Abstract—Optical Coherence Tomography Angiography (OCT-A) is a non-invasive imaging technique, and has been increasingly used to image the retinal vasculature at capillary level resolution. However, automated segmentation of retinal vessels in OCT-A has been under-studied due to various challenges such as low capillary visibility and high vessel complexity, despite its significance in understanding many eye-related diseases. In addition, there is no publicly available OCT-A dataset with manually graded vessels for training and validation. To address these issues, for the first time in the field of retinal image analysis we construct a dedicated Retinal OCT-A SEgmentation dataset (ROSE), which consists of 229 OCT-A images with vessel annotations at either centerline-level or pixel level. This dataset has been released for public access to assist researchers in the community in undertaking research in related topics. Secondly, we propose a novel Split-based Coarse-to-Fine vessel segmentation network (SCF-Net), with the ability to detect thick and thin vessels separately. In the SCF-Net, a split-based coarse segmentation (SCS) module is first introduced to produce a preliminary confidence map of vessels, and a split-based refinement (SRN) module is then used to optimize the shape/contour of the retinal microvasculature. Thirdly, we perform a thorough evaluation of the state-of-the-art vessel segmentation models and our SCF-Net on the proposed ROSE dataset. The experimental results demonstrate that our SCF-Net yields better vessel segmentation performance in OCT-A than both traditional methods and other deep learning methods.

Index Terms—Optical coherence tomography angiography, vessel segmentation, deep network, benchmark.

I. INTRODUCTION

The vasculature is an essential structure in the retina, and its morphological changes can be used not only to identify and classify the severity of systemic, metabolic, and hematologic diseases [1], but also to facilitate a better understanding of disease progression, and assessment of therapeutic effects [2]. Color fundus is the most commonly used retinal imaging technique: however, it is difficult with this method to capture microvascular structures (thin vessels and capillaries), which are surrounded in the fovea and parafovea regions, as shown in the green rectangle area of Fig. 1 (A). Fluorescein angiography and indocyanine green angiography can resolve the retinal vasculature including capillaries, but these are invasive techniques and may cause severe side effects and even death [3].

Y. Ma, H. Hao, J. Yang and Y. Zhao are with Cixi Institute of Biomedical Engineering, Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo, China. H. Fu is with Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates. J. Zhang is with Keck School of Medicine, University of Southern California, Los Angeles, US. J. Liu is with Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China. Y. Zheng is with Department of Eye and Vision Science, University of Liverpool, Liverpool, UK.

Corresponding author: Yitian Zhao (yitian.zhao@nimte.ac.cn), and Yalin Zheng (yalin.zheng@liverpool.ac.uk).

Fig. 1. Illustration of macula in color fundus image and macula-centred (green rectangle area) $3 \times 3 \text{mm}^2$ OCT-A images of the same eye: (A) color fundus, (B) superficial retinal vascular plexus, (C) deep inner retinal vascular plexus, (D) outer retina, (E) choriocapillaris and (F) the total macula vascular plexus.

In contrast, Optical Coherence Tomography Angiography (OCT-A) is a newly emerging non-invasive imaging technique, with the ability to produce high-resolution 3D images of the retinal vasculature, and has been increasingly accepted as a valuable imaging tool with which to observe retinal vessels [4], [5]. By means of OCT-A imaging technology, such as is provided by the RTVue XR Avanti SD-OCT system (Optovue, Inc, Fremont, California, USA), equipped with AngioVue software (version 2015.1.0.90), four en face angiographs were generated by the maximum projection of OCT-A flow signals within the slab, as shown in Fig. 1 (B-E): superficial and deep inner retinal vascular plexuses, outer retina, and choriocapillaris. Several works have already demonstrated that a combination of a C-scan (frontal scan, or en face) and B-scan (cross-sectional scan), enables clear vision of the superficial and deep inner retinal vascular plexuses, and the total macula plexus, as shown in Fig. 1 (B, C, and F). OCT-A enables observation of microvascular details down to capillary level, permitting quantitative assessment of the microvascular and morphological structures in the retina.

By extracting microvascular structures from different OCT-A depth layers, one can obtain their corresponding en face projections to analyze their respective variations. In particular, the microvasculature distributed within the parafovea is of great interest, as any abnormality there often indicates the presence of any of a number of diseases such as early stage glaucomatous optic neuropathy, diabetic retinopathy, and age-related macula degeneration [6]–[8]. More recently, several studies have shown that the morphological changes
of microvasculature revealed by OCT-A are associated with Alzheimer's Disease and Mild Cognitive Impairment [9]. A new avenue is thereby opened up to study the relation between the appearance of retinal vessels and various neurodegenerative diseases. Thus, automatic vessel detection and quantitative analysis of OCT-A images are of value for the early diagnosis of vascular-related diseases affecting retinal circulation, and the assessment of disease progression. However, automated vessel segmentation in OCT-A images has been explored only rarely, and remains a challenging task, despite the fact that many medical segmentation approaches - particularly deep learning based techniques [10]–[12] - have achieved great success in segmenting blood vessels.

There is no publicly available OCT-A dataset with manual vessel annotations with which to train a supervised model, which hinders the further validation of OCT-A segmentation techniques. To our best knowledge, only four automated methods have been demonstrated to segment the retinal vessels, each using its own private OCT-A dataset. Mou et al. [1] trained a deep learning method [1] with 30 OCT-A images, but the small dataset used suggests that this method may not be universally applicable across different pathological scenarios. Eladawi et al. [13] proposed an automatic method on 47 OCT-A images. Li et al. [14] more recently introduced an image projection network that can achieve 3D-to-2D vessel segmentation: they evaluated their method on 316 OCT-A images. Zhang et al. [8] set up the first 3D OCT-A microvascular segmentation approach to directly extract 3D capillary networks from OCT-A volume data for subsequent shape modeling and analysis. However, they mainly evaluated the test re-test reliability of their 3D framework on 360 OCT-A volume images, as there were no manually annotated 3D vessel networks available.

Although these methods achieve useable segmentations for OCT-A vessel analysis, the privacy of their datasets makes a unified evaluation benchmark impossible. In addition, Eladawi [13] and Li et al. [14] set up their segmentation frameworks based on the OCT-A data within a $6 \times 6 \ mm^2$ fovea-centered field of view (FOV). Previous findings by Zhang et al. [8] have shown that small capillaries play a much more important role in distinguishing different disease groups compared with relatively large vessels. It is therefore necessary to establish a dedicated OCT-A dataset focusing on more detailed capillary networks within a $3 \times 3 \ mm^2$ FOV. Fig. 2 demonstrates a comparison between $3\times3 \ mm^2$ and $6\times6 \ mm^2$ FOVs. We can clearly observe the richer capillary information graded by experts from the $3 \times 3 \ mm^2$ scans. Throughout this work, therefore, the proposed SCF-Net will be fully evaluated using the well-established $3 \times 3 \ mm^2$ OCT-A dataset for more detailed microvascular study.

The OCT-A imaging process typically produces images with a low signal-to-noise ratio (SNR) [5]. Additionally, varying vessel appearances at different depth layers, motion artefacts, and the potential existence of pathologies [8] increase the challenge for achieving accurate segmentations, particularly for densely connected capillaries, which can easily result in discontinuous segmentations. Most deep learning-based methods are region-based [15], a technique which is prone to produce imprecise and discontinuous vessel segmentation results, and existing methods do not perform well when required to detect subtle differences in microvascular networks with different vessel thicknesses and imaging depths. To this end, a coarse-to-fine vessel segmentation framework is discover edge information by paying particular attention to the regions that are not salient in high-level semantic features.

A. Contributions

In order to mitigate the issues of lacking a public retinal OCT-A dataset and effective microvascular segmentation methods, we make the following contributions in this work.

- For the first time in the retinal image analysis field, we establish a publicly available retinal OCT-A dataset, with precise manual annotations of retinal microvascular networks, in order to promote relevant research topics in the community.
- We propose a novel Split-based Coarse-to-Fine vessel segmentation network (SCF-Net) with a coarse-to-fine stage for blood vessel segmentation in OCT-A, aimed at detecting thick and thin vessels separately. In our method, a split-based coarse segmentation (SCS) module is first introduced to produce a preliminary confidence map of vessels, and a split-based refinement (SRN) module is then used to optimize towards the finer vessels, with a view to obtaining more accurate overall segmentation results.
- We give a full evaluation/benchmarking of OCT-A microvascular segmentation, both quantitatively and qualitatively. Comparative analysis shows that the proposed SCF-Net works robustly on different types of retinal images and yields accurate vessel segmentations.

To further promote developments in this field, the code, baseline models, and evaluation tools, are publicly available at https://imed.nimte.ac.cn/dataofrose.html

II. RELATED WORKS

In the past two decades, we have witnessed a rapid development of retinal blood vessel segmentation methods for
color fundus images (as evidenced by extensive reviews [2], [16]). As blood vessels are curvilinear structures distributed across different orientations and scales, the conventional vessel segmentation methods are mainly based on various filters, include Hessian matrix-based filters [17], matched filters [18], multi-oriented filters [19], symmetry filters [2], and tensor-based filters [20]. These filter-based methods aim to suppress non-vascular or non-fiber structures and image noise, and enhance the curvilinear structures, thereby simplifying the subsequent segmentation problem.

In recent years, deep learning-based methods have made significant progress in the fields of medical image segmentation. Broadly, many deep neural networks have been modified and applied for blood vessel segmentation [1], [10], [12], [21], [22] and have yielded promising results. However, the extraction of vessels from OCT-A images has been relatively unexplored. We will review and discuss the most relevant vessel segmentation works in this section.

A method based on Convolutional Neural Network (CNN) [21] was proposed to enhance training samples for better retinal vessel detection: subsequently, a Conditional Random Field (CRF) was incorporated into the CNN by Fu et al. for retinal vessel detection [10]. Wang et al. [23] applied the U-Net convolutional network [24] for retinal vessel segmentation in fundus images of pathological conditions. Xiao et al. [25] modified ResU-Net [26] by introducing a weighted attention mechanism for high-quality retinal vessel segmentation. Gu et al. [11] proposed a context encoder network (CE-Net), which consists of dense atrous convolution and residual multi-kernel pooling modules for retinal vessel image segmentation. Mou et al. [1] proposed a channel and spatial attention network (CS-Net) for curvilinear structures (including vessels in some example OCT-A images): they then applied spatial and channel attention to further integrate local features with their global dependencies adaptively. Jin et al. [27] integrated deformable convolution into the DUNet, which is designed to extract context information and enable precise localization by combining low-level features with high-level ones. Yan et al. [28] proposed a three-stage deep learning model to segment thick vessels and thin vessels separately, which achieves accurate vessel segmentation for both types of vessels. Eladawi et al. [13] introduced a joint Markov-Gibbs random field model to segment the retinal vessels in different OCT-A projection maps. Li et al. [14] presented an image projection network, which is a novel end-to-end architecture that can achieve 3D-to-2D image segmentation in OCT-A images.

III. DATASET

Our constructed Retinal OCT-Angiography vessel SEGmentation (ROSE) dataset comprises of two subsets, named as ROSE-1 and ROSE-2, which were acquired by two different devices. All the OCT-A images included in this dataset were acquired with appropriate institutional approvals. The details of ROSE dataset are as follows.

A. ROSE-1

Th ROSE-1 set consists of a total of 117 OCT-A images from 39 subjects (including 26 with Alzheimers disease (AD) and 13 healthy controls). All the OCT-A scans were captured by the RTVue XR Avanti SD-OCT system (Optovue, USA) equipped with AngioVue software, with an image resolution of 304 × 304 pixels. The scan area was 3 × 3 mm² centered on the fovea, within an annular zone of 0.6 mm-2.5 mm diameter around the foveal center. Each subject supplied en face angiographs of the superficial (SRVP), deep (DRVP), and total macula (MRVP) retinal vascular plexuses, respectively. The SRVP extended from 3 μm below the internal limiting membrane (ILM) to 15 μm below the inner plexiform layer.
Fig. 4. Example of an OCT-A image and its centerline-level manual annotations in ROSE-2.

(IPL): the DRVP extended from 15 to 70 $\mu m$ below the IPL; and the MRVP extended from 3 $\mu m$ beneath the retinal pigment epithelium (RPE) layer. These plexuses were distinguished and separated automatically, using the proprietary tool supplied with the device.

Two different types of vessel annotations were made by image experts and clinicians for the ROSE-1 dataset, and the consensus of them was then used as the ground truth:

1. **Centerline-level annotation.** The centerlines of vessels were manually traced using ImageJ software [29] by our experts on the SRVP, DRVP, and MRVP images individually, as shown in Fig. 3 A-2, B-2, and C-2.

2. **Pixel-level annotation.** We first invited an image expert to grade the complete microvascular segments with varying diameters in the SRVP and MRVP images at pixel level. Since it is difficult for a human expert to perceive the diameters of small capillaries located around the macula region, we asked the expert to grade the small capillaries at centerline-level. The combination of these different labels is defined as the final pixel-level annotation, as shown in Fig. 3 A-3, B-3, and C-3. (Note that, Fig. 3 B-3 is also the centerline-level label of the DRVP, as it is difficult to obtain pixel-level grading in this layer.)

**B. ROSE-2**

The ROSE-2 subset contains a total of 112 OCT-A images taken from 112 eyes, acquired by a Heidelberg OCT2 system with Spectralis software (Heidelberg Engineering, Heidelberg, Germany). These images are from eyes with various macula diseases. All the images in this dataset are en face angiograms of the SRVP within a $3 \times 3$ mm$^2$ area centred at the fovea. These OCT-A images were reconstructed from 512 $\times$ 512 repeated A-scans, with the Heidelberg automated real time (ART) and TruTrack system employed to reduce artefacts and noise. Each image was resized into a grayscale image with 840 $\times$ 840 pixels. All the visible retinal vessels were manually traced using an in-house program written in Matlab (Mathworks R2018, Natwick) by an experienced ophthalmologist. An example OCT-A image and its corresponding centerline-level annotation are shown in Fig. 4.

**IV. PROPOSED METHOD**

In this section, we introduce a novel Split-based Coarse-to-Fine network (SCF-Net) for retinal vessel segmentation in OCT-A images. The overall proposed pipeline has two indispensable stages, as illustrated in Fig. 5. First, we design a split-based coarse segmentation (SCS) module, which consists of separate thick- and thin-vessel segmentation networks, to produce two preliminary confidence maps. A split-based refinement (SRN) module is then used to fuse these vessel confidence maps to produce the final optimized results.

**A. Split-based Coarse Segmentation Module**

Since the ROSE dataset has both pixel-level and centerline-level vessel annotations for each OCT-A image, we design a SCS module with two components with an encoder and two decoders, to balance the importance of both pixel-level and centerline-level vessel information.

**Pixel-level vessel segmentation:** ResNet block [30], a novel ResNet-style structure with split attention module, is used as the backbone of the pixel-level vessel segmentation network. The ResNet block consists of two modules: feature-map groups with the same cardinal group index reside next to each other, and split attention into cardinal groups, as shown in top row of Fig. 6. First, the input of the module is divided into $K$ cardinal groups, and each cardinal group is further split into $R$ feature-map groups. Therefore, the total number of feature-map groups is $G = K \cdot R$. Then the $1 \times 1$, and $3 \times 3$ convolutional layers are followed by batch normalization (BN), and ReLU layers are applied to each feature-map group to obtain intermediate representations $\{U_1, U_2, \cdots, U_G\}$.

In each cardinal group, a combined representation obtained by an element-wise summation across multiple feature-map splits is denoted by $\hat{U}_c^k$, where $U_c^k \in \mathbb{R}^{H \times W \times C/K}$ for $k \in \{1, 2, \ldots, K\}$, and $H$, $W$, and $C$ are output sizes of the block. In addition, a split attention module is applied to $\hat{U}_c^k$ to obtain final representation of the cardinal group. The bottom row of Fig. 6 shows the structure of the split attention module, which includes a global pooling layer, fully connected (FC) layers with BN and ReLU layers and a Softmax layer. First, $\hat{U}_c^k$ is fed into the global average pooling to aggregate over the global context representation $s_c^k \in \mathbb{R}^{C/K}$. Thus, the $c$-th channel of $s_c^k$ is calculated as:

\[
s_c^k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \hat{U}_c^k(i,j).
\]

Two FC layers with the same number of groups of cardinal group are added after the global pooling layer, to predict the attention weights for each split. Afterwards, a weighted fusion of the cardinal group representation $V_c^k \in \mathbb{R}^{H \times W \times C/K}$ is obtained using the attention weights by a Softmax layer. In addition, each feature-map channel of $V_c^k$ is assigned a soft weight and produced using a weighted combination over splits. To this end, the $c$-th channel of $V_c^k$ is defined as:

\[
V_c^k = \sum_{i=1}^{R} a_i^k(c) U_{R(k-1)+i},
\]

where $a_i^k(c)$ denotes a (soft) assignment weight given by:

\[
a_i^k(c) = \begin{cases} 
\frac{\exp(G_i^k(s_c^k))}{\sum_{j=0}^{R} \exp(G_j^k(s_c^k))} & \text{if } R > 1, \\
1 + \exp(-G_i^k(s_c^k)) & \text{if } R = 1,
\end{cases}
\]
Fig. 5. Architecture of the proposed two stage SCF-Net network. The SCS module (coarse stage) is designed to produce two preliminary two confidence maps that segment pixel-level (thick) and centerline-level (thin) vessels, respectively. The SRN module (fine stage) is then introduced as a fusion network to obtain the final refined segmentation results.

Fig. 6. ResNeSt block with split attention module.

where mapping $\mathcal{G}_c^k$ determines the weight of each split for the $c$-th channel based on $s_k$.

The cardinal group representations are concatenated along the channel dimension: $V = \text{Concat} \{V^1, V^2, \ldots, V^K\}$. As in standard residual blocks, the final output $Y$ of the ResNeSt block is produced using a shortcut connection: $Y = V + X$, if the input and output feature-maps share the same shape. For blocks with a stride, an appropriate transformation $T$ is applied to the shortcut connection to align the output shapes: $Y = V + T(X)$. For example, $T$ can be stride convolution, or combined convolution with pooling.

**Centerline-level vessel segmentation:** Compared with pixel-level annotation, vessel annotation at centerline-level aims to grade the vessels in regions with poor contrast, more complex topological structures, and relatively smaller diameters. On one hand, considering the differences between centerline-level and pixel-level vessels, the features used for pixel-level vessel segmentation might not be suitable for centerline-level vessel segmentation. The deeper architecture might be detrimental to closer attention to low-level features, which are of great significance for centerline-level vessel segmentation. On the other hand, pixel- and centerline-level vessel segments may reveal shared features after feature extraction, due to spatial dependencies between the two types of vessel annotations. Based on the above considerations, we append several convolutional layers (with BN and ReLU) followed by an up-sampling layer in the second encoder layer of the backbone, as the decoder of the centerline-level vessel segmentation network. The outputs of the decoder are then concatenated with the feature maps from the first encoder layer of the backbone, and processed by one $1 \times 1$ convolutional layer with a Sigmoid function to achieve the centerline-level segmentation map.

**B. Split-based Refinement Module**

In order to fuse pixel-level and centerline-level vessel information from the SCS module, and further recover continuous details of small vessels, we propose a split-based refinement (SRN) module to adaptively refine the vessel prediction results. The structure of our SRN module is illustrated as the refinement stage in Fig. 5. Inspired by [31], the predicted pixel-level and centerline-level vessel maps and the original (single channel) OCT-A image are first concatenated as input to the SRN module. In the SRN module, a mini network including three convolutional layers with $3 \times 3$ kernels is designed to refine the pixel-level map derived from the coarse stage. As in the case of the coarse stage, one additional $3 \times 3$ convolutional layer is appended to the second layer of the mini network to refine the centerline-level map from the first stage. BN and ReLU layers are adopted after each convolutional layer.

The refined pixel- and centerline-level maps are then merged into a complete vessel segmentation map, by choosing the larger value from the two maps at each pixel. The detailed channel configuration of the SRN module is shown in Fig. 5.
For both pixel- and centerline-level vessel refinements, the module produces normalized \( m \times m \) local propagation coefficient maps for all the positions, formulated as:

\[
    w_{ij}^p = \frac{\exp(h_{ij}^p)}{\sum_{t=1}^{m \times m} \exp(h_{ti}^t)}, \quad p \in 1, 2, \ldots, m \times m, \tag{4}
\]

where \( h_{ij}^p \) is the confidence value at position \( i \) for its neighbor \( p \), and \( m \times m \) is the size of propagation neighbors. Finally, the local propagation coefficient vector \( w_{ij}^p \) at position \( i \) will be multiplied by the confidence maps of thick or thin vessels from the front model and aggregate to the center point to generate the refinement result, denoted as:

\[
    g_i = \sum_{p=1}^{m \times m} w_{ij}^p \cdot f_{ij}^p, \tag{5}
\]

where \( f_{ij}^p \) is the confidence vector of the neighbor \( p \) at position \( i \) from the SCS module, and \( g_i \) is the final predicted vector of position \( i \).

Note that the propagation coefficient maps can learn the spatial relationship between position \( i \) and its neighbors to refine the structure information of both pixel-level and centerline-level vessels. In addition, the final vessel map must be similar to that before refinement. To achieve this goal, the coefficient of position \( i \) should be far larger than that of its neighbors, and we adopt a reasonable method for initialization following [31].

\[\begin{array}{cccccc}
\text{image} & \text{ground truth} & \text{U-Net} & \text{CE-Net} & \text{CS-Net} & \text{backbone} \\
\text{SFC-Net} & \end{array}\]

\[\begin{array}{cccccc}
\text{ROSE-1} & \text{MRVP} & \text{SRVP} & \text{DRVP} & \text{MRVP} & \text{SRVP} \\
\text{ROSE-2} & \end{array}\]

Fig. 7. Vessel segmentation results of different methods on different layers of ROSE-1 and ROSE-2. From left to right: en face angiographs (original images), manual annotations, vessel segmentation results obtained by U-Net, CE-Net, CS-Net, backbone (ResNeSt) and the proposed method (SFC-Net), respectively.

V. EXPERIMENTS

A. Experimental Setting

The proposed method was implemented with PyTorch and run in parallel on a workstation equipped with two NVIDIA GTX 1080 (8GB) GPUs. Both the first stage and the second stage were trained with 200 epochs and with the following settings: Adam optimization with the initial learning rate of 0.0005, batch size of 2 and weight decay of 0.0001. For the first stage, We set \( r = 16 \) as the reduction ratio of the FC layers in the split attention modules, and selected mean square error (MSE) as the loss function. For the second stage, we set \( m = 3 \) as the size of aggregation neighbors, and used the Dice coefficient (Dice) as the loss function. For more stable training, we adopted a poly learning rate policy with power 0.9. For training and inference of the proposed method, the ROSE-1 subset was split into 90 images for training and 27 images for testing, and the ROSE-2 subset was split into 90 images for training and 22 images for testing. Data augmentation was conducted by randomly rotation during all training stages.

B. Evaluation Metrics

To achieve comprehensive and objective assessment of the segmentation performance of the proposed method, the following metrics are calculated and compared:

- Area Under the ROC Curve (AUC);
- Kappa score;
- Accuracy (ACC) = (TP + TN) / (TP + TN + FP + FN);
- Sensitivity (SEN) = (TP) / (TP + FN);
- False Discovery Rate (FDR) = FP / (FP + TP);
- G-mean score = \( \sqrt{\text{Sensitivity} \times \text{Specificity}} \);
- Dice coefficient (Dice) = 2 \times TP / (FP + FN + 2 \times TP);

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative. The use of specificity, defined as the number of correctly classified pixels on the
true negative class, is not adequate for the evaluation of this segmentation task, since the vast majority of pixels do not belong to vessels. Specifically, for centerline-level vessel detection in the DRVP images from the ROSE-1 and all images from the ROSE-2, a three-pixel tolerance region around the manually traced centerlines is considered a true positive, which follows the evaluating methods for extracting one pixel-wide curves in [32].

C. Performance Comparison and Analysis

We have thoroughly evaluated the proposed method over our ROSE dataset, and compared it to existing state-of-the-art segmentation methods to demonstrate the superiority of the proposed method in the segmentation of OCT-A microvasculature.

Comparison methods. In order to verify the superiority of our method, we compared our method with other state-of-the-art segmentation methods on both ROSE-1 and ROSE-2, including three conventional methods: infinite perimeter active contour (IPAC) [33], trainable COSFIRE filters [17], and curvelet denoising based optimally oriented flux enhancement (COOF) [8] for their effectiveness in detecting vessels with irregular and oscillatory boundaries; and six deep learning approaches: U-Net [24], ResU-Net [26], CE-Net [11], AG-Net [34], CS-Net [1] and three-stage networks [28]. For [33] and [17], the parameters were tuned to achieve segmentation results of all vessels in both ROSE-1 and ROSE-2 subsets. For deep learning approaches, all hyper-parameters were manually adjusted to yield the best achievable performances.

- Subjective comparisons. Fig. 7 presents the respective vessel segmentation results produced by the proposed method, backbone, and the other three selected state-of-the-art segmentation networks. We can see that U-Net achieves relatively low performance, due to its over-segmentation at regions with high density. CE-Net and CS-Net achieve better performance than U-Net. However, they are not able to preserve fine capillaries well in terms of producing weak vessel responses. In contrast, the proposed method yields more visually informative results. The benefit of the proposed method for segmentation can be observed from the representative regions (green patches). It is clear from visual inspection that our method has identified more complete and thinner vessels particularly in ROSE-1 subset (shown in purple). It achieves relatively uniform responses in both thick and thin vessels, and provides more sensitive and accurate segmentation on capillaries, as demonstrated in the segmentation results of the DRVP and MRVP layers.

In contrast, all the methods yield very similar segmentation performance in ROSE-2. Therefore, to better evaluate the performance of the proposed method, we provide quantitative results in the following subsections.

- Performance on the SRVP layer in ROSE-1. We will first evaluate the vessel segmentation performance on each of the plexus layers of the ROSE-1 subset. Table I quantifies the segmentation performance in SRVP images of the different approaches. Overall, our method achieves the best performance in terms of almost all the metrics, with the single exception that its FDR score is 0.0038 lower than that of the three-stage network [28]. Nevertheless, the proposed method is able to correctly identify the majority of vessels using our two-stage architecture.

- Performance on the DRVP layer in ROSE-1. For the DRVP images in the ROSE-1 subset, we first adopted the
TABLE V

ABLATION STUDIES OF OUR TWO-STAGE METHOD ON BOTH THE ROSE-1 AND ROSE-2 SUBSETS

| Methods                  | ROSE-1 (MRVP) | ROSE-2 |
|--------------------------|---------------|--------|
|                          | AUC | ACC | SEN | G-mean | Kappa | Dice | FDR | AUC | ACC | SEN | G-mean | Kappa | Dice | FDR |
| U-Net [24]               | 0.9039 | 0.8865 | 0.6944 | 0.8050 | 0.6308 | 0.7012 | 0.2892 | 0.8909 | 0.8313 | 0.7202 | 0.7894 | 0.5901 | 0.7211 | 0.2768 |
| ResU-Net [26]            | 0.9108 | 0.8970 | 0.7095 | 0.8188 | 0.6689 | 0.7309 | 0.2248 | 0.8942 | 0.8374 | 0.7228 | 0.7913 | 0.6042 | 0.7238 | 0.2743 |
| Backbone (ResNeSt)       | 0.9187 | 0.9064 | 0.7093 | 0.8228 | 0.6887 | 0.7461 | 0.2112 | 0.9038 | 0.8442 | 0.7415 | 0.8023 | 0.6120 | 0.7383 | 0.2639 |
| Backbone (joint learning) | 0.9363 | 0.9091 | 0.7303 | 0.8341 | 0.7003 | 0.7564 | 0.2135 | - | - | - | - | - | - | - |
| Ours                     | 0.9375 | 0.9099 | 0.7285 | 0.8338 | 0.7022 | 0.7576 | 0.2087 | 0.9046 | 0.8449 | 0.7421 | 0.8031 | 0.6137 | 0.7394 | 0.2625 |

Fig. 8. ROC curves of different methods on ROSE-1 (MRVP layer only) and ROSE-2 datasets.

ResNeSt backbone to obtain preliminary vessel segmentation results at the first stage. Afterwards, the initial segmentations are combined with their original images as inputs of the second stage for producing final vessel segmentations. Table II demonstrates the segmentation results achieved by our method and the state-of-the-art methods. Obviously, we can observe that the proposed network significantly outperforms the other methods by a large margin, with an increase of about 12.0% and 11.9% in kappa and Dice, respectively, and a reduction of about 13.2% in FDR when compared with CS-Net. These performance improvements are consistent with the segmentation results shown in the middle row of Fig. 7, where the proposed method successfully extract the small capillaries from macula regions with promising continuity and integrity, while the other methods produces relatively lower capillary responses.

- Performance on the MRVP layer in ROSE-1. Table III shows the results of using different segmentation approaches on the MRVP layers. Again, the proposed method achieves overall the best performance, with a single exception at the FDR score, where a performance score of 0.0988 is obtained using the method by Azzopardi et al. [17]. It is worth noting that the sensitivity scores obtained by the methods of Azzopardi et al. [17] and Zhao et al. [33] are only 0.5336 and 0.5829, which are significantly lower than those of all the deep learning-based methods. This shows that the conventional methods have yet to solve the problems as posed by the high degree of anatomical variations across the population, and the varying scales of vessels within an image. Moreover, motion artefacts, noise, poor contrast and low resolution in OCT-A exacerbate these problems. By contrast, deep learning-based methods extract a set of higher-level discriminative representations, which are derived from both local and global appearance features and thus can achieve better performance.

- Performance on ROSE-2. ROSE-2 only provides centerline-level manual annotation, and includes en face OCT-A images of the SRVP. Therefore, as with the DRVP images in ROSE-1 subset, for ROSE-2, our method adopts a ResNeSt backbone at the first stage to obtain the preliminary vessel segmentation results. Then, the final results is obtained at the second stage, using the original image and the preliminary results from the first stage as input. Table IV presents the performance of the different segmentation methods. It shows that our method achieves the best AUC, ACC, SEN, Kappa and Dice of 0.9046, 0.8449, 0.7421, 0.8031 and 0.7394, respectively, and also reaches almost the same FDR score by CS-Net.

Moreover, in order to illustrate the vessel segmentation performance of different methods in a more intuitive manner, we have provided ROC curves for both the ROSE-1 and ROSE-2 subsets, as shown in Fig. 8. Due to limited space, here we show only the segmentation performances on the MRVP layers in ROSE-1. Compared with the conventional methods such as the algorithms proposed by Zhao et al. [33] and Azzopardi et al. [17], deep learning based methods demonstrate their superiority in segmenting OCT-A images. This is because the introduction of excellent modules such as ResNet and attention blocks, are usually helpful in improving the AUC score of the encoder-decoder architecture. In addition, there are two reasons that our two-stage architecture achieves the best ROC curve (shown in red in both the subfigures of Fig. 8). Firstly, using ResNeSt as the backbone of the encoder-decoder architecture further strengthens performance for feature extraction, which improves the extraction of vessel information at different complexities. Secondly, the second stage can adjust local details on the basis of preliminary results from the first stage, which additionally refines the segmentation accuracy from the first stage.

VI. DISCUSSION AND CONCLUSION

A. Ablations Studies

In this paper, the proposed vessel segmentation method consists of a ResNeSt backbone, joint learning from pixel-level and centerline-level vessel segmentation, and two-stage training. In order to validate the effectiveness of these components, we carried out the following ablation studies. U-Net [24] is treated as the baseline encoder-decoder method. Then, we gradually evaluated how each of these components affect the results.

- Ablation for ResNeSt backbone. To discuss the performance of the ResNeSt backbone, we compare segmentation performance of the original U-Net, ResU-Net [26] (the modified U-Net with residual blocks in the encoder) and our proposed encoder-decoder architecture (with ResNeSt as the backbone), as shown in Table V. For both the ROSE-1 and ROSE-2 subsets, our encoder-decoder architecture with ResNeSt as the backbone achieves the best performance on
AUC, ACC, Kappa, Dice and FDR in comparison with the original U-Net and ResU-Net. This indicates that the ResNeSt backbone is superior in feature extraction, which reveals more information about vessels with different characteristics.

- **Ablation for joint learning at pixel-level and centerline-level vessel segmentation.** In addition, we compared joint learning from pixel-level and centerline-level vessel segmentation with only one segmentation branch (with ResNeSt as the backbone) of all vessels for the MRVP images in the ROSE-1 subset, so as to demonstrate the advantages of joint learning from both pixel-level and centerline-level vessel segmentation. Comparisons of both performance are illustrated in Table V. We can observe that joint learning achieves higher scores in terms of AUC, ACC, SEN, Kappa and Dice than learning from only one segmentation branch. This suggests that joint learning could help to improve both pixel-level and centerline-level vessel segmentation performance by highlighting the relevant topological distinctions between pixel-level and centerline-level vessels.

- **Ablation for two-stage training.** Furthermore, we analysed the impact of the second stage on the first stage in our two-stage procedure. At the first stage, for the ROSE-1 subset, pixel-level and/or centerline-level vessel segmentation results are treated as the preliminary vessel segmentation results, while for the ROSE-2 subset, vessel segmentation results produced by the ResNeSt backbone are treated as the preliminary vessel segmentation results. At the second stage, final vessel segmentation results are derived from both the original images and preliminary results from the first stage. Accordingly, we made a comparison between the preliminary vessel segmentation results of the first stage and final vessel segmentation results of the second stage. As illustrated in Table V, final vessel segmentation performance at the second stage for the most parts shows improvement when compared with results from the first stage. The last column of Fig. 7 also indicates that some details of microvasculature are optimized in the second stage.

### B. Clinical Evaluation

It has been suggested that the retina may serve as a window for monitoring and assessing cerebral microcirculation [35] and neurodegeneration conditions [9]. OCT imagery has been utilized to observe neurodegenerative changes occurring in the ganglion cell-inner plexiform layer (GC-IPL) thickness and retinal nerve fiber layer (RNFL) thickness of AD and MCI patients [36]. Recently, contributions of vascular biomarkers such as length, density and tortuosity, to the diagnosis of MCI and AD are increasingly recognized [36], [37]. OCT-A, as an extension of OCT, that can provide in vivo, noninvasive visualization of the retinal vessels in different layers. With the simultaneous occurrence of both neurodegeneration and microvascular changes in the brain, many studies [35], [38], [39] have suggested that the macula microvasculature may provide vital information on the changes in the cerebral microcirculation during the subclinical phase. Changes in the retinal capillary network may indicate the onset and progression of retinopathy, and fractal dimension (FD) analysis may offer deeper insights into retinal vascular disease than other geometric measures.

In this work, we performed an FD analysis by applying a box-counting method named Fraclab [40] on the segmentation results of the SRVP, DRVP, and MRVP images using the proposed method. The box-plots in Fig. 9 show the statistical analysis results on the ROSE-1 dataset, including 39 images of normal and 78 images of AD patients. It can be observed that the AD group has reduced FD in the SRVP, MRVP and DRVP when compared with the control group. Student’s t-test was employed to assess the differences between the AD and control groups and results confirmed that the differences between the AD and control participants are significant in the SRVP (p = 0.004 < 0.05), MRVP (p = 0.007 < 0.05) and DRVP (p = 0.028 < 0.05), respectively. These results are consistent with the previous findings that retinal microvascular changes may reflect neurodegenerative changes [9], [36].

![Fig. 9. Descriptive results of the FD of AD patients and controls in the SRVP, DRVP and MRVP, respectively.](image)

### C. Conclusions

In this paper, we have presented a novel Retinal OCT-A SEgmentation dataset (ROSE) dataset, which is a large, carefully designed and systematically manually-annotated dataset for vessel segmentation in OCT-A images. To our best knowledge, this is the first OCT-A dataset released to the research community for the vessel segmentation task. It contains two sub-sets, where the images were acquired by two different devices. All the vessels were manually annotated by human experts at either centerline level and/or pixel level. All of the images contained in the dataset were eventually used for clinical diagnostic purposes. To ensure the utmost protection of patient privacy, the identities of all patients have been removed and cannot be reconstructed. We plan to keep growing the dataset with more challenging situations and various types of eye and neurodegenerative diseases, such as diabetic retinopathy and Parkinson’s disease.

In addition to the new dataset, we further proposed a novel two-stage framework for vessel segmentation in OCT-A images. In the first stage, a split-based coarse segmentation (SCS) module has been designed to achieve the preliminary vessel segmentation results: a ResNeSt block is used as the backbone of the encoder-decoder framework. In the second stage, a split detail refinement network (SRN) module has been adopted to improve the vessel segmentation results by utilizing both the original images and the preliminary results from
the first stage. The experimental results on the ROSE dataset show that our vessel segmentation approach outperforms other state-of-the-art methods, and the sub-analysis on AD shows the great potential of exploring retinal microvascular-based analysis for the diagnosis of various neurodegenerative diseases. Finally, our ROSE dataset and code of the proposed segmentation network are publicly available at https://timed.nlmte.ac.cn/dataofrose.html

REFERENCES

[1] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, et al., "Cs-net: Channel and spatial attention network for curvilinear structure segmentation," in MICCAI, pp. 721–730, 2019.

[2] Y. Zhao, Y. Zheng, Y. Liu, Y. Zhao, L. Luo, S. Yang, et al., "Automatic 2D/3D vessel enhancement in multiple modality images using a weighted symmetry filter," IEEE Trans. Med. Imaging, vol. 37, no. 2, pp. 438–450, 2018.

[3] G. T. Witmer, G. Parlitis, S. Patel, and S. Kiss, "Comparison of ultra-widefield fluorescence angiography with the heidelberg spectralis noncontact ultra-widefield module versus the ots optomap," Clinical Ophthalmology, vol. 7, pp. 389–394, 2013.

[4] R. A. Leitgeb, "En face optical coherence tomography: a technology review," Biomed. Opt. Express, vol. 10, no. 5, pp. 2177–2201, 2019.

[5] Y. Jia, O. Tan, J. Tokayer, B. Potsaid, Y. Wang, J. J. Liu, et al., "Split-spectrum amplitude-decorrelation angiography with optical coherence tomography," Opt. Express, vol. 20, no. 4, pp. 4710–4725, 2012.

[6] Y. Zhao, Y. Zheng, Y. Liu, J. Yang, Y. Zhao, D. Chen, et al., "Intensity and compactness enabled saliency estimation for leakage detection in diabetic and malarial retinopathy," IEEE Trans. Med. Imaging, vol. 36, no. 1, pp. 51–63, 2016.

[7] J. Zhang, A. H. Kashani, and Y. Shi, "3D surface-based geometric and topological quantification of retinal microvasculature in oct-angiography via reeb analysis," in MICCAI, pp. 57–65, Springer, 2019.

[8] J. Zhang, Y. Qiao, M. Sarabi, M. Khansari, J. Gahm, A. Kashani, et al., "3D shape modeling and analysis of retinal microvasculature in oct-angiography images," IEEE Trans. Med. Imaging, vol. 39, no. 5, pp. 1335–1346, 2020.

[9] Y. Jin, D. S. Grewal, A. C. Thompson, B. W. Polascik, C. Dunn, J. R. Burke, et al., "Retinal microvascular and neurodegenerative changes in alzheimer’s disease and mild cognitive impairment compared with control participants.," Ophthalmology Retina, vol. 3, no. 6, pp. 489–499, 2019.

[10] 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 MICCAI, pp. 132–139, Springer, 2016.

[11] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, et al., "CE-Net: Context Encoder Network for 2D Medical Image Segmentation," IEEE Trans. Med. Imaging, vol. 38, no. 10, pp. 2281–2292, 2019.

[12] 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. Imaging, vol. 39, no. 5, pp. 1392–1403, 2020.

[13] N. Eladawii, M. Elmogy, O. Helmy, A. Aboelfetouh, A. M. Riad, H. Sandhu, et al., "Automatic blood vessels segmentation based on different retinal maps from oct scans," Comput. Biol. Med., vol. 89, pp. 150–161, 2017.

[14] M. Li, Y. Chen, Z. Ji, K. Xie, S. Yuan, Q. Chen, et al., "Image projection network: 3d to 2d image segmentation in oct images," IEEE Trans. Med. Imaging, vol. PP, pp. 1–1, 05 2020.

[15] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, "Et-net: Ageneric edge-attention guidance network for medical image segmentation," in MICCAI, pp. 442–450, 2019.

[16] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. T. Owen, et al., "Blood vessel segmentation methodologies in retinal images - a survey," Comput. Meth. Programs Biomed., vol. 108, no. 1, pp. 407–433, 2012.

[17] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable costfire filters for vessel delineation with application to retinal images," Med. Image Anal., vol. 19, no. 1, pp. 46–57, 2015.

[18] J. Zhang, B. Dashbozorg, E. J. Bekkers, J. P. W. Pluim, R. Duits, and B. M. T. H. Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," IEEE Trans. Med. Imaging, vol. 35, no. 12, pp. 2631–2644, 2016.