MFI-Net: Multiscale Feature Interaction Network for Retinal Vessel Segmentation

Yiwen Ye, Chengwei Pan, Yicheng Wu, Shuqi Wang, and Yong Xia, *Member, IEEE*

***Abstract*—Segmentation of retinal vessels on fundus images plays a critical role in the diagnosis of micro-vascular and ophthalmological diseases. Although being extensively studied, this task remains challenging due to many factors including the highly variable vessel width and poor vessel-background contrast. In this paper, we propose a multiscale feature interaction network (MFI-Net) for retinal vessel segmentation, which is a U-shaped convolutional neural network equipped with the pyramid squeeze-and-excitation (PSE) module, coarse-to-fine (C2F) module, deep supervision, and feature fusion. We extend the SE operator to multiscale features, resulting in the PSE module, which uses the channel attention learned at multiple scales to enhance multiscale features and enables the network to handle the vessels with variable width. We further design the C2F module to generate and re-process the residual feature maps, aiming to preserve more vessel details during the decoding process. The proposed MFI-Net has been evaluated against several public models on the DRIVE, STARE, CHASE_DB1, and HRF datasets. Our results suggest that both PSE and C2F modules are effective in improving the accuracy of MFI-Net, and also indicate that our model has superior segmentation performance and generalization ability over existing models on four public datasets.**

***Index Terms*—Retinal vessel segmentation, pyramid squeeze-and-excitation, multiscale feature interaction, fundus images.**

Manuscript received 23 August 2021; revised 8 February 2022 and 7 May 2022; accepted 5 June 2022. Date of publication 13 June 2022; date of current version 9 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grants 62171377 and 62141605, and in part by Key Research and Development Program of Shaanxi Province under Grant 2022GY-084. *(Y. Ye and C. Pan contributed equally to this work.) (Corresponding author: Yong Xia.)*

Yiwen Ye and Yong Xia are with the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China (e-mail: ywye@mail.nwpu.edu.cn; yxia@nwpu.edu.cn).

Chengwei Pan is with the Institute of Artificial Intelligence, Beihang University, Beijing 100191, China (e-mail: pancw@buaa.edu.cn).

Yicheng Wu is with the Faculty of Information Technology, Monash University, Clayton, VIC 3800, Australia (e-mail: yicheng.wu@monash.edu).

Shuqi Wang is with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: wuwangshuqi@outlook.com).

Digital Object Identifier 10.1109/JBHI.2022.3182471

I. Introduction

RETINAL vessel segmentation on fundus images is a fundamental and critical step in the diagnosis and risk assessment of micro-vascular and ophthalmological diseases, such as arteriosclerosis, diabetic retinopathy, glaucoma, and stroke [1]. Manual segmentation is labor-intensive, time-consuming, and prone to operator bias, especially when the operator has insufficient expertise and concentration. Therefore, automated retinal vessel segmentation is of great value to accelerate and improve the clinical practice. This task, however, is challenging due to four reasons: 1) the retinal vessel with central light reflection is prone to be split into two parallel ones, 2) the vessel network has a highly complex morphological structure, 3) the width of vessels varies greatly, ranging from 1 pixel to almost 20 pixels [2], and 4) the contrast between thin vessels (particularly capillaries) and the background is extremely poor (see Fig. 1).

Automated retinal vessel segmentation has been extensively studied. Early research focuses mainly on the design of matching filters [3], [4] and the construction of multiscale models [5]. Although these methods can detect wide and high-contrast vessels, they usually fail to identify capillaries. With the success of deep learning in computer vision, deep convolutional neural networks (DCNNs) have been constructed for retinal vessel segmentation. To achieve improved performance, three techniques are generally used [6]. First, thick vessels and thin vessels are treated respectively by multiple segmentation networks [7] or in multiple stages of a segmentation process [8]. Second, the coarse-to-fine strategy is used in a two-stage approach, where a fundus image is segmented, and the segmentation result is then further refined. This technique has a proven track record, particularly in the segmentation of thin vessels [9]–[11]. Third, sampling and skip connections are used to transfer the feature maps from the encoder to decoder, and multiscale feature extraction [10], [12], [13] and attention learning [14]–[16] are adopted to strengthen image features. Despite their improved performance, these methods still suffer from a limited ability to reconstruct the details of vessels, particularly the structures of capillaries.

In this paper, we propose a multiscale feature interaction network (MFI-Net) for retinal vessel segmentation on fundus images. We design the pyramid squeeze-and-excitation (PSE) module and coarse-to-fine (C2F) module to encourage the interaction among the features extracted at multiple scales, and incorporated both modules, together with deep supervision and feature fusion, into a U-shaped backbone. The PSE module uses

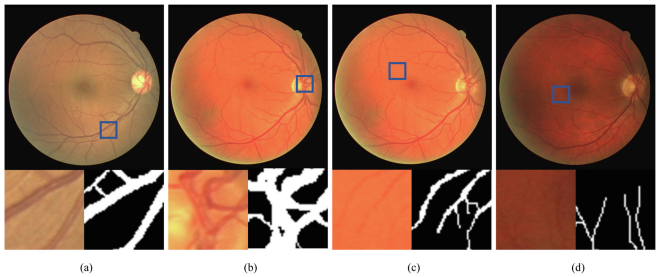

Fig. 1. Challenges of retinal vessel segmentation on fundus images: (a) Vessels with central light reflection, (b) thick vessels, (c) thin vessels with low contrast, (d) capillaries with extremely low contrast. For better visualization, the image patch in each rectangle is enlarged and displayed, together with the corresponding ground truth, beneath each fundus image.

the channel attention learned at multiple scales to enhance multi-scale features, and thus enables the network to handle adaptively the vessels with variable width. The C2F module is constructed to generate and re-process the residual feature maps, aiming to alleviate the impact of the semantic gap between the encoder and decoder and reduce the loss of vessel details during the decoding process. The proposed MFI-Net has been evaluated against several segmentation models on four public fundus image datasets, including the DRIVE, STARE, CHASE_DB1, and HRF. Our results suggest that MFI-Net achieves better performance than all competing models in retinal vessel segmentation.

The contributions of this work are two-fold. First, we devise the PSE module, which applies the element-wise attention learning to the global features obtained at multiple scales, to boost the image representation ability of our MFI-Net and enable our MFI-Net to handle adaptively the vessels with variable width. Second, we devise the C2F module to complement the feature maps produced by the decoder by the inconsistent semantic information generated by a subtraction operation. Our MFI-Net is novel in introducing inconsistent semantic information into the neural network to boost the performance of vessel segmentation.

II. Related Work

A. Retinal Vessel Segmentation

Recent retinal vessel segmentation methods are mostly based on deep learning, and can be roughly divided into four categories.

First, thick vessels and thin vessels should be treated separately. Wu *et al.* [7] divided a fundus image into wide-vessel regions, middle-vessel regions, and capillary regions according to the vessel width, and designed three UNets to segment three classes of vessels, respectively. Yan *et al.* [8] proposed a three-stage segmentation network, in which the segmentation processes is divided into three stages, including thick vessel segmentation, thin vessel segmentation, and vessel fusion.

Second, retinal vessels can be segmented in a coarse-to-fine way. In our previous work [9], we proposed the multiscale network follow network (MS-NFN) model, in which the initial retinal vessel segmentation results produced by the front network is further refined by the followed network. Mou *et al.* [10] also proposed a method that includes a dense dilated network for initial segmentation and a probability regularized walk algorithm for refinement.

Third, many methods focus on improving the sampling and skip connection operations. Li *et al.* [14] proposed the full attention network (FANet) to learn rich features on vessels and leveraged multiscale information to improve segmentation results. To jointly leverage the global and local information, Wang *et al.* [13] designed a two-channel encoding model for vessel segmentation, in which a context channel block uses several convolutions with different kernel sizes to expand the receptive fields, and a spatial channel block adopts larger kernels to preserve more spatial information. Wu *et al.* [12] proposed a scale and context-sensitive network (SCS-Net) to deal with the highly variable scales and complex anatomical structures of retinal vessels. In SCS-Net, the scale-aware feature aggregation (SFA) module and adaptive feature fusion (AFF) module are used to extract multiscale semantic information and fuse the high- and low-level features. Recently, Yuan *et al.* [16] constructed AACA-MLA-D-UNet, which consists of the dropout dense block, adaptive atrous channel attention, and multi-level attention, to utilize the low-level details and complementary information encoded in different layers for retinal vessel segmentation.

Alternatively, the information on vessel boundaries is also explored to further improve the segmentation performance. Zhang *et al.* [17] proposed a boundary enhancement and feature denoising (BEFD) module, which uses the Sobel edge detector to collect the edge prior information and enhance the segmentation boundary in an unsupervised manner. Xu *et al.* [18] proposed a model to achieve retinal vessel segmentation and vessel centerline extraction simultaneously and designed a multi-scaled cross-task aggregation module to fuse the features extracted for both tasks.

Although deep learning methods are superior to traditional ones, they still suffer from limited performance in the segmentation of vessels with small width and low contrast. In this paper, we introduce the PSE model and C2F module to a U-shaped network, aiming to segment adaptively the vessels with variable width and preserve more vessel details.

B. Multiscale Features and Attention Learning

Multiscale features can characterize the objects of interests with variable sizes. To extract multiscale features, SPP-Net [19] first divides the feature maps of a candidate region into multiple patches with different sizes and then performs max-pooling on each patch. PSPNet [20] employs the pyramid pooling to aggregate the information at different scales to form the global context. The atrous spatial pyramid pooling (ASPP) module used in DeepLab [21] employs multiple atrous convolutions [22] with different dilation rates to capture the contextual information. The context encoder network (CE-Net) [23], which uses the pretrained ResNet [24] block for feature extraction and a newly proposed dense atrous convolution block and a residual multi-kernel pooling block for context extraction, captures more high-level information and preserves spatial information for

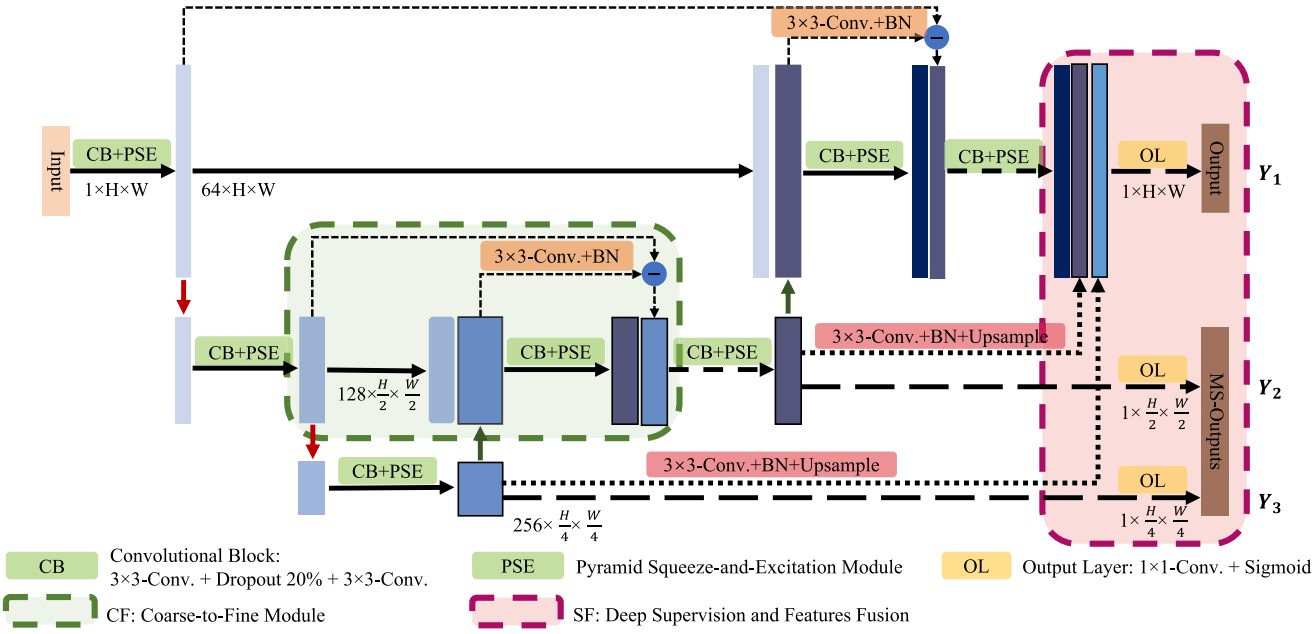

Fig. 2. Diagram of our MFI-Net. MFI-Net has a U-like encoder-decoder structure with the PSE module, C2F module, deep supervision, and feature fusion. The PSE module enables the network to handle the vessels with variable width, and the C2F module reduces the loss of vessel details during the decoder process.

2D medical image segmentation. Inspired by DenseNet [25], MSD-Net [26] and MSI-MFNet [27] combine multiscale feature maps and dense connections to produce high-level feature representations and reuse the existing features produced by prior layers. SKNet [28] utilizes a dynamic selection mechanism to adaptively adjust the size of convolutional kernels according to the input and then fuses their outputs. To alleviate the semantic gap between multiscale features, PIPO-FAN [29] adopts an equal convolutional depth module, which constrains the features fused at each level to pass through the same number of convolutional layers.

The attention mechanism learns a weighting scheme for relevant features, highlighting more informative channels and / or spatial regions. The convolutional block attention module (CBAM) [30] and spatial group-wise enhance (SGE) module [31] introduce both channel and spatial attention.NLNet [32] uses the self-attention mechanism to gain long-range dependency. Cao *et al.* [33] incorporated SE-Net into MLNet in a query-independent way and thus proposed GCNet to reduce the computational cost while maintaining the segmentation accuracy.

It is nowadays widely recognized that both multiscale features and attention learning can boost the discriminatory power of feature maps, leading to improved image segmentation performance. Hence, we combine both techniques in our PSE module for accurate retinal vessel segmentation.

III. Method

The proposed MFI-Net has a U-like encoder-decoder structure with the PSE module and C2F module. The PSE module uses multiscale information extraction and attention learning to enable our network to focus adaptively on the vessels with variable width. The C2F module targets at bridging the semantic gap between the encoder and decoder and preserving vessel details during the decoding process. Meanwhile, deep supervision and feature fusion are applied to the outputs of decoder blocks, aiming to further improve the segmentation accuracy. The architecture of our MFI-Net was illustrated in Fig. 2. We now delve into the details of each part.

A. PSE Module

The PSE module extends the SE operation [34] to multiple scales using a patch-level pyramid design. As shown in Fig. 3, the first row (highlighted in yellow) is the SE block, and the other two rows (highlighted in green and blue) are the extended SE blocks at different scales. Let the input feature map be denoted by $F_{in} \in \mathbb{R}^{C_{in}\times H\times W}$. The workflow of PSE consists of five steps (see Fig. 3). First, average pooling is applied repeatedly to F_{in} to generate the feature maps at different scales in a manner with no learnable parameters, each being denoted by $F_i \in \mathbb{R}^{C_{in}\times k_i\times k_i}, i \in \{1, 2, 3\}$. Second, each feature map F_i is processed by two convolutional layers. The first layer, followed by the ReLU activation, reduces the channel number by a preset proportion ∂, whereas the second layer restores the channel number. Third, each processed feature map is restored to its original size via duplication and concatenation, and then transformed element-wisely into an attention map using the Sigmoid function. Fourth, each attention map is applied to the input F_{in} via element-wise multiplication. Finally, three attention weighted feature maps are concatenated and further processed by a convolutional layer with 3×3 kernels and batch normalization.

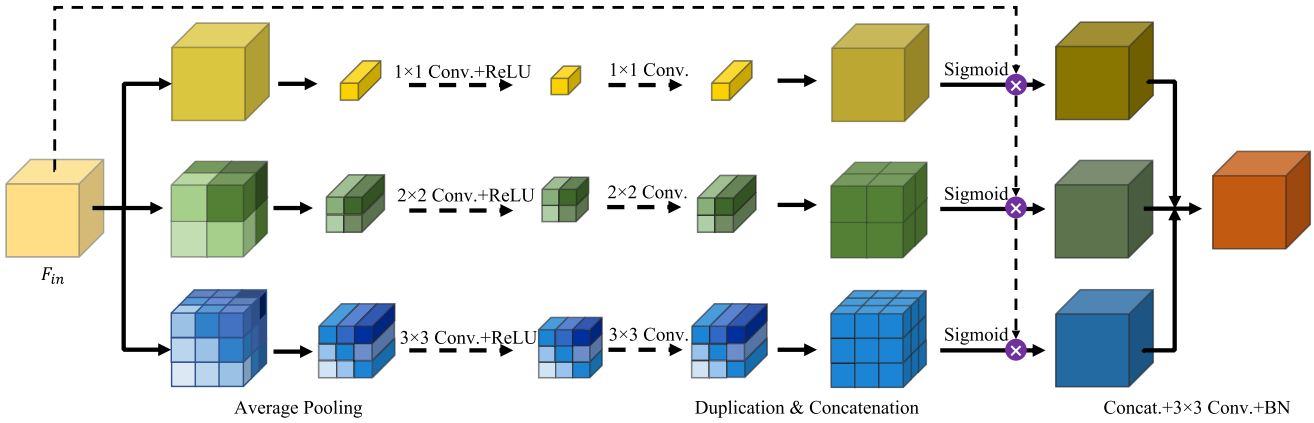

Fig. 3. Architecture of our PSE module. This module contains three parallel paths for attention learning at three scales, and thus is able to handle the vessels with variable width.

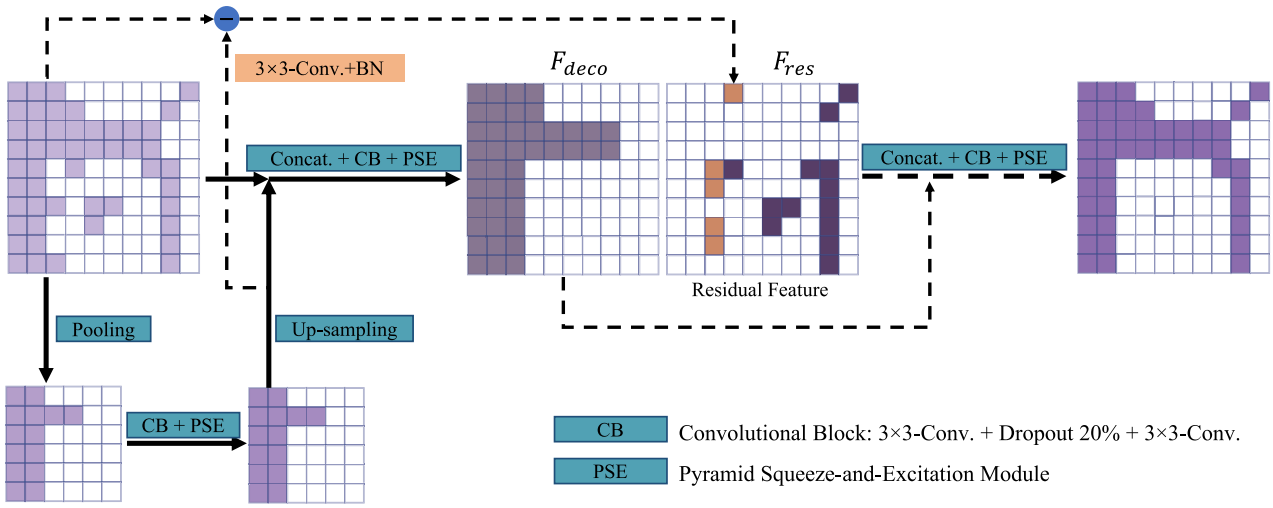

Fig. 4. Diagram of our C2F module. The C2F module adds expanded paths to the decoder to process the inconsistent semantic features that exist between the encoder and decoder and thus reduces the loss of vessel details during the decoding process.

B. C2F Module

In UNet, the skip connections directly pass features from the encoder to decoder and hence have less ability to bridge the semantic gap between these two modules. The amendment this, the C2F module improves the decoder in three steps (see Fig. 4). First, the up-sampled feature maps are first concatenated with the feature maps from the encoder and then processed by a convolutional block and the PSE module to produce F_{deco}. Next, the up-sampled feature maps are also processed by a convolutional layer with batch normalization and then subtracted from the feature maps produced by the encoder to generate residual feature maps F_{res}. Finally, the output of the C2F module is available by concatenating F_{deco} and F_{res} and subsequently passing it through a convolutional block and the PSE module. With the help of F_{res}, the C2F module has the ability to extract the inconsistent semantic features that exist between the encoder and decoder and then adds them into the decoding process via a concatenation operation to enhance this information. The inconsistent features F_{res} act as a complement to F_{deco}, contributing to MFI-Net's ability to extract more vessel details and alleviate background noise. In this way, the decoder can reduce the loss of vessel details during the decoding process.

C. Deep Supervision

We adopt the deep supervision technique to strengthen the vessel segmentation ability of our MFI-Net. As shown in Fig. 2, MFI-Net can generate segmentation results at three scales, and hence the segmentation loss can be defined as follows

$$\mathbb{L} = \sum_{i=1}^{3} \lambda_i \times \mathbb{L}_{BCE}(Y_i, Y_i'), \tag{1}$$

where $\mathbb{L}_{BCE}$ represents the binary cross-entropy function, Y_i and Y_i' are the segmentation result and ground truth at scale

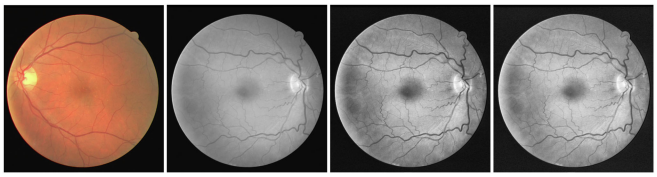

Fig. 5. Workflow of fundus image pre-processing. From left to right: A fundus image and the results of normalization, CLAHE, and brightness correction, respectively.

i, respectively, and λ_i is a weighting parameter. The ground truths at other scales are obtained by down-sampling the original ground truth via max-pooling.

D. Features Fusion

To use multiscale information in an effective way, we fuse the multiscale feature maps generated by different decoder blocks and use them jointly to predict the vessel probability map (see Fig. 2). Specifically, the outputs of all decoder blocks (except for the last one) are first fed to a convolutional layer with batch normalization, then restored to the input size via upsampling, and finally concatenated with the output of the last decoder block to generate the vessel probability maps.

E. Implementation Details

Pre-processing: For this study, each fundus image was pre-processed in three steps (see Fig. 5). First, since the fundus images in our datasets have highly variable hue and illumination, we converted each color image into gray-scale, normalized the intensity value of each pixel to be of zero mean and unit standard deviation, and linearly mapped the normalized intensity values to the range [0, 255]. Second, we applied the contrast limited adaptive histogram equalization (CLAHE) method [35] to each normalized fundus image to improve the contrast of all pixels evenly. Third, we employed data augmentation techniques, including flipping images along the horizontal axis or vertical axis and rotating images with an angle selected from $[0°, 90°, 180°, 270°]$, to expand the training dataset eight times.

Implementation: We implement our MFI-Net on NVIDIA GeForce RTX 2080 Ti using the Pytorch [36]. We leveraged the max-pooling layer for down-sampling and the bilinear interpolation for up-sampling. In the output layer, we employed a 1×1 convolutional layer to reduce the number of channels and used the Sigmoid function to generate the probability output. Except for that, all activation functions were set to ReLU.

In the training stage, we randomly extracted 1000 patches from each image in the DRIVE, STARE, and CHASE_DB1 datasets and 2000 patches from each image in the HRF dataset. The patch size is 48×48 on the DRIVE and STARE datasets, 96×96 on the CHASE_DB1 dataset, and 120×120 on the HRF dataset. We adopted the mini-batch Adam algorithm as the optimizer and set the batch size to 64 for DRIVE and STARE, 32 for CHASE_DB1, and 16 for HRF. We set the learning rate to 0.001, the maximum number of epochs to 60, and the weighting parameters λ_1, λ_2, and λ_3 to 1, 2/3, and 1/3, respectively. The parameters used in the PSE module were set as follows: $k_1 = 1$, $k_2 = 2$, $k_3 = 3$, and the proportion $\partial = 16$.

In the test stage, the same data pre-processing procedure was applied to each test image. We use a sliding window whose size is the same as the corresponding patch size to extract image patches on each test image. Along each direction, the step is set to 5 on the DRIVE, STARE, and CHASE_DB1 datasets and 20 on the HRF dataset. Thus, each pixel may appear in multiple patches, and its probability of belonging to retinal vessels is the average probability obtained on those patches. Finally, we applied the threshold of 0.5 to binarize retinal vessel probability maps. Moreover, the test time augmentation was used to make the segmentation results more robust to randomness in all experiments. Specifically, we applied the aforementioned augmentation operations (rotation and flipping) to each test image, used the trained model to segment each of eight augmented copies, and averaged the pixel-wise probabilities of the eight results to form the final probability map.

Evaluation Metrics: To quantitatively analyze the segmentation performance of our MFI-Net and competing models, we adopted five metrics, including the area under receiver operating characteristic (AUC), accuracy (ACC), sensitivity (SE), specificity (SP), and F1-score (F1).Among them, AUC, ACC, and F1 measure the overall performance, while SE and SP reflect the ability of the segmentation model to detect vessel pixels and background pixels, respectively. For this study, the performance metrics are calculated inside the FoV mask for original size images and on all pixels for resized images.

IV. Datasets

Four public fundus image datasets were used for this study. An overview of them was given in Table I.

The DRIVE dataset [43] contains 40 fundus images, 20 for training and others for test. These images were taken from 40 diabetic patients aged 25–90 years old. Among them, 33 cases have no pathological manifestations, and others have mild early diabetic retinopathy. Each image has a size of 584×565 and is saved in the JPEG format. Each training image has one manually annotated vessel segmentation result as the ground truth, while each test image has two manual segmentation results. We chose the annotation provided by the first observer as the ground truth for each test image.

The STARE dataset [44] contains 20 fundus images captured at 35° FOV. Each image has a size of 700×605, and was annotated manually by two experts. We also chose the first observer's annotation as the ground truth.

The CHASE_DB1 dataset [45] contains 28 fundus images captured from the eyes of 14 children at 30° FOV. Each image has a size of 999×960 and was annotated manually by two experts. Similarly, the annotation provided by the first observer was treated as the ground truth.

The HRF dataset [46] contains 45 fundus images of size 3504×2336, which can be grouped into three categories captured from healthy patients, glaucoma patients, and diabetic retinopathy patients, respectively. Each category has 15 images. Each image is equipped with binary segmentation ground truth.

TABLE I
AN OVERVIEW OF OUR FUNDUS IMAGE DATASETS USED FOR THIS STUDY

Dataset	Quantity	Train-test	Resolution	FoV
DRIVE	40	20-20	565×584 & 512×512	✓
STARE	20	10-10 & 4-Fold & Leave-one-Out & 10-Fold	700×605 & 512×512	×
CHASE_DB1	28	20-8 & 4-Fold	999×960 & 512×512	×
HRF	45	15-30 & 30-15	3504×2336 & 1024×1024	✓

TABLE II
PERFORMANCE OF OUR MFI-NET AND 13 COMPETING MODELS ON THE DRIVE DATASET. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED IN **RED** AND CYAN, RESPECTIVELY

Train-test	Methods	Year	AUC	ACC	SP	SE	F1
20-20	Wu [7]	2018	0.9793	0.9562	0.9830	0.7721	-
	Wu [9]	2018	0.9807	0.9567	0.9819	0.7844	-
	Yan [8]	2018	0.9750	0.9538	0.9820	0.7631	-
	Wu [37]	2019	0.9821	0.9578	0.9802	0.8038	-
	Wang [38]	2019	0.9772	0.9567	0.9816	0.7940	0.8270
	Wang [15]	2020	0.9823	0.9581	0.9813	0.7991	0.8293
	Xu [18]	2020	0.9821	0.9571	0.9750	**0.8339**	**0.8319**
	Li [39]	2020	0.9774	0.9557	0.9799	0.7890	0.8192
	Li [40]	2020	0.9813	0.9574	**0.9831**	0.7791	0.8218
	Wu [41]	2020	0.9830	**0.9582**	0.9813	0.7996	0.8295
	Yuan [16]	2021	0.9827	0.9581	0.9805	0.8046	0.8303
	Our MFI-Net	-	**0.9836**	0.9581	0.9790	0.8170	0.8315
	Zhang [42]*	2020	0.9863	0.9695	0.9846	0.8151	-
	Wu [12]*	2021	0.9837	0.9697	0.9838	**0.8289**	-
	Our MFI-Net*	-	**0.9884**	**0.9699**	**0.9847**	0.8166	0.8249

*Resizing each image to 512×512.

TABLE III
PERFORMANCE OF OUR MFI-NET AND EIGHT COMPETING MODELS ON THE STARE DATASET. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED IN **RED** AND CYAN, RESPECTIVELY

Train-test	Methods	Year	AUC	ACC	SP	SE	F1
10-10	Wang [48]	2019	0.9704	0.9538	0.9722	0.7914	-
	Guo [49]	2020	**0.9885**	0.9674	0.9857	**0.8109**	**0.8424**
	Li [39]	2020	0.9721	0.9581	0.9808	0.7536	0.7826
	Our MFI-Net	-	**0.9885**	**0.9689**	**0.9889**	0.7917	0.8319
4-Fold	Xu [18]	2020	0.9881	0.9664	0.9802	**0.8463**	**0.8384**
	Our MFI-Net	-	**0.9887**	**0.9678**	**0.9845**	0.8194	0.8320
Leave-one-Out	Yan [8]	2019	0.9833	0.9638	0.9857	0.7735	-
	Wang [15]	2020	0.9881	0.9673	0.9844	0.8186	0.8379
	Guo [49]	2020	0.9859	0.9649	0.9848	0.8020	0.8237
	Yuan [16]	2021	0.9864	0.9665	**0.9870**	0.7914	0.8276
	Our MFI-Net	-	**0.9897**	**0.9687**	0.9854	**0.8220**	**0.8396**
10-Fold	Wu [12]*	2021	0.9877	0.9736	0.9839	0.8207	-
	Our MFI-Net*	-	**0.9919**	**0.9753**	**0.9874**	**0.8237**	0.8280

*Resize images to 512×512.

TABLE IV
PERFORMANCE OF OUR MFI-NET AND 11 COMPETING MODELS ON THE CHASE_DB1 DATASET. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED IN **RED** AND CYAN, RESPECTIVELY

Train-test	Methods	Year	AUC	ACC	SP	SE	F1
20-8	Wu [9]	2018	0.9825	0.9637	**0.9847**	0.7538	-
	Yan [8]	2018	0.9776	0.9607	0.9806	0.7641	-
	Wu [37]	2019	0.9860	0.9661	0.9814	0.8132	-
	Wang [38]	2019	0.9812	0.9661	0.9821	0.8074	0.8037
	Wang [15]	2020	0.9871	0.9670	0.9813	0.8239	0.8191
	Li [40]	2020	0.9851	0.9655	0.9823	0.7970	0.8073
	Guo [49]	2020	0.9854	0.9652	0.9841	0.7757	0.8080
	Yuan [16]	2021	0.9874	0.9673	0.9801	**0.8402**	**0.8248**
	Our MFI-Net	-	**0.9879**	**0.9675**	0.9806	0.8388	0.8246
	Wu [12]*	2021	0.9867	0.9744	0.9839	**0.8365**	-
	Our MFI-Net*	-	**0.9916**	**0.9770**	**0.9868**	0.8335	0.8205
4-fold	Xu [18]	2020	0.9873	0.9651	0.9780	**0.8508**	0.8270
	Wu [41]	2020	0.9894	0.9688	**0.9880**	0.8003	-
	Our MFI-Net	-	**0.9897**	**0.9693**	0.9847	0.8317	**0.8424**

*Resizing each image to 512×512.

In the DRIVE and HRF datasets, each image is equipped with a binary field of view (FOV) mask. Following [9], [47], we manually generated the FOV mask for each image on the STARE and CHASE_DB1 datasets.

V. RESULTS

A. Comparing to Existing Methods

Performance on DRIVE: Table II gives the performance of the proposed MFI-Net and 13 competing methods on the DRIVE dataset. When using the images of the original size (565×584), our MFI-Net achieves the highest AUC of 0.9836, which is 0.06% higher than the second-best performance [41], and the second-highest ACC, SE, and F1. It is worth noting that, although obtaining the highest SE and F1, this method in [18] has relatively low SP and ACC, indicating the risk of over-segmentation due to lack of sensitivity to background pixels. By contrast, our MFI-Net achieves the most balanced SE and SP, guaranteeing low over-segmentation and under-segmentation. When using the resized images (512×512), our MFI-Net also has competitive performance, obtaining the highest AUC, ACC, and SP, notably with 0.19% higher than the second-best method [42] in terms of AUC.

Performance on STARE: Table III gives the performance of the proposed MFI-Net and eight competing methods on the STARE dataset. Since this dataset has no official data split, these competing methods were evaluated using different validation schemes, including using the first 10 images for training and other 10 images for test, four-fold cross-validation, ten-fold cross-validation, and leave-one-out. Meanwhile, Wu *et al.* [12] resized the images from 700×605 to 512×512 during the pre-processing step. To make a fair comparison, we adopted the performance of competing methods from the literature and evaluated our MFI-Net against each of them using the same validation scheme and image size. It shows in Table III that, no matter using which validation scheme, our MFI-Net achieves the highest AUC and ACC and the highest or second highest SE and F1, suggesting the superiority of our MFI-Net again.

Performance on CHASE_DB1: Table IV gives the performance of the proposed MFI-Net and eight competing methods on the CHASE_DB1 dataset. Similarly, due to the lack of official data split, these competing methods were evaluated either using the first 20 images for training and others for test, or using the four-fold cross-validation scheme. We adopted the performance of competing methods from the literature and evaluated our MFI-Net against each of them using the same validation scheme and image size. It shows that, when using the first validation scheme, our MFI-Net achieves the highest AUC and ACC, as well as the second highest SE and F1 in both the cases of with

TABLE V
PERFORMANCE OF OUR MFI-NET AND FOUR COMPETING MODELS ON THE HRF DATASET. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED IN **RED** AND CYAN, RESPECTIVELY

Train-test	Methods	Year	AUC	ACC	SP	SE	F1
15-30	Jin [50]	2019	0.9831	0.9651	0.9874	0.7464	-
	Wang [15]	2020	0.9837	0.9654	0.9843	**0.7803**	**0.8074**
	Our MFI-Net	-	**0.9858**	**0.9668**	**0.9884**	0.7544	0.8054
30-15	Soomro [51]	2019	0.978	0.962	0.962	**0.829**	-
	Our MFI-Net	-	**0.9857**	**0.9672**	**0.9867**	0.7740	0.8115
	Wu*[12]	2021	0.9842	0.9687	0.9823	0.8114	-
	Our MFI-Net*	-	**0.9880**	**0.9713**	**0.9836**	**0.8236**	0.8159

*Resize images to 1024×1024.

and without the resized operation (512×512 and 999×960). It shows in Table IV that, when using the first validation scheme, our MFI-Net achieves the highest AUC and ACC, as well as the second highest SE and F1, no matter resizing each image from 999×960 to 512×512 or not. When using the four-fold cross-validation, our MFI-Net achieves the highest AUC and ACC, as well as the second highest SE and F1 in both the cases of with and without the resized operation (512×512 and 999×960). Meanwhile, our MFI-Net achieves the highest AUC, ACC, and F1 with 0.9897, 0.9693, and 0.8424, respectively, and also achieves the second highest SP and SE in the four-fold cross-validation. Comparing the second best value, the improvements of AUC, ACC, and F1 are 0.03%, 0.05%, and 1.54%, respectively. It reveals that our MFI-Net outperforms eight competing methods on the CHASE_DB1 dataset, which is consistent with the conclusion obtained on DRIVE and STARE.

Performance on HRF: Table V gives the performance of the proposed MFI-Net and four competing methods on the HRF dataset. Similarly, there is no official data split, and two empirical data splits were used. When using the first five images in each category for training and others for test, our MFI-Net obtains the highest AUC of 0.9858, highest ACC of 0.9668, and the second-highest SE and F1. Our MFI-Net beats the second-best method [15] by 0.21% in AUC and 0.14% in ACC. When using the first ten images in each category for training and others for test, our MFI-Net achieves substantially improved AUC and ACC than the two competing methods, which use either the original image size of 3504×2336 [51] or reduced image size of 1024×1024 [12].

Statistical Test: We performed the statistical tests to demonstrate the significance of the performance gains achieved by our MFI-Net over four existing methods [16], [37], [40], [41], whose sourcecode is available. Specifically, we reproduced the results of those four methods on DRIVE, STARE using the "10-10" data split, and CHASE_DB1 using the "20-8" data split. Since not all results satisfy the assumption of variance equality, we applied the non-parametric Wilcoxon signed-rank test [52] to the AUC values of those methods. The obtained p-values are shown in Table VI. It is clear that all p-values are less than 0.05, indicating that the performance gains (in terms of AUC) of our MFI-Net over those four methods are statistically significant.

Summary: Our MFI-Net achieves the highest or second highest SE for all data split of four datasets and the highest or second highest SP under most data splits of the four datasets. Since SE and SP reflect the ability of a segmentation method to detect the vessel and background, respectively, and there is a trade-off between them, these results suggest our MFI-Net can simultaneously reduce the over-segmentation and under-segmentation of retinal vessels. It is in line with the observation that our MFI-Net achieves the highest or second highest AUC, ACC, and F1 on all four datasets. Therefore, we believe our MFI-Net is able to produce better segmentation performance than existing methods, setting the new state of the art in retinal vessel segmentation.

B. Ablation Studies

The proposed MFI-Net is a U-shaped network equipped with the PSE module, C2F module, deep supervision, and feature fusion. Let the joint use of deep supervision and feature fusion be denoted as the SF module. To verify the effectiveness of these three modules, we conducted ablation studies on the DRIVE dataset. The segmentation performance of the baseline UNet and the baseline with any one, two, or three modules was displayed in Table VII. It shows that incorporating the PSE, C2F, or SF module alone into the baseline improves AUC by 0.32%, 0.25%, and 0.26% and improves ACC by 0.17%, 0.09%, and 0.13%, respectively. Interestingly, using any two modules can further improve the AUC to 0.9832. Finally, with all three modules, the proposed MFI-Net achieves the highest AUC of 0.9836 and the highest ACC, SP, and F1. It suggests that the proposed PSE, C2F, and SF modules are effective in promoting the performance of a segmentation network. The p-values of the non-parametric Wilcoxon signed-rank test [52] on AUC are also provided in Table VII. It shows that all p-values are less than 0.05, suggesting that the performance gain achieved by our MFI-Net over its variants is significant.

C. Visualization of Segmentation Results

For a qualitative comparison, we chose one fundus image from each dataset and displayed four images, the pre-processed version, the segmentation results produced by the baseline UNet and proposed MFI-Net, and the ground truth in Fig. 6. To visualize the details of vessels and segmentation results, we selected two regions (highlighted with yellow boxes) on each image and displayed the enlarged versions beneath each image. It shows that the results produced by the proposed MFI-Net are more similar to the ground truth than those produced by UNet. Particularly, our MFI-Net can segment tiny vessels with improved accuracy on those low-contrast regions. More important, our MFI-Net achieves better connectivity of retinal vessels (highlighted with red arrows), which is critical for subsequent analysis and real clinical settings.

Since the segmented vessels around the diseased area are prone to discontinuousness, we also visualized the segmentation results with diseased fundus images containing abnormalities including bleeding, exudation, glaucoma, and microaneurysms in Fig. 7 . Similarly, we selected one region (highlighted with a yellow box) on each image and displayed the enlarged region at the bottom-right corner of each image. It shows in the enlarged regions that 1) in the bleeding region, both the baseline UNet

TABLE VI
RESULTS OF THE STATE-OF-THE-ART METHODS ON THE THREE DATASETS INCLUDING DRIVE, STARE, AND CHASE_DB1. THE P-VALUES OF THE NON-PARAMETRIC WILCOXON SIGNED-RANK TEST ON AUC ARE GIVEN

Dataset/Train-test	Methods	Year	AUC	ACC	SP	SE	F1	p-value (AUC)
DRIVE/20-20	Wu [37]	2019	0.9821	0.9578	0.9802	0.8038	0.8277	3.38×10^{-4}
	Li [40]	2020	0.9822	0.9576	0.9828	0.7830	0.8232	1.94×10^{-3}
	Wu [41]	2020	0.9830	0.9582	0.9813	0.7996	0.8295	1.51×10^{-3}
	Yuan [16]	2021	0.9827	0.9577	0.9816	0.7958	0.8265	1.71×10^{-3}
	Our MFI-Net	-	0.9836	0.9581	0.9790	0.8170	0.8315	-
STARE/10-10	Wu [37]	2019	0.9875	0.9657	0.9936	0.7062	0.7969	5.06×10^{-3}
	Li [40]	2020	0.9866	0.9642	0.9905	0.7345	0.8083	5.06×10^{-3}
	Wu [41]	2020	0.9877	0.9673	0.9846	0.8110	0.8270	5.06×10^{-3}
	Yuan [16]	2021	0.9876	0.9669	0.9925	0.7310	0.8096	9.34×10^{-3}
	Our MFI-Net	-	0.9885	0.9689	0.9889	0.7917	0.8319	-
CHASE_DB1/20-8	Wu [37]	2019	0.9860	0.9661	0.9814	0.8132	0.8118	1.17×10^{-2}
	Li [40]	2020	0.9866	0.9660	0.9831	0.7955	0.8100	1.17×10^{-2}
	Wu [41]	2020	0.9869	0.9662	0.9807	0.8237	0.8160	1.17×10^{-2}
	Yuan [16]	2021	0.9850	0.9651	0.9813	0.8057	0.8112	1.17×10^{-2}
	Our MFI-Net	-	0.9879	0.9675	0.9806	0.8388	0.8246	-

TABLE VII
RESULTS OF ABLATION STUDIES ON DRIVE DATASET: PERFORMANCE OF BASELINE UNET WITH NO, ANY ONE, ANY TWO, OR THREE MODULES. THE P-VALUES FOR THE NON-PARAMETRIC WILCOXON SIGNED-RANK TEST ON AUC ARE GIVEN. THE BEST RESULTS ARE HIGHLIGHTED IN **RED**

PSE	C2F	SF	AUC	ACC	SP	SE	F1	p-value (AUC)
			0.9798	0.9560	0.9784	0.8047	0.8224	8.86×10^{-5}
		✓	0.9824	0.9573	0.9836	0.7771	0.8201	6.81×10^{-4}
	✓		0.9823	0.9569	0.9786	0.8105	0.8264	8.86×10^{-5}
✓			0.9830	0.9577	0.9811	0.8000	0.8273	8.86×10^{-5}
	✓	✓	0.9832	0.9574	0.9833	0.7822	0.8229	2.51×10^{-2}
✓		✓	0.9832	0.9577	0.9816	0.7962	0.8267	8.03×10^{-3}
✓	✓		0.9832	0.9575	**0.9853**	0.7689	0.8208	1.94×10^{-3}
✓	✓	✓	**0.9836**	**0.9581**	0.9790	**0.8170**	**0.8315**	-

and proposed MFI-Net produce serious over-segmentation; 2) in the exudation region, our MFI-Net detects one vessel but misses a minor one, whereas UNet produces a completely wrong result; 3) in the glaucoma region and microaneurysms region, the segmentation results of MFI-Net are very similar to the ground truth, substantially superior to the results of UNet.

VI. DISCUSSION

A. Effects of Residual Feature Maps

Two pairs of the decoder feature map F_{deco} and residual feature map F_{res} generated for one sample from the DRIVE dataset were visualized in Fig. 8. The decoder feature map F_{deco} is produced by concatenating the upsampled feature maps with the output of the corresponding encoder and then through a convolution block. We compressed the feature maps along the channel dimension and selected the maximum value over all channels for each pixel. It shows that F_{res} can complement F_{deco}, since some lost features in F_{deco} are enhanced in F_{res} (see the magenta box). In addition, F_{res} can better describe the width of blood vessels compared with F_{deco} (see the blue box), while F_{deco} can retain more vascular structures compared with F_{res} (see the yellow box). Therefore, in the C2F module, we concatenated F_{deco} and F_{res} to jointly participate in the decoding process, aiming to exploit the inconsistent features embedded in F_{res} and further reduce the loss of vessel details during the decoding process.

TABLE VIII
RESULTS OF MFI-NET WITH DIFFERENT FEATURE INTERACTION MODULES ON THE DRIVE DATASET. THE P-VALUES FOR THE NON-PARAMETRIC WILCOXON SIGNED-RANK TEST ON AUC ARE GIVEN

Methods	AUC	ACC	SP	SE	F1	p-value (AUC)
MFI-Net w/ AG [53]	0.9834	0.9578	0.9789	0.8148	0.8300	4.00×10^{-2}
MFI-Net w/ CAM [54]	0.9831	0.9569	0.9857	0.7619	0.8172	3.59×10^{-3}
MFI-Net w/ AFF [12]	0.9835	0.9578	0.9808	0.8021	0.8279	1.11×10^{-2}
MFI-Net w/ C2F	0.9836	0.9581	0.9790	0.8170	0.8315	-

To further demonstrate the superiority of our C2F module over other multi-scale layer/feature's interaction operations used in existing methods, we attempted to replace the C2F module in MFI-Net with the AG module from Attention U-Net [53], the CAM module from LA-Net [54], and the AFF module from SCS-Net [12], respectively. We conducted comparative experiments on the DRIVE dataset and reported the results in Table VIII. It reveals that MFI-Net with C2F obtains the highest AUC and ACC. It also shows that the p-values calculated based on AUC are all less than 0.05, indicating that the performance gain over each of the three competing methods in terms of AUC is significant. More importantly, MFI-Net with C2F gets the highest SE and F1, which confirms our insight that the inconsistent features generated by our C2F can be used as a

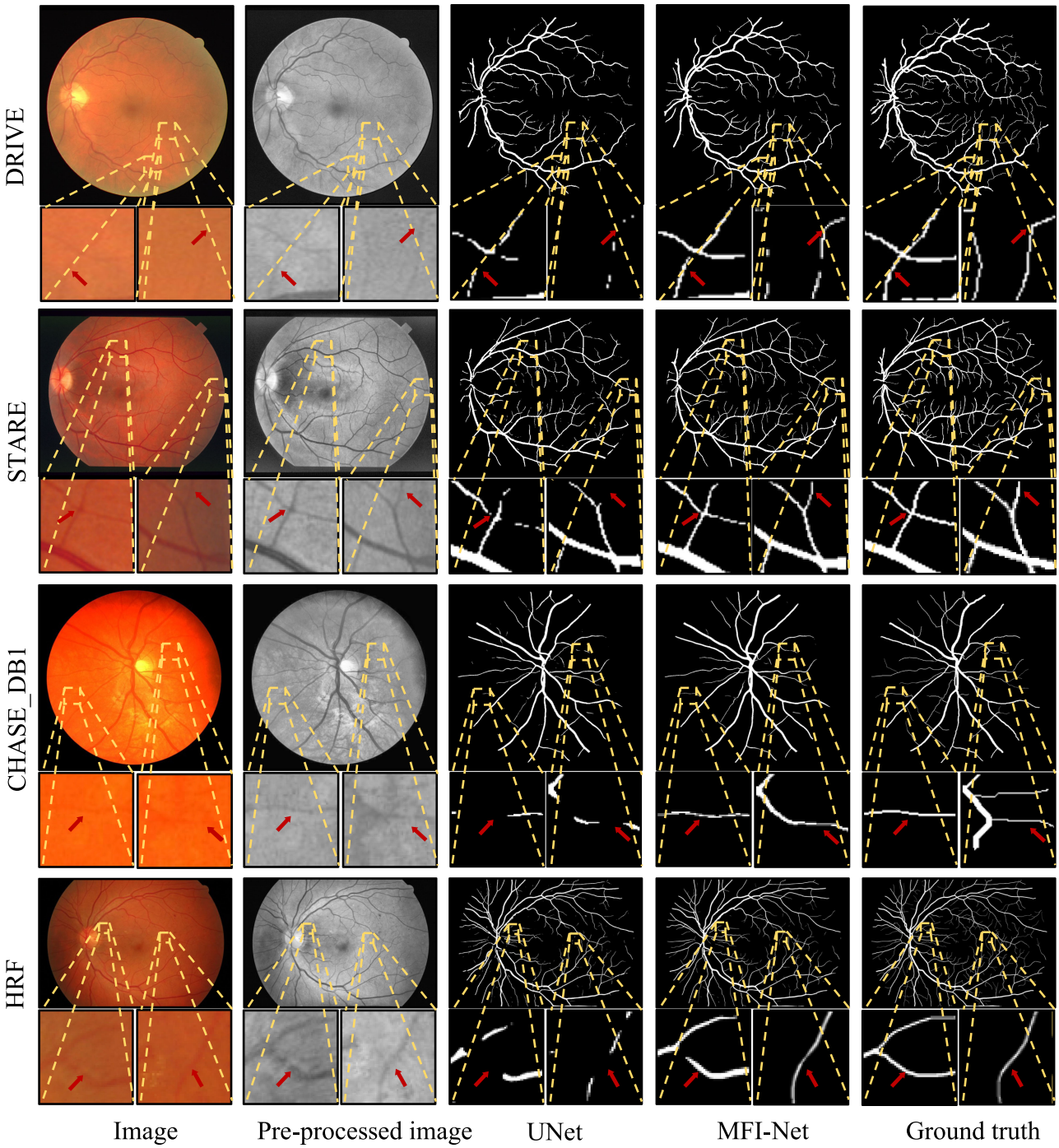

Fig. 6. Visualization of segmentation results: (1^{st} Column: Four fundus images from DRIVE, STARE, CHASE_DB1, and HRF, respectively; 2^{nd} Column: pre-processed images; 3^{rd} Column: Results of UNet; 4^{th} Column: Results of our MFI-Net; and 5^{th} Column: Ground truth). Two image patches marked by yellow rectangles are enlarged and displayed beneath each image for better visualization. Differences between the results of two models are highlighted with red arrows.

complement to improve the MFI-Net's ability to extract more vessel details.

B. Parameter Settings

In the proposed MFI-Net, there are two groups of major hyperparameters, *i.e.*, the number of compressed channels ∂ and scale parameters k_i used in the PSE module. To verify the impact of both hyperparameters on the segmentation performance, we conducted extra experiments on the DRIVE dataset.

First, since the minimum number of feature channels in MFI-Net is 64, we set the value of ∂ to 1, 2, 4, 8, 16, 32, and 64 in turn and plotted the obtained AUC and ACC versus ∂ in Fig. 9. It shows that, with the increase of ∂, both AUC and

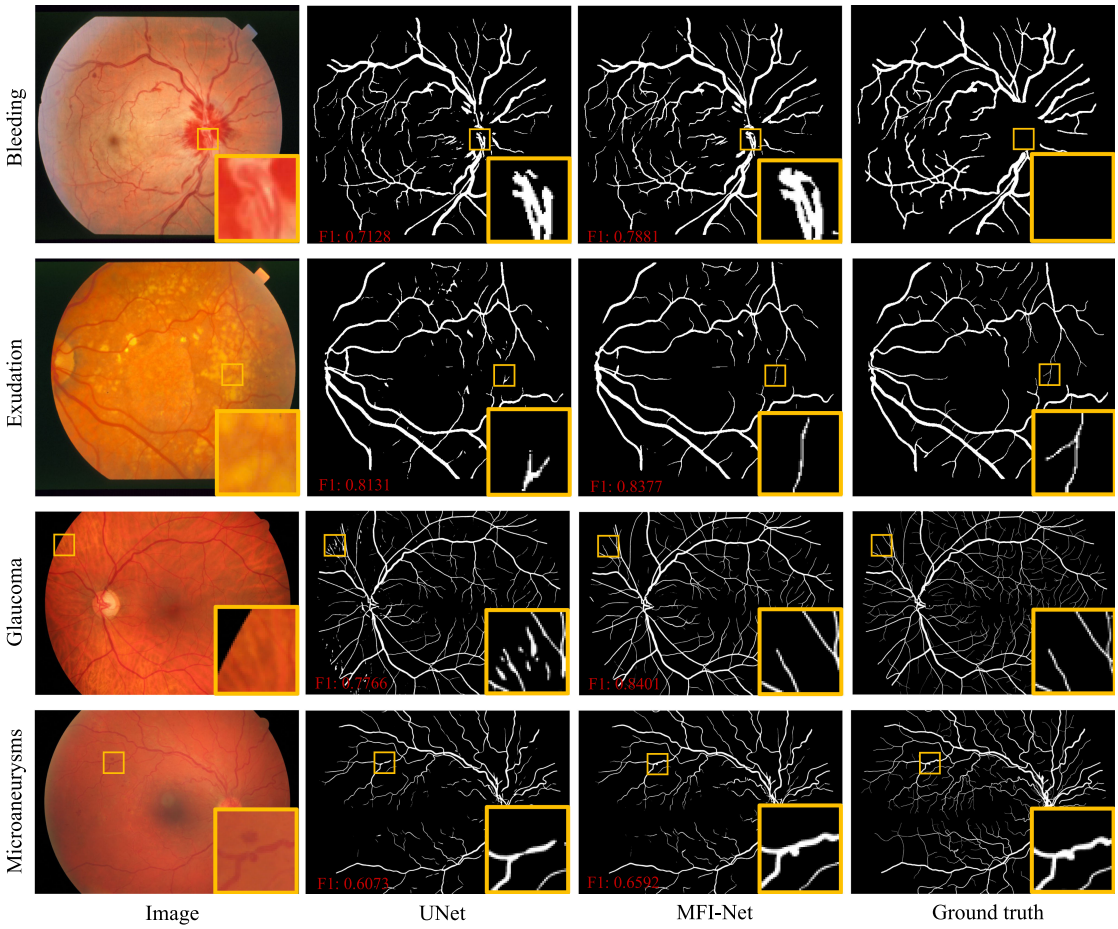

Fig. 7. Visualization of segmentation results with diseased fundus images containing abnormalities including bleeding, exudation, glaucoma, and microaneurysms (1^{st} Column: Four fundus images; 2^{nd} Column: Results of UNet; 3^{rd} Column: Results of our MFI-Net; and 4^{th} Column: Ground truth). On each image, a rectangular region (highlighted with a yellow box) is enlarged and displayed at the bottom-right corner.

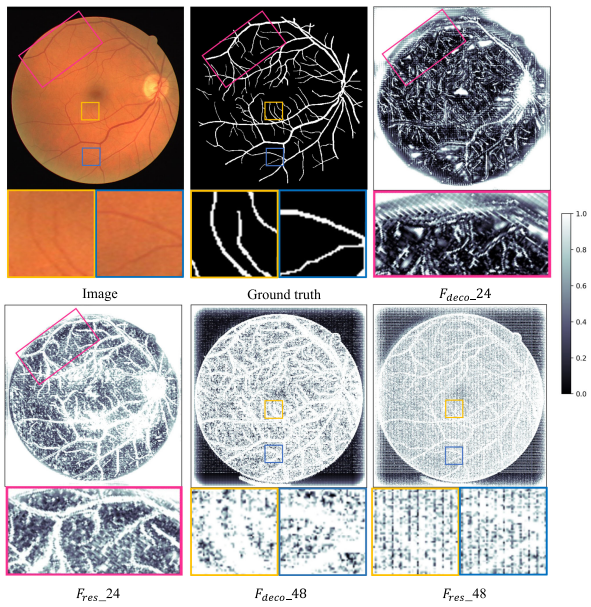

Fig. 8. Two pairs of feature maps F_{deco} and F_{res} in the visualization C2F module. F_m_n represents the visualization of the F_m feature maps with scale n. The blocks of images with different colors represent different regions and are enlarged for better visualization.

TABLE IX
THE PERFORMANCE OF DIFFERENT VALUES OF THE k_1, k_2, AND k_3 ON THE DRIVE DATASET. THE BEST RESULTS ARE HIGHLIGHTED IN **RED**

k_1, k_2, k_3	AUC	ACC	SP	SE	F1
1,2,3	**0.9836**	**0.9581**	0.9790	**0.8170**	**0.8315**
1,2,4	0.9835	0.9581	0.9817	0.7994	0.8287
1,2,6	0.9831	0.9574	**0.9844**	0.7738	0.8212
1,3,4	0.9832	0.9577	0.9826	0.7893	0.8252
1,3,6	0.9835	0.9578	0.9842	0.7787	0.8236
1,4,6	0.9835	0.9578	0.9839	0.7808	0.8241
2,3,4	0.9834	0.9577	0.9835	0.7834	0.8242
2,3,6	0.9835	0.9579	0.9816	0.7977	0.8276
2,4,6	0.9835	0.9579	0.9809	0.8020	0.8281
3,4,6	0.9835	0.9579	0.9814	0.7990	0.8277

ACC first improve sharply, then maintain high values, and finally drop dramatically. Considering that a higher value of ∂ means fewer model parameters and less computational complexity. We suggest setting ∂ to 16.

Second, k_1, k_2, and k_3 determine the scale of feature maps after global average pooling. We selected the values of the k_1, k_2, and k_3 from [1, 2, 3, 4, 6]. The results were shown in Table IX. It shows that, when setting k_1, k_2, and k_3 to 1, 2, and 3, respectively,

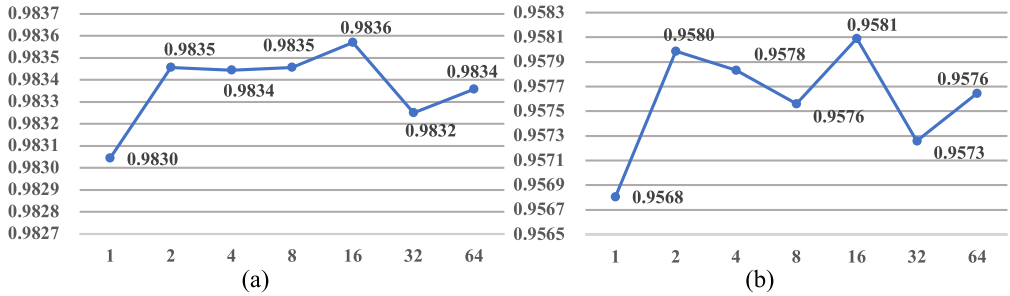

Fig. 9. Plots of (a) AUC and (b) ACC values obtained by applying our MFI-Net to the DRIVE dataset versus the hyperparameter ∂.

TABLE X
PERFORMANCE OF THREE COMPETING METHODS AND OUR MFI-NET ON THE DRIVE AND STARE DATASETS WHEN USING THE CROSS-TRAINING STRATEGY. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED IN **RED** AND CYAN, RESPECTIVELY

Training	Testing	Models	Year	AUC	ACC	SP	SE
DRIVE	STARE	Jin [50]	2019	0.9571	0.9474	0.9759	0.7000
		Wu [41]	2020	0.9635	0.9540	0.9785	0.7378
		Guo [49]	2020	0.9689	**0.9558**	**0.9841**	0.7100
		Our MFI-Net	2021	**0.9747**	0.9550	0.9741	**0.7805**
STARE	DRIVE	Jin [50]	2019	0.9718	0.9481	**0.9914**	0.6505
		Wu [41]	2020	0.9761	**0.9538**	0.9881	0.7187
		Guo [49]	2020	0.9685	0.9499	0.9801	**0.7410**
		Our MFI-Net	2021	**0.9762**	**0.9538**	0.9867	0.7313

our MFI-Net achieves the best performance. Hence, we adopted these settings in our experiments.

C. Generalization

The generalization ability is highly desirable for a computer-aided diagnosis system. We adopted the cross-training scheme [41] to verify the generalization ability of the proposed MFI-Net. Specifically, we first trained MFI-Net on the DRIVE training set and tested it on the STARE dataset without fine-tuning. Then, we trained MFI-Net on the STARE dataset and tested it on the DRIVE test set without fine-tuning. We also compared our MFI-Net with three existing methods and reported the results in Table X. It shows that MFI-Net achieves the highest AUC and SE and the second-highest ACC when testing on the STARE dataset, and achieves the highest AUC and ACC and the second-highest SE when testing on the DRIVE test set. These results indicate that our MFI-Net has a stronger generalization ability than the three competing models on the retinal vessel segmentation task.

However, comparing to the results in Table II, we found that the cross-training causes a substantial drop of the SE value on the DRIVE dataset. It may attribute to the fact that the STARE dataset has much fewer annotations of thin vessels than the DRIVE dataset, leading to the insufficiently learned ability of MFI-Net to extract thin and low-contrast vessels.

VII. CONCLUSION

In this paper, we propose MFI-Net for retinal vessel segmentation on fundus images and evaluate it against several existing methods on four public datasets. Our results indicate that the proposed MFI-Net has superior performance and generalization ability over existing ones in retinal vessel segmentation. The results of ablation studies also suggest that the PSE module we designed is effective in making our model focus adaptively on the vessels with variable width, and the C2F module we created is effective in preserving vessel details during the decoding process. In our future work, we plan to further enhance the sensitivity of the network in extracting retinal vessel structures via exploring the information about the neighborhood of each vessel pixel.

REFERENCES

[1] M. R. K. Mookiah *et al.*, "A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification," *Med. Image Anal.*, vol. 68, 2021, Art. no. 101905.
[2] J. Mo and L. Zhang, "Multi-level deep supervised networks for retinal vessel segmentation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 12, pp. 2181–2193, 2017.
[3] B. S. Y. Lam and H. Yan, "A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields," *IEEE Trans. Med. Imag.*, vol. 27, no. 2, pp. 237–246, Feb. 2008.
[4] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. Pluim, R. Duits, and B. M. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2631–2644, Dec. 2016.
[5] B. Al-Diri, A. Hunter, and D. Steel, "An active contour model for segmenting and measuring retinal vessels," *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1488–1497, Sep. 2009.
[6] T. Li *et al.*, "Applications of deep learning in fundus images: A review," *Med. Image Anal.*, vol. 69, 2021, Art. no. 101971.
[7] Y. Wu, Y. Xia, and Y. Zhang, "Deep classification and segmentation model for vessel extraction in retinal images," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis.*, 2018, pp. 250–258.
[8] Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE J. Biomed. health Informat.*, vol. 23, no. 4, pp. 1427–1436, Jul. 2019.
[9] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "Multiscale network followed network model for retinal vessel segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2018, pp. 119–126.
[10] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao, and J. Liu, "Dense dilated network with probability regularized walk for vessel detection," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1392–1403, May 2020.
[11] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *Proc. Int. Conf. Med. image Comput. Comput.- Assist. intervention*, 2016, pp. 132–139.
[12] 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, 2021, Art. no. 102025.

[13] B. Wang, S. Wang, S. Qiu, W. Wei, H. Wang, and H. He, "CSU-net: A context spatial u-net for accurate blood vessel segmentation in fundus images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1128–1138, Apr. 2021.

[14] 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.

[15] 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.

[16] Y. Yuan, L. Zhang, L. Wang, and H. Huang, "Multi-level attention network for retinal vessel segmentation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 1, pp. 312–323, Jan. 2022.

[17] M. Zhang, F. Yu, J. Zhao, L. Zhang, and Q. Li, "BEFD: Boundary enhancement and feature denoising for vessel segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2020, pp. 775–785.

[18] R. Xu *et al.*, "Joint extraction of retinal vessels and centerlines based on deep semantics and multi-scaled cross-task aggregation," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2722–2732, Jul. 2021.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "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.

[20] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2017, pp. 2881–2890.

[21] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "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.

[22] G. Papandreou, I. Kokkinos, and P.-A. Savalle, "Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2015, pp. 390–399.

[23] Z. Gu *et al.*, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2016, pp. 770–778.

[25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2017, pp. 4700–4708.

[26] G. Huang, D. Chen, T. Li, F. Wu, L. van der Maaten, and K. Weinberger, "Multi-scale dense networks for resource efficient image classification," in *Proc. Int. Conf. Learn. Representations*, 2018.

[27] T. S. Sheikh, Y. Lee, and M. Cho, "Histopathological classification of breast cancer images using a multi-scale input and multi-feature network," *Cancers*, vol. 12, no. 8, 2020, Art. no. 2031.

[28] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 510–519.

[29] X. Fang and P. Yan, "Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3619–3629, Nov. 2020.

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

[31] X. Li, X. Hu, and J. Yang, "Spatial group-wise enhance: Improving semantic feature learning in convolutional networks," *CoRR*, 2019.

[32] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2018, pp. 7794–7803.

[33] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 1971–1980.

[34] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2018, pp. 7132–7141.

[35] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Proc. Graph. gems*, 1994, pp. 474–485.

[36] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.

[37] Y. Wu *et al.*, "Vessel-net: Retinal vessel segmentation under multi-path supervision," in *Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2019, pp. 264–272.

[38] B. Wang, S. Qiu, and H. He, "Dual encoding u-net for retinal vessel segmentation," in *Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2019, pp. 84–92.

[39] H. Li *et al.*, "Mau-net: A retinal vessels segmentation method," in *2020 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2020, pp. 1958–1961.

[40] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2020, pp. 3656–3665.

[41] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "NFN+ : A novel network followed network for retinal vessel segmentation," *Neural Netw.*, vol. 126, pp. 153–162, 2020.

[42] S. Zhang, H. Fu, Y. Xu, Y. Liu, and M. Tan, "Retinal image segmentation with a structure-texture demixing network," in *Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2020, pp. 765–774.

[43] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," in *Med. Imag. 2004: Imag. Process. Int. Soc. Opt. Photon.*, 2004, pp. 648–656.

[44] A. 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.

[45] M. M. Fraz *et al.*, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012.

[46] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *Int. J. Biomed. Imag.*, vol. 2013, 2013, Art. no. 154860.

[47] J. V. Soares, J. J. 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.

[48] C. Wang, Z. Zhao, Q. Ren, Y. Xu, and Y. Yu, "Dense u-net based on patch-based learning for retinal vessel segmentation," *Entropy*, vol. 21, no. 2, 2019, Art. no. 168.

[49] S. Guo, "Dpn: Detail-preserving network with high resolution representation for efficient segmentation of retinal vessels," *J. Ambient Intell. Humanized Comput.*, pp. 1–14, 2021.

[50] 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, 2019.

[51] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, L. Zheng, and M. Paul, "Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation," *Expert Syst. with Appl.*, vol. 134, pp. 36–52, 2019.

[52] F. Wilcoxon, "Individual Comparisons by Ranking Methods," in *Breakthroughs in Statistics*. Berlin, Germany: Springer, 1992, pp. 196–202.

[53] O. Oktay *et al.*, "Attention u-net: Learning where to look for the pancreas," in *Proc. Int. Conf. Med. Imag. Deep Learn.*, 2018.

[54] F. Uslu, M. Varela, G. Boniface, T. Mahenthran, H. Chubb, and A. A. Bharath, "LA-net: A multi-task deep network for the segmentation of the left atrium," *IEEE Trans. Med. Imag.*, vol. 41, no. 2, pp. 456–464, Feb. 2022.