IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 April 2023, accepted 30 April 2023, date of publication 8 May 2023, date of current version 15 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3273597

RESEARCH ARTICLE

AFFD-Net: A Dual-Decoder Network Based on Attention-Enhancing and Feature Fusion for Retinal Vessel Segmentation

XIANG ZIJIAN[1], NING CHUNYU[1], LI MINGYE[2,3], MA KAIZHENG[1], SHI LEMIN[1], WANG WEI[3,4], AND YE GUANSHI[5]

[1]School of Life Science and Technology, Changchun University of Science and Technology, Changchun 130022, China
[2]Department of Information Systems and Business Analytics, RMIT University, Melbourne, VIC 3001, Australia
[3]School of Computing and Information Systems, The University of Melbourne, Parkville, VIC 3010, Australia
[4]Department of Software Systems and Cybersecurity, Monash University, Clayton, VIC 3800, Australia
[5]School of Electrical and Information Engineering, Jilin Agricultural Science and Technology University, Jilin 132109, China

Corresponding authors: Ning Chunyu (ningcy@cust.edu.cn) and Ye Guanshi (yeguanshi@163.com)

This work was supported by the Science and Technology Development Project of Jilin Province, China, under Grant 20220101123JC and Grant 20200404219YY.

ABSTRACT The morphological characteristics of retinal vessels in the fundus serve as the primary basis for diagnosing and assessing the risk of ophthalmic diseases. An effective segmentation scheme for retinal vessels can aid in the early diagnosis and treatment of these diseases, as well as help prevent their progression. However, accurate vessel segmentation is challenging due to the low contrast of fundus images and the complexity of the vessels' morphological structure. To address the low sensitivity and poor generalization ability of the existing methods in vascular extraction, a Dual-decoder Network based on Attention-enhancing and multi-scale Feature Fusion (AFFD-Net) is proposed. AFFD-Net inherits the codec concept of U-Net. To improve the performance of our U-Net model, we made two modifications. Firstly, we reduced the number of convolution kernel filters in each layer, thereby significantly reducing the number of training parameters. This helps to avoid overfitting and improves the model's ability to generalize. Secondly, we added a Multi-scale Feature Extraction (MFE) module and an M/A intermediate decoder to enhance the model's sensitivity. MFE is designed as the first encoding unit of AFFD-Net to obtain rich vascular features in the complex anatomical background and adapt to the large-scale variations of vessels. The M/A intermediate decoder is composed of the Multi-scale Feature Fusion (MFF) module and the Attention-enhancing Hybrid Feature Fusion (AHFF) module. The MFF module integrates deep semantic information and shallow spatial information to ensure that the features at each scale in the middle layer are fully utilized. The AHFF module adaptively fuses the hybrid features at different scales to generate two feature descriptors with different focuses which can improve the expressiveness of the model. AFFD-Net is evaluated on three public databases including DRIVE, STARE, and CHASE_DB1, and the sensitivity values obtained are 84.19%, 84.58%, and 82.62%, respectively. It has higher sensitivity and better generalization ability than other state-of-the-art methods. Compared with classical networks including U-Net, U-Net++, and U-Net3+, AFFD-Net has fewer parameters and higher segmentation accuracy. Our proposed segmentation model exhibits superior performance across a range of metrics, indicating its promising potential for practical applications.

INDEX TERMS Deep learning, feature extraction, image segmentation, neural network, retinal vessel.

I. INTRODUCTION

It is estimated that more than 418 million people worldwide suffer from glaucoma, diabetic retinopathy, age-related macular degeneration (AMD), and other blinding ophthalmic

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

diseases [1]. The early symptoms of these ophthalmic diseases are not obvious and are often ignored by patients, which results in aggravation or even blindness [2]. In clinical practice, ophthalmologists can use morphological features of retinal vessels, such as branching pattern, angle, curvature, width, and length, to determine and evaluate ocular diseases [3]. For example, choroidal neovascularization is the diagnostic basis for AMD, and the diameter of thick and thin vessels is an important indicator for diagnosing microaneurysms [4]. Therefore, the accurate segmentation of retinal vessels is of great significance for the prevention and treatment of ophthalmic diseases.

The segmentation methods of retinal vessels are divided into manual extraction and automatic computer extraction. The former is not only time-consuming, but also affected by subjective factors. While the latter can overcome these problems very well. However, due to the complex structure of retinal blood vessels [5], the low contrast between capillaries and the background, and the uneven illumination in the acquired images [6], the extraction of retinal blood vessels is still a challenging task.

Based on this background, a Dual-decoder Network (AFFD-Net) based on Attention-enhancing and multi-scale Feature Fusion is proposed in this paper for segmenting fundus vessels automatically. The model inherits the encoding and decoding ideas of traditional U-Net [7]. Three interrelated modules are proposed to improve the accuracy and sensitivity of vessel segmentation and reduce the computation complexity. The first is the Multi-scale Feature Extraction (MFE) module which responds to the feature at each scale fully. The second is the Multi-scale Feature Fusion (MFF) module which enhances the relationship between features at each stage. The third is the Attention-enhancing Hybrid Feature Fusion (AHFF) module that emphasizes the perception of tiny vessels. The combination of the three modules makes AFFD-Net more capable of capturing vascular features and the generalization ability of the model is also stronger.

To sum up, the main contributions of this paper are as follows: (1) MFE module that adopts an equivalent receptive field can adapt to the large-scale variations of blood vessels and obtain sufficient vascular information to improve the expressiveness of the network. (2) M/A intermediate decoder formed by the MFF and AHFF modules can make full use of the context of various scale to emphasize the learning of prominent components while suppressing the influence of irrelevant noise and strengthening the perception of micro vessels. (3) Compared with U-Net and its classical variations, AFFD-Net not only exhibits better segmentation performance, but also reduces the number of parameters significantly.

The rest of the paper is organized as follows. Section II reviews the existing methods and strategies for retinal vessel segmentation. Section III presents the description of the principle and framework of the proposed model. Section IV describes the datasets, evaluation metrics and experimental setting used in this paper. Section V analyzes and discusses the performance of the model in detail. The last section shows the conclusions and triggers future work.

II. RELATED WORK

A. RETINAL VESSEL SEGMENTATION

Retinal vessel segmentation is essentially a binary classification problem. Each pixel in the image is assigned to either the vessel pixel or the background pixel [8]. The current segmentation methods for vessel are mainly categorized into two main types: unsupervised and supervised.

The unsupervised approaches construct segmentation model by using filter responses or model-based techniques without any label information [9]. For example, Zana et al. [10] proposed the algorithm combining mathematical morphology with curvature evaluation. Hoover et al. [11] designed an unsupervised method based on a two-dimensional filter, which used both local and global vascular features to segment vessels. Azzopardi et al. [12] achieved direction selectivity by computing the weighted geometric mean, constructed filter bank, and performed thresholding in order to segment vessel tree. Neto et al. [13] utilized spatial correlation and probability statistics to obtain coarse vessel segmentation maps, then refined vessels through curvature analysis and morphological reconstruction. Sazak et al. [14] introduced a bowler-hat transform technique which used morphological features to extract vessel-like structures. Zhao et al. [15] proposed a method based on vascular network topological properties to segment veins and arteries in fundus images. The method adopted the concept of dominant set clustering to transform the classification of blood vessels into a clustering problem. Tariq et al. [16] classified retinal vessels into large or small vessels, then processed them according to their unique characteristics using an enhancement detecting filters that captured tiny vessels through a directional filter bank.

The supervised approaches are to construct a predictive model to complete the segmentation using the labeled information from professional doctors. They are further divided into shallow learning-based methods and deep learning-based methods [17]. In shallow learning-based methods, the basis for feature classification needs to be defined empirically. Marin et al. [18] calculated a 7-D vector consisting of a combination of gray-level and moment invariant-based features for vascular pixel representation and realized the classification of vascular pixel by a neural network. Fraz et al. [19] constructed an ensemble system of bagged and boosted decision trees, which obtained a feature vector by utilizing orientation analysis, morphological transformation, line strength measurement, and Gabor filter. They completed the vascular segmentation of healthy and pathological retinal images. Orlando et al. [20] proposed a fully connected conditional random field model for extracting the thin and elongated structures of blood vessels, and the parameters of the model were acquired by a structured output support vector machine automatically. Srinidhi et al. [21] decomposed the complex vessel tree into multiple sub-trees through local information, then a random forest classifier was taken to train a set of handcrafted features in order to classify these sub-trees into arteries and veins. Topta et al. [22] created an 18-dimensional feature vector for each pixel from five different feature groups, then the vector was fed into the artificial neural network for training.

Shallow learning-based methods require human intervention because of the definition of classification basis, so subjective factors will lead to certain bias in the classification results. In recent years, with the continuous improvement of computer performance, deep learning has been widely used in retinal blood vessel segmentation tasks. The segmentation results are more accurate because of its ability to automatically learn image features from massive datasets. Jiang et al. [23] used fully convolutional network combined with transfer learning to achieve retinal vessel segmentation, overcoming the problem of insufficient data. Yan et al. [17] improved U-Net model and presented a new segment-level loss structure to emphasize the consistency of thick and thin vessels during the training process by combining segment-level loss and pixel-level loss. Jin et al. [9] utilized deformable convolutions to capture various shapes and scales by adaptively adjusting the receptive fields according to vascular characteristics. Henda et al. [24] replaced the convolution layer in the U-Net with the LCM layer. LCM layer was composed of a 3 × 3 depth wise convolution and a 1 × 1 convolution, which greatly reduced the number of parameters while ensuring the feature extraction capability. Wang et al. [25] suggested a three-decoder network for coarse and fine vessels and vessel contours and fused all decoder outputs to obtain the final segmentation. Li et al. [26] proposed a U-shaped network based on soft attention. The network adopted a dual-direction attention module to build global dependencies and utilized selective kernel units instead of standard convolutions to generate multi-scale features. Ghosh et al. [27] designed a novel model which utilized a ranking support vector machine (rSVM) to extract features, then input the features into the core CNN framework. The model reduced the training cost and achieved good vessel segmentation results. Zhao et al. [28] presented a new nested U-shaped attention network that connected encoder-decoder branches through nested skip-connection pyramid architecture to extract rich retinal details. Morano et al. [29] suggested a new loss function named "binary cross-entropy by 3" according to the vessel types, and adopted a fully convolutional neural network (FCN) to segment arteries, veins and vascular tree. Zheng et al. [30] presented a self-attention U-shaped network (SAUNet) to roughly localize the retinal vascular structure, then cascaded SAUNet with a residual self-attention U-Net to improve the expression of local features. Zhang et al. [31] proposed the Bridge-Net model, which adopted U-Net to obtain the corresponding features of patches and RNN to fuse probability maps so that the target region contained rich context information.

Recent published works in this field have shown that segmentation models for retinal vessel are mainly designed based on U-Net. The classical variants of U-Net include Res-UNet [32], Dense-UNet [33], Attention-UNet [34], nnUNet [35], LadderNet [36], R2U-Net [37], U-Net++ [38], U-Net3+ [39], etc. Res-UNet and Dense-UNet improve the utilization of features through residual connection and dense connection respectively. Attention-UNet adds attention mechanism to U-Net to increase the model's learning for regions of interest. LadderNet adopts multiple pairs of encoder-decoder branches, and has skip connections between every pair of adjacent decoder and decoder branches in each level to alleviate the pressure on high-level representation of details in the model. R2U-Net uses recurrent residual convolutional units to accumulate features to ensure better feature representation for segmentation tasks. The nnUNet refines the segmentation results through U-Net cascading. Although these classical variants have good performance in segmentation tasks, they do not make good use of multi-scale features. In order to make up for this deficiency, U-Net++, U-Net3+ have been proposed successively. From the perspective of network structure, U-Net++ and U-Net3+ both improve the utilization of features at each scale. However, U-Net++ up-samples features at each scale in the encoding process separately and continuously, and features of the same scale are also reused repeatedly. These lead to a huge number of parameters and overfitting when training on complex or small size datasets. U-Net3+ improves the model's utilization of features by concatenating codec features, but the number of feature channels in the decoding stage are multiplication. Although the authors alleviate the surge in the number of parameters by reducing the number of convolution layers per decoding unit, this also reduces the interdependencies of the feature map channel of the decoder, and is unable to effectively filter out invalid features transmitted from shallow layers.

To address the shortcomings of these classical U-Net variants, AFFD-Net is proposed in this paper. In AFFD-Net, the M/A intermediate decoder composed of MFF and AHFF modules can compensate the semantic difference between codecs. By iteratively applying the MFF module, AFFD-Net can acquire intermediate layers' features at various scales and compress the number of feature channels simultaneously, which overcomes the problem of channel multiplication caused by the concatenation and upsampling module used in U-Net. Therefore, AFFD-Net not only has higher segmentation accuracy, but also is far lower than classical U-Net variants in the number of parameters and time consuming.

B. MULTI-SCALE CONTEXT EXTRACTION

In the image processing tasks, features of single-scale receptive field cannot fully express all kinds of features. Effective extraction and utilization of context at each scale can significantly improve the model performance. Therefore, He et al. [40] extended the bag-of-words model and proposed spatial pyramid pooling (SPP), which enabled the network to extract features at different scales by inputting images with different proportions. Tong et al. [41] applied this idea to the retinal vascular segmentation and achieved multi-scale feature extraction by feeding original images of different sizes into different layers. Chen et al. [42] proposed atrous spatial pyramid pooling (ASPP) bases on SPP, which used atrous convolution to capture multi-scale context information. Zheng et al. [30] and Wu et al. [43] adopted the idea to the task of retinal vessel segmentation and obtained good results. Szegedy et al. [44] proposed the model named inception V1, in which the inception module designed convolution of three different scales to implement multi-scale feature extraction. Yan [3], Mohamed [45], Yan [17], Tang et al. [46] also adopted similar ideas to realize the multi-scale feature extraction of retinal vessels. In addition, Jin et al. [9] obtained deformable receptive fields by utilizing deformable convolution to capture vessel information at various scales. Hu et al. [47] adopted a five-stage network model to obtain multi-scale features. Li et al. [26] adopted selective kernel (SK) units instead of standard convolutions to obtain multi-scale features produced by soft attention. Referring to the idea of replacing large convolutional kernels by small ones proposed by Simonyan [48], the MFE module is designed by improving the inception structure. MFE achieves the multi-scale feature extraction by using multiple 3 × 3 convolutional kernels in series. This module can obtain rich scale receptive fields and capture multi-scale context information while reducing the number of parameters.

C. FEATURE FUSION

The shallow and deep layers in the deep segmentation model yield different semantic information. Shallow features are used for detail reconstruction and deep ones for the locating of the whole framework. In order to improve the model accuracy, it is necessary to integrate the features of different depths. The obvious characteristic of the classic U-Net model is adopting skip connection to achieve the fusion of the shallow and deep features. Subsequently, Yan [17] and Xiang et al. [49] improved the traditional U-Net network by using jump connections and achieved good vascular segmentation effect. To compensate for the semantic gap between deep and shallow features, Zhou et al. [38] proposed the U-Net++ model which up-sampled the feature maps of different layers and implemented the fusion of the features by connection operation. Then Zhao et al. [28] introduced the channel attention mechanism in a similar way to fuse different layer features. Although the model achieved good results for blood vessel segmentation, the number of parameters was very large. It is worth mentioning that Lin et al. [50] proposed the FPN (feature pyramid networks) model in order to enhance the detection for small objects, which embedded the low-level features into high-level ones. Similarly, this idea has also been applied to the field of retinal blood vessel segmentation. Hu et al. [47] up sampled the features of different levels at each stage and fused the features through element summation. Wu et al. [43] performed sigmoid operations on the feature maps obtained from decoding units at each level, then the results at all levels were embedded into high-level features in the order from low to high. Besides, Yan [3], Wang [25] et al. obtained the segmentation maps of thick and thin blood vessels and adopted concatenation operation and various convolution weighting to achieve feature fusion and refinement. In order to capture the complex anatomical semantics of retinal vessels and improve the segmentation accuracy of the model, two feature fusion mechanisms are proposed in this paper. One is to merge the features of middle layers by continuously up-sampling and aggregating the context information. The other is to realize adaptive fusion according to the importance of features.

D. ATTENTION MECHANISM

The introduction of attention mechanisms has led to their increasing adoption in U-Net networks. Currently, most attention modules in U-Net are based on spatial or channel attention models such as squeeze-and-excitation (SE) [51], convolutional block attention module (CBAM) [52], efficient channel attention (ECA) [53], and the attention gate [54].

Inspired by the SE and the CBAM module, Zhao et al. [55] proposed a spatial attention mechanism and a channel attention mechanism, respectively. According to the task characteristics, they introduced the spatial attention mechanism to the back end of the first encoding unit and the back end of the last decoding unit of U-Net, and integrates the channel attention mechanism to the back end of the last encoding unit. Trinh et al. [56] proposed a new spatial channel attention mechanism based on the attention gate and CBAM module. The mechanism is the same as CBAM in that both Maxpooling and Avgpooling are used to acquire aggregation features in the spatial and channel dimensions. The difference is that Trinh et al. use 7 × 7 convolution to fuse these acquired aggregation features. Guo et al. [57] improved the SE module and proposed the Modified Efficient Channel Attention (MECA) module. On this basis, they introduced the residual idea and designed the Channel Attention Double Residual Block (CADRB) module. This module is used to build deeper networks to obtain more complex semantic features. Li et al. [58] integrated the attention gate and SE module to design a new channel attention mechanism and incorporated it into U-Net as part of the skip connection. Zhang [59] et al. added a set of dynamic coefficients to the attention gate to scale the weight coefficients of the spatial information of the attention features. Chen et al. [60] proposed a new spatial attention mechanism with a structure similar to the attention gate. However, the mechanism uses only encoded features to obtain feature descriptors.

The AHFF module proposed in this paper is based on the attention gate, and uses the feature aggregation idea of Maxpooling in CBAM. However, the input of AHFF is the multi-scale features acquired from MFF module, the attention feature map obtained contains richer semantic information.

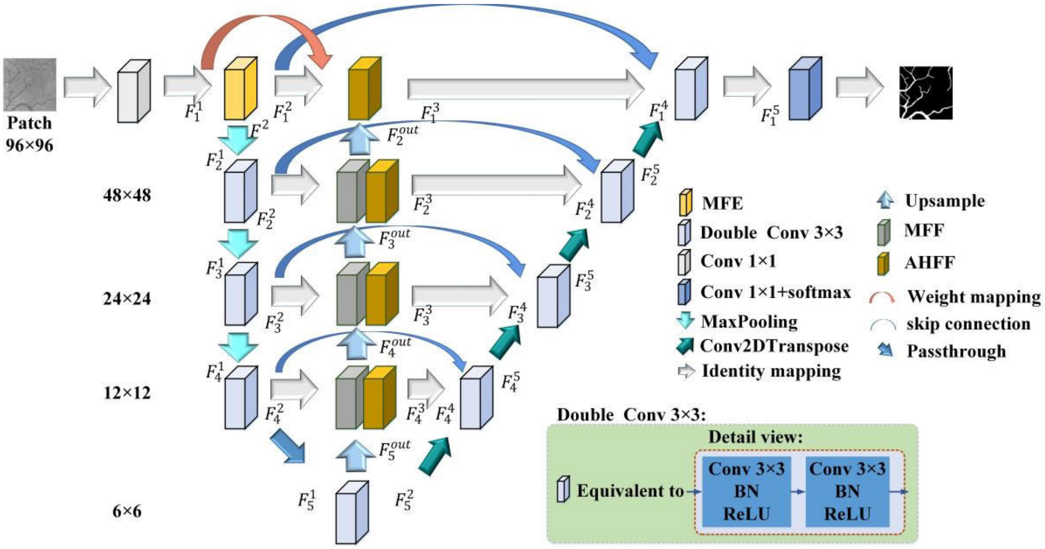

FIGURE 1. Architecture of AFFD-Net.

III. METHODOLOGY

A. MODEL ARCHITECTURE

The proposed AFFD-Net model is an improvement based on U-Net. First, the first encoding unit of U-Net is replaced by the MFE module to extract rich vascular features in a complex anatomical background. Secondly, the M/A intermediate decoder formed by MFF module and AHFF module is integrated into the proposed AFFD-Net model. The purpose of MFF is to aggregate features at each scale of the middle layers, to provide semantic information at different depths for AHFF, and to enhance the localization of the vessel contour. AHFF module guides the adaptive fusion of hybrid features while obtaining the maximum response of micro vessels to produce accurate segmentation results. Finally, the passthrough layer takes the place of the pooling layer to ensure that the semantic features extracted by deep network are not lost due to down-sampling. Fig.1 illustrates the overall architecture of the proposed AFFD-Net model.

The original retina images need to be preprocessed and then cut into patches of size 96 × 96 before inputting into the model. In Fig.1, $F_n^1$and F_n^2 denote the input and output feature of the encoding unit respectively. F_n^3 is the output of the M/A intermediate decoder. F_n^4 and F_n^5 are the input and output of the decoding unit respectively. The subscript $n \in [1, 5]$ represents the nth scale, the resolutions of the feature maps from the 1st to the 5th scale are 96 × 96, 48 × 48, 24 × 24, 12 × 12, 6 × 6, and the number of feature channels are 32, 64, 64, 64, and 128, respectively. F_n^{out} is the multi-scale hybrid feature produced by MFF module, where $n \in [2, 4]$ represents the nth scale, and the number of channels for each scale is 128. The feature F_5^{out} is equivalent to F_5^2, and it is not a hybrid feature.

B. MFE MODULE

The width of blood vessels varies from one pixel to a dozen pixels. The traditional U-Net model has a relatively fixed receptive field and responds varyingly to features of different sizes, which can lead to detection errors for too-thick and too-thin vessels or fragment segmentation [43]. To overcome this problem, Mohamed et al. [45] employed three convolution kernels of different sizes to achieve multi-scale context extraction. Wu et al. [43] utilized atrous convolution to obtain features under different receptive fields. However, convolution of different sizes does perform well in segmentation and it significantly increases the number of parameters. The atrous convolution although it can theoretically expand the receptive field, suffers from the problem of feature loss [61]. In order to solve the shortcomings of the above two methods and improve the extraction ability for vessels of different sizes, the proposed MFE module adopts the idea of reusing the features acquired by each convolutional layer, and stacks multiple small convolutional kernels to obtain different receptive fields, which can enhance the recognition for vessels of various sizes and finally achieve multi-scale feature extraction.

As shown in Fig.2, the MFE simulates 5 × 5 and 7 × 7 kernel through three 3 × 3 convolution kernels, and the equivalent receptive fields are shown in the gray dashed box. Since MFE is designed to increase the response to small vessels, which are expressed almost exclusively in shallow layers, the MFE is only used at the beginning of the encoder. Meanwhile, in order to avoid feature confusion and affect the model convergence, the next encoding unit only uses the feature map F^2under 7 × 7 receptive field.

C. MFF MODULE

The traditional U-Net only integrates the feature maps of the same resolution through skip connection, and thus a large number of intermediate layer features are neglected. For vessel segmentation, shallow features contain rich spatial information crucial for tiny vessel localization and supplementation of detail information, but lack the necessary

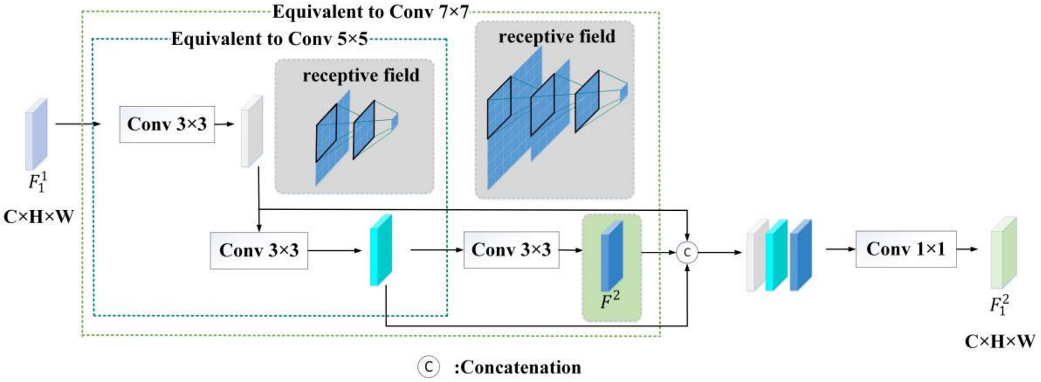

FIGURE 2. The flowchart of proposed MFE module.

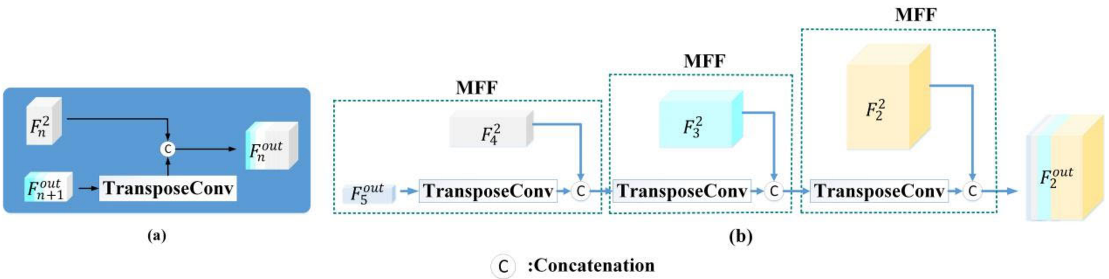

FIGURE 3. (a) The flowchart of proposed MFF module. (b) The acquisition process of hybrid features.

semantic information to precisely separate vessel contours. The deep features act on the contrary. To improve the segmentation potential of U-Net, we propose the MFF module which fuses semantic features of different scales at each stage through feature reuse and provides abundant contextual information for the AHFF module. Fig.3 (a) shows the structure of it. The hybrid feature for each scale F_n^{out} is given as:

$$F_n^{out} = H[F_n^2, \mu(F_{n+1}^{out})] \tag{1}$$

where$\mu(\ )$ denotes the transpose convolution and $H[,]$ the concatenation operation. The feature mapF_{n+1}^{out}should be performed up-sampling once to make its size consistent with the encoder feature F_n^2. The upsampling method of transpose convolution adopted by MFF not only restores the feature map size, but also compresses the number of feature channels. Therefore, it can avoid the multiplication phenomenon of feature channel number caused by the iteration of the module.

As illustrated in Fig.3 (b), the final hybrid feature F_2^{out} obtained after layer by layer fusion of MFF contains the intermediate layer features, in which the proportion of the number of channels for F_2^2, F_3^2, F_4^2, and F_5^2 is 4:2:1:1 (Here, $F_5^2 = F_5^{out}$, see section III. A.).

D. AHFF MODULE

To enhance the perception for microvascular, the AHFF module (Fig.4) is proposed which consists of the attention gate, the attention-enhancing module, and the detail supplement unit.

1) ATTENTION GATE BASED ON ADAPTIVE MULTI-SCALE FEATURE FUSION

Different from the traditional attention gate, the proposed fuses the hybrid feature F_{n+1}^{out} and F_n^2 adaptively according to the importance of features, and generates the feature descriptor M_1 $\in \mathrm{R}^{1\times \mathrm{H}\times \mathrm{W}}$ (where 1 represents the number of channels, H the height, and W the width). M_1can simulate the correlation of spatial features among channels and emphasize the focus on the overall structure of vascular tree. $M_1$is formulated by Eq.(2).

$$M_1 = \mu(\sigma(f^{1\times1}(\delta(w_1 \cdot f^{2\times2}(F_n^2 \oplus w_2 \cdot \mu_1(f^{1\times1}(F_{n+1}^{out}))))) \tag{2}$$

where $f^{1\times1}$ denotes a 1×1 convolution, $f^{2\times2}$ denotes a 2×2 convolution with stride $=2$, $\mu_1(\ )$ is a transpose convolution with kernel of 3 and stride $=1$, $\oplus$ denotes the element-wise sum, $\delta(\ )$ and $\sigma(\ )$ denote ReLU activation function and sigmoid function respectively, $\mu(\ )$ denotes up-sampling, w_1 and $w_2$are weights ($0< \ w \ <1$) which are continuously updated through backpropagation during the training process.

2) ATTENTION-ENHANCING MODULE

The proposed attention-enhancing module obtains the maximum response of spatial location at channel dimension

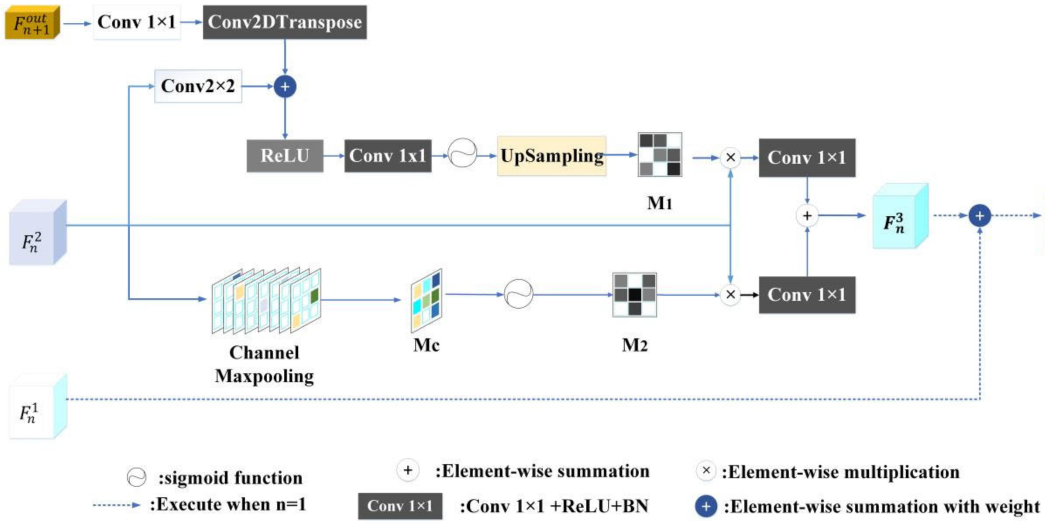

FIGURE 4. The flowchart of proposed AHFF module.

by squeezing the encoded feature F_n^2, and produces the 2-D attention coefficient M_c. $M_c$is formulated by Eq.(3). Then the sigmoid function is adopted to construct the feature descriptor $M_2 \in \mathrm{R}^{1\times \mathrm{H}\times \mathrm{W}}$. M_2enlarges the features of thin vessels which makes the model focus on the perception of it while not affecting the expression of thick vessels. $M_2$is calculated as follow:

$$M_c(i,j) = \max\left\{F_n^{21}(i,j), F_n^{22}(i,j), \ldots\ldots, F_n^{2C_2}(i,j)\right\}, \quad 1 \le i \le H, \quad 1 \le j \le W \tag{3}$$

$$M_2 = \sigma(M_c) = 1\big/(1+\exp(M_c)) \tag{4}$$

where $\mathrm{F}_\mathrm{n}^{2\mathrm{C}_2}$ (i, j) denotes the probability value of the feature map F_n^2 at (i, j) on the C_2 channel.

The feature descriptors M_1 and M_2 emphasize the semantic information of F_n^2, and produce two sets of feature maps with different focus which are fused to form F_n^3. F_n^3 is defined as:

$$F_n^3 = \hat{f}^{1\times1}(F_n^2 \otimes M_1) \oplus \hat{f}^{1\times1}(F_n^2 \otimes M_2),\ \mathrm{n} \in [2, 4] \tag{5}$$

where $\otimes$ is the element-wise multiplication, $\hat{f}^{1\times1}$is a 1 × 1 convolution with ReLU activation function and batch normalization. When n ∈ [2, 4], F_n^3 is fed into the decoder. When $n = 1$, F_n^3 is obtained by the detail supplement unit.

3) DETAIL SUPPLEMENT UNIT

In order to compensate for the loss of detail features due to network depth and enhance the ability of the model to reconstruct tiny vessels, the proposed detail supplement unit integrates the detail information and attention feature map by weight mapping. Since the details are fully expressed only in the shallow layer, this unit is only executed when $n = 1$. The calculation of F_n^3 is as follows:

$$\begin{aligned} F_n^3 &= (w_3 \cdot (\hat{f}^{1\times1}(F_n^2 \otimes M_1) \oplus \hat{f}^{1\times1}(F_n^2 \otimes M_2))) \oplus (w_4 \cdot F_n^1) \\ &= (w_3 \cdot F_n^3) \oplus (w_4 \cdot F_n^1),\ n = 1 \end{aligned} \tag{6}$$

where $w_3$and w_4 are weights ($0< w <1$) which are continuously updated through backpropagation during the training process.

E. PASSTHROUGH LAYER

The 2×2 Max Pooling used at the bottom layer of the encoder in traditional U-Net will lose three-quarters of the semantic information of the deep layers. These missing features cannot be further abstracted, which influences the accurate localization of the vascular skeleton. To solve the problem, the proposed segmentation model replaces this pooling layer with the Passthrough layer [62], which connects the high and low resolution elements by stacking the adjacent elements to different channels (rather than spatial locations). Therefore, the model can retain the deep semantic features completely. Fig.5 illustrates the diagram of passthrough layer.

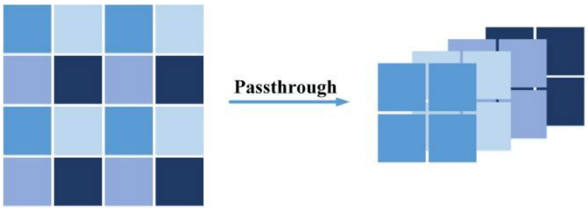

FIGURE 5. The schematic diagram of passthrough layer.

IV. EXPERIMENT SETTING

A. DATASETS

We evaluate the proposed model on three published datasets: STARE, DRIVE and CHASE_DB1. STARE consists of 20 fundus images provided by the University of California, out of which 10 are abnormal and the remaining 10 are normal. Each image has a spatial resolution of 700 × 605. Manual annotations of the vascular by two experts' manual

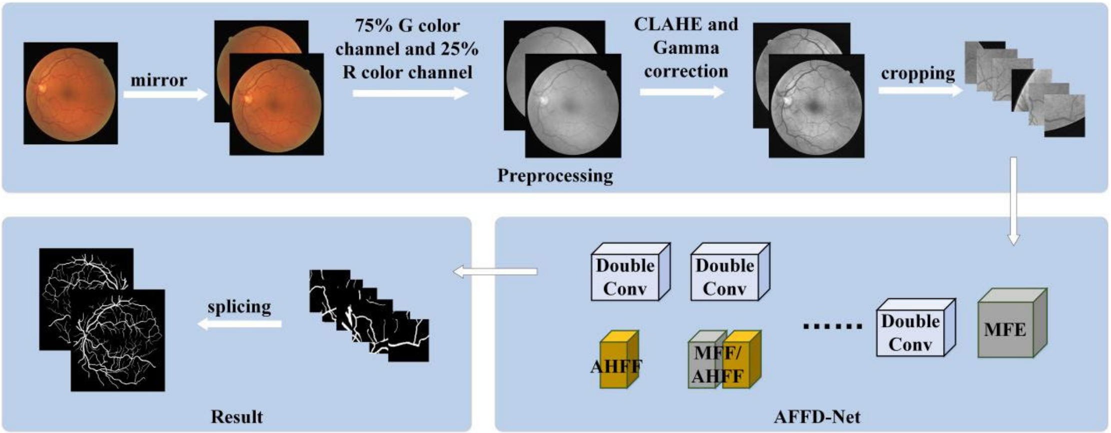

FIGURE 6. The detailed flow of segmentation task.

are available, and the annotations of the first expert are used as the ground truth. The 5-fold cross validation method is used for dividing the dataset into training and testing sets. DRIVE comprises 40 colored fundus images collected from a diabetic retinopathy screening program in the Netherlands. Among them, 7 images show early diabetes retinopathy, and the remaining 33 have no pathological manifestations. Each image has a spatial resolution of 565 × 584. In this study, the first 20 images are used for testing and the remaining 20 for training. CHASE_DB1 comprises 28 fundus images of 14 school-age children provided by Kingston University London. The resolution of each image is 999 × 960. Two experts manually annotated the vascular, and the annotations of the first expert are used as the ground truth. The first 20 images are used for training and the remaining 8 images for testing. Table 1 provides further details on these datasets.

TABLE 1. Datasets details.

Dataset	DRIVE	STARE	CHASE_DB1
Quantity	40	20	28
Train-test split	20-20	15-5	20-8
Resolution	565×584	700×605	999×960
Format	TIF	PPM	JPG
Train total patches	20000	15000	20000
Test total patches	184240	61710	156565

B. DETAILED PROCESS OF RETINAL VESSEL SEGMENTATION

The overall flow of the proposed segmentation scheme is shown in Fig.6. The retinal image is preprocessed to form 96×96 patches, which are fed into AFFD-Net model for training. Then all patches are spliced into the final segmentation result.

The preprocessing block includes the following operations: expanding the data in the training set twice by image mirror; fusing the green and red channels in the proportion of 75% and 25%; enhancing the contrast between vessels and background by contrast-limited adaptive histogram equalization (CLAHE) and gamma correction (gamma=1.2); cropping each image in training set into 500 patches of 96 × 96 randomly, while cropping image in testing set from left to right and top to bottom.

C. IMPLEMENTATION DETAILS

The AFFD-Net model is trained on Tensorflow framework, and Table 2 shows the hyper-parameters for segmentation. The formulas of the loss function and the learning rate decay are as follows:

$$\begin{aligned} &binary\_crossentropy\left(y,\hat{y}\right) \\ &\quad = -\sum_{i=1}^{outputsize}\left(y_i\cdot\log\hat{y}_i+(1-y_i)\cdot\log(1-\hat{y}_i)\right) \end{aligned} \tag{7}$$

$$lr = lr * \frac{1}{1 + decay * iterations} \tag{8}$$

where y_{i} represents label value and $\hat{y}_i$ predictive value, *decay* denotes decay rate and *iterations*for the number of iterations.

TABLE 2. Hyper-parameters for segmentation.

Hyper-parameters	Value
Epoch	20
Batch size	25
Optimizer	Adam
Initial learning rate	0.001
decay	0.01

D. EVALUATION METRICS

Five metrics, including accuracy (ACC), sensitivity (SE), specificity (SP), F1-score (F1) and area under the receiver

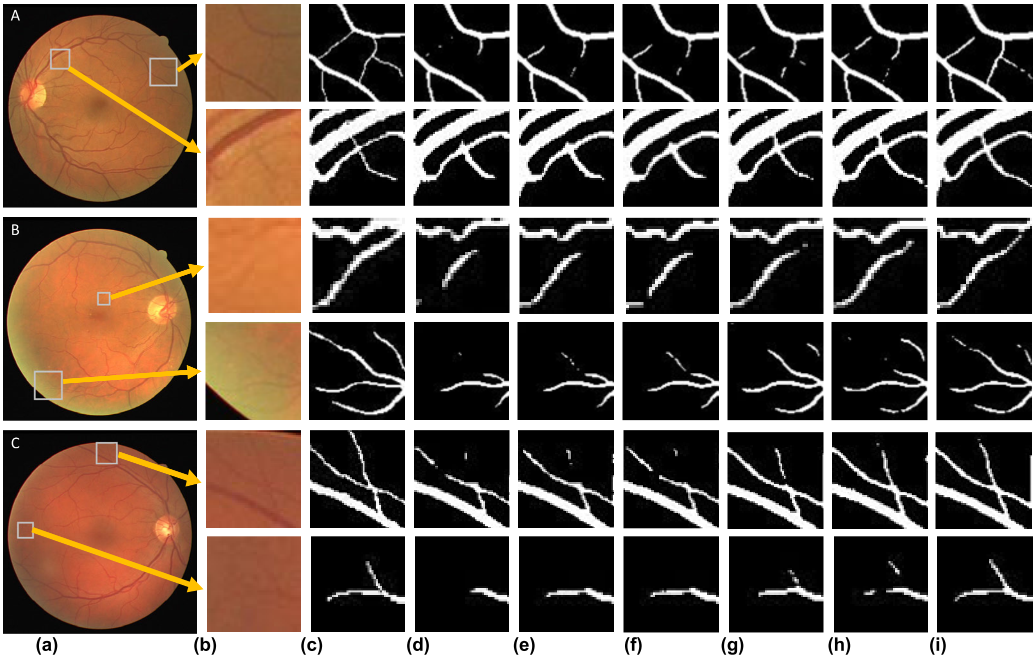

FIGURE 7. Comparison of details in ablation studies. (a) original image, (b) detailed view, (c) ground truth, (d) baseline, (e) baseline+M, (f) baseline+P, (g) baseline+M/A, (h) baseline+M+M/A, (i) baseline+M+M/A+P.

operating characteristic curve (AUC), are employed to evaluate the performance of the proposed AFFD-Net. The formulas are as follow:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

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

$$SP = \frac{TN}{TN + FP} \tag{11}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

where FN, FP, TN and TP are derived from the confusion matrix, and denotes false negative, false positive, true negative and true positive respectively.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. ABLATION STUDIES

The AFFD-Net model represents an improvement upon the widely-used U-Net architecture. Firstly, to reduce the number of training parameters without sacrificing performance, the number of convolutional filters in each layer of the U-Net architecture has been decreased. This modified U-Net serves as the baseline for the AFFD-Net model. Afterward, it replaces its first encoding unit with the MFE (denoted as baseline+M), and adds the intermediate decoder consisting of the MFF and AHFF modules (denoted as baseline+M+M/A). At the same time, to ensure that the deep abstract semantic features are not lost due to downsampling, the passthrough layer is introduced to obtain the final AFFD-Net model (denoted as baseline+M+M/A+P). To verify the effectiveness of the proposed modules, ablation experiments are conducted on DRIVE dataset in this section. Fig.7 exhibits the visual illustration of test examples, and Table 3 shows the quantitative comparisons of these methods. Among them, baseline+M/A represents that only the M/A intermediate decoder is added to the baseline network, and baseline+P represents adding the passthrough layer only.

1) EFFECTIVENESS OF mFE MODULE

The MFE module is designed to extract multi-scale context information to obtain richer information on vessel details.

Comparing the columns (d) and (e) in Fig.7, it can be seen that the baseline can hardly identify the tiny vessels with low-contrast. However, the expression of baseline+M on tiny vessels has been significantly improved, with metrics SE, ACC, F1, and AUC, increased by 0.51%, 0.02%, 0.20%, and 0.07%, respectively. The fact proves that MFE can improve the recognition sensitivity of model to different scale vessels, and play a positive role in retinal vessel extraction task.

2) EFFECTIVENESS OF M/A INTERMEDIATE DECODER

The M/A intermediate decoder aims to focus on the prominent components of the target and enhance the learning ability of microvessels. Comparing the two columns (d) and (g) in Fig.7, it is obvious that baseline+M/A can get more complete vascular structure, and can even identify some vessels hard to find by naked eye. The metrics SE, ACC, F1, and AUC, increased by 2.45%, 0.12%, 1.17% and 0.30%, respectively. Columns (e) and (h) of Fig.7 show that baseline+M+M/A

TABLE 3. Result indicators of ablation studies.

Method	SE(%)	SP(%)	ACC(%)	F1(%)	AUC(%)
baseline	81.06±0.21	98.08±0.14	96.59±0.04	80.63±0.24	98.02±0.08
baseline+M	81.57±0.16	98.06±0.09	96.61±0.06	80.83±0.18	98.09±0.07
baseline+P	81.68±0.18	**98.12±0.25**	96.65±0.04	80.97±0.29	98.11±0.06
baseline+M/A	83.51±0.26	98.02±0.11	96.71±0.04	81.80±0.23	98.32±0.05
baseline+M+M/A	84.02±0.19	98.05±0.13	96.75±0.05	81.89±0.31	98.36±0.07
baseline+M+M/A+P	**84.19±0.24**	**98.12±0.21**	**96.81±0.07**	**82.17±0.43**	**98.41±0.08**

TABLE 4. Performance (mean ± standard deviation) of our model on three datasets.

Dataset	SE(%)	SP(%)	ACC(%)	F1(%)	AUC(%)
DRIVE	84.19±0.24	98.12±0.21	96.81±0.07	82.17±0.43	98.41±0.08
STARE	84.58±8.18	98.45±0.44	97.40±0.22	83.16±2.57	98.79±0.68
CHASE_DB1	82.62±0.31	98.34±0.16	97.44±0.09	80.86±0.26	98.76±0.05

makes a considerable improvement in vascular perception, and the above metrics increased by 2.45%, 0.14%, 1.06%, 0.27%, respectively. It can be concluded that the M/A intermediate decoder can reduce the semantic difference between feature maps and improve the model's perception for tiny vessels. It is quite effective to recognize and locate the vascular structure.

3) EFFECTIVENESS OF PASSTHROUGH LAYER

The two columns (d) and (f) in Fig.7 demonstrate that baseline+P is more accurate in the extraction of blood vessels, and the metrics are improved by 0.62%, 0.03%, 0.06%, 0.34%, 0.09%, respectively. It can be seen from Fig.7 (h) and (i) that the vessels extracted by baseline+M+ M/A+P are more coherent, and the perception of fuzzy small vessels is also stronger. In addition, the case of the second row of group A indicates the model is more accurate in locating the main vessels and closer to ground truth.

The ablation experiments show that each proposed module is effective and can improve the overall performance of the network. Compared with the baseline network, the metrics (SE, SP, ACC, F1, and AUC) of AFFD-Net have an increase of 3.13%, 0.03%, 0.22%, 1.54%, and 0.39%, respectively. This indicates that AFFD-Net outperforms the baseline in segmentation and has a stronger ability to discriminate vascular.

B. SEGMENTATION PERFORMANCE ON EACH DATASET

The average values of the performance metrics by AFFD-Net on three datasets are presented in Table 4. The AUC of AFFD-Net under the optimal conditions reaches 98.49%, 99.47%, and 98.81%, respectively, and the ROC curves are shown in Fig.8.

Fig.9 shows the segmentation results obtained by AFFD-Net on three datasets. The pixels with wrong segmentation are marked. Blue indicates that the vascular pixel is judged as background (FN) and red indicates the background pixel is judged as vascular (FP). As can be observed, AFFD-Net to correctly classify all pixels despite excelling at vascular extraction. It is worth noting that the blue pixels are mostly from the challenging regions of vessels such as microvessles and cross vessels, where even the manual labelling of blood vessels by experts may vary.

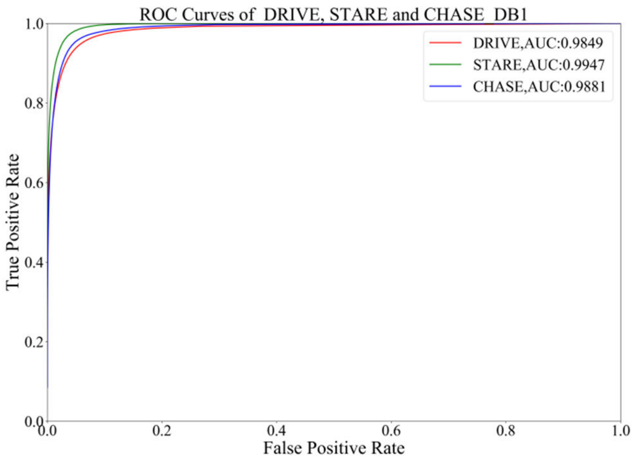

FIGURE 8. The ROC curves on DRIVE, STARE and CHASE_DB1 datasets in the best case.

C. COMPARISON AGAINST EXISTING METHODS

To demonstrate the superiority of the proposed AFFD-Net model, this section compares it with other state-of-the-art methods. The statistics on the three datasets are summarized in Table 5, 6, and 7.

On the DRIVE dataset (Table 5), the ACC, SE, and AUC values of AFFD-Net achieves 96.81%, 84.19%, and 98.41%, respectively, and higher than the suboptimal values: ΔACC$=$ 0.05% (Zheng et al. [30]), ΔSE $=$ 0.19% (Topta et al. [22]), ΔAUC $=$ 0.18% (Wang et al. [25]). Although the SP value is 0.37% lower than the optimal value (Morano et al. [29]), the other metrics far exceed it, which indicates the superiority of the proposed model in terms of overall performance.

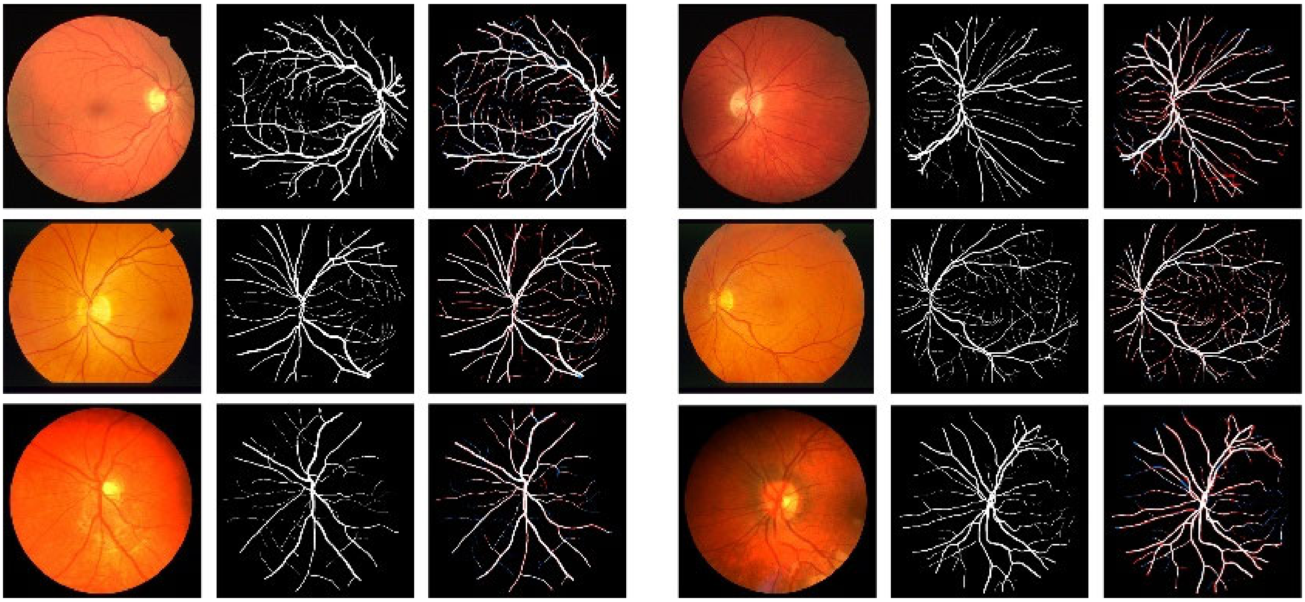

FIGURE 9. Segmentation results on three datasets. The first column to the third column of each group of images: Fundus images, ground-truth and segmentation results respectively. The first row to the third row: DRIVE, STARE and CHASED_DB1 datasets respectively.

TABLE 5. Performance of our model for retinal vessel segmentation on DRIVE dataset.

Model	Year	ACC(%)	SE(%)	SP(%)	AUC(%)
Li et al. [63]	2015	95.27	75.69	98.16	97.38
Orlando et al. [20]	2017	—	78.97	96.84	—
Zhang et al. [64]	2017	94.66	78.61	97.12	97.03
Hu et al. [47]	2018	95.33	77.72	97.93	97.59
Yan et al. [17]	2018	95.42	76.53	98.18	97.52
Jin et al. [9]	2019	95.66	79.63	98.00	98.02
Yan et al. [3]	2019	95.38	76.31	98.20	97.50
Sazak et al. [14]	2019	95.90	71.80	98.10	94.60
Wang et al. [25]	2020	95.81	79.91	98.13	98.23
Xiang et al. [49]	2021	95.68	79.21	98.10	98.06
Yang et al. [65]	2021	95.79	83.53	97.51	—
Morano et al. [29]	2021	95.45	75.42	**98.49**	97.81
Topta et al. [22]	2021	96.18	84.00	97.16	—
Tariq et al. [15]	2022	96.10	81.25	97.63	—
Zheng et al. [30]	2022	96.76	83.40	98.10	97.58
Our	2022	**96.81**	**84.19**	98.12	**98.41**

TABLE 6. Performance of our model for retinal vessel segmentation on STARE dataset.

Model	Year	ACC(%)	SE(%)	SP(%)	AUC(%)
Li et al. [63]	2015	96.28	77.26	98.44	98.79
Orlando et al. [20]	2017	—	76.80	97.38	—
Zhang et al. [64]	2017	95.47	78.82	97.29	97.40
Hu et al. [47]	2018	96.32	75.43	98.14	97.51
Yan et al. [17]	2018	96.12	75.81	98.46	98.01
Jin et al. [9]	2019	96.41	75.95	**98.78**	98.32
Yan et al. [3]	2019	96.38	77.35	98.57	98.33
Sazak et al. [14]	2019	96.20	73.00	97.90	96.20
Wang et al. [25]	2020	96.73	81.86	98.44	**98.81**
Xiang et al. [49]	2021	96.78	83.52	98.23	98.75
Yang et al. [65]	2021	96.26	79.46	98.21	—
Ghosh et al. [27]	2021	96.38	84.16	98.14	—
Topta et al. [22]	2021	94.56	63.08	98.24	—
Tariq et al. [15]	2022	95.86	80.78	97.21	—
Zheng et al. [30]	2022	97.28	83.04	98.62	98.76
Our	2022	**97.40**	**84.58**	98.45	98.79

On the STARE dataset (Table 6), the ACC and SE values of our model are highest, 97.40% and 84.58%, respectively. Although the AUC value is slightly lower than Wang et al. [25] (98.79% Vs 98.81%), it can be seen that the performance of ours on DRIVE (Table 5) and CHASE_DB1 (Table 7) is better than that of [25], which proves AFFD-Net is more stable.

On the CHASE_DB1 dataset (Table 7), all the metrics are the optimal values, and higher than the suboptimal method (Wang et al. [25]): ΔACC=0.74%, ΔSE=0.23%, ΔSP=0.21% and ΔAUC=0.05%, which indicates that AFFD-Net can better implement the pixel-level vessel segmentation task.

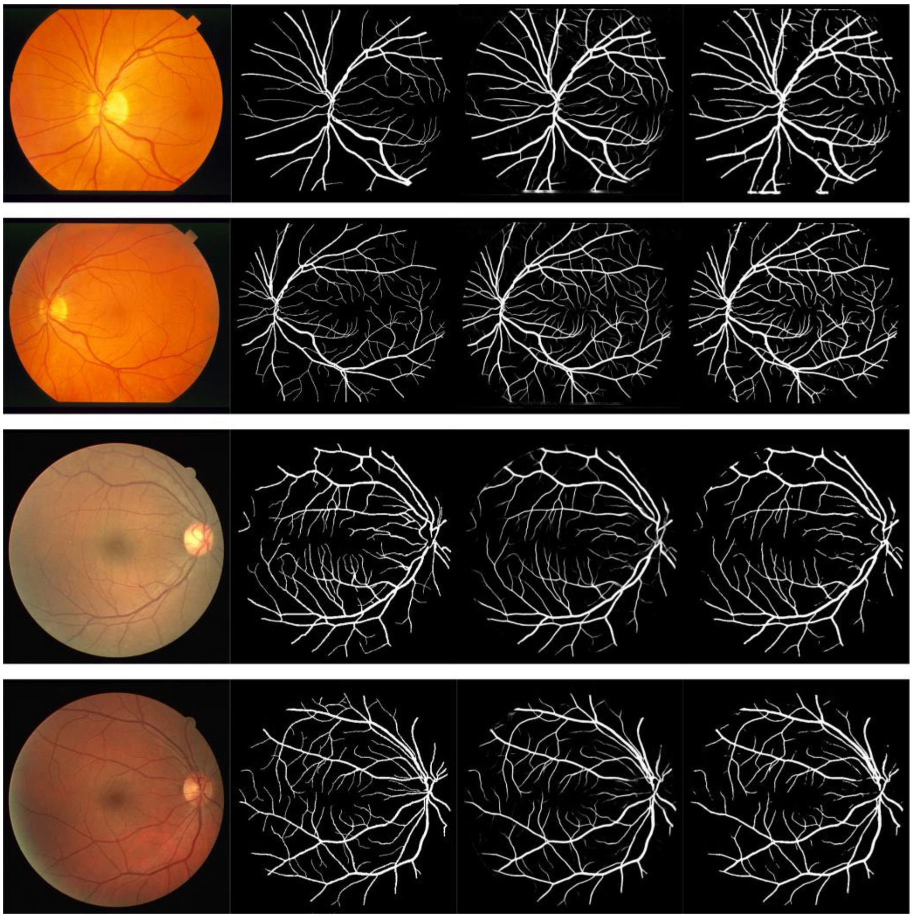

FIGURE 10. Segmentation results of cross-dataset evaluation. The first column to the last column: Fundus images, ground-truth, segmentation possibility map and segmentation results. The first row to the second row: train on DRIVE but test on STARE, train on STARE but test on DRIVE.

TABLE 7. Performance of our model for retinal vessel segmentation on CHASE_DB1 dataset.

Model	Year	ACC(%)	SE(%)	SP(%)	AUC(%)
Li et al. [63]	2015	95.81	75.07	97.93	97.16
Orlando et al. [20]	2017	—	72.77	97.12	—
Zhang et al.[64]	2017	95.02	76.44	97.16	97.06
Yan et al. [17]	2018	96.10	76.33	98.09	97.81
Jin et al. [9]	2019	96.10	81.55	97.52	98.04
Yan et al. [3]	2019	96.07	76.41	98.06	97.76
Wang et al. [25]	2020	96.70	82.39	98.13	98.71
Xiang et al. [49]	2021	96.35	78.18	98.19	98.10
Yang et al. [65]	2021	96.32	81.76	97.76	—
Tariq et al. [15]	2022	95.78	80.12	97.30	—
Our	2022	**97.44**	**82.62**	**98.34**	**98.76**

D. GENERALIZATION

In order to verify the generalization ability of AFFD-Net, the cross-dataset evaluation between DRIVE and STARE is performed. Table 8 shows the performance of AFFD-Net and other end-to-end methods. Fig.10 illustrates the segmentation results of our model obtained from cross-dataset evaluation, and Fig.11 plots the ROC curve.

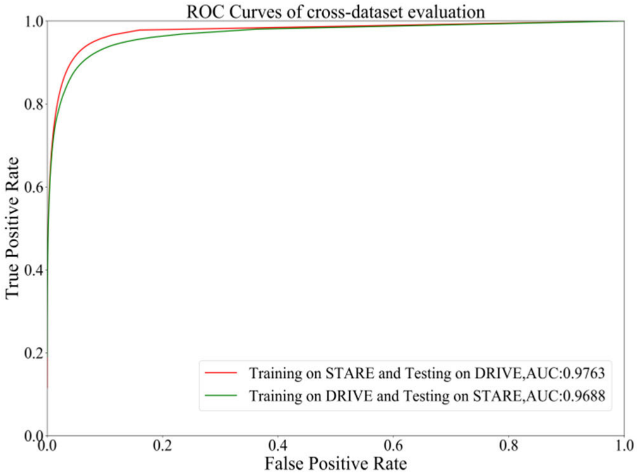

FIGURE 11. The ROC curves of cross-dataset evaluation.

TABLE 8. Performance of various methods in cross-dataset validation.

Training	Testing	Models	ACC(%)	SE(%)	SP(%)	AUC(%)
DRIVE	STARE	Soares et al. [66]	93.97	—	—	—
		Ricci et al. [67]	94.64	—	—	—
		Marın et al. [18]	94.48	—	—	—
		Roychowdhury et al. [68]	94.94	—	—	—
		Liskowski et al. [8]	94.16	—	—	96.05
		Li et al. [63]	94.86	72.73	98.10	96.77
		Zhang et al. [64]	94.47	—	—	95.93
		Yan et al. [17]	94.94	72.92	98.15	95.99
		Wang et al. [69]	94.95	—	—	—
		Yan et al. [3]	95.80	73.19	**98.40**	96.78
		Our	**96.17**	**81.70**	97.35	**96.88**
STARE	DRIVE	Soares et al. [66]	93.27	—	—	—
		Ricci et al. [67]	92.66	—	—	—
		Marin et al. [18]	95.26	—	—	—
		Roychowdhury et al. [68]	95.10	—	—	—
		Liskowski et al. [8]	95.05	—	—	95.95
		Li et al. [63]	95.45	70.27	98.28	96.71
		Zhang et al. [64]	94.88	—	—	96.76
		Yan et al. [17]	95.69	**72.11**	98.4	97.08
		Wang et al. [69]	95.73	—	—	—
		Yan et al. [3]	94.44	70.14	98.02	95.68
		Our	**96.67**	71.70	**99.04**	**97.63**

TABLE 9. Parameter quantity, time cost, and AUC of each model on DRIVE dataset.

Model	Trainable params	Total params	Time	Test_AUC
U-Net[7]	7,765,442	7,771,330	**299ms**	98.02%
nnUNet [35]	15,004,416	15,006,442	671ms	98.29%
LadderNet [36]	15,520,138	15,520,138	753ms	97.93%
R2U-Net [37]	24,091,716	24,091,716	945ms	97.84%
U-Net++[38]	9,035,438	9,041,634	458ms	98.24%
U-Net3+ [39]	6,744,034	6,750,562	830ms	98.31%
AFFD-Net	**1,777,016**	**1,780,208**	378ms	**98.41%**

According to Table 8, AFFD-Net gets well performance with the highest ACC and AUC value. For training on DRIVE and testing on STARE, the metric SE of ours is particularly outstanding and much higher than other methods. Combined with the visualization of the first line in Fig.10, it can be indicated that AFFD-Net has a stronger ability to extract blood vessels, even for the microvasculars. For training on STARE and testing on DRIVE, the SP value reaches 99.04% and higher than other methods, which indicates that our model can classify non-vascular pixels accurately. Therefore, the conclusion can be drawn that AFFD-Net has a better generalization ability and can still obtain more accurate segmentation results even when the difference between the training and test sets is large.

E. PERFORMANCE ON CHALLENGING CASES

The challenges of blood vessel segmentation are mainly manifested in the following four aspects: difficulty of extracting tiny vessels, fusion of parallel vessels, indistinct edges at the intersection of crossing vessels, and interference of lesions. Fig.12 visualizes the segmentation results of AFFD-Net in various types of challenges. It can be observed that AFFD-Net can well overcome the segmentation challenges posed by crossing and parallel vessels, and even tiny vessels with very low contrast can be extracted accurately. In the case of lesion interference, AFFD-Net can also locate the vessels definitely while ignoring the lesion, again demonstrating the superiority of the proposed model for vessel extraction.

F. CAUSE ANALYSIS OF SPECIFICITY DECLINE

As can be seen from Table 5, Table 6 and Table 7, compared with other state-of -the-art methods, ACC and SE of AFFD-Net are optimal, while SP on DRIVE and START datasets are slightly less than the optimal values. This fact indicates that AFFD-Net is weaker in judging the background

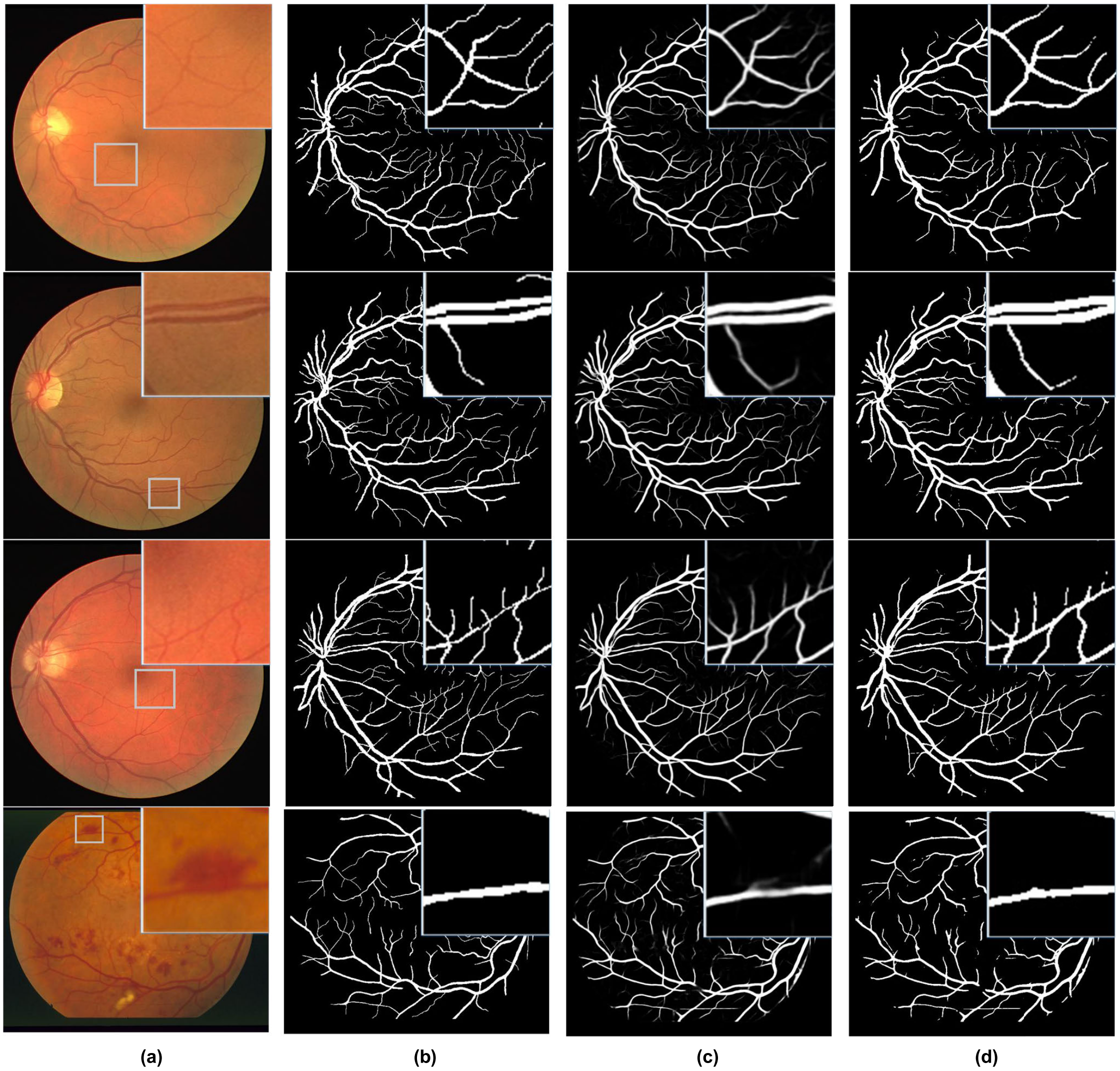

FIGURE 12. Segmentation results on challenging areas. The first row to fourth row: crossing vessels, parallel vessels, tiny vessels and focal interference. From left to right column: (a) original fundus images, (b) ground-truth, (c) segmentation possibility map and (d) binary segmentation map.

pixels. The statistics in Table 3 indicate that this phenomenon can be primarily attributed to the introduction of the MFE module and the M/A intermediate decoder. MFE combines the features under the three receptive fields and transmits them to the decoder. Although the features under 3 × 3 receptive field contains abundant detail information, some of the information is irrelevant because the features are filtered only by a layer of convolution. Most of these irrelevant information is misjudged as vascular pixel at the decoding stage. The M/A intermediate decoder is composed of MFF and AHFF. AHFF module uses Channel Maxpooling to select the maximum channel response of spatial information, resulting in some indistinguishable noise patches being judged as blood vessels as well.

G. COMPUTATION COMPLEXITY

Table 9 AFFD-Net, U-Net and classical variants of U-Net in terms of parameter quantity, time cost per each iteration and AUC value. The parameters of AFFD-Net are much lower than those of other methods, and the AUC value is the best. Moreover, the consumption time of each iteration is lower than that of U-Net++, LadderNet, R2U_Net, nnUNet and U-Net3+. Because of involving some complex matrix operations, such as the channel maxpooling in passthrough layer and AHFF module, AFFD-Net consumes a little more time than U-Net. However, the average time spent by AFFD-Net to complete an iteration is 378ms, which means that it only takes 6048s (1.68h) to train a model with high accuracy on the DRIVE dataset. This fact indicates that the proposed model is highly efficient and has potential for clinical use.

VI. CONCLUSION

The AFFD-Net model proposed in this paper has been improvement of the U-Net. The first encoding unit of U-Net is replaced with the MFE module which implements multi-scale feature extraction by stacking multiple small convolution kernels to obtain abundant spatial location information, so as to strengthen the model's response to vessels of various sizes. Then, a M/A intermediate decoder consisting of MFF and AHFF is added to the model. The MFF fuses semantic features of different scales at different stages through feature reuse to enhance the positioning capability of vessel contours. While the AHFF integrates the mixed features of different scales according to their importance and produces two feature descriptors with different focuses to increase the attention to tiny vessels. Finally, the passthrough layer takes the place of the last maximum pooling layer which can preserve the deep semantic features as much as possible.

The AFFD-Net has been tested on three public fundus image datasets: DRIVE, STARE and CHASE_DB1. The average AUC reaches 98.41%, 98.79%, and 98.76%, and the average SE reaches 84.19%, 84.58% and 82.62% for each dataset. In the cross-dataset validation between DRIVE and STARE, the AUC value reaches 96.88% and 97.63%, respectively. In the computational complexity validation, the number of training parameters of AFFD-Net is only 1777016, which is much lower than that of the classical U-Net variants, but the segmentation accuracy is the highest. The above subjected to both qualitative and quantitative evaluation, demonstrating high sensitivity, strong generalization ability, and low computational complexity. These results effectively address the problem of microvessel recognition. Overall, the model's comprehensive performance is superior and holds significant clinical practical value.

The proposed AHFF module does enhance the perception of tiny vessels, however, it also led to a decrease in specificity. Future research endeavors could aim to further enhance the performance of AFFD-Net by improving the feature filtration ability at the end of the decoder, filtering the noise interference caused by the AHFF module. Additionally, exploring alternative pre-processing methods, such as Minpooling filtering [70], to enhance the contrast between vessels and background could help reduce the difficulty of the segmentation task. Finally, investigating the problem of adaptive threshold setting for the segmentation possibility map could help resolve the conflicting relationship between SE and SP.

VII. AUTHOR CONTRIBUTIONS

Conceptualization, Ning Chunyu; methodology, Xiang Zijian and Li Mingye; software, Xiang Zijian and Ning Chunyu; formal analysis, Shi Lemin; validation, Ma Kaizheng and Shi Lemin; investigation, Li Mingye and Ma Kaizheng; data curation, Ye Guanshi; writing—original draft preparation, Xiang Zijian; writing—review and editing, Li Mingye and Wang Wei; visualization, Wang Wei; supervision, Ning Chunyu; funding acquisition, Ning Chunyu; resources, Ye Guanshi; project administration, Xiang Zijian and Ning Chunyu. All authors read and approved the final manuscript.

REFERENCES

[1] R. Bashshur and C. Ross, "World report on vision," *Int. J. Eye Banking*, vol. 8, no. 3, 2020.

[2] T. Li, W. Bo, C. Hu, H. Kang, H. Liu, K. Wang, and H. Fu, "Applications of deep learning in fundus images: A review," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101971, doi: 10.1016/j.media.2021.101971.

[3] Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1427–1436, Jul. 2019, doi: 10.1109/JBHI.2018.2872813.

[4] E. Golkar, H. Rabbani, and A. Dehghani, "Hybrid registration of retinal fluorescein angiography and optical coherence tomography images of patients with diabetic retinopathy," *Biomed. Opt. Exp.*, vol. 12, no. 3, pp. 1707–1724, 2021, doi: 10.1364/BOE.415939.

[5] S. Lian, L. Li, G. Lian, X. Xiao, Z. Luo, and S. Li, "A global and local enhanced residual U-Net for accurate retinal vessel segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 3, pp. 852–862, May 2021, doi: 10.1109/TCBB.2019.2917188.

[6] Y. Wu, Y. Xia, and Y. Song, "Multiscale network followed network model for retinal vessel segmentation," presented at the Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-030-00934-2_14.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," presented at the Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-24574-4_28.

[8] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016, doi: 10.1109/TMI.2016.2546227.

[9] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019, doi: 10.1016/j.knosys.2019.04.025.

[10] F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1010–1019, Jul. 2001, doi: 10.1109/83.931095.

[11] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000, doi: 10.1109/42.845178.

[12] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable COSFIRE filters for vessel delineation with application to retinal images," *Med. Image Anal.*, vol. 19, no. 1, pp. 46–57, Jan. 2015, doi: 10.1016/j.media.2014.08.002.

[13] L. C. Neto, G. L. B. Ramalho, J. F. S. R. Neto, R. M. S. Veras, and F. N. S. Medeiros, "An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images," *Expert Syst. Appl.*, vol. 78, pp. 182–192, Jul. 2017, doi: 10.1016/j.eswa.2017.02.015.

[14] C. Sazak, C. Nelson, and B. Obara, "The multiscale bowler-hat transform for blood vessel enhancement in retinal images," *Pattern Recognit.*, vol. 88, pp. 739–750, Apr. 2019, doi: 10.1016/j.patcog.2018.10.011.

[15] Y. Zhao et al., "Retinal vascular network topology reconstruction and artery/vein classification via dominant set clustering," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 341–356, Feb. 2020, doi: 10.1109/TMI.2019.2926492.

[16] T. M. Khan, M. A. U. Khan, N. U. Rehman, K. Naveed, I. U. Afridi, S. S. Naqvi, and I. Raazak, "Width-wise vessel bifurcation for improved retinal vessel segmentation," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103169, doi: 10.1016/j.bspc.2021.103169.

[17] Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1912–1923, Sep. 2018, doi: 10.1109/TBME.2018.2828137.

[18] D. Marín, A. Aquino, M. E. Gegundez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Trans. Med. Imag.*, vol. 30, no. 1, pp. 146–158, Jan. 2011, doi: 10.1109/TMI.2010.2064333.

[19] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012, doi: 10.1109/TBME.2012.2205687.

[20] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 16–27, Jan. 2017, doi: 10.1109/TBME.2016.2535311.

[21] C. L. Srinidhi, P. Aparna, and J. Rajan, ''Automated method for retinal artery/vein separation via graph search metaheuristic approach,'' *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2705–2718, Jun. 2019, doi: 10.1109/TIP.2018.2889534.

[22] B. Toptaş and D. Hanbay, ''Retinal blood vessel segmentation using pixel-based feature vector,'' *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, Art. no. 103053, doi: 10.1016/j.bspc.2021.103053.

[23] Z. Jiang, H. Zhang, Y. Wang, and S.-B. Ko, ''Retinal blood vessel segmentation using fully convolutional network with transfer learning,'' *Comput. Med. Imag. Graph.*, vol. 68, pp. 1–15, Sep. 2018, doi: 10.1016/j.compmedimag.2018.04.005.

[24] H. Boudegga, Y. Elloumi, M. Akil, M. H. Bedoui, R. Kachouri, and A. B. Abdallah, ''Fast and efficient retinal blood vessel segmentation method based on deep learning network,'' *Comput. Med. Imag. Graph.*, vol. 90, Jun. 2021, Art. no. 101902, doi: 10.1016/j.compmedimag.101902.

[25] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, ''Hard attention net for automatic retinal vessel segmentation,'' *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3384–3396, Dec. 2020, doi: 10.1109/JBHI.2020.3002985.

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

[27] S. K. Ghosh and A. Ghosh, ''A novel retinal image segmentation using rSVM boosted convolutional neural network for exudates detection,'' *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102785, doi: 10.1016/j.bspc.2021.102785.

[28] R. Zhao, Q. Li, J. Wu, and J. You, ''A nested U-shape network with multi-scale upsample attention for robust retinal vascular segmentation,'' *Pattern Recognit.*, vol. 120, Dec. 2021, Art. no. 107998, doi: 10.1016/j.patcog.2021.107998.

[29] J. Morano, Á. S. Hervella, J. Novo, and J. Rouco, ''Simultaneous segmentation and classification of the retinal arteries and veins from color fundus images,'' *Artif. Intell. Med.*, vol. 118, Aug. 2021, Art. no. 102116, doi: 10.1016/j.artmed.2021.102116.

[30] Z. Huang, M. Sun, Y. Liu, and J. Wu, ''CSAUNet: A cascade self-attention U-shaped network for precise fundus vessel segmentation,'' *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103613, doi: 10.1016/j.bspc.2022.103613.

[31] Y. Zhang, M. He, Z. Chen, K. Hu, X. Li, and X. Gao, ''Bridge-net: context-involved U-Net with patch-based loss weight mapping for retinal blood vessel segmentation,'' *Expert Syst. Appl.*, vol. 195, Jun. 2022, Art. no. 116526, doi: 10.1016/j.eswa.2022.116526.

[32] X. Xiao, L. Shen, Z. Luo, and S. Li, ''Weighted res-UNet for high-quality retina vessel segmentation,'' presented at the 9th Int. Conf. Inf. Technol. Med. Educ., 2018, doi: 10.1109/ITME.2018.00080.

[33] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, ''Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal,'' *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 568–576, Feb. 2020, doi: 10.1109/JBHI.2019.2912935.

[34] O. Oktay, J. Schlemper, and L. L. Folgoc, ''Attention U-Net: Learning where to look for the pancreas,'' in *Proc. Comput. Vis. Pattern Recognit.*, 2018, pp. 1–10, doi: 10.48550/arXiv.1804.03999.

[35] F. Isensee, J. Petersen, and A. Klein, ''nnU-Net: Self-adapting framework for U-Net-based medical image segmentation,'' in *Proc. Comput. Vis. Pattern Recognit.*, 2018, pp. 1–11, doi: 10.1038/s41592-020-01008-z.

[36] J. Zhuang, ''LadderNet: Multi-path networks based on U-Net for medical image segmentation,'' in *Proc. Comput. Vis. Pattern Recognit.*, 2018, pp. 1–4, doi: 10.48550/arXiv.1810.07810.

[37] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, ''Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation,'' in *Proc. Comput. Vis. Pattern Recognit.*, 2018, pp. 1–12, doi: 10.48550/arXiv.1802.06955.

[38] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, ''UNet++: Redesigning skip connections to exploit multiscale features in image segmentation,'' *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020, doi: 10.1109/TMI.2019.2959609.

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

[40] H. Kaiming and X. Yu, ''Spatial pyramid pooling in deep convolutional networks for visual recognition,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015, doi: 10.48550/arXiv.1406.4729.

[41] H. Tong, Z. Fang, Z. Wei, Q. Cai, and Y. Gao, ''SAT-Net: A side attention network for retinal image segmentation,'' *Appl. Intell.*, vol. 51, no. 8, pp. 5146–5156, 2021, doi: 10.1007/s10489-020-01966-z.

[42] L. C. Chen, G. Papandreou, and I. Kokkinos, ''DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.48550/arXiv.1409.4842.

[43] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, ''SCS-Net: A scale and context sensitive network for retinal vessel segmentation,'' *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102025, doi: 10.1016/j.media.2021.102025.

[44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, ''Going deeper with convolutions,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[45] M. Chala, B. Nsiri, M. H. El yousfi Alaoui, A. Soulaymani, A. Mokhtari, and B. Benaji, ''An automatic retinal vessel segmentation approach based on convolutional neural networks,'' *Expert Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115459, doi: 10.1016/j.eswa.2021.115459.

[46] X. Tang, B. Zhong, J. Peng, B. Hao, and J. Li, ''Multi-scale channel importance sorting and spatial attention mechanism for retinal vessels segmentation,'' *Appl. Soft Comput.*, vol. 93, Aug. 2020, Art. no. 106353, doi: 10.1016/j.asoc.2020.106353.

[47] K. Hu, Z. Zhang, X. Niu, Y. Zhang, C. Cao, F. Xiao, and X. Gao, ''Retinal vessel segmentation of color fundus images using multi-scale convolutional neural network with an improved cross-entropy loss function,'' *Neurocomputing*, vol. 309, pp. 179–191, Oct. 2018, doi: 10.1016/j.neucom.2018.05.011.

[48] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' *Comput. Sci.*, pp. 1–12, Nov. 2014, doi: 10.48550/arXiv.1409.1556.

[49] X. Li, Y. Jiang, M. Li, and S. Yin, ''Lightweight attention convolutional neural network for retinal vessel image segmentation,'' *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1958–1967, Mar. 2021, doi: 10.1109/TII.2020.2993842.

[50] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, ''Feature pyramid networks for object detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125, doi: 10.1109/CVPR.2017.106.

[51] J. Hu, L. Shen, and G. Sun, ''Squeeze-and-excitation networks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[52] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, ''CBAM: Convolutional block attention module,'' in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, 2018, pp. 3–19.

[53] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, ''ECA-Net: Efficient channel attention for deep convolutional neural networks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, vol. 7, no. 4, pp. 11534–11542.

[54] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, ''Attention gated networks: Learning to leverage salient regions in medical images,'' *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.

[55] P. Zhao, J. Zhang, W. Fang, and S. Deng, ''SCAU-Net: Spatial-channel attention U-Net for gland segmentation,'' *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 670, Jul. 2020.

[56] T. L. B. Khanh, D.-P. Dao, N.-H. Ho, H.-J. Yang, E.-T. Baek, G. Lee, S.-H. Kim, and S. B. Yoo, ''Enhancing U-Net with spatial-channel attention gate for abnormal tissue segmentation in medical imaging,'' *Appl. Sci.*, vol. 10, no. 17, p. 5729, Aug. 2020.

[57] C. Guo, M. Szemenyei, Y. Hu, W. Wang, W. Zhou, and Y. Yi, ''Channel attention residual U-Net for retinal vessel segmentation,'' presented at the Int. Conf. Acoust., Speech Signal Process., Jun. 2021.

[58] W. Li, S. Qin, F. Li, and L. Wang, ''MAD-UNet: A deep U-shaped network combined with an attention mechanism for pancreas segmentation in CT images,'' *Med. Phys.*, vol. 48, no. 1, pp. 329–341, 2021, doi: 10.1002/mp.14617.

[59] Y. Zhang, X. Cai, Y. Zhang, H. Kang, X. Ji, and X. Yuan, ''TAU: Transferable attention U-Net for optic disc and cup segmentation,'' *Knowl.-Based Syst.*, vol. 213, no. 11, Feb. 2021, Art. no. 106668, doi: 10.1016/j.knosys.2020.106668.

[60] R. Chen, H. Zhang, and J. Liu, ''Multi-attention augmented network for single image super-resolution,'' *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108349, doi: 10.1016/j.patcog.2021.108349.

[61] W. Panqu, C. Pengfei, and Y. Ye, ''Understanding convolution for semantic segmentation,'' presented at the 6th IEEE Winter Conf. Appl. Comput. Vis., Mar. 2018, doi: 10.48550/arXiv.1702.08502.

[62] J. Redmon and A. Farhadi, ''YOLO9000: Better, faster, stronger,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 7263–7271, doi: 10.1109/CVPR.2017.690.

[63] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, ''A cross-modality learning approach for vessel segmentation in retinal images,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 109–118, Jan. 2016, doi: 10.1109/TMI.2015.2457891.

[64] J. Zhang, Y. Chen, E. Bekkers, M. Wang, B. Dashtbozorg, and B. M. T. H. Romeny, ''Retinal vessel delineation using a brain-inspired wavelet transform and random forest,'' *Pattern Recognit.*, vol. 69, pp. 107–123, Sep. 2017, doi: 10.1016/j.patcog.2017.04.008.

[65] L. Yang, H. Wang, Q. Zeng, Y. Liu, and G. Bian, ''A hybrid deep segmentation network for fundus vessels via deep-learning framework,'' *Neurocomputing*, vol. 448, pp. 168–178, Aug. 2021, doi: 10.1016/j.neucom.2021.03.085.

[66] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, ''Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification,'' *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006, doi: 10.1109/TMI.2006.879967.

[67] E. Ricci and R. Perfetti, ''Retinal blood vessel segmentation using line operators and support vector classification,'' *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1357–1365, Oct. 2007, doi: 10.1109/TMI.2007.898551.

[68] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, ''Blood vessel segmentation of fundus images by major vessel extraction and subimage classification,'' *IEEE J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 1118–1128, May 2015, doi: 10.1109/JBHI.2014.2335617.

[69] X. Wang, X. Jiang, and J. Ren, ''Blood vessel segmentation from fundus image by a cascade classification framework,'' *Pattern Recognit.*, vol. 88, pp. 331–341, Apr. 2019, doi: 10.1016/j.patcog.2018.11.030.

[70] B. Graham, ''Kaggle diabetic retinopathy detection competition report,'' University of Warwick, Coventry, U.K., Tech. Rep. 1ST in this competition, 2015. [Online]. Available: https://github.com/btgraham/SparseConvNet/tree/kaggle_Diabetic_Retinopathy_competition

XIANG ZIJIAN received the B.S. degree in software engineering from Huainan Normal University, Huainan, China, in 2020. He is currently pursuing the M.S. degree with the School of Life Science and Technology, Changchun University of Science and Technology, Changchun, China. His research interests include biomedical image processing and deep learning applications.

NING CHUNYU received the M.S. degree in computer application technology from Jilin University, Changchun, China, in 2005. She is currently an Associate Professor with the School of Life Science and Technology, Changchun University of Science and Technology. Her work has been published in journals, such as *IET Image Processing* and *Chinese Optics*. Her research interests include artificial intelligence, pattern recognition, and biomedical signal and image processing.

LI MINGYE is currently pursuing the Ph.D. degree with the Department of Information Systems and Business Analytics, RMIT University. He is a Sessional Lecturer and a Project Supervisor with the School of Computing and Information Systems and the Department of Management and Marketing, The University of Melbourne. His main research theme is AI-enabled digital transformation. His work has been published in journals, such as *The Computer Journal* and presented at conferences, such as Australasian Conference on Information Systems.

MA KAIZHENG received the B.S. degree in automation from Dalian Maritime University, Dalian, China, in 2019. He is currently pursuing the M.S. degree with the School of Life Science and Technology, Changchun University of Science and Technology, Changchun, China. His research interests include biomedical image processing and deep learning applications.

SHI LEMIN received the B.S. and M.S. degrees in software engineering from the Changchun University of Science and Technology, Changchun, China, in 2012 and 2017, respectively. He is currently a Lecturer with the School of Life Science and Technology, Changchun University of Science and Technology. His research interests include biomedical signal processing and machine learning.

WANG WEI is currently pursuing the Ph.D. degree with the Department of Software Systems and Cybersecurity, Monash University. She is a Tutor and a Project Supervisor with the School of Computing and Information Systems, The University of Melbourne. Her main research interests include software engineering and digital health. Her work has been published in journals, such as the *International Journal of Environmental Research and Public Health*.

YE GUANSHI received the B.S. degree in computer software from the Jilin University of Technology, Changchun, China, in 1998, and the Ph.D. degree in world economic from Northeast Normal University, Changchun, in 2018. He is currently a Professor with the School of Electrical and Information Engineering, Jilin Agricultural Science and Technology University. His research interests include artificial intelligence and data mining.