morphology. The generalization experiments also offer some valuable insights that can help us improve the segmentation methods, enhance their accuracy and consistency across different datasets. To evaluate the generalization performance of our method, we chose three public datasets, DRIVE, CHASEDB1 and HRF, as our test subjects. They contain various retinal images and allow us to evaluate how our method performs under different conditions. The basic information of the two public datasets is shown in Table 7, and the experimental results on them are displayed in Tables 8, 9 and 10.

From the data in the table, it is obvious that in terms of segmentation accuracy for the three publicly available datasets, SLOA-HGC outperforms the segmentation models developed in recent years in a number of evaluation metrics, in only slightly inferior to the other models in terms of the background pixel segmentation ratio and Hausdorff distance (0.9858, 0.9910, and 501.277). However, compared to the other networks, the SLOA-HGC network model has a lower FPS metric, which means that the number of images detected per unit of time is lower, and the performance of the hardware needs to be taken into account in the future application of the network to actual retinal vessel segmentation. Taken together, SLOA-HGC has better performance than the current mainstream retinal vessel segmentation networks on the retinal vessel segmentation task, while maintaining no significant reduction in FPS.



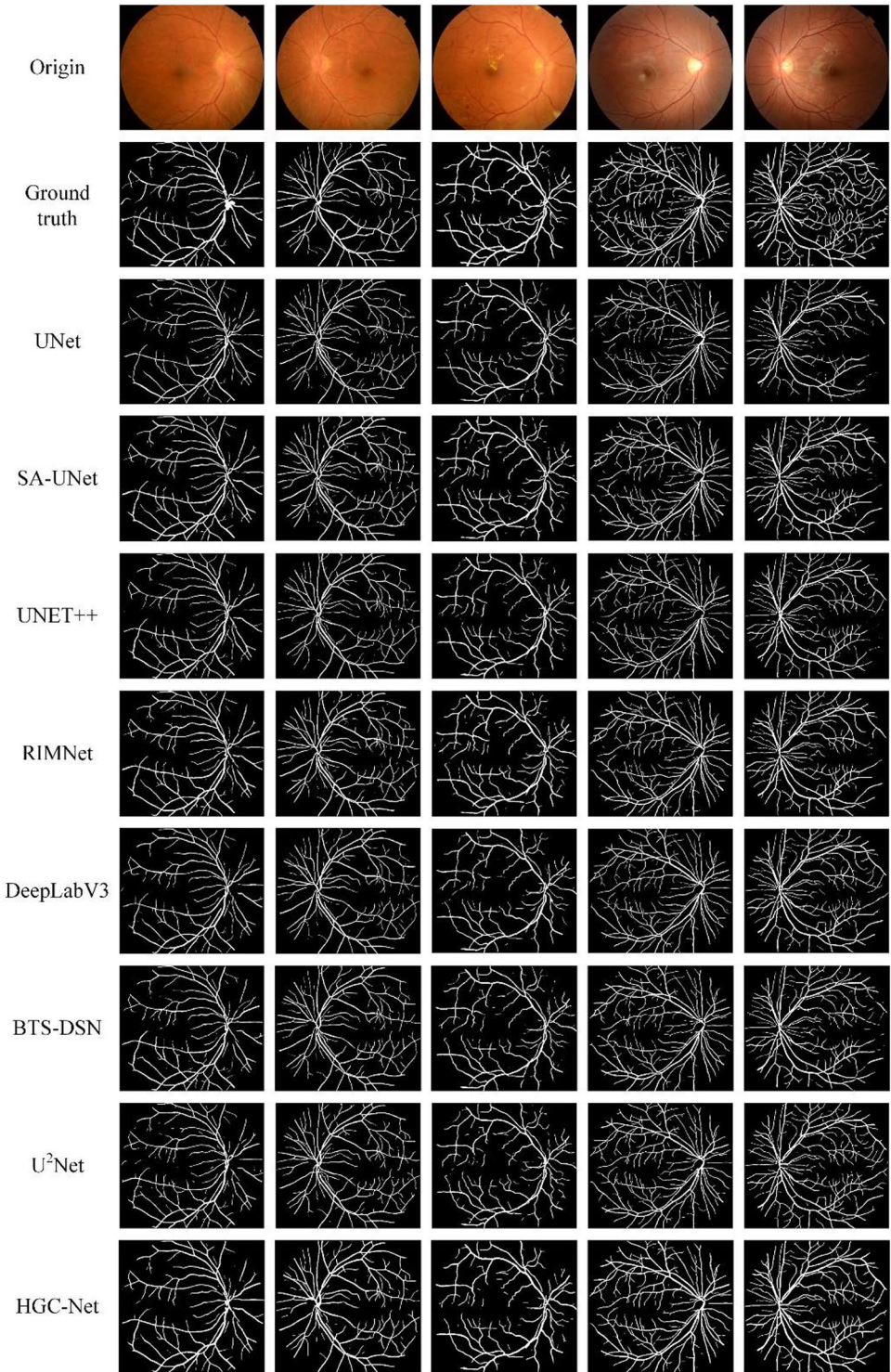


Fig. 11. Comparison of results of different methods.

Dataset	Quantity	Trian-test split	Resolution	Format	FoV mask
DRIVE	40	20–20	565×584	.tiff	√
CHASEDB1	28	20–8	999×960	.jpg	×
HRF	45	30–15	3504×2336	.jpg	√

Table 7. Overview of the three public datasets.

Model	MIoU	Dice	ACC	HD	FPS
UNet	0.720 ± 0.017	0.659 ± 0.007	0.952 ± 0.002	81.623 ± 4.383	12.799 ± 0.148
SA-UNet	0.723 ± 0.015	0.662 ± 0.010	0.952 ± 0.010	86.405 ± 5.639	13.619 ± 0.117
UNET++	0.751 ± 0.025	0.710 ± 0.022	0.955 ± 0.003	94.507 ± 7.181	11.853 ± 0.232
RIMNet	0.747 ± 0.009	0.701 ± 0.021	0.957 ± 0.002	87.038 ± 8.854	<b>14.265 ± 0.431</b>
SegR-Net	0.736 ± 0.031	0.686 ± 0.035	0.954 ± 0.007	81.209 ± 6.704	13.781 ± 0.337
BTS-DSN	0.731 ± 0.029	0.677 ± 0.027	0.953 ± 0.001	87.094 ± 6.188	13.761 ± 0.231
U <sup>2</sup> Net	0.746 ± 0.016	0.701 ± 0.019	0.955 ± 0.005	95.252 ± 7.346	9.754 ± 0.103
SLOA-HGC	<b>0.761 ± 0.027</b>	<b>0.725 ± 0.021</b>	<b>0.957 ± 0.004</b>	<b>75.538 ± 6.034</b>	10.757 ± 0.112

**Table 8.** Performance of different models on the DRIVE dataset. Significant values are in bold.

Model	MIoU	Dice	ACC	HD	FPS
UNet	0.741 ± 0.039	0.683 ± 0.030	0.964 ± 0.004	264.201 ± 4.383	12.833 ± 0.451
SA-UNet	0.751 ± 0.028	0.700 ± 0.005	0.965 ± 0.002	236.520 ± 5.639	11.807 ± 0.361
UNET++	0.752 ± 0.008	0.701 ± 0.016	0.9650 ± 0.001	253.966 ± 7.181	10.276 ± 0.120
RIMNet	0.758 ± 0.007	0.711 ± 0.006	0.966 ± 0.002	276.962 ± 8.854	<b>12.859 ± 0.212</b>
SegR-Net	0.760 ± 0.018	0.708 ± 0.008	0.965 ± 0.003	249.776 ± 6.704	10.309 ± 0.489
BTS-DSN	0.748 ± 0.014	0.695 ± 0.025	0.965 ± 0.004	227.326 ± 6.188	12.761 ± 0.292
U <sup>2</sup> Net	0.744 ± 0.022	0.692 ± 0.025	0.961 ± 0.006	255.537 ± 7.346	9.295 ± 0.232
SLOA-HGC	<b>0.769 ± 0.027</b>	<b>0.728 ± 0.019</b>	<b>0.967 ± 0.003</b>	<b>218.765 ± 4.034</b>	10.476 ± 0.360

**Table 9.** Performance of different models on the CHASEDB1 dataset. Significant values are in bold.

Model	Miou	Dice	ACC	HD	FPS
UNet	0.703 ± 0.005	0.626 ± 0.011	0.952 ± 0.002	767.217 ± 5.064	5.394 ± 0.267
SA-UNet	0.722 ± 0.020	0.658 ± 0.026	0.955 ± 0.008	802.238 ± 7.555	4.681 ± 0.389
UNET++	0.727 ± 0.018	0.666 ± 0.017	0.956 ± 0.007	576.168 ± 8.069	3.682 ± 0.264
RIMNet	0.707 ± 0.010	0.631 ± 0.017	0.954 ± 0.002	<b>462.498 ± 5.349</b>	<b>4.832 ± 0.209</b>
DeepLabV3	0.718 ± 0.019	0.657 ± 0.012	0.950 ± 0.005	614.745 ± 8.894	4.007 ± 0.205
BTS-DSN	0.717 ± 0.016	0.649 ± 0.028	0.955 ± 0.001	658.788 ± 9.313	4.598 ± 0.501
U <sup>2</sup> Net	0.735 ± 0.011	0.679 ± 0.023	0.958 ± 0.007	567.485 ± 8.611	2.575 ± 0.469
SLOA-HGC	<b>0.741 ± 0.024</b>	<b>0.689 ± 0.031</b>	<b>0.958 ± 0.007</b>	501.277 ± 7.102	3.264 ± 0.308

**Table 10.** Performance of different models on the HRF dataset. Significant values are in bold.

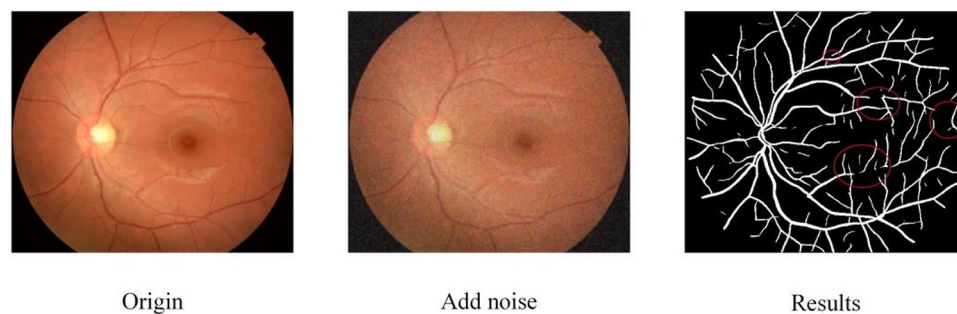
## Discussion and conclusions

### Discussion

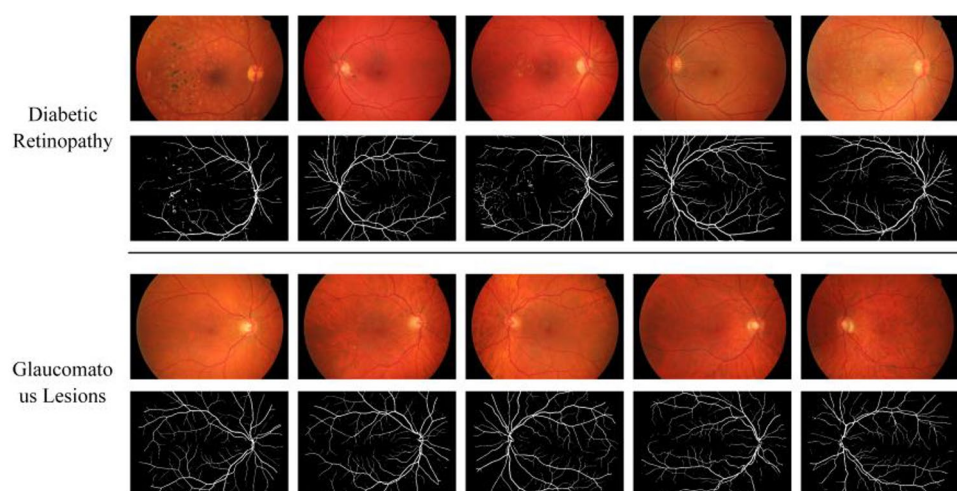
In this paper, a retinal vessel dataset provided by Zhuzhou Sansanyi Aier Eye Hospital has been established, which includes retinal vessel images with different topologies. The performance of the proposed SLOA-HGC for retinal vessel segmentation is verified by comparing several experimental sets. The SLOA-HGC is composed of three core modules, CAS, HAP, and GFCReLU, and optimized training using SLOA. This method aims to design a network with an adaptive mechanism to address the challenges of segmenting retinal blood vessels with microfine size, blurring edge texture, and irregular distribution. The comparison of the data in the table shows that the method proposed in this paper has good performance.

However, it is worth mentioning that although our model outperforms existing models in segmenting retinal blood vessels, there are still cases of segmentation errors in some special cases (e.g., adding noise to the prediction image). The segmentation error of SLOA-HGC with added noise is shown in Fig. 12. The noise interferes with the vessel boundary, making it hard for SLOA-HGC to segment accurately, causing segmentation errors. The noise also alters the pixel values, blurring or distorting the vessel edges. SLOA-HGC cannot differentiate them well, resulting in segmentation errors. The noise also affects the threshold selection of the network, mislabeling some non-vascular regions as vascular ones, causing segmentation errors.

And we selected 5 photos with diabetic retinopathy (diabetic retinopathy) and 5 photos with glaucomatous lesions (glaucomatous lesions) from the HRF dataset for segmentation prediction. These cases contain different degrees of lesions such as diabetic retinopathy, and retinopathy caused by glaucoma. Specific segmentation results are shown in Fig. 13. By showing the segmentation results of these challenging cases, we can better evaluate the performance and adaptability of the network in dealing with complex lesions.



**Fig. 12.** Effect of adding noise to an image on segmentation results.



**Fig. 13.** Two disease segmentation examples.

## Conclusion

Retinal vascular segmentation is a key technique in the diagnosis and treatment of various ophthalmic and cardiovascular diseases. It provides precise information about the vascular structure of the retina and helps physicians assess the severity and progression of the disease. However, retinal vascular segmentation faces several challenges, such as extraction of microscopic vessels, blurred edge texture, irregular vessel distribution, and unstable network training. To address these challenges, we propose a new U-shaped segmentation network, SLOA-HGC, which achieves better performance in retinal vessel segmentation.

- Ablation experiments showed that CAS, HAP, GFCReLU, and SLOA were more effective for retinal vessel segmentation with +1.72%, +2.68%, +2.27%, and +1.6% MIoU, respectively. With the same experimental setup, SLOA-HGC improved MIoU, DICE coefficient, and accuracy by 5.55%, 6.6%, and 0.88%, respectively, compared to UNet.
- Comparative experiments on our home-made RVSD dataset show that our MIoU is 80.61%, DICE coefficient is 78.97%, accuracy is 94.83%, and FPS is 12.541, which compares to U<sup>2</sup>Net, which is one of the best performers in the network in recent years, and has higher accuracy (+2.25%, +1.89%, +0.27%) and FPS is also ahead. The experimental results show that the method has obvious advantages in extracting microvessels and vessel edges as well as localizing blood vessels. It can be effectively applied to retinal vessel segmentation to aid in the diagnosis of ophthalmic diseases.
- Generalization experiments on the three mainstream public datasets DRIVE, CHASEDB1, and HRF show that the SLOA-HGC model outperforms existing models in several key evaluation metrics. The overall performance of the model is strong, although there are slight shortcomings in the background pixel segmentation ratio and Hausdorff distance. In addition, the results of the generalization experiments of the model emphasize its potential for practical applications, especially in important areas such as ophthalmic disease diagnosis.

In this study, we present an innovative retinal vessel segmentation network (SLOA-HGC). This approach efficiently extracts and integrates retinal vessel features from different scales and levels, significantly enhancing the consistency of same-category predictions. After extensive experimental validation, SLOA-HGC outperforms all comparative methods in key performance metrics such as MIoU, Accuracy (ACC), DICE coefficient, and



Sensitivity, demonstrating its superior segmentation capability and generalization. These features make SLOA-HGCNet a powerful tool to support the diagnosis of ophthalmic diseases, providing accurate and reliable auxiliary information for identifying a wide range of ocular diseases associated with retinal vascular structures. Going forward, we plan to apply SLOA-HGCNet to a wider range of medical image segmentation scenarios and pursue technological innovations to further enhance the robustness and computational efficiency of our approach.

## Data availability

We have uploaded the dataset at <https://github.com/ZhouGuoXiong/Retinal-blood-vessels-for-segmentation>. If you need to use this data, please contact [edalin@163.com](mailto:edalin@163.com).

Received: 12 March 2024; Accepted: 30 December 2024

Published online: 06 January 2025

## References

1. Frazao, L. B., Theera-Umpon, N. & Auephanwiriyakul, S. Diagnosis of diabetic retinopathy based on holistic texture and local retinal features. *Inf. Sci.* **475**, 44–66. <https://doi.org/10.1016/j.ins.2018.09.064> (2019).
2. Di Marco, E., Ciancimino, L., Cutrera, R., Bagnato, G. & Bagnato, G. F. A literature review of hypertensive retinopathy: systemic correlations and new technologies. *Eur. Rev. Med. Pharmacol. Sci.* **26**, 6424–6443 (2022).
3. Gao, S., Li, Y., Zhang, J., Zhang, Y. & Wang, X. Automatic arteriosclerotic retinopathy grading using four-channel with image merging. *Comput. Methods Programs Biomed.* **208**, 106274. <https://doi.org/10.1016/j.cmpb.2021.106274> (2021).
4. Khan, K. B., Khan, M. A. & Khan, S. A. A review of retinal blood vessels extraction techniques: challenges, taxonomy, and future trends. *Pattern Anal. Appl.* **22**, 767–802. <https://doi.org/10.1007/s10044-018-0740-8> (2019).
5. Orujov, F., Mammadova, S. & Hajiye, T. Fuzzy based image edge detection algorithm for blood vessel detection in retinal images. *Appl. Soft Comput.* **94**, 106452. <https://doi.org/10.1016/j.asoc.2020.106452> (2020).
6. Hartnett, M. E. Pathophysiology and mechanisms of severe retinopathy of prematurity. *Ophthalmology* **122**, 200–210. <https://doi.org/10.1016/j.ophtha.2014.07.050> (2015).
7. Kuppusamy, P., Basha, M. M. & Hung, C. L. Retinal blood vessel segmentation using random forest with Gabor and Canny edge features. In *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, pp. 1–4. <https://doi.org/10.1109/ICSTSN2022.2022.0001> (IEEE, 2022).
8. Dong, H. & Wei, L. Vessels segmentation base on mixed filter for retinal image. In *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 187–191. <https://doi.org/10.1109/CISP-BMEI50418.2020.00049> (IEEE, 2020).
9. Lam, B. S. Y. & Yan, H. A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields. *IEEE Trans. Med. Imaging* **27**, 237–246. <https://doi.org/10.1109/TMI.2007.906785> (2008).
10. Tuba, E., Mrkela, L. & Tuba, M. Retinal blood vessel segmentation by support vector machine classification. In *2017 27th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 1–6. <https://doi.org/10.1109/RADIOELEK.2017.7937588> (IEEE, 2017).
11. Wang, C., Xu, R., Xu, S. et al. DA-Net: Dual branch transformer and adaptive strip upsampling for retinal vessels segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2022*, pp. 528–538 (Springer Nature, 2022).
12. Chen, D. et al. PCAT-UNet: UNet-like network fused convolution and transformer for retinal vessel segmentation. *PLoS One* **17**(1), e0262689 (2022).
13. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, pp. 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28) (Springer International Publishing, 2015).
14. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965> (IEEE, 2015).
15. Li, G., Jin, D., Yu, Q. & Qi, M. IB-TransUNet: Combining information bottleneck and transformer for medical image segmentation. *J. King Saud Univ. Comput. Inf. Sci.* **35**(3), 249–258. <https://doi.org/10.1016/j.jksuci.2023.02.012> (2023).
16. Akbar, A. S., Fatchah, C. & Suciati, N. Single level UNet3D with multipath residual attention block for brain tumor segmentation. *J. King Saud Univ. Comput. Inf. Sci.* **34**(6, Part B), 3247–3258. <https://doi.org/10.1016/j.jksuci.2022.03.022> (2022).
17. Jin, L. 3AU-Net: triple attention U-net for retinal vessel segmentation. In *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, pp. 612–615. <https://doi.org/10.1109/ICCASIT50830.2020.00015> (IEEE, 2020).
18. Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B. & Fan, C. SA-UNet: Spatial attention U-Net for retinal vessel segmentation. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 1236–1242. <https://doi.org/10.1109/ICPR48806.2021.9412987> (IEEE, 2021).
19. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 5998–6008 (2017).
20. Dosovitskiy, A., Beyer, L., Kolesnikov, A. et al. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
21. Sultana, F., Sufian, A. & Dutta, P. RIMNet: Image magnification network with residual block for retinal blood vessel segmentation. In *2022 IEEE Region 10 Symposium (TENSYP)*, pp. 1–6. <https://doi.org/10.1109/TENSYP2022.2022.0001> (IEEE, 2022).
22. Guo, S. et al. BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation. *Int. J. Med. Inform.* **126**, 105–113. <https://doi.org/10.1016/j.ijmedinf.2019.03.006> (2019).
23. Puro, D. G. Pathophysiology of pericyte-containing retinal microvessels: roles of ion channels and transporters. In *Ocular Transporters in Ophthalmic Diseases and Drug Delivery: Ophthalmology Research* (ed. Tombran-Tink, J.) 127–137 (Humana Press, 2008). [https://doi.org/10.1007/978-1-59745-210-6\\_9](https://doi.org/10.1007/978-1-59745-210-6_9).
24. Christodoulidis, A. et al. A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images. *Comput. Med. Imaging Graph.* **52**, 28–43 (2016).
25. Niemeijer, M., Staal, J., Van Ginneken, B. et al. Comparative study of retinal vessel segmentation methods on a new publicly available database. In *Medical Imaging 2004: Image Processing* 5370, pp. 648–656 (SPIE, 2004).
26. Bek, T. Arterial oxygen saturation in neovascularizations in proliferative diabetic retinopathy. *Retina* **38**(12), 2301–2308. <https://doi.org/10.1097/IAE.0000000000001873> (2018).
27. Petersen, L. & Bek, T. The oxygen saturation in vascular abnormalities depends on the extent of arteriovenous shunting in diabetic retinopathy. *Investig. Ophthalmol. Vis. Sci.* **60**(12), 3762–3767. <https://doi.org/10.1167/iops.19-27365> (2019).
28. Guo, C., Szemenyei, M., Hu, Y. et al. Channel attention residual U-Net for retinal vessel segmentation. In *ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1185–1189. <https://doi.org/10.1109/ICASSP39728.2021.9413627> (IEEE, 2021).



29. Cao, H. et al. Swin-Unet: Unet-like pure transformer for medical image segmentation. arXiv preprint [arXiv:2105.05537](https://arxiv.org/abs/2105.05537) (2021).
30. Laibacher, T., Weyde, T. & Jalali, S. M2u-net: Effective and efficient retinal vessel segmentation for real-world applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* <https://doi.org/10.1109/CVPRW.2019.00123> (2019).
31. Liang, L. et al. Retinal vessel segmentation algorithm integrating Mobile Vit and inverted gated autoencoder. *J. Beijing Univ. Aeronaut. Astronaut.* <https://doi.org/10.1007/s11804-023-00123-4> (2023).
32. Liu, M. et al. AA-WGAN: Attention augmented Wasserstein generative adversarial network with application to fundus retinal vessel segmentation. *Comput. Biol. Med.* **158**, 106874. <https://doi.org/10.1016/j.compbiomed.2023.106874> (2023).
33. Samuel, P. M. & Veeramalai, T. VSSC Net: Vessel Specific Skip chain Convolutional Network for blood vessel segmentation. *Comput. Methods Programs Biomed.* **198**, 105769. <https://doi.org/10.1016/j.cmpb.2020.105769> (2020).
34. Ren, K. et al. An improved U-net based retinal vessel image segmentation method. *Heliyon* **10**, e11187. <https://doi.org/10.1016/j.heliyon.2022.e11187> (2022).
35. Tang, S. et al. W-Net: A boundary-aware cascade network for robust and accurate optic disc segmentation. *iScience* **27**(1), 108247. <https://doi.org/10.1016/j.isci.2023.108247> (2024).
36. Zhang, M. et al. Augmented transformer network for MRI brain tumor segmentation. *J. King Saud Univ. Comput. Inf. Sci.* **36**(1), 101917. <https://doi.org/10.1016/j.jksuci.2024.101917> (2024).
37. Cheng, D., Meng, G., Cheng, G. & Pan, C. SeNet: Structured edge network for sea-land segmentation. *IEEE Geosci. Remote Sens. Lett.* **14**, 247–251. <https://doi.org/10.1109/LGRS.2016.2631128> (2017).
38. Huang, Z. et al. CCNet: criss-cross attention for semantic segmentation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 603–612. <https://doi.org/10.1109/ICCV.2019.00072> (IEEE, 2019).
39. Goyal, A. et al. Non-deep networks. In *Advances in Neural Information Processing Systems* 35, pp. 6789–6801 (NeurIPS, 2022).
40. He, K., Zhang, X., Ren, S. & Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1904–1916. <https://doi.org/10.1109/TPAMI.2015.2389824> (2015).
41. Singh, P. et al. HetConv: Heterogeneous kernel-based convolutions for deep CNNs. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4830–4839. <https://doi.org/10.1109/CVPR.2019.00496> (IEEE, 2019).
42. Han, K. et al. GhostNet: More features from cheap operations. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1577–1586. <https://doi.org/10.1109/CVPR42600.2020.00162> (IEEE, 2020).
43. Chen, X. et al. Symbolic discovery of optimization algorithms. arXiv preprint [arXiv:2302.06675](https://arxiv.org/abs/2302.06675) (2023).
44. Xue, J. & Shen, B. A novel swarm intelligence optimization approach: sparrow search algorithm. *Syst. Sci. Control Eng.* **8**(1), 22–34 (2020).
45. Sousa, R. C. et al. Large tunneling magnetoresistance enhancement by thermal anneal. *Appl. Phys. Lett.* **73**(22), 3288–3290 (1998).
46. Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. J. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 850–863. <https://doi.org/10.1109/34.232073> (1993).
47. Woo, S. et al. CBAM: convolutional block attention module. arXiv preprint [arXiv:1807.06521](https://arxiv.org/abs/1807.06521) (2018).
48. Hou, Q. et al. Coordinate attention for efficient mobile network design. arXiv preprint [arXiv:2103.02907](https://arxiv.org/abs/2103.02907) (2021).
49. Liu, L. et al. On the variance of the adaptive learning rate and beyond. arXiv preprint [arXiv:1908.03265](https://arxiv.org/abs/1908.03265) (2019).
50. Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980), 2014.
51. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. arXiv preprint [arXiv:1711.05101](https://arxiv.org/abs/1711.05101) (2017).
52. Marini, F. & Walczak, B. Particle swarm optimization (PSO). A tutorial. *Chemom. Intell. Lab. Syst.* **149**, 153–165 (2015).
53. Zhou, Z. et al. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11. [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1) (Springer, 2018).
54. Ryu, J. et al. SegR-Net: A deep learning framework with multi-scale feature fusion for robust retinal vessel segmentation. *Comput. Biol. Med.*, 107132. <https://doi.org/10.1016/j.compbiomed.2023.01.002> (2023).
55. Qin, X. et al. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit.* **106**, 107404. <https://doi.org/10.1016/j.patcog.2020.107404> (2020).

## Acknowledgements

We are grateful to all members of the Zhuzhou Sansanyi Aier Eye Hospital for providing the retinal vascular dataset and assistance with the course of this research.

## Author contributions

Zerui Liu: Methodology, Writing—original draft, Conceptualization, Writing—review and editing, Resources. Weisi Dai: Methodology, Conceptualization, Writing—review and editing. Wenke Zhu: Visualization, Resources. Lin Li: Validation, Project administration, Supervision, Funding acquisition. Zewei Liu: Language editing, Software. Linan Hu: Resources. Lin Chen: Resources. Lixiang Sun: Resources.

## Funding

This work was support by Changsha Municipal Natural Science Foundation (Grant No. kq2014160); in part by the National Natural Science Foundation in China (Grant No. 61703441); in part by the key projects of Department of Education Hunan Province (Grant No. 19A511); in part by Hunan Key Laboratory of Intelligent Logistics Technology (Grant No. 2019TP1015); in part by the National Natural Science Foundation of China (Grant No. 61902436).

## Declarations

## Competing interests

The authors declare no competing interests.

## Ethical statement

All methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were approved by the Medical Ethics Committee of Zhuzhou Sansanyi Aier Eye Hospital. Informed consent was obtained from all subjects and their legal guardian.

## Additional information

**Correspondence** and requests for materials should be addressed to Z.Y. or L.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024