Full-Resolution Network and Dual-Threshold Iteration for Retinal Vessel and Coronary Angiograph Segmentation

Wentao Liu, *Graduate Student Member, IEEE*, Huihua Yang, *Member, IEEE*, Tong Tian, Zhiwei Cao, Xipeng Pan, Weijin Xu, Yang Jin, and Feng Gao

***Abstract*—Vessel segmentation is critical for disease diagnosis and surgical planning. Recently, the vessel segmentation method based on deep learning has achieved outstanding performance. However, vessel segmentation remains challenging due to thin vessels with low contrast that easily lose spatial information in the traditional U-shaped segmentation network. To alleviate this problem, we propose a novel and straightforward full-resolution network (FR-UNet) that expands horizontally and vertically through a multiresolution convolution interactive mechanism while retaining full image resolution. In FR-UNet, the feature aggregation module integrates multiscale feature maps from adjacent stages to supplement high-level contextual information. The modified residual blocks continuously learn multiresolution representations to obtain a pixel-level accuracy prediction map. Moreover, we propose the dual-threshold iterative algorithm (DTI) to extract weak vessel pixels for improving vessel connectivity. The proposed method was evaluated on retinal vessel datasets (DRIVE, CHASE_DB1, and STARE) and coronary angiography datasets (DCA1 and CHUAC). The results demonstrate that FR-UNet outperforms state-of-the-art methods by achieving the highest Sen, AUC, F1, and IOU on most of the above-mentioned datasets with fewer parameters, and that DTI enhances vessel connectivity while greatly improving sensitivity. The code is available at: https://github.com/lseventeen/FR-UNet.**

***Index Terms*—Vessel segmentation, full-resolution network, dual threshold iteration, model pruning.**

Manuscript received 9 September 2021; revised 4 April 2022 and 6 June 2022; accepted 28 June 2022. Date of publication 5 July 2022; date of current version 9 September 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0102600 and in part by the National Natural Science Foundation of China under Grant 62002082.

Wentao Liu, Zhiwei Cao, Weijin Xu, and Yang Jin are with the School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: liuwentao@bupt.edu.cn; czw@bupt.edu.cn; xwj1994@bupt.edu.cn; jinyangmail@bupt.edu.cn).

Huihua Yang is with the School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China (e-mail: yhh@bupt.edu.cn).

Tong Tian is with the State Key Laboratory of Structural Analysis for Industrial Equipment, School of Aeronautics and Astronautics, Dalian University of Technology, Dalian 116024, China (e-mail: tiantong@mail.dlut.edu.cn).

Xipeng Pan is with the Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China, and also with the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China (e-mail: pxp201@guet.edu.cn).

Feng Gao is with the Department of Interventional Neuroradiology, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China (e-mail: 13910172189@163.com).

Digital Object Identifier 10.1109/JBHI.2022.3188710

I. Introduction

BLOOD vessels are a crucial part of the circulatory system that keeps the human body running normally. In clinical practice, doctors diagnose diseases (e.g., diabetic retinopathy, macular edema, and arteriosclerosis) using the morphology of vessels and perform surgical planning and navigation depending on the structure and location of vessels [1], [2]. For example, the geometric characteristics of retinal vessels, such as the vessel diameter, branch angle, and branch length, can be utilized for early diagnosis and effective monitoring of retinal pathology [3]. In addition, vessel paths in the digital subtraction image are regarded as an indicator of guidewire movement when doctors perform vascular interventional treatment [4]. Due to imaging equipment limitations and inherent characteristics of biological tissues, the obtained initial medical images of vessels cannot intuitively reflect accurate structural information, necessitating a specialist to manually segment vessel images, which is time-consuming and subjective. As a result, automated vessel segmentation technology is essential and has recently become a hot topic in computer-aided medical diagnosis.

Early studies on vessel segmentation focused on hand-crafted features [5], [6], filtering-based models [7], and statistical models [8]. These methods aim to enhance the boundary gradient, remove undesired background information, and filter image noise, thereby simplifying the segmentation problem into a mathematical optimization problem with a fixed pattern. Benefiting from the influence of data-driven and innovation in computing equipment, deep learning has been widely researched and applied in the medical image analysis field due to its excellent representation learning ability. Extensive research [9] has shown that the vessel segmentation performance of deep learning-based methods is superior to the other approaches mentioned above. In particular, after the milestone network, U-Net [10], was proposed, all sorts of outstanding variant models for vessel segmentation emerged. Despite the fact that these works achieved

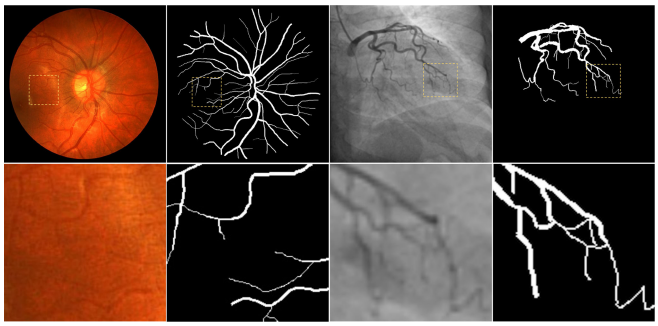

Fig. 1. Vessel images and the corresponding ground truths in the public datasets. The first and second columns are retinal vessel images from the DRIVE dataset, and the third and fourth columns are coronary angiography images from the DCA1 dataset. The second row is enlarged detailed views of the first row.

impressive segmentation results according to evaluation metrics, there are still undetected vessel pixels in segmentation maps, resulting in poor vessel connectedness visually.

Traditionally, a segmentation network based on the encoder-decoder architecture obtains high-level feature representations by stacked convolution and down-sampling and utilizes mirrored operations with skip-connections to recover the original resolution feature maps for pixel-level supervised learning. Although the down-sampling shrinks the feature map to increase the receptive field and decrease the amount of calculation, it sacrifices high-resolution spatial information, which is terrible for vessel segmentation. Vessel images are different from other medical images such as the heart and cells, where vessels are relatively thin with low pixel proportion and contrast, especially in capillaries (see Fig. 1 for further details). It's difficult for networks to learn the semantic information of small-scale vessels since the valuable spatial information is diminished after down-sampling.

Motivated by the above issues, we recommend a full-resolution UNet (FR-UNet) with a multiresolution convolution interactive mechanism, in which the stacked convolution at the image resolution continuously learns full-resolution representations for pixel-accurate segmentation. Simultaneously, the feature aggregation module fuses the feature maps from the up-sampling and down-sampling from the adjacent stages, and the results are finally aggregated into the full-resolution stage to increase the receptive field and supplement the high-level contextual information. The full-resolution feature maps were maintained throughout the entire process to alleviate the loss of spatial information. Furthermore, many works have concentrated on designing network architecture while ignoring how to accurately produce binary segmentation results from prediction probability maps. In this context, by taking into account the neighborhood relationships between the pixels in prediction maps, we utilize a dual-threshold iterative algorithm to extract more vessel pixels and enhance the vascular connectivity.

There are three main contributions of this article:

- We propose a novel full-resolution UNet (FR-UNet) for vessel segmentation, which consists of interactive multiresolution convolution layers and continuously learns full-resolution representations to alleviate the loss of spatial information.
- We propose a feature aggregation module and embed it in FR-UNet before each convolution block to aggregate feature maps from up-sampling and down-sampling, which can effectively extract multiscale contextual information.
- We propose a dual-threshold iterative algorithm to improve the connectivity of vessels that gradually extracts weak vessel pixels from the probability map and innovatively introduces a quantitative evaluation method for vessel connectivity by calculating the number of connected components.

The rest of this article is organized as follows. Section II reviews existing vessel segmentation methods based on deep learning and high/full-resolution network models. In Section III, we elaborate on the proposed FR-UNet and dual-threshold iteration method. Section IV introduces the datasets, the evaluation metrics, and the implementation details of the experiments. In Section V, we analyze the comprehensive performance of FR-UNet through comparison with state-of-the-art methods, ablation studies, and model pruning. Moreover, we verify the effectiveness of dual-threshold iteration in optimizing vessel connectivity. Finally, Section VI concludes this article.

II. Related Work

A. Vessel Segmentation Based on Deep Learning

Early image segmentation methods based on deep learning divided an image into patches and predict the center pixel of the patches using a network composed of convolutional layers and fully connected layers. The same was true for vessel segmentation [11]. However, this approach requires considerable calculations, and the size of the patches limits the receptive field of representation learning for the center pixel. Afterwards, the fully convolutional network, FCN [12], was proposed to solve this problem. In particular, U-Net [10], which is composed of a symmetrical encoder-decoder with skip-connections, was proposed and became a landmark network in the field of medical image segmentation. Inspired by U-Net, a variety of variant network models for vessel segmentation have been proposed. Reza *et al.* [13] takes full advantage of U-Net, bi-directional ConvLSTM, and the mechanism of dense convolutions to propose Bi-directional ConvLSTM U-Net for medical image segmentation. To further integrate local features of vessel image with global dependencies adaptively, Mou *et al.* [14] added a dual self-attention mechanism consisting of spatial attention and channel attention between the encoder and decoder. Wu *et al.* [15] proposed SCS-Net to capture multiscale contextual information and promote feature fusion at different levels to obtain more semantic representations. Although these networks achieve excellent vessel segmentation performance, they only apply additional modules at the lowest scale stages, neglecting high-resolution detailed information.

Besides the typical U-shaped architecture, several noval neural network architectures have also achieved success in vessel

segmentation. Pearl *et al.* [16] proposed VSSC Net that uses two-vessel extraction layers with added supervision on top of the base VGG-16 network to improve segmentation accuracy for retinal vessel and coronary angiogram segmentation. The architecture does not adopt a conventional symmetrical design, and the overall results are not competitive in evaluation metrics. Kamran *et al.* [17] proposed RV-GAN, a new multiscale generative architecture for accurate retinal vessel segmentation, which uses two generators and two multiscale autoencoding discriminators for better microvessel localization and segmentation. RV-GAN exhibits strong vessel segmentation ability, but the low sensitivity in the literature implies that it can't extract thin vessels well enough. In addition, there are several works devoted to vessel segmentation research beyond the network architecture. Zhou *et al.* [18] designed the pipeline of synthesizing noisy labels and proposed a Study Group Learning (SGL) scheme to boost the performance of model trained with imperfect labels. However, these strategies greatly increase training time. In FR-UNet, inspired by these works, we employ efficient network architecture designs such as multiscale fusion and deep supervision.

B. High/full-Resolution Network

High/full-resolution representation learning can extract more detailed information with greater accuracy, which plays an essential role in semantic segmentation. Pohlen *et al.* [19] proposed full-resolution residual networks for semantic segmentation in street scenes, which combined the multiscale context with pixel-level precision by coupling the convolution of the image resolution. Sun *et al.* [20], [21], [22] proposed multiple high-resolution networks (HRNet) for pose estimation, detection, and segmentation. The HRNet maintains high-resolution representations throughout the process by connecting high-to-low-resolution convolutions in parallel and has become one of the popular backbones of network architecture design due to its powerful high-resolution feature learning ability. UNet++[23], [24] redesigns with rich skip connections to aggregate features of varying semantic scales at the decoder subnetworks, leading to a highly flexible feature fusion scheme. The first stage of the network, composing multiple full-resolution dense convolution blocks, efficiently integrates semantic representations of different depths using deep supervision. Previous research has shown that full- or high-resolution convolutional networks perform exceptionally well on a variety of image analysis tasks. However, there are still problems such as redundant skip connections, simple feature aggregation patterns, and heavy computational burden in those networks. We focus on solving these problems for better vessel segmentation performance.

III. Methodology

In this section, we propose a simple but powerful CNN architecture, FR-UNet, which produces precise predicted probability maps for vessel segmentation. Besides, the dual-threshold iterative algorithm is introduced to determine whether the pixels of the probability maps belong to a vessel. The details are described below.

A. Full Resolution UNet

1) Multiresolution Convolution Interactive Mechanism: Fig. 2(a) shows the architecture of FR-UNet. The network expands horizontally and vertically by up-sampling, down-sampling, and convolution, similar to the structure of UNet++. However, we removed the dense connections at each stage and introduced a multiresolution convolution interactive mechanism to achieve information interaction between adjacent stages. The shallow stage can provide more refined semantic information. On the other hand, the deep stages supplement high-level contextual information and increase the local receptive field of the feature maps. In order to reduce the parameters, we merge only the feature maps of adjacent stages compared to HRNet's full parallel connection, which increases the diversity and effectiveness of feature aggregation using multiple kernel sizes and atrous convolutions. Each stage aggregates the feature maps of adjacent locations in parallel expansions and learns hierarchical representations. Unlike the network based on the traditional encoder-decoder architecture, the first stage of FR-UNet continuously integrates high-level contextual information while maintaining the original resolution. In view of the above-mentioned characteristics of architecture, FR-UNet is suitable for dense prediction, especially for small targets that easily lose spatial information in down-sampling layers, i.e., thin vessels a few pixels in width.

The composition of FR-UNet is straightforward, mainly including the feature aggregation modules, residual blocks [25], up-sampling, and down-sampling. The essence of up-sampling and down-sampling is a convolution block consisting of a convolution (Conv) layer, a batch normalization (BN) layer [26], and a LeakyReLU activation function with a negative slope of 0.1 in sequence. In detail, the down-sampling is a 2×2 Conv with a stride of 2, where the number of channels is halved and the spatial size is doubled. The Conv layer in the up-sampling is a 2×2 deconvolution with a stride of 2, which doubles the number of feature map channels and halves the spatial size. Starting at 32, the number of channels of FR-UNet is gradually doubled from top to bottom.

2) Feature Aggregation Module: The feature aggregation module concatenates feature maps from the previous residual block as well as up-sampling and down-sampling of adjacent stages. The operation is defined as:

$$c_{ij} = \begin{cases} \left[\mathcal{D}\left(x_{(i-1,j-1)}\right), \mathcal{U}\left(x_{(i+1,j-1)}\right), \left(x_{(i,j-1)}\right)\right], & \text{case 1} \\ \left[\mathcal{D}\left(x_{(i-1,j-1)}\right), \left(x_{(i,j-1)}\right)\right], & \text{case 2} \\ \left[\mathcal{U}\left(x_{(i+1,j-1)}\right), \left(x_{(i,j-1)}\right)\right], & \text{case 3} \end{cases} \tag{1}$$

where $\mathcal{D}()$ and $\mathcal{U}()$ denote down-sampling and up-sampling, respectively. [] denotes the concatenation operation. As shown in Fig. 3, $x_{i,j}$ is a stacked feature map node, where i and j are defined as the rows and columns of the network, respectively. According to the node location, the feature fusion methods are divided into three cases. The nodes $x_{i,j-1}$ are combined with the up-sampling output of $x_{i-1,j-1}$ and the down-sampling output of $x_{i+1,j-1}$ respectively, or both. This is followed a 1×1 convolution, a 3×3 convolution, and a 3×3 atrous convolution

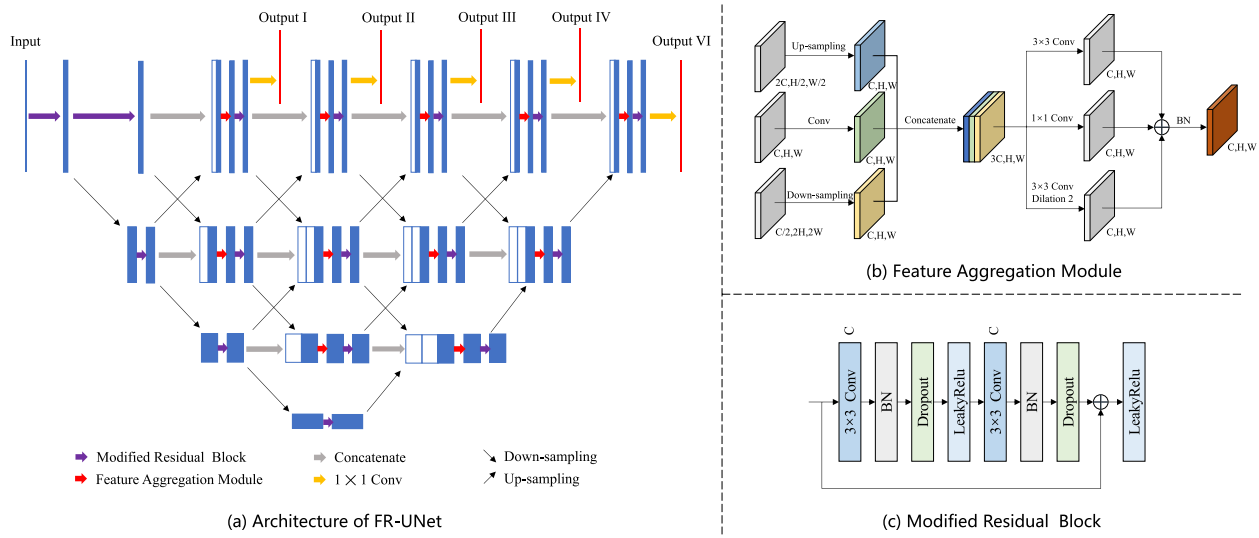

Fig. 2. (a) Architecture of FR-UNet. The network composition contains feature aggregation modules, residual blocks, up-sampling layers, and down-sampling layers. Deep supervision is introduced by adding the loss function at multiple semantic levels in the first stage. (b) The feature aggregation module integrates the up-sampling and down-sampling results of adjacent stages by concatenating and parallel multi-convolution. (c) The modified residual block adds a dropout layer after each BatchNorm layer to reduce overfitting.

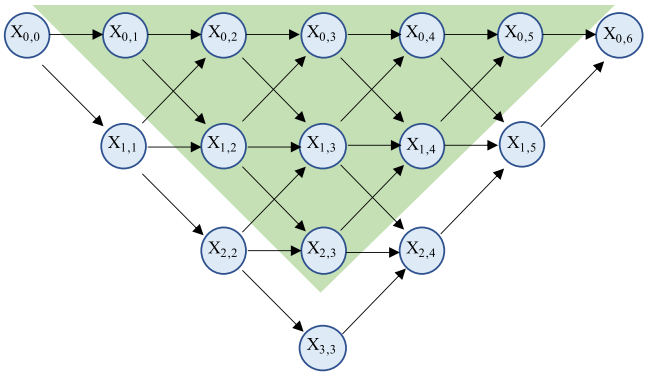

Fig. 3. The feature map node in FR-UNet. $x_{i,j}$ is a stacked feature map node, where i and j are defined as the rows and columns of the network, respectively.

with a dilated rate of 2. The output channel of each convolution is reduced to the current stage, the feature maps are obtained by adding the outputs of three different convolution modes, and a BN layer follows (see Fig. 2(b)), which is given as follows:

$$f_{ij} = \mathcal{B}\left(\mathcal{C}_{1\times1}(c_{ij}) + \mathcal{C}_{3\times3}(c_{ij}) + \mathcal{C}_{3\times3d2}(c_{ij})\right), \quad (2)$$

where $\mathcal{B}()$ and $\mathcal{C}()$ denote a BN layer and a Conv operation respectively. The modified residual block which adds a dropout layer with a dropout rate of 20% after each BN layer to reduce overfitting, is used for representation learning (see Fig. 2(c)).

3) Deep Supervision: Owing to stacked convolution and feature aggregation, FR-UNet can learn deep representation information. However, shallow full-resolution feature maps can provide fine low-level spatial representations, which should not be overlooked. As a result, we recommend using deep supervision at multiple semantic levels [23] in the first stage of the network. As depicted in Fig. 2(a), the feature maps of the last few convolution blocks in the first stage generate a probability map using a 1×1 Conv, respectively, i.e., outputs I-VI. The predicted probability result is the weighted sum of these probability maps. The calculation is:

$$y = \sum_{k=1}^{M} \gamma_k y_k, \quad (3)$$

where M is the number of outputs for deep supervision. γ_k denotes output weight in different semantic levels and $\sum_{k=1}^{M} \gamma_k = 1$. Moreover, the binary cross-entropy is introduced as the loss function on the probability map, which is calculated as:

$$\mathcal{L}(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)\right], \quad (4)$$

where y and $\hat{y}$ indicate predicted probability and ground truth of i^{th} image, respectively; N denotes the batch size.

B. Dual-Threshold Iteration

The network predicts the probability that each pixel belongs to a vessel. Generally, for two classes of segmentation problems, the segmentation result is determined by the threshold of the predicted probability map $p \in (0, 1)$, which is commonly 0.5 [27]. Nevertheless, there are inevitably some weak vessel pixels whose scores are slightly smaller than the threshold on the probability map. To address this issue, considering the neighborhood relationship between pixels, we propose a dual-threshold iterative algorithm to extract these pixels effectively. First of all, we set a high threshold γ_h and a low threshold γ_l. The probability map produces a strong segmentation map l using γ_h, in which $\hat{y} \geq \gamma_h$ pixels are set as 1, and the others are reserved. Then, each pixel in the strong segmentation map is traversed to check whether it is between high and low thresholds $(\gamma_h > \hat{y} \geq \gamma_l)$,

Algorithm 1: Dual-threshold Iteration.

Input: Probability map $\hat{y}$ with $H \times W$, high threshold γ_h, and low threshold γ_l
Output: Binary map p

```
for i = 0 to H − 1 do
    for j = 0 to W − 1 do
        if ŷ_{i,j} ≥ γ_h then
            ŷ_{i,j} = 1
        end
    end
end
do
    p = ŷ;
    for i = 0 to H − 1 do
        for j = 0 to W − 1 do
            if γ_h > ŷ_{i,j} ≥ γ_l and (ŷ_{i±1,j±1} == 1 or
              ŷ_{i±1,j} == 1 or ŷ_{i,j±1} == 1) then
                ŷ_{i,j} = 1
            end
        end
    end
while ŷ ≠ p;
for i = 0 to H − 1 do
    for j = 0 to W − 1 do
        if p_{i,j} < γ_h then
            p_{i,j} = 0
        end
    end
end
```

and there is a 1-value in the neighborhood. If the conditions are satisfied, the pixel is set as 1. The strong segmentation map is continuously iterated in this way until there is no change after the iteration. Finally, the remaining non-1 pixels are set as 0 to obtain the final result. We summarize the procedure of our proposed approach in Algorithm 1.

IV. Experiments

A. Database

The **DRIVE** [28] dataset was obtained from a diabetic retinopathy screening program in the Netherlands. It contains 40 retinal images with segmentation annotations, 33 of which show no signs of diabetic retinopathy and 7 of which show mild early diabetic retinopathy. Each image was captured using 8 bits per color plane at 565 × 584 pixels. These retinal images are officially divided into a training set and a test set, each containing 20 images.

The **CHASE_DB1** [29] is a dataset for retinal vessel segmentation that contains 28 999 × 960 pixel color retina images that were collected from 14 school children's left and right eyes. Each image is manually annotated by two independent experts for segmentation, and the first is usually used as the ground truth. Following the previous work [30], [31], the first 20 images are used for training, and the remaining 8 images are used for testing.

The **STARE** [32] comprises 20 retinal fundus images, half of which contain signs of pathologies. Since there is no uniform division for the STARE dataset, it was experimented with a 10-fold cross-validation method [15], i.e., taking 18 images as the training samples and the remaining images as the test samples. Specifically, we repeat this process 10 times until the entire dataset is covered, which can reduce the deviation as much as possible and ensure the reliability of the experimental results.

The **DCA1** [33] dataset consists of 134 X-ray coronary angiograms and their corresponding ground-truth images outlined by an expert cardiologist. The entire image database is provided by the Mexican Social Security Institute, UMAE T1-León. Each angiogram is a 300 × 300 pixel gray-scale image in PGM format. The database was divided into two subsets: the training set consists of 100 images, and the remaining 34 angiograms compose the test set.

The **CHUAC** [29] dataset is composed of 30 189 × 189 pixel coronary angiography images taken by the CHUAC Haemodynamics Unit. For each image, a 512 × 512 pixel ground truth was determined by the doctors at the hospital using a manual image editing tool. In order to make the size of the image and ground truth consistent, we enlarged the angiography images to 512 × 512 pixels using bilinear interpolation. Following previous work, the first 20 images are used for training and the remaining 10 images are used for testing [16].

B. Evaluation Metrics

We calculated the area under the receiver operating characteristic curve (AUC) by comparing the segmentation results predicted with the corresponding ground truths and we evaluated the accuracy (Acc), sensitivity (Sen), specificity (Spe), F1 score (F1), and intersection over union (IOU) on the binary segmentation maps obtained by the threshold. They are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}$$

$$Sen = \frac{TP}{TP + FN}, \tag{6}$$

$$Spe = \frac{TN}{TN + FP}, \tag{7}$$

$$F1 = \frac{2TP}{2TP + FP + FN}, \tag{8}$$

$$IOU = \frac{TP}{TP + FP + FN}, \tag{9}$$

where true positives (TP) and true negatives (TN) denote the numbers of correctly segmented vascular pixels and nonvascular pixels, respectively; false positives (FP) and false negatives (FN) denote the numbers of incorrectly segmented vascular pixels and nonvascular pixels, respectively.

C. Implementation Details

We implemented the proposed FR-UNet with PyTorch and conducted experiments with a single GeForce RTX 3090 GPU. The Adam [34] algorithm with a weight decay of 1e-5 and

TABLE I
COMPARISON OF FR-UNET WITH THE STATE-OF-THE-ART METHODS ON DRIVE AND CHASE_DB1

Methods	Time	Param (M)	DRIVE						CHASE_DB1					
			Acc	Sen	Spe	AUC	F1	IOU	Acc	Sen	Spe	AUC	F1	IOU
U-Net [10]	2015	7.76	0.9678	0.8057	0.9833	0.9825	0.8141	0.6864	0.9743	0.7650	**0.9884**	0.9836	0.7898	0.6526
UNet++ [23]	2018	9.05	0.9679	0.7891	0.9850	0.9825	0.8114	0.6827	0.9739	0.8357	0.9832	0.9881	0.8015	0.6688
Attention U-Net [35]	2018	8.73	0.9662	0.7906	0.9831	0.9774	0.8039	0.6721	0.9730	0.8384	0.9820	0.9848	0.7964	0.6617
HRNet [20]	2019	9.64	0.9704	0.8040	0.9864	0.9869	0.8265	0.7043	0.9758	0.8443	0.9847	0.9902	0.8148	0.6875
CS-Net [14]	2019	8.40	0.9632	0.8170	0.9854	0.9798	0.8039	0.7017	0.9742	0.8400	0.9832	0.9881	0.8042	0.6725
AG-Net [30]	2019	-	0.9692	0.8100	0.9848	0.9856	-	0.6965	0.9743	0.8186	0.9848	0.9863	-	0.6669
RVSeg-Net [31]	2020	5.20	0.9681	0.8107	0.9845	0.9817	-	-	0.9726	0.8069	0.9836	0.9833	-	-
SCS-Net [15]	2021	-	0.9697	0.8289	0.9838	0.9837	-	-	0.9744	0.8365	0.9839	0.9867	-	-
VSSC Net [16]	2021	8.05	0.9627	0.7827	0.9821	0.9789	-	-	0.9633	0.7233	0.9865	0.9706	-	-
SGL [18]	2021	15.53	0.9705	**0.8380**	0.9834	0.9886	0.8316	-	**0.9771**	0.8690	0.9843	**0.9920**	0.8271	-
RV-GAN [17]	2021	14.81	**0.9790**	0.7927	**0.9969**	0.9887	**0.8690**	-	0.9697	0.8199	0.9806	0.9914	**0.8957**	-
FR-UNet	2021	5.72	0.9705	0.8356	0.9837	**0.9889**	0.8316	**0.7120**	0.9748	**0.8798**	0.9814	0.9913	0.8151	**0.6882**

TABLE II
COMPARISON OF FR-UNET WITH THE STATE-OF-THE-ART METHODS ON DCA1 AND CHUAC

Methods	Time	Param (M)	DCA1						CHUAC					
			Acc	Sen	Spe	AUC	F1	IOU	Acc	Sen	Spe	AUC	F1	IOU
U-Net [10]	2015	7.76	0.9758	0.7816	0.9866	0.9879	0.7735	0.6307	0.9784	0.5881	**0.9940**	0.9614	0.6768	0.5115
Unet++ [23]	2018	9.05	0.9761	0.7954	0.9862	0.9884	0.7786	0.6375	**0.9812**	0.6687	0.9937	0.9707	0.7323	0.5777
Attention U-Net [35]	2018	8.73	0.9755	0.7986	0.9853	0.9855	0.7748	0.6324	0.9800	0.6526	0.9913	0.9506	0.7154	0.5569
HRNet [20]	2019	9.64	0.9777	0.8007	**0.9876**	0.9899	0.7919	0.6554	0.9811	0.7456	0.9906	0.9870	0.7526	0.6033
CS-Net [14]	2019	8.40	0.9763	0.7895	0.9867	0.9889	0.7790	0.6380	0.9796	0.6735	0.9918	0.9747	0.7171	0.5589
VSSC Net [16]	2021	8.05	0.9700	0.7728	0.9809	0.9831	-	-	0.9721	0.7892	0.9797	0.9757	-	-
FR-UNet	2021	5.72	**0.9788**	**0.8248**	0.9875	**0.9924**	**0.8022**	**0.6708**	0.9803	**0.8171**	0.9868	**0.9908**	**0.7601**	**0.6151**

an initial learning rate of 1e-4 is used to optimize the model parameters. The learning rate is gradually reduced by the cosine annealing algorithm over 40 epochs, and we use the last epoch of results for testing. Moreover, we set empirically the weight parameters γ_k to 0.2. In the training phase, the image is preprocessed as follows:

1) We convert the color retinal images in the DRIVE, CHASE_DB1, and STARE datasets to grayscale images.
2) The images in all datasets are normalized uniformly.
3) A 48 $\times$ 48 sliding window with a stride of 6 is used to extract patches from vessel images to increase the quantity of training data.
4) We performed data augmentation by horizontal flipping, vertical flipping, and [90,180,270] degree rotation randomly in the training phase to increase the diversity of the data and reduce overfitting.

However, the input is a full-size image in the testing phase without patch extraction.

V. RESULTS

A. Comparisons With the State-of-The-Art Methods

We conducted vessel segmentation experiments on the most popular networks, including U-Net [10], UNet++[23], Attention U-Net [35], CS-Net [14], and HRNet [20]. In order to overcome the limitations of the number of down-samplings using a 48 $\times$ 48 patch, we removed the one-quarter down-samplings at the beginning of HRNet. The full resolution was maintained, similar to FR-UNet, while other structures remained unchanged. In addition, we also compared the segmentation results of state-of-the-art vessel segmentation methods in the literature, including AG-Net [30], RVSeg-Net [31], SCS-Net [15], VSSC Net [16], SGL [18], and RV-GAN [17].

TABLE III
COMPARISON OF FR-UNET WITH THE STATE-OF-THE-ART METHODS ON STARE

Methods	Acc	Sen	Spe	AUC	F1	IOU
U-Net [10]	0.9730	0.7850	0.9884	0.9875	0.8118	0.6856
UNet++ [23]	0.9734	0.7909	0.9883	0.9884	0.8150	0.6902
Attention U-Net [35]	0.9730	0.7804	**0.9887**	0.9878	0.8106	0.6839
HRNet [20]	0.9745	0.8127	0.9878	0.9893	0.8278	0.7088
CS-Net [14]	0.9735	0.7926	0.9882	0.9885	0.8159	0.6912
SCS-Net [15]	0.9736	0.8207	0.9839	0.9877	-	-
FR-UNet	**0.9752**	**0.8327**	0.9869	**0.9914**	**0.8330**	**0.7156**

Table I and III provide the qualitative results of vessel segmentation for retinal vessel datasets DRIVE, CHASE_DB1, and STARE. It is obviously apparent from these tables that FR-UNet is competitive with state-of-the-arts by achieving the best overall

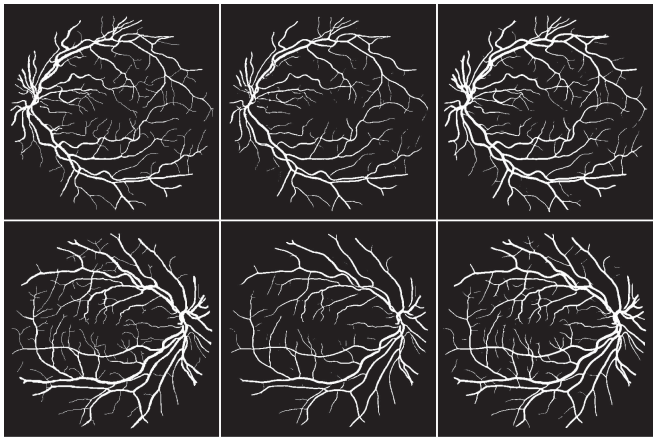

Fig. 4. Comparison of RV-GAN and FR-UNet segmentation results on the DRIVE dataset.

performance with the highest AUC (0.9889 and 0.9914) on DRIVE and STARE, and the third (0.9913, only 0.07% lower than the first SGL) on CHASE_DB1. Additionally, it is worth mentioning that the Sen score obtained by FR-UNet is much higher than other methods on CHASE_DB1 and STARE. They have respectively increased by 1.08 and 1.2% compared with the second scores obtained by SGL and SCS-Net on DRIVE and CHASE_DB1. Many studies [15], [18] point out that better sensitivity indicates the model is able to extract more thin vessels and boundary pixels, which demonstrates that FR-UNet has a great ability to extract microvascular structures. From all the results, FR-UNet, SGL, and RV-GAN achieved the highest scores on some evaluation metrics, respectively. Among them, RV-GAN, as a GAN-based segmentation architecture, achieves the highest Acc, Spe, and F1 on DRIVE and the highest F1 on CHASE_DB1, but Sen (0.7927 and 0.8199) is relatively low compared with that of SGL (0.8380 and 0.8690) and FR-UNet (0.8356 and 0.8798) on DRIVE and CHASE_DB1. As a result, it ignores many microvascular structures, as shown in Fig 4. More importantly, the parameter amount of RV-GAN, consisting of coarse and fine generators and discriminator, is 14.81 M, which is about 2.59 times that of FR-UNet. In addition, it can be found from the evaluation metrics on DRIVE and CHASE_DB1 that the performance of FR-UNet and SGL is extremely close. However, SGL adopts model cascade and cross-validation-based pseudo label generation strategies, which greatly increases model complexity and training time compared to FR-UNet.

The results of segmentation on the coronary angiograph datasets DCA1 and CHUAC are shown in Table II. FR-UNet obtained the highest AUC, Acc, Sen, F1 and IOU on DCA1 and the highest AUC, Sen, F1 and IOU on CHUAC. However, unlike other comprehensive evaluation indicators, the Spe score was low. Sen and Spe indicate the proportion of positive and negative pixels correctly predicted. When extracting more weak vessel pixels, FR-UNet introduces some false positives, i.e., nonvessel pixels are mispredicted. We consider that the model further improves the extraction ability of weak vessels while inevitably increasing the tolerance of false positives. In addition, the number of FR-UNet's parameters is 5.72 M, which is only slightly more than RVSeg-Net among all the methods in Table I. Compared with state-of-the-art methods, our FR-UNet is still a relatively light-weight network, which demonstrates that the improvement in FR-UNet performance is not attributable to model complexity.

TABLE IV
ABLATION STUDY OF FR-UNET WITH DIFFERENT NETWORK CONFIGURATIONS IN VESSEL SEGMENTATION (BASELINE: RES-UNET, MRCI: MULTI-RESOLUTION CONVOLUTION INTERACTIVE MECHANISM, FAM: FEATURE AGGREGATION MODULE, AND DS:DEEP SUPERVISION)

Methods	Acc	Sen	Spe	AUC	F1	IOU
baseline	0.9688	0.8021	0.9848	0.9828	0.8184	0.6927
baseline+MRCI	0.9709	0.7991	**0.9874**	0.9881	0.8279	0.7064
baseline+MRCI+FAM	0.9707	0.8245	0.9857	0.9882	0.8299	0.7093
baseline+MRCI+DS	**0.9711**	0.8101	0.9865	0.9884	0.8306	0.7102
baseline+MRCI+FAM+DS	0.9705	**0.8356**	0.9837	**0.9889**	**0.8316**	**0.7120**

We further visualize the vessel segmentation results, including those of UNet++, Attention U-Net, CS-Net, and FR-UNet, as shown in Fig. 5. The retinal vessel image contains many thin vessels. Hence, we zoom in on the image details for a clearer visualization, as shown in the highlighted rectangular regions. From the first subgraph of DRIVE and the two subgraphs of CHASE_DB1 and STARE, we can observe that FR-UNet detects more thin vessel pixels with low contrast than UNet++, Attention U-Net, and CS-Net. On the other hand, as shown in the second subfigure on the DRIVE, FR-UNet is superior with fewer false segmentations compared with other methods. In addition, FR-UNet also achieves significant segmentation effects on coronary angiography images from the DCA1 and CHUAC datasets, which are the closest to the ground truth compared to other methods due to retaining most of the spatial information of vessels. The aforementioned comparisons show that the FR-UNet has a strong capability for vessel segmentation.

B. Ablation Studies

As shown in Fig. 2(a), FR-UNet can be regarded as an encoder-decoder network consisting of a multi-resolution convolution interactive mechanism (MRCI, see the green triangle region in Fig. 3), feature aggregation module (FAM), and deep supervision (DS). To validate the effectiveness of the components, we conducted ablation studies to evaluate how each component affects the results, using the DRIVE dataset as an example. Res-UNet [36] served as the baseline of the experiment, we employed the residual block to replace the original block of U-Net and utilized a 1×1 convolution followed by a BN layer to halve the number of channels after skip-connections, aiming to apply residuals in the decoder. We gradually added the above components to the baseline for ablation studies. All of the experiments were performed using the same hyperparameter configuration. The results are shown in Table IV. We observe that the multiresolution convolution interactive mechanism (referred to as 'Baseline+MRCI') significantly improved the overall

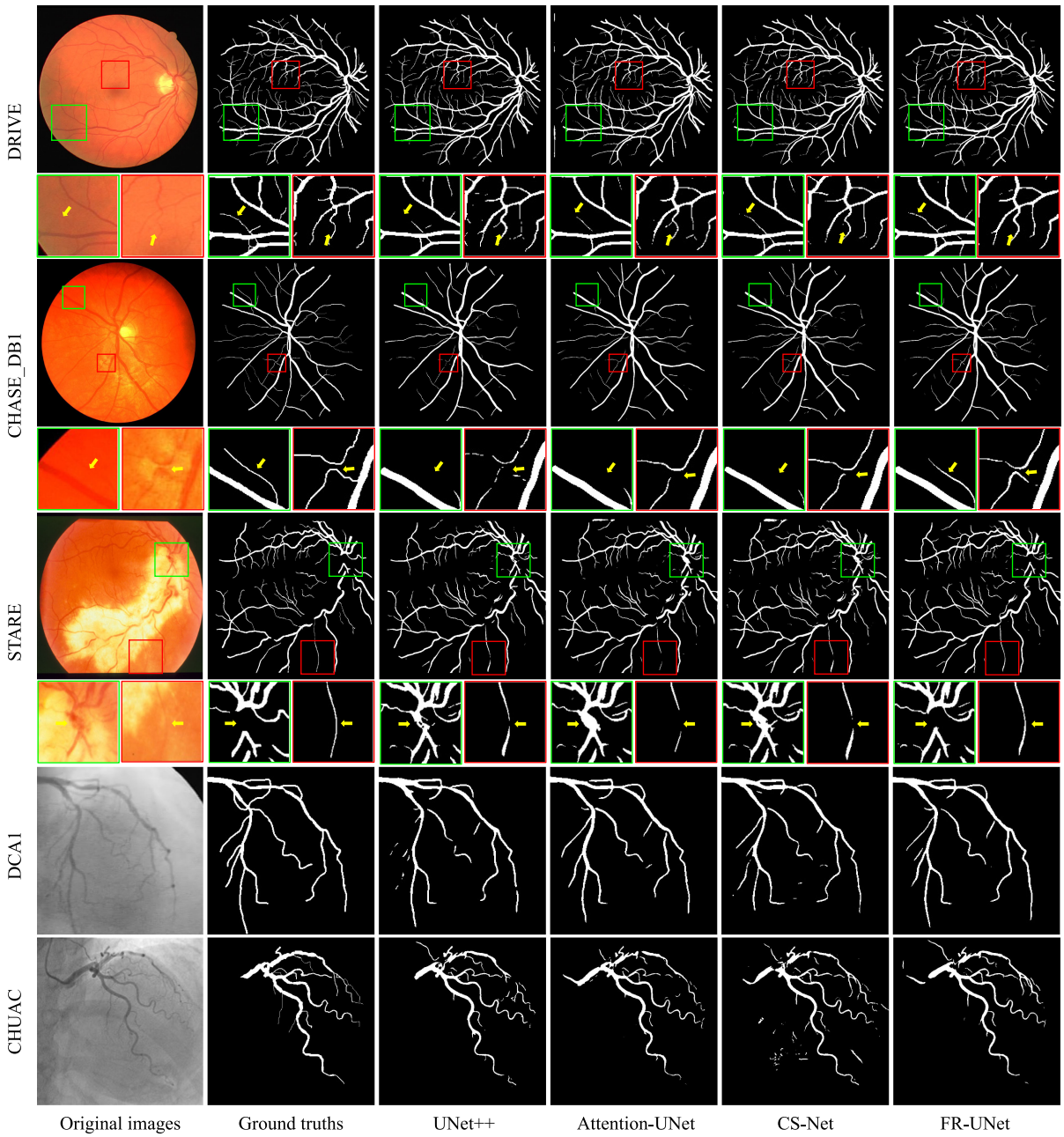

Fig. 5. Visualization of vessel segmentation results. The images from the first row to the last row viewed vertically are from the DRIVE, CHASE_DB1, STARE, DCA1, and CHUAC datasets, respectively. From a horizontal perspective, the first column to the last column are the original images, ground truths, and segmentation results of UNet++, Attention U-Net, CS-Net, and FR-UNet, respectively. The DRIVE, CHADE_DB1, and STARE datasets contain many thin vessels, and we zoom in on the image details for clearer visualization, as shown in the highlighted rectangular regions.

performance compared with Res-UNet. The Acc, Spe, AUC, F1, and IOU increased by 0.21, 0.26, 0.53, 0.95, and 1.37%, respectively. It is worth noting that Spe achieved the highest score, but Sen slightly decreased by 0.3%. The experimental results show that the vessel segmentation performance is enhanced due to high-resolution representation learning and interactive fusion of contextual information.

Res-UNet combined with a multiresolution convolution interactive mechanism constitutes the main body of FR-UNet.We add feature aggregation modules and deep supervision to this architecture, respectively. The experimental results are described as 'Baseline+MRCI+FAM' and 'Baseline+MRCI+DS' in Table IV. The feature aggregation module replaces the 1×1 convolution with multiple convolutions in parallel and improves the Sen, AUC, F1, and IOU scores. In particular, 'Baseline+MRCI+FAM' outperforms 'Baseline+MRCI' with an improvement of 2.54% in Sen. We believe that the strategy aggregates richer representation information before the residual block, which benefits vessel pixels with extremely imbalanced categories. As a practical module, deep supervision is widely used in various popular network architecture. We apply it to multilevel full-resolution feature maps for supervised learning. Compared

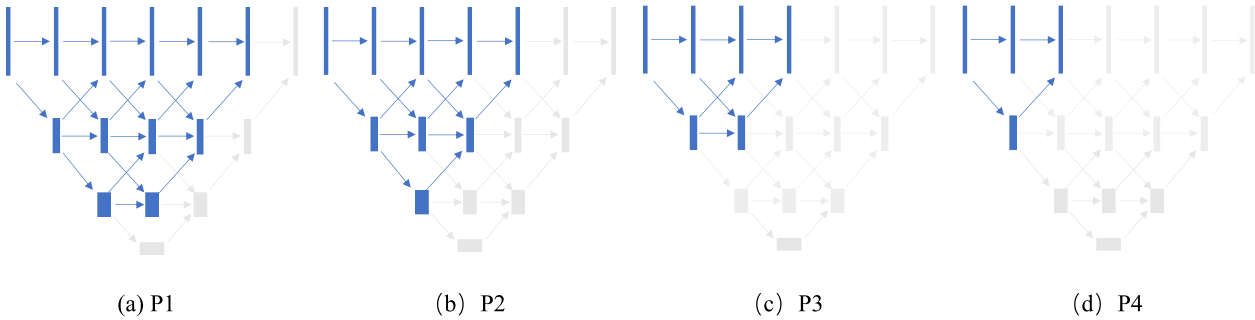

Fig. 6. Different levels of model pruning on FR-UNet. The blue convolutional nodes on the P1-P4 subgraphs constitute the sub-networks, and the gray convolutional nodes are removed on FR-UNet.

TABLE V
PRUNING SUBNETWORK PERFORMANCE COMPARISON USING TFTS AND TTS. (TFTS:TRAINING USING THE FULL FR-UNET AND TESTING USING SUBNETWORKS, AND TTS:BOTH TRAINING AND TESTING USING INDEPENDENT SUBNETWORKS.)

Methods	Param (M)	TFTS						TTS					
		Acc	Sen	Spe	AUC	F1	IOU	Acc	Sen	Spe	AUC	F1	IOU
FR-UNet-Full	5.72	0.9705	**0.8356**	0.9837	**0.9889**	**0.8316**	**0.7120**	0.9705	**0.8356**	0.9837	**0.9889**	**0.8316**	**0.7120**
FR-UNet-P1	2.60	**0.9707**	0.8069	0.9865	0.9882	0.8285	0.7072	**0.9711**	0.7961	0.9879	0.9884	0.8284	0.7071
FR-UNet-P2	1.23	0.9685	0.7358	0.9909	0.9852	0.8038	0.6720	0.9709	0.7922	**0.9881**	0.9880	0.8267	0.7046
FR-UNet-P3	0.49	0.9653	0.6737	**0.9933**	0.9815	0.7727	0.6295	0.9703	0.8053	0.9861	0.9873	0.8260	0.7036
FR-Unet-P4	0.19	0.9618	0.6408	0.9926	0.9708	0.7462	0.5952	0.9694	0.7815	0.9876	0.9856	0.8174	0.6911

with 'Baseline+MRCI,' the all-around performance was improved by deep supervision, especially the Acc, which achieved the highest score. Finally, we further embed both FAM and DS into the baseline (referred to as 'Baseline+MRCI+FAM+DS') to verify the superposition effect of all components. As shown in Table IV, this method achieved the highest Sen, AUC, F1, and IOU in the ablation study, which shows the effectiveness of FR-UNet for blood vessel segmentation. However, Spe is the lowest, and the reason was analyzed in the last subsection. In summary, the results demonstrate that the strategy of FR-UNet, which maintains full-resolution representation learning and superimposes deep supervision and feature aggregation modules, is qualified for blood vessel segmentation.

C. Model Pruning

As an essential valuation metric for model performance, network parameters affect training and inference time. Although we deliberately decreased the number of channels and simplified skip-connections to alleviate this problem, FR-UNet still has 5.72 million parameters. Therefore, we propose a compromise solution to reduce the number of model parameters through model pruning. FR-UNet is generated by iteratively embedding convolution nodes on the periphery of the initial encoder-decoder network. We sequentially pruned the outermost nodes of FR-UNet to obtain four subnetworks P1, P2, P3, and P4 with different pruning levels. As shown in Fig. 6, the blue convolutional nodes in the P1-P4 subgraph constitute the subnetwork, and the gray convolutional nodes are removed in FR-UNet. As a result, the horizontal and vertical depths of subnetworks P1 to P4 are gradually reduced, but the segmentation performance may degrade. We evaluate the model's performance on the DRIVE dataset in two ways: 1) training using the full FR-UNet and testing using subnetworks (TFTS), in which the network parameters are trained using the full FR-UNet model, and the pruned subnetwork gives the corresponding training weights for testing; 2) both training and testing in subnetworks (TTS), where the subnetwork is trained and tested in isolation without any interaction with other networks.

As shown in Table V, the number of parameters is gradually reduced by at least 50% from FR-UNet to P4, but the vessel segmentation performance gradually degrades according various evaluation metrics. Specifically, the number of parameters of the P1 subnet is 2.60 M, which is approximately 45% that of the full FR-UNet. However, the overall performance is close to that of full FR-UNet regardless of TFTS or TTS, and even P1's Acc and Spe scores are slightly superior to those of the full FR-UNet. Compared to the results of each pruning subnetwork of TFTS and TTS, the vessel segmentation outcome of TTS is significantly better than that of the same subnetwork of TFTS in P2, P3, and P4. Moreover, the performance gap between the subnetworks of the TTS is smaller. Due to the different principles of the two methods, we believe that the evaluation results are reasonable. TFTS and TTS are suitable for different situations: 1) To meet the inference time requirements, TFTS can be deployed directly for a pruned model trained by full FR-UNet without retraining. 2) TTS is suitable for more flexible situations where subnetworks with different pruning levels are trained and tested as needed to achieve a trade-off between inference time and segmentation performance.

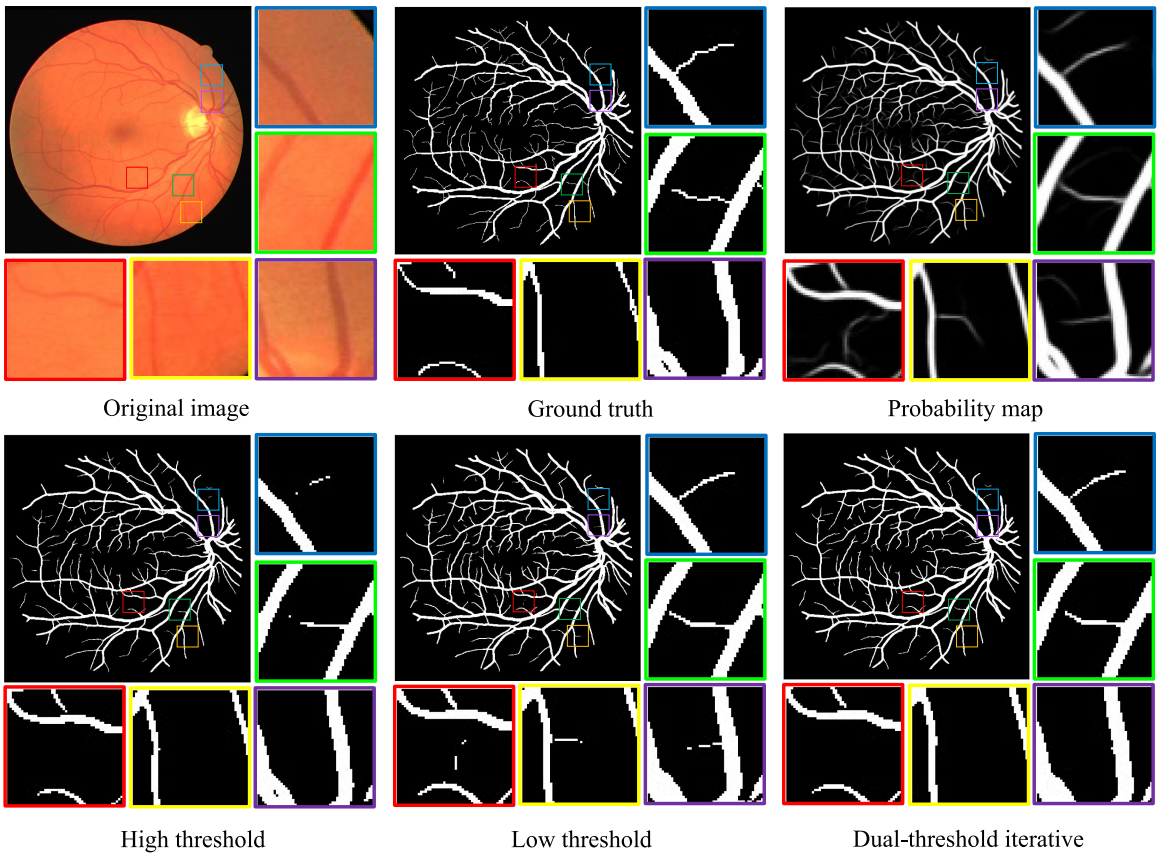

Fig. 7. Visualization of dual-threshold iteration and single-threshold results on DRIVE dataset. The first row of images from left to right are the original images, ground truths, and the probability maps predicted by FR-UNet; the second row is the binary map obtained by applying high threshold, low threshold, and dual-threshold iteration to the probability map. To show the details, we enlarge the five highlighted rectangular regions in the images.

D. Comparison of Threshold Methods

In order to estimate the proposed dual-threshold iterative method, we separately applied single-threshold (ST) and DTI algorithms to the probability map predicted by the network model to obtain a binary segmentation map. For the predicted probability map $p \in (0, 1)$, we set the high threshold at 0.5 and empirically determined the low threshold at 0.3. The visualization result is shown in Fig. 7. We observe enlarged subgraphs in which the thin vessels in the original image have extremely low contrast and are difficult for ordinary people to recognize. Nevertheless, the experts annotated the thin vessels in the blue and green subgraphs but did not annotate them in the red, yellow, and purple subgraphs. The probability map shows that the gray value difference between the two types of vessels predicted by the network model is slight. It is observed from the blue and green subgraphs that the thin vessels extracted by the high threshold are not complete, whereas the DTI preserves more details closer to the ground truth. Furthermore, the low threshold produced oversegmentation in the red, yellow, and green subgraphs. However, unlike the low threshold, DTI did not produce oversegmentation due to its iterative optimization mechanism that emphasizes the neighborhood relationships between pixels. It can be seen from the histogram of the evaluation metrics in

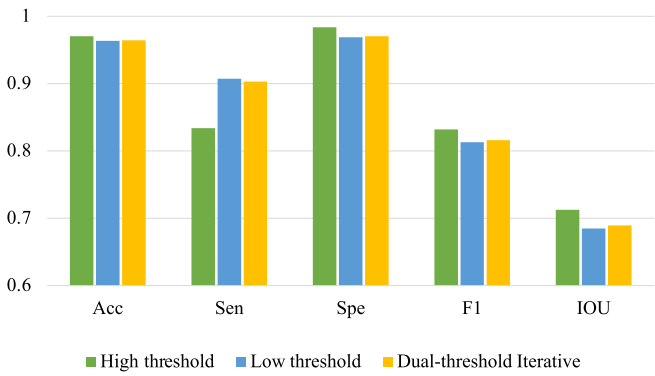

Fig. 8. The performance comparison of single-threshold and dual-threshold iterative methods on the same probability map. The green, blue, and yellow histograms denote the high threshold, low threshold, and dual-threshold iteration, respectively.

Fig. 8 that DTI is significantly higher than the high threshold on Sen while the others are low, and DTI is slightly better than the low threshold in overall performance. Therefore, the low threshold extracts more thin vessel pixels than the high threshold, but more false-negative pixels are introduced. In this regard, DTI effectively alleviates this problem. Considering the connectivity

TABLE VI
STATISTICS OF CONNECTED COMPONENTS WHERE THE RESULTS ARE PRODUCED BY SINGLE-THRESHOLD ($\gamma = 0.5$, 0.4 AND 0.3) AND DUAL-THRESHOLD ITERATION METHODS ($\gamma_h, \gamma_l = 0.5$ AND 0.4 AND $\gamma_h, \gamma_l = 0.5$ AND 0.3) WITH DIFFERENT PARAMETERS ON THE DRIVE DATASET

Network models	VCA↓				
	ST			DTI	
	0.5	0.4	0.3	0.5+0.4	0.5+0.3
Unet [10]	19.76	16.33	15.26	9.94	7.26
Unet++ [23]	18.86	14.38	**12.81**	9.46	6.79
Attention U-Net [35]	20.68	15.39	14.56	10.04	7.23
HRNet [20]	19.59	**14.20**	13.73	8.04	5.81
CSNet [34]	25.50	20.09	16.89	9.55	6.34
FR-UNet	**18.44**	15.29	13.16	**5.84**	**3.31**

of vessel structures, DTI gradually explores the weak vessel pixels between the high and low thresholds, starting with the vessel pixels extracted from the high threshold. DTI introduces fewer false negative pixels than low thresholds and visually improves vessel connectivity.

In addition to the evaluation of pixel-level dense prediction, the connectivity analysis of vessel segmentation is also critical. However, there are few works in the literature on the quantitative evaluation of vascular connectivity. For this, we innovatively calculate the number of connected components for connectivity analysis. Theoretically, the smaller the number of connected components on the map, the better the connectivity of the vessels. The characteristics of biological tissues mean that vessels are connected and cannot exist independently. Nevertheless, the vessel annotation has breakpoints, resulting in a ground truth with multiple connected components. According to the statistics, the total number of connected components of 20 ground truths is 80 on the DRIVE test set, which is a standard to measure connectivity. The vessel connectivity assessment (VCA) is calculated as:

$$VCA = \frac{1}{N}\sum_{k=1}^{N}\frac{\mathcal{L}(pre)}{\mathcal{L}(gt)}, \tag{10}$$

where pre and gt denote the binarized prediction map and ground truth, respectively. $\mathcal{L}()$ denotes the calculation of the number of connected components, and N is the number of vessel images in the test set, which is 20 for DRIVE. Theoretically, the lower the value of VCA, the better the connectivity of vessels. We applied ST and DTI to the probability maps predicted by different network models and calculated the VCA. It can be seen from the data in Table VI that the VCA calculated by DTI is significantly lower than that of ST among all networks. For example, the number of connected components is the lowest when $\gamma = 0.3$ in ST (VCA $= 13.16$) of FR-UNet. In contrast, VCA is only 3.31 in DTI when $\gamma_h = 0.5$ and $\gamma_l = 0.3$, which is approximately 25.15% of the best result of the single threshold method. It demonstrates how the dual-threshold iteration method achieves excellent vessel connectivity when binarizing the vessel prediction map. Furthermore, from the VCA results of each network, our FR-UNet achieves the best connectivity, which further demonstrates that it can extract more vessels.

VI. CONCLUSION

In this article, a full-resolution UNet for vessel segmentation is proposed. In the network, the parallel convolutional layers horizontally extend to learn full-resolution representations continuously, and the feature aggregation module aggregates multiscale representations to interact information with different resolutions and supplement high-level contextual representations to form full-resolution feature maps. The architecture of FR-UNet is simple but effective and can be pruned at different levels to meet efficient application scenarios. The evaluation metrics show that FR-UNet has achieved outstanding comprehensive performance compared with other state-of-the-art methods. Moreover, the visualized segmentation results of retinal vessels and coronary angiography illustrate FR-UNet's powerful vessel extraction capabilities. In addition, we introduced dual-threshold iteration instead of a single-threshold method to extract weak vessel pixels further to improve blood vessel connectivity. In the future, we will continue to explore this work and extend it to 3D vessel segmentation such as magnetic resonance angiography and computed tomography angiography.

REFERENCES

[1] W. Xu, H. Yang, M. Zhang, X. Pan, W. Liu, and S. Yan, "DecNet: A dual-stream edge complementary network for retinal vessel segmentation," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2021, pp. 1595–1600.

[2] R.-Q. Li et al., "Real-time multi-guidewire endpoint localization in fluoroscopy images," *IEEE Trans. Med. Imag.*, vol. 40, no. 8, pp. 2002–2014, Aug. 2021.

[3] T. A. Soomro et al., "Deep learning models for retinal blood vessels segmentation: A review," *IEEE Access*, vol. 7, pp. 71696–71717, 2019.

[4] U. Mitrović, B. Likar, F. Pernuš, and ž. Špiclin, "3D–2D registration in endovascular image-guided surgery: Evaluation of state-of-the-art methods on cerebral angiograms," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 2, pp. 193–202, 2018. [Online]. Available: https://doi.org/10.1007/s11548-017-1678-2

[5] M. Javidi, H.-R. Pourreza, and A. Harati, "Vessel segmentation and microaneurysm detection using discriminative dictionary learning and sparse representation," *Comput. Methods Programs Biomed.*, vol. 139, pp. 93–108, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260716305454

[6] M. Javidi, A. Harati, and H. Pourreza, "Retinal image assessment using bi-level adaptive morphological component analysis," *Artif. Intell. Med.*, vol. 99, 2019, Art. no. 101702.

[7] C. Wang et al., "Tensor-cut: A tensor-based graph-cut blood vessel segmentation method and its application to renal artery segmentation," *Med. Image Anal.*, vol. 60, 2020, Art. no. 101623.

[8] S. Kalaie and A. Gooya, "Vascular tree tracking and bifurcation points detection in retinal images using a hierarchical probabilistic model," *Comput. Methods Programs Biomed.*, vol. 151, pp. 139–149, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260716314201

[9] D. Jia and X. Zhuang, "Learning-based algorithms for vessel tracking: A review," *Computerized Med. Imag. Graph.*, Vol. 89, Apr. 2021, Art. no. 101840 .

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[11] 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. [Online]. Available: https://doi.org/10.1109/TMI.2016.2546227

[12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440. [Online]. Available: https://doi.org/10.1109/CVPR.2015.7298965

[13] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional convLSTM U-net with densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, Seoul, Korea (South), Oct. 27 28, 2019, pp. 406–415, [Online]. Available: https://doi.org/10.1109/ICCVW.2019.00052

[14] L. L. Mou et al., "Cs-Net: Channel and spatial attention network for curvilinear structure segmentation," in *Proc. 22nd Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2019, vol. 11764, pp. 721–730. [Online]. Available: https://doi.org/10.1007/978-3-030-32239-7_80

[15] 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. [Online]. Available: https://doi.org/10.1016/j.media.2021.102025

[16] P. M. Samuel and T. Veeramalai, "VSSC Net: Vessel specific skip chain convolutional network for blood vessel segmentation," *Comput. Methods Programs Biomed.*, vol. 198, 2021, Art. no. 105769. [Online]. Available: https://doi.org/10.1016/j.cmpb.2020.105769

[17] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, K. M. Sanders, and S. A. Baker, "RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervention*, 2021, pp. 34–44.

[18] Y. Zhou, H. Yu, and H. Shi, "Study group learning: Improving retinal vessel segmentation trained with noisy labels," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2021, pp. 57–67.

[19] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3309–3318. [Online]. Available: https://doi.org/10.1109/CVPR.2017.353

[20] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Computer Vision Foundation / IEEE, Jun. 16–20, 2019, pp. 5693–5703.

[21] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Computer Vis. Pattern Recognit.*, 2019, pp. 5693–5703.

[22] J. Wang et al., "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021.

[23] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet : A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, D. Stoyanov et al. Eds. Cham, Switzerland: Springer, 2018, pp. 3–11.

[24] Z. Zhou, M. M. R. 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, Jan. 2020. [Online]. Available: https://doi.org/10.1109/TMI.2019.2959609

[25] 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. [Online]. Available: https://doi.org/10.1109/CVPR.2016.90

[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, 2015, vol. 37, pp. 448–456. [Online]. Available: http://proceedings.mlr.press/v37/ioffe15.html

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

[28] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004. [Online]. Available: https://doi.org/10.1109/TMI.2004.825627

[29] A. Carballal et al., "Automatic multiscale vascular image segmentation algorithm for coronary angiography," *Biomed. Signal Process. Control.*, vol. 46, pp. 1–9, 2018. [Online]. Available: https://doi.org/10.1016/j.bspc.2018.06.007

[30] S. Zhang et al., "Attention guided network for retinal image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2019, pp. 797–805.

[31] W. Wang, J. Zhong, H. Wu, Z. Wen, and J. Qin, "RVSeg-Net: An efficient feature pyramid cascade network for retinal vessel segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2020, pp. 796–805.

[32] 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.

[33] F. Cervantes-Sanchez, I. Cruz-Aceves, A. Hernandez-Aguirre, M. A. Hernandez-Gonzalez, and S. E. Solorio-Meza, "Automatic segmentation of coronary arteries in x-ray angiograms using multiscale analysis and artificial neural networks," *Appl. Sci.*, vol. 9, no. 24, 2019, Art. no. 5507. [Online]. Available: https://www.mdpi.com/2076-3417/9/24/5507

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds., San Diego, CA, USA, May 7–9, 2015. [Online]. Available: http://arxiv.org/abs/1412

[35] J. Schlemper et al., "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207 2019.

[36] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-UNet for high-quality retina vessel segmentation," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ.*, 2018, pp. 327–331.