Applied Intelligence (2023) 53:15027–15044
https://doi.org/10.1007/s10489-022-04252-2

Retinal vessel segmentation based on self-distillation and implicit neural representation

Jia Gu[1] · **Fangzheng Tian**[1] · **Il-Seok Oh**[1,2]

Accepted: 7 October 2022 / Published online: 8 November 2022

Abstract
Segmenting retinal blood vessels from retinal images is a crucial step in ocular disease diagnosis. It is also one of the most important applications and research in ophthalmic image analysis. However, the contrast between the retinal vessels and background in fundus images is low. The size and shape of retinal vessels vary significantly, and the width of some small vessels is often below 10 pixels or even 1 pixel. Moreover, some blood vessels are discontinuous owing to illumination, which complicates the segmentation of retinal blood vessels. To address these problems, this paper innovatively proposes a novel retinal vessel segmentation network framework based on self-distillation and implicit neural representation, which predicts retinal vessels in two stages. First, the self-distillation method extracts the main features of retinal images using the properties of Vision Transformer (ViT) to obtain preliminary images for the blood vessel segmentation. Second, implicit neural representation improves the resolution of the original retinal image, and the details of blood vessels are enhanced through the texture enhancement module to obtain accurate results of the blood vessel segmentation. Furthermore, we adopted an improved centerline dice (clDice) loss function to constrain the topology of blood vessels. We experimented on two benchmark retinal datasets (i.e., Drive and Chase) to quantitatively and qualitatively analyze the proposed method. The results indicate that the proposed outperformed the mainstream baseline. The visual segmentation results also show that this method can segment thin blood vessels more accurately and ensure the continuity of blood vessels.

Keywords Retinal vessel segmentation · Neural networks · Self-distillation · Implicit neural representation · Texture enhancement

1 Introduction

The retina is the only part of the body with non-invasively observable blood vessels; therefore, retinal images provide a non-invasive way to detect diabetes [1, 2], hypertension [3, 4], glaucoma [5, 6], and other eyes or related body diseases. Accurate segmentation of blood vessels in retinal images is a prerequisite for doctors to predict these diseases because morphological information, such as branch, thickness, curvature, and length of retinal vessels are important indicators for the diagnosis of the mentioned diseases [7]. However, manual segmentation of retinal vessels is time-consuming and requires the expertise of ophthalmologists. Automatic segmentation of retinal blood vessels has received extensive attention from computer science and medical communities due to improvements in computing power. Automatic segmentation of fundus vessels reduces the burden on ophthalmologists and human segmentation errors.

In the past ten years, retinal vessel segmentation has become a hot topic in medical image segmentation, and hundreds of various segmentation techniques have been published. Traditionally, features are extracted from pixels of retinal images [8] by some artificially designed filters (e.g., Gabor [9] and Gaussian-based filters [10]) [11]. These features were then clustered to segment the blood vessels in the retinal images. Because of the complex topology of fundus blood vessels, many researchers have designed enhancement filters based on the Hessian matrix to capture the unique morphological features of blood vessels [12]. In recent years, with further improvements in computing power, deep learning has gradually become the mainstream technology for retinal vessel segmentation.

✉ Il-Seok Oh
isoh@jbnu.ac.kr

[1] Division of Computer Science and Engineering, Jeonbuk National University, 567 Baekje-daero, Jeonju, 54896, Republic of Korea

[2] Center for Advanced Image Information Technology, Jeonbuk National University, 567 Baekje-daero, Jeonju, 54896, Republic of Korea

Deep learning can automatically learn rich features from manually annotated images. U-Net [13] is the most representative network structure in medical image analysis. U-Net connects the encoder and decoder with a "skip connection", which effectively combines the shallow and deep features of the image. Because of this structure, U-Net exhibits superior performance in medical image segmentation. Therefore, many novel architectures have been developed based on U-Net. A convolutional neural network (CNN) based on the encoder-decoder structure filters out some useful spatial details in the convolution and pooling operations. Simultaneously, in an earlier coding process of the CNN, the context information cannot be fully obtained because of the extremely small receptive field. Therefore, many researchers have begun using Transformers with powerful attention mechanisms for image segmentation [14]. These methods regard an image as a patch-based sequence, allowing the network to focus on contextual information of the entire image at the beginning and with global attention [15]. Because the CNN effectively extracts semantic information of the image, and the Transformer has a contextual attention mechanism, some researchers have combined ViT and U-Net for medical image segmentation [16, 17].

Retinal vessel segmentation remains a challenging task because of the blurriness of retinal images, evident vessel differences, and low capillary pixels. Therefore, there have been many new studies on retinal vessel segmentation every year. In the past two years, in the task of retinal blood vessel segmentation, some researchers have focused on network innovation [18–22], using multi-scale methods to solve the problem of fundus blood vessel segmentation. Some researchers have focused on the preprocessing of retinal images to improve the accuracy of fundus vessels [23–26]. In addition, some researchers have achieved better fundus vessel segmentation by combining deep learning with traditional algorithms [27–29]. Researchers have proposed various solutions to different challenges. First, the contrast between the foreground and background in the retinal blood vessel image is low, making the blood vessels in the retinal image ambiguity [20, 30, 31]. In addition, the segmentation of blood vessels due to hemorrhage, exudate, and some pathological areas in retinal images brings certain noise [32]. Yin et al. [20] used a Frangi filter to obtain the domain-invariant features. Palanivel et al. [33] adopted Holder's exponent to minimize noise. Zhang et al. [34] enhanced the detection of the vessel edges with a Sobel edge detector. Refs. [30–32] used novel network structure to retain the main features in the image through learning. However, traditional filtering methods can weaken the blood vessel contrast, whereas learning-based methods only allow the network to learn certain objects under supervision, making the network less attentive to the important parts of the entire image [35]. To address this problem, we used the self-distillation method to extract the main features of retinal images using the ViT properties. This method adopts self-supervised training, which ensures that the network is unaffected by noise and focuses only on the important parts of the entire image. Second, the size and shape of retinal vessels vary significantly and have different shapes and appearances. The size of the retinal blood vessels usually varies between 1 and 20 pixels [20, 32], which makes it difficult to segment capillaries in retinal images. Most of the previous methods use multi-resolution images and multi-scale networks [18–22] to address this issue. The low-resolution large-scale part focuses on global information and the high-resolution small-scale part focuses on detailed information. However, the multi-resolution images do not exceed the resolution of the original image; even being attentive to the details is constrained by the limited original image resolution. To solve this problem, we then used implicit neural representation to make the image appear in a continuous way, thereby enlarging the resolution of the original image and making the width of the capillaries larger. This method that has not yet been considered for fundus vessel segmentation before. Furthermore, considering blurred capillary boundaries [21], Zhao et al. [36] combined global pixel loss and local matting loss to deal with blurred pixels around capillaries, which is not robust enough. In this paper, using a texture enhancement module to capture texture-related information of blood vessels to enhance the texture details of blood vessels in retinal images. Finally, due to the factors of illumination, low contrast between the foreground and background, etc., the blood vessels are displayed discontinuously in the image. Disconnected vessels can completely alter the hemodynamics. To address this problem, we used the clDice loss function [37] to guarantee the topology of the blood vessels.

In summary, the main contributions of this paper are summarized as follows:

1. We designed a novel retinal vessel segmentation network framework based on self-distillation and implicit neural representation that predicts retinal vessels in two stages.
2. Through self-distillation, the ViT properties are used to obtain the main features of the retinal images, and the influence of noise in the original image is removed.
3. Implicit neural representation enlarges the resolution of the original image, the texture enhancement module highlights the texture details of blood vessels, and the clDice loss function guarantees the predicted topology of the blood vessels.
4. The experimental results show that our method achieves state-of-the-art performance on the Drive [38] and Chase [39] datasets.

The remainder of this paper is organized as follows. Section 2 summarizes the latest work on retinal vessel segmentation and medical image segmentation. Section 3 introduces the system framework and principle, including the structure of the method and details of the algorithm. Section 4 presents experimental results and comparisons with other methods. Section 5 concludes the paper.

2 Related work

This section summarizes the analysis of fundus blood vessels in the past 2 years. Depending on the research method, we divided it into three cases: innovation in the network, preprocessing of datasets, and traditional algorithms combined with deep learning. We also introduce some recent related work on medical image segmentation.

Multi-scale context extraction of retinal images Lin et al. [18] proposed a high-resolution representation network with multi-path scale for automatic retinal vessel segmentation, which comprehensively considered features from the different resolution levels and the inner-level, aiming to improve the performance of extracting retinal vessels. Jiang et al. [19] proposed a multi-scale residual attention network called MRA-UNet, where multi-scale input enables the network to learn information at different scales, thereby increasing the robustness of the network. Yin et al. [20] designed a deep fusion network (DF-Net), including multi-scale fusion, feature fusion, and classifier fusion, for multi-source vessel image segmentation, where the multi-scale fusion module allows the network to detect vessels at different scales. Jiang et al. [21] proposed a multi-resolution fusion input network (MFI-Net) model. The multi-resolution input module in MFI-Net avoids the loss of shallow coarse-grained feature information by extracting local and global feature information at different resolutions. Moreover, the novel multi-resolution and multi-scale fusion input module enhance the accuracy of boundary segmentation by extracting more and richer feature information in the shallow layer. Kamran [22] proposed a new multi-scale generative architecture, RV-GAN, and introduced a new weighted feature matching loss for accurate retinal vessel segmentation. The network alleviates the problem of successive resolution loss in the encoding stage and the inability to recover the lost information in the decoding stage. Furthermore, it alleviates the limitation of auto-encoding-based segmentation methods for extracting retinal microvascular structures. Multi-scale context extraction is beneficial for the network to focus on the global and local features of fundus images at different resolutions to better segment capillaries in the retina. However, the multi-scale operation reduces the resolution, which causes the fundus blood vessel images lose a lot of capillaries while reducing the resolution and focusing on the global information. Even the extraction of details is based on the original resolution of the image. This study used the method of implicit neural representation to enlarge the resolution of the original image and the details in the image. Compared with the above methods, our advantage is to provide richer details of capillaries.

Retinal image data preprocessing Boudegga et al. [23] proposed a novel deep learning architecture by extending the lightweight convolutional module, characterized by lower complexity relative to standard convolutions. Preprocessing was performed to improve the image quality and enhance the contrast of the retinal vessels. Also, data augmentation was proposed to transform and crop the images to guarantee the robustness of training. Yu et al. [24] proposed a supervised retinal vessels extraction scheme using constrained-based nonnegative matrix factorization and three dimensional modified attention U-Net architecture, which performs Gaussian filtering and Gamma correction on the green channel of retinal image to suppress background noise and adjust the contrast of image. Sun et al. [25] proposed two new data augmentation modules, namely, channel-wise random Gamma correction and channel-wise random vessel augmentation, to study the problem of robust retinal vessel segmentation from the perspective of data augmentation. Navee et al. [26] proposed an efficient unsupervised vessel segmentation strategy as a step to accurately classify eye diseases from noisy fundus images. To that end, an ensemble block matching 3D speckle filter is proposed for removal of unwanted noise leading to improved detection. The proposed denoiser minimizes noise to discover tiny blood vessels, so that more tiny vessels can be detected. The purpose of retinal preprocessing is to filter the noise in the image and highlight the features of blood vessels. The proposed method utilizes the properties of ViT and unsupervised method of knowledge distillation to enable the network to automatically acquire the core features considered by the computer.

Traditional algorithms and deep learning are combined for retinal vessel segmentation Dash et al. [27] proposed an illumination normalization technique to enhance the blood vessels by jointly combining the morphological filter, differential filter, and homomorphic filter to make it more robust under illumination variations and also minimizing computational time. Tavakoli et al. [28] proposed an automatic unsupervised retinal vessel segmentation method based on hybrid methods. The algorithm initially applies a preprocessing step using morphological operators to enhance the vessel tree structure against a non-uniform image background. The main processing applies the Radon transform to overlapping windows, followed by vessel

validation, vessel refinement, and vessel reconstruction to achieve the final segmentation. Toptaş et al. [29] proposed a method for retinal blood vessel analysis using classical methods, with five different feature groups used for feature extraction. These feature groups are edge detection, morphological, statistical, gradient, and Hessian matrix. An 18-D feature vector is created for each pixel. This feature vector is given to the artificial neural network for training. The advantage of traditional methods is that they are interpretable and more robust. They are often used for preprocessing or post-processing. Learning-based methods can obtain more abstract semantic information in images. Combining the two facilitates the application of knowledge in different domains.

The latest research on medical image segmentation Ding et al. [40] proposed a region-aware fusion network for incomplete multi-modal brain tumor segmentation, while introducing a new region-aware fusion module and a segmentation-based regularization. This network is able to adaptively and efficiently utilize different combinations of multimodal data for tumor segmentation. Yang et al. [41] proposed a new automated machine learning algorithm, TAutoML, which not only searches for the best neural architecture, but also finds the best combination of hyperparameters and data augmentation strategies simultaneously. Zhang et al. [42] proposed a dynamic on-demand network, which can be trained on partially labeled datasets for multi-organ and tumor segmentation. Reiß et al. [43] proposed an approach based on a new formulation of deep supervision and student-teacher model and allows for easy integration of different supervision signals. Viedma et al. [44] explored an instance segmentation method based on region proposal architecture, and examined the importance of adequate hyperparameter selection. Gour et al. [45] designed a residual learning-based 152-layered convolutional neural network for breast cancer histopathology image classification. Due to the self-attention mechanism of Transformer, it exhibits superior performance in the field of image segmentation. Some researchers started using Transformer to realize the task of (medical) image segmentation. Strudel et al. [14] modeled the global context in the first layer and the entire network based on the ViT for better image segmentation. Sagar [15] proposed to take three different scale feature maps as input, and designed three continuous Transformer structures by means of jumpers, realizing the segmentation task of multiple medical image datasets. Refs. [16, 17] combined Transformer with U-shaped architecture to enhance the functionality and flexibility of traditional encoder-decoder architecture. Wang et al. [46] proposed the pyramid vision transformer that combines the advantages of CNN and Transformer, eliminating the need for convolution and reducing the computation of large feature maps. Different from the methods based on ViT or the combination of ViT and CNN, the proposed method does not take ViT as the backbone network. This method pays more attention to the properties of ViT under the self-attention mechanism, and uses the properties to effectively extract the main features in the fundus image to achieve the purpose of removing noise.

The novelty of the proposed method is that, first, unlike a multi-scale network using different downsampling low-resolution operation methods, this study uses the implicit neural representation to enlarge the resolution of the original image, and then achieve the purpose of enlarging the details. In contrast to traditional denoising methods, this study combines the properties of ViT and knowledge distillation to automatically obtain the core features in images in an unsupervised way. Finally, the texture enhancement module highlights the texture details of vessels and constrains the continuity of vessel predictions.

3 System framework

3.1 System framework overview

This paper designs a novel retinal vessel segmentation network framework based on self-distillation and implicit neural representation, which predicts retinal vessels in two stages. As shown in Fig. 1, in the first stage, the self-distillation method removes irrelevant information from the original image using the properties of ViT, and the main features of the retinal image are obtained. The original image and the main feature image are fused, and the initial blood vessel segmentation result of the fundus image is obtained through the first-level network. In the second stage, the implicit neural representation is used to double the resolution of the original fundus image, and the texture enhancement module captures texture-related information of blood vessels to enhance the texture details of blood vessels in retinal images. Then, the high-resolution images, enhanced blood vessel images and initial segmentation results of retinal blood vessels obtained by the first-level network are sent to the second-level network for fine fundus vessel segmentation. We randomly erase the fundus blood vessels with thickness less than 1 pixel in ground truth to supervise the output of the first-level network, so that the first-level network has the ability to search for small blood vessels. The processed ground truth image is used to supervise the secondary network. In our system, in addition to using traditional loss functions such as MSE, we improve the clDice loss function for constraining the topology of vessels.

We further elaborate on the details of the system framework in three subsections. Section 3.2 mainly

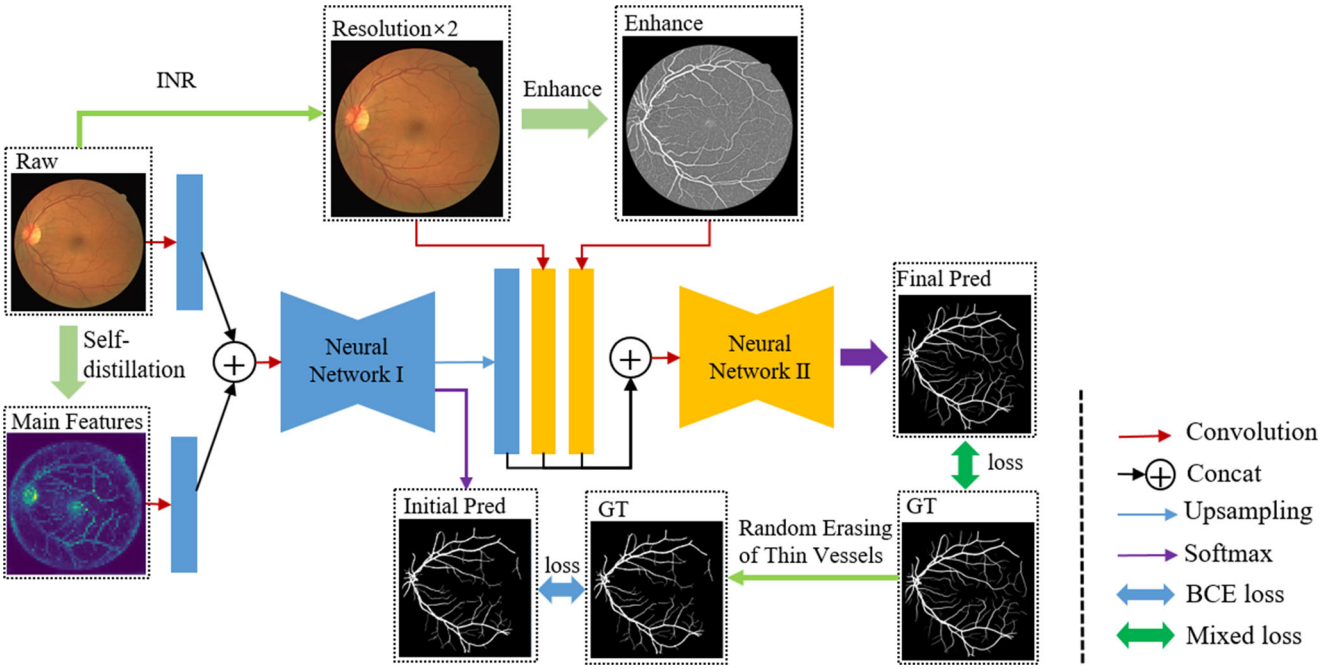


Fig. 1 System framework of the retinal vessel segmentation based on self-distillation and implicit neural representation

introduces obtaining the main features in the fundus image through the self-distillation module and completing the initial segmentation of the fundus blood vessels. Section 3.3 introduces the final segmentation of the fundus vessels through implicit neural representation and texture enhancement. Section 3.4 mainly introduces the loss function applied in this paper.

3.2 Initial vessel segmentation based on self-distillation

3.2.1 Self-distillation module

Compared with a CNN, the most prominent feature of ViT is its attention mechanism. The combination of self-supervised methods and ViT enables the model to automatically obtain the layout in the image, and the boundaries of the objects. Although ViT has various versions, the basic ViT has been extensively studied by Ref. [35]. Based on the experience of previous research work and the properties of Transformer, we utilize a self-distillation module to remove noise in the fundus image to obtain the main features in the image.

Figure 2 shows that the self-distillation module is similar to the popular self-supervised framework and knowledge distillation [47], which consists of two identical network architectures. These two networks can be divided into guidance network and studying network according to their functions. We input a set of data into the guidance network and the studying network respectively through different random transformations, and make a difference based on the output features of the two networks, so as to form self-supervision to update the parameters of the network. The difference between the two networks is that only the parameters of the studying network are updated in the process of back propagation. The guidance network only performs forward prediction without back propagation. The weight of the guidance network is updated by copying the parameters of the studying network. The ultimate goal of self-distillation is to make these two networks similar in output. The backbone of both networks is composed of ViT [48].

Given an input image I, both the studying network and the guidance network output a K-dimensional feature distribution. The final feature distributions of the two networks are normalized with the softmax function. The specific mathematical model is expressed as follows,

$$F_H(I_i) = \frac{exp[H(I_i)/r]}{\sum_{j=1}^{K}[H(I_j)/r]}, \tag{1}$$

where I represents the input image, F represents the output of the network, H represents the studying network S_t or the guidance network G_u, and r is a parameter that controls the degree of softmax sharpening.

In the process of self-distillation module training, the cross-entropy loss function is used to learn the feature distribution, and its mathematical form is

$$Loss = -F_{S_t}(I)log[F_{G_u}(I)]. \tag{2}$$

Fig. 2 Self-distillation module

During the prediction process, only the main features of the input images are obtained using the studying network.

3.2.2 Network for the first-stage retinal vessel segmentation

In the first stage, the network architecture for the initial segmentation of retinal vessels is shown in Fig. 3. First, the main features of the fundus image are obtained using the self-distillation module. Subsequently, the original fundus image and main feature image are sent to the V-shaped network at the original resolution, and preliminary results of the retinal blood vessel segmentation are obtained.

As shown in Fig. 3, the network adopts a simple V-shaped structure. First, the data in the original fundus image and the main feature image are transformed into similar distributions using 1×1 convolution, and 8-channel feature maps are obtained respectively. Then they are concatenated into a 16-channel feature map and sent to the downsampling module. The encoder consists of five feature extraction encoders. Each encoder has two layers. Each layer contains a 3×3 convolution operation, a batch normalization (BN) operation, and a parametric rectified linear unit (PReLU) activation operation. In the entire network, all activation functions use PReLU. The specific definition of the PReLU is expressed as follows,

$$PReLU(x_i) = \begin{cases} x_i, & if \quad x_i > 0 \\ a_i x_i, & if \quad x_i \leq 0 \end{cases}, \tag{3}$$

where x_i is the tensor that the function inputs to different channels, and a_i is its corresponding parameter. In PReLU, each channel has an activation function with different parameters. In the network training process, a_i is dynamic, and its change process is expressed as follows,

$$\Delta a_i := \mu \Delta a_i + \delta \frac{\partial \varepsilon}{\partial a_i}, \tag{4}$$

where μ is momentum, δ is the learning rate, and ε is the objective function. The use of PReLU reduces the risk of overfitting while hardly increasing the calculation cost.

To accelerate the convergence of the network, the two layers of each encoder are fused to form a residual structure. A max pooling layer is used between the encoders to reduce the resolution of the feature maps. The purpose of using max pooling instead of 2×2 convolution is to avoid overfitting to some extent by losing some information.

The upsampling module also contains five decoders. In the first four upsampling modules, each module is cascaded with its corresponding downsampling module by using skip connections. Each decoder consists of two layers, and each layer contains the 3×3 convolution kernel with a stride of 1, BN, and PReLU. A 2×2 deconvolution kernel with a stride of 2 is used between each upsampling module to enlarge the feature map resolution. After the fourth upsampling module, the resolution of the feature map reaches the same resolution as the original fundus image. The purpose of the fifth upsampling module is to further upscale the resolution of the feature maps in preparation for the final retinal vessel segmentation in the second stage. We convert the fourth upsampled feature map into the segmentation probabilities of the foreground and background using 1×1 convolution and softmax to obtain preliminary vessel segmentation.

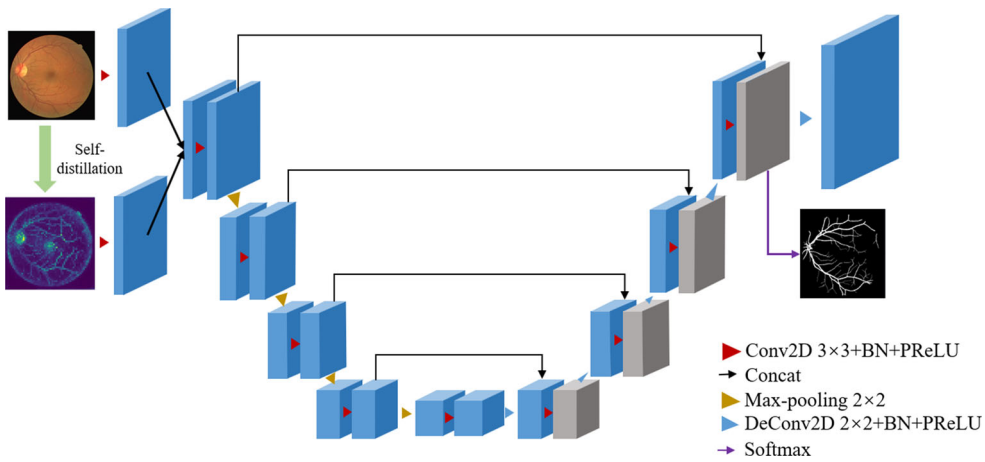


Fig. 3 Network architecture for the retinal vessel segmentation in the first stage

3.2.3 Supervision way of the initial vessel segmentation

There are many capillaries smaller than one pixel in retinal images, and it is complicated to manually label small blood vessels in the retina. Therefore, in the labeling process, the labeler ignores many capillaries [49]. In the first stage of initial fundus blood vessel segmentation prediction, what is wanted is to make this initial fundus blood vessel segmentation model have strong blood vessel prediction ability, and this process is somewhat similar to the process of de-occlusion in computer vision. Many researchers use deliberate occlusion of the ground truth to make the model achieve the goal of predicting invisible regions [50]. This paper also adopts this method for training. The method randomly erases some small blood vessels in the ground truth and then uses the processed ground truth to supervise the network in the first stage, to achieve our goal. Specifically, the first step is to find the skeleton of the blood vessel and use these skeletons to locate the centerline of the blood vessel. The second step is to expand the blood vessel according to the centerline to the entire labeled area, so as to calculate the thickness of the blood vessel. The final step is to analyze the thickness of blood vessels and randomly erase 30% of the annotation content for blood vessels with 1 pixel or less.

3.3 Final vessel segmentation based on implicit neural representation and texture enhancement module

3.3.1 High-resolution fundus image acquisition based on implicit neural representation

There are many capillaries smaller than one pixel in retinal images, which complicates the automatic segmentation by computers. Current images discretely store data in the form of pixels. Our visual world is presented in a continuous form. Recently emerging implicit neural representation can continuously represent image information [51]. The continuous representation of image information can make the image break the limitation of resolution. Therefore, we use this technology to enlarge the original fundus image to maintain a high fidelity while enlarging the details.

In implicit neural representation, each continuous image is a feature map of a 2D image. The entire image shares the decoding function f. Suppose that the parameter of the decoding function is w, and then its mathematical model is expressed as follows,

$$c = f_w(v, x), \tag{5}$$

where c is the predicted RGB value, f is the decoding function, and x is the coordinate of the continuous image. For continuous images, the RGB value at coordinate x_i can be defined as

$$RGB(x_i) = f_w(v^*, x_i - r^*), \tag{6}$$

where v^* is the closest possible encoding position to x_i and r^* is the coordinate of the possible encoding position. As shown in Fig. 4, there are three possible encoding positions, and the numbers are 3, 5, and 8, respectively.

Through this implicit continuous representation, theoretically, an image can be enlarged without limitations. In the processing of this paper, considering the specific effect and computational cost of computer hardware, we upscale the original retina image by a factor of two.

3.3.2 Unsupervised texture enhancement module

To highlight the texture details of blood vessels in retinal images, we perform texture enhancement operations on the enlarged retinal images in an unsupervised manner. Consistent with the previous unsupervised method [49], two cascaded thin V-shaped networks are designed, somewhat similar to a scaled-down version of our two-stage network framework for the entire system. The difference is that the two thin V-shaped networks in the texture enhancement module are completely consistent, and their structure is shown in Fig. 5. The purpose of the texture enhancement module is to generate high-contrast enhancement maps in the first-level V-shaped network in a computer automatically. Without any form of supervision in the first-level network, the difference between the prediction results of the second-level network and the ground truth is calculated in a supervised manner. The weights in the first-level network are updated simultaneously by means of gradient descent. The first-level network obtains the model that can enhance the image texture without supervision. In the prediction process, the image enhancement module only refers to the first-level unsupervised network that has been trained by the second-level network.

3.3.3 Network for second-stage retinal vessel segmentation

As shown in Fig. 6, the second-stage network takes the enlarged original fundus image, the texture-enhanced image and the output of the fifth decoder of the first-stage network as input. The network architecture and hyperparameters are set almost the same as the first-stage network. The difference is that atrous spatial pyramid pooling (ASPP) is added to the last encoder part of the second-stage network to enlarge the receptive field and capture multi-scale context information.

ASPP is similar to the pooling operation, and its purpose is the same as the ordinary pooling layer, to extract features as much as possible. The structure of the ASPP is shown

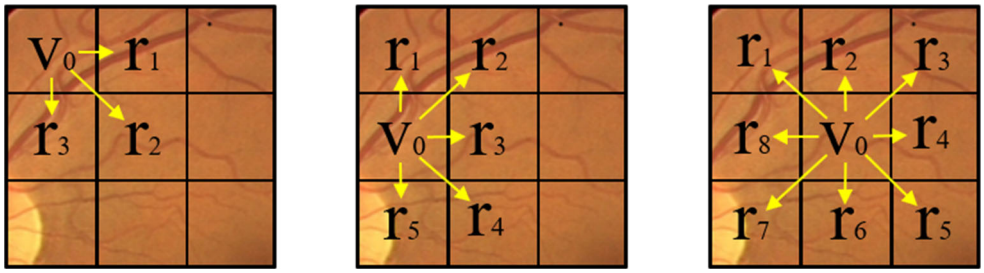


Fig. 4 Representation of continuous image coding

in Fig. 6, which consists of a 1×1 convolution, pooling pyramid, and ASPP pooling. And free multi-scale feature extraction is achieved by customizing the expansion factor of each layer of the pooling pyramid.

3.4 Loss function

This study uses a combination of multiple loss functions to optimize the model.

clDice loss Retinal vascular structure has high connectivity. To maintain the topology of the predicted retinal vessel segmentation, this paper introduces a novel topology preserving loss function clDice for tubular structure segmentation to accurately segment retinal vessels. The core of clDice is to use the skeleton of the blood vessel as a constraint to maintain the topological structure of the blood vessel through topological accuracy and topological sensitivity. The mathematical form of topological accuracy and topological sensitivity is

$$T_{prec}(S_P, G_t) = \frac{|S_P \cap G_t|}{|S_P|}, \tag{7}$$

$$T_{sen}(S_P, P_{re}) = \frac{|S_G \cap P_{re}|}{|S_G|}, \tag{8}$$

where G_t and P_{re} represent the ground truth masks and predicted retinal vessel images, respectively. S_G, S_P denote the vessel skeletons extracted from the ground truth masks and predicted retinal vessel images, respectively.

The mathematical form of the clDice loss function is further determined by topological accuracy and topological sensitivity as

$$L_{clDice} = 2 * \frac{T_{prec} * T_{sen}}{T_{prec} + T_{sen}}. \tag{9}$$

Mean-squared error (MSE) loss MSE refers to the mean of the squared distance between the predicted retinal vessels P_{re} and ground truth masks G_t. The mathematical model is

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}[G_t(i) - P_{re}(i)]^2, \tag{10}$$

where i represents a single retinal image and n is the number of images.

Binary cross entropy (BCE) loss BCE is a loss function of binary classification. In this paper, the retinal images are divided into two categories: blood vessels and background. The mathematical model is

$$\begin{aligned} L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n}[&G_t(i)log(P_{re}(i)) \\ &+ (1 - G_t(i))log(1 - P_{re}(i))], \end{aligned} \tag{11}$$

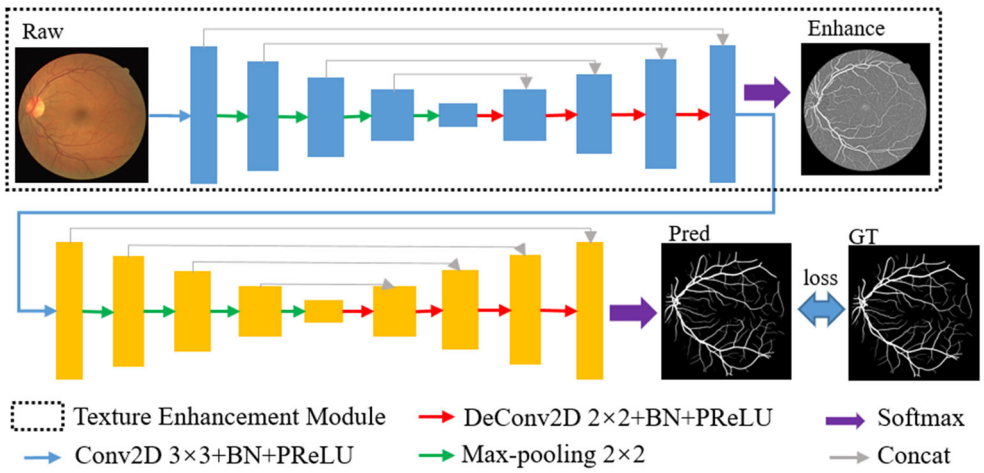


Fig. 5 Texture enhancement module architecture

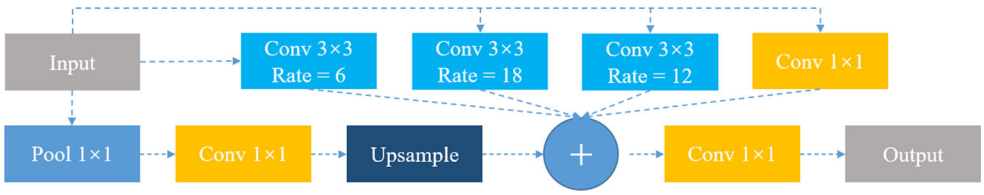


Fig. 6 Structure of ASPP in the second stage

where i represents a single retinal image and n is the number of images. G_t and P_{re} represent the ground truth masks and predicted retinal vessel images, respectively. When $G_t(i)$ is 0, the first half of the equation is 0, and $P_{re}(i)$ needs to be as 0 as possible to make the value of the second half smaller. When $G_t(i)$ is 1, the second half is 0, and $P_{re}(i)$ needs to be as 1 as possible to make the value of the first half smaller, to make the prediction results achieve the effect of ground truth as much as possible.

Dice loss The Dice loss function extracts features from specific regions and evaluates pixel-level similarity. In short, Dice calculates the "intersection of union" between the output predicted segmented image and the real segmented image, and measures the overlap between the real and predicted images. Dice can be defined as

$$Dice(G_t, P_{re}) = 2\frac{\sum_{i=1,j=1}^{\Omega} \left|G_{t_{i,j}} \cap P_{re_{i,j}}\right|}{\sum_{i=1,j=1}^{\Omega} \left|G_{t_{i,j}} + P_{re_{i,j}}\right|}, \tag{12}$$

where $G_{t_{i,j}}$ and $P_{re_{i,j}}$ represent the pixels of the real segmented image and the predicted segmented image at i, j position, and Ω represents the entire image area. The value of Dice is between 0 and 1. If Dice is 1, the predicted segmentation result is the same as the actual segmentation result. Dice loss can be defined as follows,

$$L_{Dice} = 1 - Dice(G_t, P_{re}). \tag{13}$$

Total variation (TV) loss The TVloss reduces noise to reduce the difference between adjacent pixel values in the predicted image and then maintains the smoothness of the predicted image. The mathematical form of TV loss is

$$L_{TV} = \sum_{i=1,j=1}^{\Omega} [(P_{re_{i,j-1}} - P_{re_{i,j}})^2 + (P_{re_{i+1,j}} - P_{re_{i,j}})^2]^{\frac{\beta}{2}}, \tag{14}$$

where P_{re} represents the predicted retinal vessel image, i, j represent the pixel coordinates of the P_{re} image, and the default value of β is 2. This loss function promotes spatial smoothness in the image.

This study balances these loss functions with the same weight, so the final loss function is

$$Loss = L_{clDice} + L_{MSE} + L_{BCE} + L_{Dice} + L_{TV}. \tag{15}$$

4 Experiment

4.1 Dataset

We evaluate our proposed method on the two most representative public datasets: Drive [38] and Chase [39]. Figure 7 shows the sample images from the two datasets, including images of raw retinal vessels, ground truth annotated by doctors, and mask maps.

The Drive dataset was obtained from the Dutch Diabetic Retinopathy Screening Project. A total of 40 subjects were selected from 400 subjects with diabetes between the ages of 25 and 90. Among them, 33 cases had no signs of diabetic retinopathy, and 7 cases had signs. It consists of 40 retinal fundus vessel images, corresponding ground truth images and corresponding mask images. The size of the image is 565 × 584 pixels, and these images were taken by Canon camera in a 45 degree field of view (FOV).

The Chase dataset consists of 28 retinal fundus vessel images, corresponding ground truth images, and corresponding mask images. The size of the image is 960 × 999 pixels, collected from the left and right eyes of 14 children. These images were taken by Nidek camera in a 30 degree FOV [52], and the binary FOV mask and segmentation ground truth were obtained by hand-crafted methods. Unlike the Drive images, the sample images in the Chase dataset have fewer blood vessels visible in the images due to uneven illumination.

4.2 Performance evaluation indicators

To evaluate our model, we compared the segmentation results with the corresponding ground truth, using sensitivity, F1 score, specificity, accuracy, and area under the ROC curve (AUC) as evaluation indicators. If pixels are correctly classified as objects or background, they are marked as true positive (TP) or true negative (TN), respectively. Meanwhile, pixels misclassified as objects or backgrounds are marked as false positive (FP) or false negative (FN), respectively.

The sensitivity is defined by the equation (16), which represents the percentage of pixels in the segmentation result image that correctly segment blood vessels. Sensitivity reflects the proportion of unsegmented vessel pixels missed in the segmentation results. The closer the sensitivity is to 1.0, the lower the proportion of

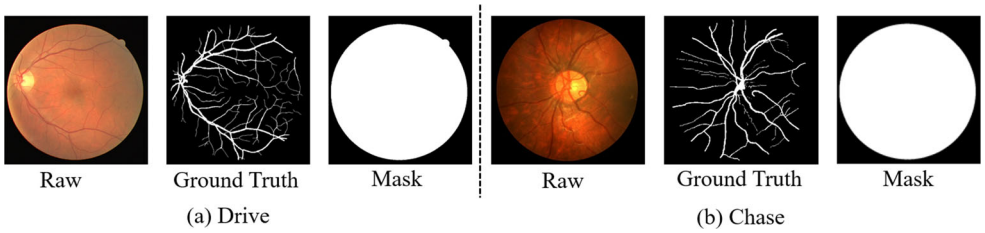


Fig. 7 Sample images from the Drive and Chase datasets

unsegmented blood vessels and the better the segmentation effect.

$$Sensitivity = \frac{TP}{TP + FN}. \tag{16}$$

F1 is a common metric in binary classification models that considers both the precision and sensitivity of the classification model. A higher F1 score indicates better segmentation.

$$F1 = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity}, \tag{17}$$

where precision is defined as

$$Precision = \frac{TP}{TP + FP}. \tag{18}$$

Specificity is defined in equation (19), which represents the proportion of blood vessel pixels incorrectly segmented in the image of the segmentation result. Specificity reflects the proportion of mis-segmented pixels.

$$Specificity = \frac{TN}{TN + FP}. \tag{19}$$

Accuracy is defined in equation (20), which represents the percentage of correctly segmented pixels (including blood vessel pixels and background pixels) in the entire segmentation map, reflecting the overall segmentation accuracy.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}. \tag{20}$$

In retinal vessel segmentation, only 9%-14% of the pixels belong to blood vessels, while the other pixels are considered as background pixels. Therefore, the accuracy can reflect the segmentation results to a certain extent, but cannot accurately evaluate the performance of the segmentation methods. The AUC can measure segmentation performance. The AUC value is higher, the better the segmentation results.

4.3 Implementation details, running speed, and cost

Implementation details This study implemented a training framework using Pytorch deep learning framework, and performed data augmentation on the training images, including flipping, color enhancement, brightness changes, and random clipping. Considering the resources of the computer hardware, training was performed in the form of patches for both datasets. The size of each patch is 256 × 256 pixels. We trained six NVIDIA RTX 2080 Ti for a total of 15 epochs. The batch size for each iteration is 32. We applied a zero-mean Gaussian random initialization of the network weights, employed Adam for network optimization, set the learning rate to 10^{-4}, and used the learning rate decay method to reduce the learning rate by half every two epochs. Consistent with most previous methods, we divided the dataset according to the criteria given by the dataset. For the Drive dataset, we used 20 retinal images for training and 20 for testing. For the Chase dataset, 20 images were used for training and 8 for testing.

Running speed and cost We used a single NVIDIA RTX 2080 Ti for inference testing. The proposed method takes on average 0.0517 s to infer an image, and about 3547 MB of Video Memory is required for the system operation.

4.4 Experimental results

This section presents the experimental visualization of the proposed system. For quantitative analysis, we compare the advanced methods from the past three years and analyze them in the next section. The segmentation results of retinal vessels predicted by our framework are shown in Fig. 8. From the overall prediction results in the second column, the method can completely predict blood vessels in the retina. The predicted results are partially enlarged, and it can be seen from the enlarged details that our method can predict capillaries smoothly, and the segmented vessels are smooth. The last column in the figure shows the false negative and false positive results predicted by our method, in which the number of false positive is significantly more than that of false negative. This is mainly because we supervise the network with deliberate occlusion in the first stage of the framework, which allows the network to have a better ability to search for capillaries.

In addition, the intermediate results of our system are output to prove the rationality of the proposed method. As shown in Fig. 9, the self-distillation module retains only the core features of the image and filters out most of

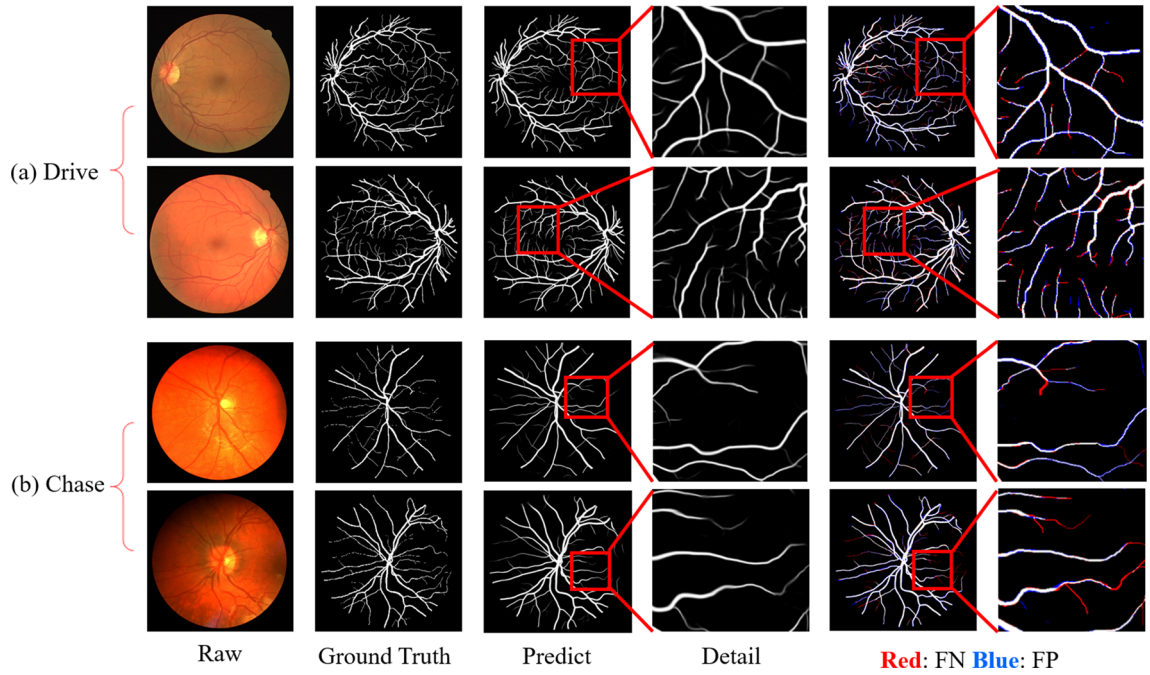


Fig. 8 Experimental results of our system

the background information. Implicit neural representation improves the resolution of the blood vessels in the image, thus making the blood vessels clearer. The texture enhancement module improves the contrast between the blood vessels and background, making the texture of the blood vessels more prominent.

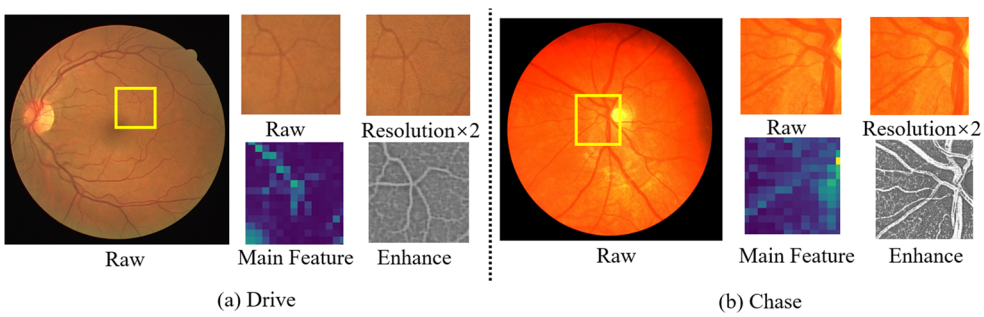
Figure 10 shows the variation of accuracy during training. As can be seen from the figure, the accuracy of the proposed method for the test set does not fluctuate significantly with the increase of the training epoch, which shows that our method does not appear overfitting.

4.5 Comparison with other advanced methods

4.5.1 Evaluation results on the drive dataset

In this study, a quantitative analysis was conducted, and the proposed method was compared with advanced methods published in the past 3 years. DF-Net [20] adopted multi-scale fusion and traditional methods for multi-source fundus blood vessel segmentation, feature fusion module provided the network with more tiny blood vessel structure information, and recovered the information loss caused by the downsampling operation. Palanivel et al. [33] used the Holder exponent to quantify the local regularity of retinal vessels after computing Gabor wavelet transforms at different scales. This form of computing the multifractal representation of vessels minimizes noise and enhances the vessels during segmentation. Zhuo et al. [53] exploited size-invariant feature maps and dense connections to improve the learning ability of CNNs. Zhou et al. [54] proposed an improved line detector for rapidly extracting the main structures of blood vessels, applying a hidden Markov model to efficiently detect vessel centerlines including thin vessels. Zhao et al. [36] proposed a new loss function in vessel segmentation that combines global pixel loss and local matting loss to deal with blurry pixels that are usually located around the boundaries of small vessels. Convert the segmentation to matting task to improve vessel segmentation performance. BEFD-Unet [34] introduced a

Fig. 9 Intermediate output results of the system



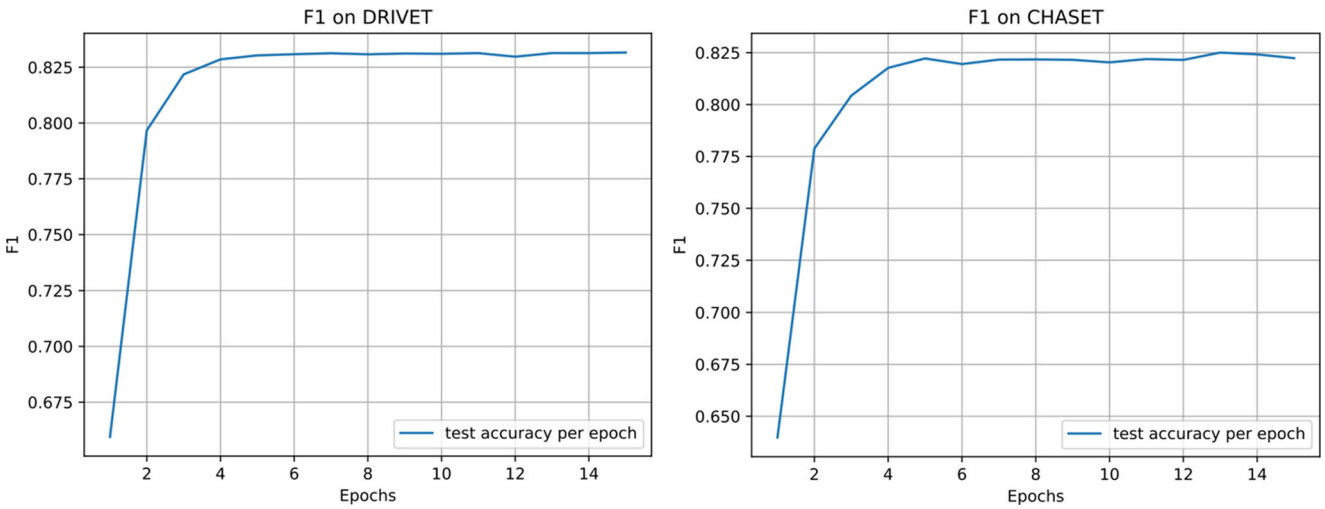


Fig. 10 Variation curve of accuracy during training

Sobel edge detector, and the network can obtain additional edge priors. The proposed boundary enhancement and feature denoising (BEFD) module promotes the ability of the network to extract boundary information in semantic segmentation. Li et al. [55] proposed a CNN combining the basic U-Net and an attention module, the latter used to capture global information and enhance features by placing it in a feature fusion process. To ensure the fairness of the experiment, the data presented in Table 1 were derived from the original paper. The table shows that our method achieves the best performance on indicators other than the specificity index, which is inferior to that of DF-Net. The proposed method achieves such results mainly because of the amplification of the details in images using implicit neural representations. Texture enhancement and noise filtering are more conducive to the segmentation of fundus blood vessels, particularly the segmentation of capillaries. DF-Net is difficult to predict small capillaries due to the inherent resolution.

To compare and evaluate the segmentation results of the different methods more intuitively, Fig. 11 shows the comparison result with the visualization of the DF-Net method, where the visualization experiment results of DF-Net are derived from the original paper. DF-Net includes three modules: scale fusion, feature fusion, and classifier fusion. The scale fusion module fuses multi-scale images, which is beneficial to detect multi-scale blood vessels. The feature fusion module contains the Frangi filter and enhances the feature map of the neural network by injecting more blood vessel information, which can prevent the spatial information loss of thin blood vessels caused by the downsampling operation. As can be seen from the figure, the proposed method can predict more capillaries. In addition, our predicted vessels are smoother. The main reason is that the multi-scale fusion used by DF-Net is the fusion of the same fundus image at different resolutions, which are obtained by continuously reducing the resolution. Reducing the resolution increases the receptive field while ignoring details in the image. Our proposed method upscales the resolution of the original image, thereby enlarging the details of capillaries, making it easier for the neural network to identify capillaries, and then predicting clearer and smoother capillaries. Moreover, the use of the clDice loss function also makes the predicted blood vessels have a more reasonable topology, ensuring the continuity of the blood vessels.

4.5.2 Evaluation results on the chase dataset

The evaluation results for the Chase dataset are presented in Table 2. Our method can improve the sensitivity by 0.23% and F1 by 0.1%, respectively. In addition, compared with the recently published DF-net, it has also won in most indicators. Therefore, we can conclude that the proposed method is sufficiently robust to apply to predictions using different retinal datasets.

To compare and evaluate the segmentation results of the different methods more intuitively, the comparison results with the partial visualization of DF-Net in the Chase dataset are shown in Fig. 12. Our method can better predict capillaries.

4.6 Ablation experimental results and analysis

To prove the performance of each component in the proposed method for retinal vessel segmentation, this study is consistent with the work in Refs. [20, 36, 55]. Ablation experiments were performed on the Drive dataset. The results of the ablation experiments are shown in Table 3. In the ablation experiments, our network architecture is not

Table 1 Experimental comparison results based on Drive dataset

Method	Year	Sensitivity	F1	Specificity	Accuracy	AUC
Palanivel et al. [33]	2020	0.7375	–	0.9788	0.9480	0.9590
Zhuo et al. [53]	2020	–	0.8178	–	0.9537	0.9754
Zhou et al. [54]	2020	0.7262	0.7786	0.9803	0.9475	–
Zhao et al. [36]	2020	0.8329	0.8229	0.9767	–	–
IterNet [56]	2020	0.7791	0.8218	0.9831	0.9574	0.9813
BEFD-Unet [34]	2020	0.8215	0.8267	0.9845	0.9701	0.9867
AA-Unet [57]	2020	0.7941	0.8216	0.9798	0.9558	0.9847
MRA-UNet [19]	2020	0.8353	0.8293	0.9828	0.9698	0.9873
Yin et al. [58]	2020	0.7614	–	0.9837	0.9604	0.9846
Li et al. [55]	2020	0.7921	–	0.9810	0.9568	0.9806
CTF-Net [59]	2020	0.7849	0.8241	0.9813	0.9567	–
Pyramid U-Net [31]	2021	0.8213	–	0.9807	0.9615	0.9815
SA-UNet [60]	2021	0.8212	0.8263	0.9840	0.9698	0.9864
RCED-Net [61]	2021	0.8252	-	0.9787	0.9649	–
DF-Net [20]	2022	0.7733	–	**0.9853**	0.9623	0.9871
Our method	2022	**0.8357**	**0.8320**	0.9847	**0.9706**	**0.9885**

The bold numbers are best results

changed. "✓" indicates that this module is included in our framework.

Table 3 shows that after adding the self-distillation module and implicit neural representation module respectively, the performance is improved. However, the improvement effect after adding the self-distillation module is not as obvious as the experimental result after adding the implicit neural representation. Since the self-distillation module removes most of the information, the implicit neural representation enlarges all details in the image. In addition, when the implicit neural representation and texture enhancement module are used concurrently, the performance is further improved, which indicates that the texture enhancement module helps the model to better learn the regions of blood vessels. The experimental results achieve the best performance when all modules are used simultaneously. Table 3 proves that each module of the proposed method has positive effect.

To more intuitively analyze the function of each component, the ablation experiment is qualitatively analyzed. The

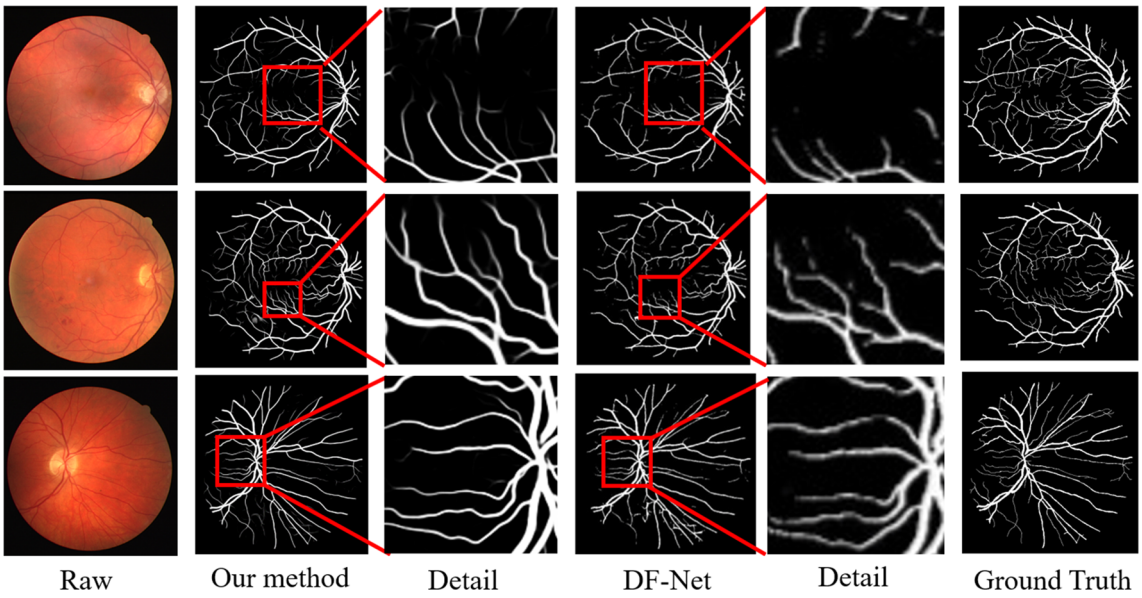


Fig. 11 Comparison of the visual results of the retinal vessel segmentation based on Drive dataset

Table 2 Experimental comparison results based on Chase dataset

Method	Year	Sensitivity	F1	Specificity	Accuracy	AUC
Palanivel [33]	2020	0.7237	-	0.9703	0.9459	0.9592
IterNet [56]	2020	0.7969	0.8072	0.9881	0.976	0.9899
SA-UNet [60]	2020	0.8573	0.8153	0.9835	0.9755	0.9905
MRA-UNet [19]	2020	0.8324	0.8127	0.9854	0.9758	0.9899
Yin et al. [58]	2020	0.7993	-	0.9868	0.9783	0.9869
Li et al. [55]	2020	0.7818	-	0.9819	0.9635	0.9810
AA-UNet [57]	2020	0.8176	0.7892	0.9704	0.9608	0.9865
Pyramid U-Net [31]	2021	0.8035	-	0.9787	0.9639	0.9832
RCED-Net [61]	2021	0.8440	-	0.9810	0.9772	-
DF-Net [20]	2022	0.8316	-	**0.9885**	0.9812	0.9901
Our method	2022	**0.8596**	**0.8250**	0.9849	**0.9824**	**0.9917**

The bold numbers are best results

visualization results are shown in Fig. 13. The details of some parts are enlarged in the figure (the enlarged part is the area of the red rectangular box in the ground truth). In the enlarged details, the white part represents the predicted result, the red part indicates false negatives, and the blue part indicates false positives. The effect of each component on the prediction performance of the model can be visually displayed through the enlarged details. Compared to the basic method, the self-distillation module can predict more blood vessels. Because implicit neural representation enlarges the details of capillaries, it is more accurate in the prediction of capillaries. After combining the texture enhancement module, the prediction of capillaries becomes clearer. Finally, after using all the modules, the visualization results achieve the best performance, in which FN and FP are significantly reduced.

5 Conclusion and future work

Retinal vessels are the only vessels that can be observed non-invasively. Morphological changes, such as changes in blood vessel diameter and thickness, can directly reflect some diseases of body. The segmentation of fundus blood vessels, especially the accurate analysis of branches of deep blood vessels and small capillaries, has important clinical significance. Therefore, this paper proposes a network framework based on self-distillation and implicit neural representation to segment blood vessels in retinal images. The self-distillation method uses the properties of ViT to extract the main features of the retinal image, which makes the network focus more on the important information in the entire image and solve the influence of the noise in the retinal image on the segmentation results. The implicit

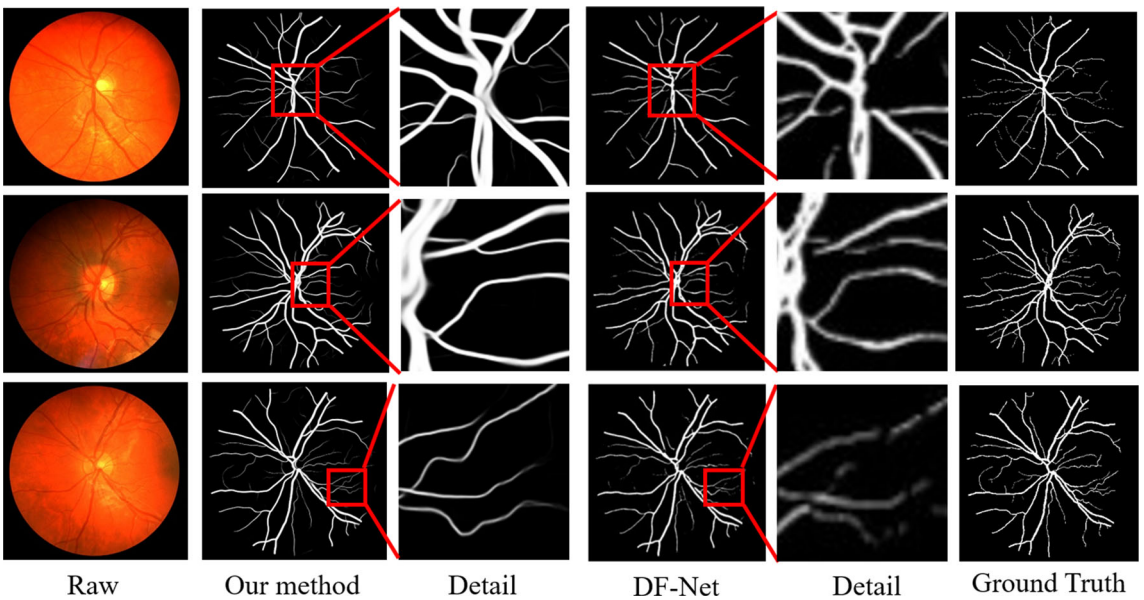


Fig. 12 Comparison of the visualization results of the retinal blood vessel segmentation based on the Chase dataset

Table 3 Ablation experiment results (where "✓" indicates that this module is included

Order	Raw	Self-distillation	Implicit neural representation	Texture enhancement	F1	Sensitivity	Specificity	Accuracy	AUC
1	✓				0.8285	0.8332	0.9818	0.9671	0.9859
2	✓	✓			0.8289	0.8336	0.9820	0.9677	0.9863
3	✓		✓		0.8301	0.8342	0.9828	0.9689	0.9873
4	✓		✓	✓	0.8315	0.8356	0.9845	0.9703	0.9881
5	✓	✓	✓	✓	0.8320	0.8357	0.9847	0.9706	0.9885

neural representation is used to enlarge the original retinal image, so as to better capture the capillaries. The texture enhancement module highlights the texture structure of blood vessels, solving the problem of blurred capillary boundaries. Finally, the clDice loss function constrains the topology of retinal vessels. However, owing to the complex design of the system, which increases the cost of computer hardware resources while improving the accuracy of fundus blood vessel segmentation, it is necessary to consider further compressing the network model in the future.

The greatest advantage of the proposed method is that the implicit neural representation enlarges the resolution of retinal images and the capillary's details. Thus, this method can segment capillaries more effectively than existing methods. In addition, the proposed method increases the cost of computer hardware resources because of its complex structure, and the self-distillation module ignores some details in the image when filtering the noise to obtain the main features of the image. The compressed network model must be further considered in future work. In the

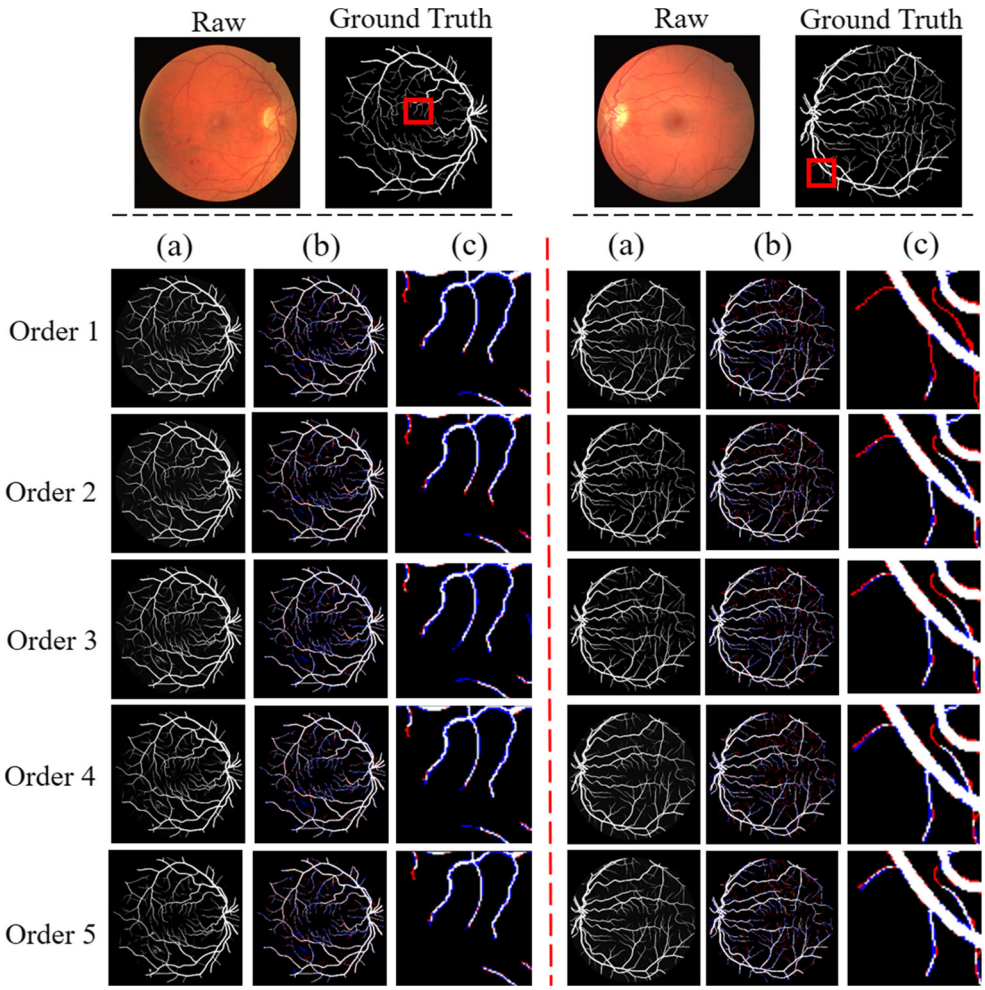


Fig. 13 Comparison of the visualization results of the ablation experiments. Column (a) shows the results predicted by different components, column (b) adds false positives and false negatives (Red:FN Blue:FP) to the predictions, and column (c) shows the details of the method in column (b) image. (The enlarged area is the area in the red box in ground truth)

selfdistillation part, we choose ViT as the main framework of the self-distillation module based on the experience of previous work. In the future, we will explore the influence of different versions of Transformer on the main features obtained from self-distillation.

References

1. Ur Rehman M, Abbas Z, Khan SH, Ghani SH et al (2018) Diabetic retinopathy fundus image classification using discrete wavelet transform. In: 2018 2nd international conference on engineering innovation (ICEI). IEEE, pp 75–80
2. Sahlsten J, Jaskari J, Kivinen J, Turunen L, Jaanio E, Hietala K, Kaski K (2019) Deep learning fundus image analysis for diabetic retinopathy and macular edema grading. Sci Rep 9(1):1–11
3. Syahputra M, Amalia C, Rahmat R, Abdullah D, Napitupulu D, Setiawan M, Albra W, Andayani U et al (2018) Hypertensive retinopathy identification through retinal fundus image using backpropagation neural network. In: Journal of physics: conference series. IOP Publishing, vol 978, p 012106
4. Nagpal D, Panda SN, Malarvel M (2021) Hypertensive retinopathy screening through fundus images-a review. In: 2021 6th international conference on inventive computation technologies (ICICT). IEEE, pp 924–929
5. Singh LK, Garg H, Khanna M, Bhadoria RS et al (2021) An enhanced deep image model for glaucoma diagnosis using feature-based detection in retinal fundus. Med Bio Eng Comput 59(2):333–353
6. Shabbir A, Rasheed A, Shehraz H, Saleem A, Zafar B, Sajid M, Ali N, Dar SH, Shehryar T (2021) Detection of glaucoma using retinal fundus images: a comprehensive review. Math Biosci Eng 18(3):2033–2076
7. Jin Q, Meng Z, Pham TD, Chen Q, Wei L, Su R (2019) Dunet: a deformable network for retinal vessel segmentation. Knowl-Based Syst 178:149–162
8. Vidya BS, Chandra E (2019) Entropy based local binary pattern (elbp) feature extraction technique of multimodal biometrics as defence mechanism for cloud storage. Alexandria Eng J 58(1):103–114
9. Oloumi F, Rangayyan RM, Oloumi F, Eshghzadeh-Zanjani P, Ayres FJ (2007) Detection of blood vessels in fundus images of the retina using gabor wavelets. In: 2007 29th annual international conference of the ieee engineering in medicine and biology society. IEEE, pp 6451–6454
10. Niemeijer M, Staal J, van Ginneken B, Loog M, Abramoff MD (2004) Comparative study of retinal vessel segmentation methods on a new publicly available database. In: Medical imaging 2004: image processing. International Society for Optics and Photonics, vol 5370, pp 648–656
11. You X, Peng Q, Yuan Y, Cheung Y-m, Lei J (2011) Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. Pattern Recognit 44(10-11):2314–2324
12. Lesage D, Angelini ED, Bloch I, Funka-Lea G (2009) A review of 3d vessel lumen segmentation techniques: models, features and extraction schemes. Med Image Anal 13(6):819–845
13. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
14. Strudel R, Garcia R, Laptev I, Schmid C (2021) Segmenter: transformer for semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 7262–7272
15. Sagar A (2021) Vitbis: vision transformer for biomedical image segmentation. In: Clinical image-based procedures, distributed and collaborative learning, artificial intelligence for combating COVID-19 and secure and privacy-preserving machine learning: 10th workshop, CLIP 2021, second workshop, DCL 2021, first workshop, LL-COVID19 2021, and first workshop and tutorial, PPML 2021, held in conjunction with MICCAI 2021, strasbourg, France, 27 September and 1, 2021 October. Proceedings, pp 34–45
16. Lin A, Chen B, Xu J, Zhang Z, Lu G, Zhang D (2022) Ds-transunet: dual swin transformer u-net for medical image segmentation. IEEE Trans Instrument Measure
17. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, Roth HR, Xu D (2022) Unetr: transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 574–584
18. Lin Z, Huang J, Chen Y, Zhang X, Zhao W, Li Y, Lu L, Zhan M, Jiang X, Liang X (2021) A high resolution representation network with multi-path scale for retinal vessel segmentation. Comput Methods Prog Biomed 208:106206
19. Jiang Y, Yao H, Wu C, Liu W (2020) A multi-scale residual attention network for retinal vessel segmentation. Symmetry 13(1):24
20. Yin P, Cai H, Wu Q (2022) Df-net: deep fusion network for multi-source vessel segmentation. Inform Fusion 78:199–208
21. Jiang Y, Wu C, Wang G, Yao H-X, Liu W-H (2021) Mfi-net: a multi-resolution fusion input network for retinal vessel segmentation. Plos one 16(7):0253056
22. Kamran SA, Hossain KF, Tavakkoli A, Zuckerbrod SL, Sanders KM, Baker SA (2021) Rv-gan: segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 34–44
23. Boudegga H, Elloumi Y, Akil M, Bedoui MH, Kachouri R, Abdallah AB (2021) Fast and efficient retinal blood vessel segmentation method based on deep learning network. Comput Med Imaging Graph 90:101902
24. Yu Y, Zhu H (2021) Retinal vessel segmentation with constrained-based nonnegative matrix factorization and 3d modified attention u-net. EURASIP J Image Video Process 2021(1):1–21
25. Sun X, Fang H, Yang Y, Zhu D, Wang L, Liu J, Xu Y (2021) Robust retinal vessel segmentation from a data augmentation perspective. In: International workshop on ophthalmic medical image analysis. Springer, pp 189–198
26. Naveed K, Abdullah F, Madni HA, Khan MA, Khan TM, Naqvi SS (2021) Towards automated eye diagnosis: an improved retinal vessel segmentation framework using ensemble block matching 3d filter. Diagnostics 11(1):114
27. Dash S, Senapati MR, Sahu PK, Chowdary P (2021) Illumination normalized based technique for retinal blood vessel segmentation. Int J Imaging Syst Technol 31(1):351–363
28. Tavakoli M, Mehdizadeh A, Pourreza Shahri R, Dehmeshki J (2021) Unsupervised automated retinal vessel segmentation based on radon line detector and morphological reconstruction. IET Image Process
29. Toptaş B, Hanbay D (2021) Retinal blood vessel segmentation using pixel-based feature vector. Biomedical Signal Process Cont 103053:70
30. Wang J, Zhao Y, Qian L, Yu X, Gao Y (2021) Ear-net: error attention refining network for retinal vessel segmentation. In: 2021 Digital Image Computing: Techniques and Applications (DICTA). IEEE, pp 1–7

31. Zhang J, Zhang Y, Xu X (2021) Pyramid u-net for retinal vessel segmentation. In: ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 1125–1129
32. Wu H, Wang W, Zhong J, Lei B, Wen Z, Qin J (2021) Scs-net: a scale and context sensitive network for retinal vessel segmentation. Med Image Anal 102025:70
33. Palanivel DA, Natarajan S, Gopalakrishnan S (2020) Retinal vessel segmentation using multifractal characterization. Appl Soft Comput 106439:94
34. Zhang M, Yu F, Zhao J, Zhang L, Li Q (2020) Befd: boundary enhancement and feature denoising for vessel segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 775–785
35. Caron M, Touvron H, Misra I, Jégou H, Mairal J, Bojanowski P, Joulin A (2021) Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 9650–9660
36. Zhao H, Li H, Cheng L (2020) Improving retinal vessel segmentation with joint local loss by matting. Pattern Recogn 107068:98
37. Shit S, Paetzold JC, Sekuboyina A, Ezhov I, Unger A, Zhylka A, Pluim JP, Bauer U (2021) Menze, BH: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 16560–16569
38. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B (2004) Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imaging 23(4):501–509
39. Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, Barman SA (2012) An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Trans Biomed Eng 59(9):2538–2548
40. Ding Y, Yu X, Yang Y (2021) Rfnet: region-aware fusion network for incomplete multi-modal brain tumor segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 3975–3984
41. Yang D, Myronenko A, Wang X, Xu Z, Roth HR, Xu D (2021) T-automl: automated machine learning for lesion segmentation using transformers in 3d medical imaging. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 3962–3974
42. Zhang J, Xie Y, Xia Y, Shen C (2021) Dodnet: learning to segment multi-organ and tumors from multiple partially labeled datasets. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1195–1204
43. Reiß S, Seibold C, Freytag A, Rodner E, Stiefelhagen R (2021) Every annotation counts: multi-label deep supervision for medical image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9532–9542
44. Viedma IA, Alonso-Caneiro D, Read SA, Collins MJ (2022) Oct retinal and choroidal layer instance segmentation using mask r-cnn. Sensors 22(5):2016
45. Gour M, Jain S, Sunil Kumar T (2020) Residual learning based cnn for breast cancer histopathological image classification. Int J Imaging Syst Technol 30(3):621–635
46. Wang W, Xie E, Li X, Fan D-P, Song K, Liang D, Lu T, Luo P, Shao L (2021) Pyramid vision transformer: a versatile backbone for dense prediction without convolutions. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 568–578
47. Hinton G, Vinyals O, Dean J et al (2015) Distilling the knowledge in a neural network. vol 2(7), arXiv:1503.02531
48. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S et al (2020) An image is worth 16x16 words: transformers for image recognition at scale. arXiv:2010.11929
49. Zhou Y, Yu H, Shi H (2021) Study group learning: improving retinal vessel segmentation trained with noisy labels. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 57–67
50. Zhan X, Pan X, Dai B, Liu Z, Lin D, Loy CC (2020) Self-supervised scene de-occlusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3784–3792
51. Chen Y, Liu S, Wang X (2021) Learning continuous image representation with local implicit image function. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 8628–8638
52. Owen CG, Rudnicka AR, Mullen R, Barman SA, Monekosso D, Whincup PH, Ng J, Paterson C (2009) Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program. Investig Ophthalmol Vis Sci 50(5):2004–2010
53. Zhuo Z, Huang J, Lu K, Pan D, Feng S (2020) A size-invariant convolutional network with dense connectivity applied to retinal vessel segmentation measured by a unique index. Computer Methods Programs Biomedicine 196:105508
54. Zhou C, Zhang X, Chen H (2020) A new robust method for blood vessel segmentation in retinal fundus images based on weighted line detector and hidden markov model. Computer Methods Programs Biomedicine 187:105231
55. Li X, Jiang Y, Li M, Yin S (2020) Lightweight attention convolutional neural network for retinal vessel image segmentation. IEEE Trans Indust Inform 17(3):1958–1967
56. Li L, Verma M, Nakashima Y, Nagahara H, Kawasaki R (2020) Iternet: retinal image segmentation utilizing structural redundancy in vessel networks. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 3656–3665
57. Lv Y, Ma H, Li J, Liu S (2020) Attention guided u-net with atrous convolution for accurate retinal vessels segmentation. IEEE Access 8:32826–32839
58. Yin P, Yuan R, Cheng Y, Wu Q (2020) Deep guidance network for biomedical image segmentation. IEEE Access 8:116106–116116
59. Wang K, Zhang X, Huang S, Wang Q, Chen F (2020) Ctf-net: retinal vessel segmentation via deep coarse-to-fine supervision network. In: 2020 IEEE 17th international symposium on biomedical imaging (ISBI). IEEE, pp 1237–1241
60. Guo C, Szemenyei M, Yi Y, Wang W, Chen B, Fan C (2021) Sa-unet: spatial attention u-net for retinal vessel segmentation. In: 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, pp. 1236–1242
61. Khan TM, Alhussein M, Aurangzeb K, Arsalan M, Naqvi SS, Nawaz SJ (2020) Residual connection-based encoder decoder network (rced-net) for retinal vessel segmentation. IEEE Access 8:131257–131272

Jia Gu received the master's degree from the School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China, in 2021. She is currently pursuing the doctor's degree with the Division of Computer Science and Engineering, Jeonbuk National University, Jeonju, South Korea. Her research interests include computer vision and medical image processing.

Fangzheng Tian received the master's degree from the School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China, in 2021. He is currently pursuing the doctor's degree with the Division of Computer Science and Engineering, Jeonbuk National University, Jeonju, South Korea. His research interests include computer vision and human body estimation.

Il-Seok Oh received the B.S. degree in computer engineering from Seoul National University, South Korea, in 1984, and the Ph.D. degree in computer science from KAIST, South Korea, in 1992. He is currently a Professor with the Division of Computer Science and Engineering, Jeonbuk National University, Jeonju, South Korea. He was a Visiting Scientist with CENPARMI, Concordia University, Canada, and UCI, USA. He is the author of the books Pattern Recognition, Computer Vision, Machine Learning, and Artificial Intelligence with Python (in Korean Language). His research interests include computer vision and machine learning.