

RAVIR: A Dataset and Methodology for the Semantic Segmentation and Quantitative Analysis of Retinal Arteries and Veins in Infrared Reflectance Imaging

Ali Hatamizadeh, Hamid Hosseini, Niraj Patel, Jinseo Choi, Cameron C. Pole, Cory M. Hoeferlin, Steven D. Schwartz and Demetri Terzopoulos

Abstract—The retinal vasculature provides important clues in the diagnosis and monitoring of systemic diseases including hypertension and diabetes. The microvascular system is of primary involvement in such conditions, and the retina is the only anatomical site where the microvasculature can be directly observed. The objective assessment of retinal vessels has long been considered a surrogate biomarker for systemic vascular diseases, and with recent advancements in retinal imaging and computer vision technologies, this topic has become the subject of renewed attention. In this paper, we present a novel dataset, dubbed RAVIR, for the semantic segmentation of Retinal Arteries and Veins in Infrared Reflectance (IR) imaging. It enables the creation of deep learning-based models that distinguish extracted vessel type without extensive post-processing. We propose a novel deep learning-based methodology, denoted as SegRAVIR, for the semantic segmentation of retinal arteries and veins and the quantitative measurement of the widths of segmented vessels. Our extensive experiments validate the effectiveness of SegRAVIR and demonstrate its superior performance in comparison to state-of-the-art models. Additionally, we propose a knowledge distillation framework for the domain adaptation of RAVIR pretrained networks on color images. We demonstrate that our pretraining procedure yields new state-of-the-art benchmarks on the DRIVE, STARE, and CHASE_DB1 datasets.

Dataset link: <https://ravirdataset.github.io/data>

Index Terms—Retinal Image Analysis, Deep Learning, Semantic Segmentation, Vascular Width Estimation, Ophthalmology.

I. INTRODUCTION

The retina and its vasculature are directly visible due to the optically clear media of the human eye; thus, the ability to image the retina rapidly and non-invasively using a variety of modalities provides a window into vital organs such as the brain, heart, and kidneys. Analysis of the retinal vasculature offers a unique avenue to monitor and assess

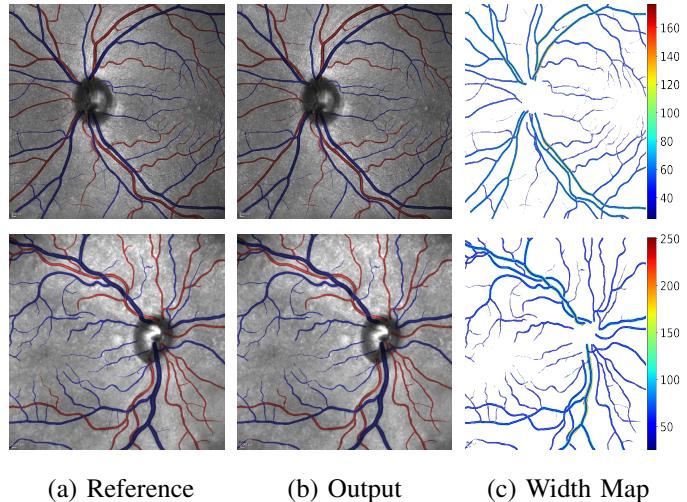


Fig. 1. The RAVIR dataset enables semantic understanding of retinal vessels in IR imaging. (a) Reference labels. (b) SegRAVIR semantic segmentation output. (c) Pixel-wise vessel width measurement. Red and blue masks denote artery and vein classes, respectively. Width measurement map presented in microns.

changes associated with systemic diseases such as hypertension, diabetes, and neurodegenerative disorders. The sequelae of these conditions, including stroke, coronary heart disease, and dementia, represent major causes of morbidity and mortality in the developed world [1]–[3]. To date, most classification schemes for retinal vascular changes due to these conditions, particularly in the early stages of disease, depend on the human assessment of qualitative morphological changes [4], [5].

We and others hypothesize that the assessment of retinal vasculature changes can elicit biomarkers associated with early, even asymptomatic, disease states [6]. For example, hypertension is known to cause structural changes in the macro and micro vasculature of vital organs throughout the body, including the brain, heart, and kidneys. As retinal vasculature is both functionally and pathologically linked to cerebral vasculature, it exhibits similar morphological changes [7]. However, early structural changes to the vasculature are inconspicuous, and detection and analysis of these early changes is often difficult due to the non-quantitative modalities used in routine clinical practice, including direct examination, fundus photography, and angiography.

Ali Hatamizadeh and Hamid Hosseini are co-primary authors.

Corresponding author: Ali Hatamizadeh (ahatamiz@cs.ucla.edu).

Ali Hatamizadeh is with the Computer Science Department, Samueli School of Engineering, University of California, Los Angeles, CA, USA.

Hamid Hosseini, Niraj Patel, Jinseo Choi, Cameron C. Pole, Cory M. Hoeferlin and Steven D. Schwartz are with the Stein Eye Institute, Department of Ophthalmology, David Geffen School of Medicine, University of California, Los Angeles, CA, USA.

Demetri Terzopoulos is with the Computer Science Department, Samueli School of Engineering, University of California, Los Angeles, CA, USA, and is a co-founder of VoxelCloud, Inc., Los Angeles, CA, USA.

This work was supported in part by an unrestricted grant from Research to Prevent Blindness (RPB) to the UCLA Department of Ophthalmology.

The accessibility of the retinal vasculature to multimodal imaging provides the opportunity to quantitatively assess prognosis, risk, and response to treatment. Infrared Reflectance (IR) imaging is regarded as an ideal modality to evaluate the retinal vasculature due to several advantages including the ability to capture images through a non-dilated pupil, increased patient comfort due to lack of visible light flash, and better penetration through opaque media. Additionally, previous studies have found a substantial correlation in image resolution and utility between IR and color imaging modalities [8].

While systemic etiologies affect both arterial and venous vasculature, the morphological changes between the two may differ significantly. In the early stages of hypertensive retinopathy, for example, increased blood pressure triggers autoregulatory mechanisms that result in diffuse narrowing of the arterial vessels. If the elevated pressure persists, a compensatory thickening of the arterial walls occurs, which can be clinically observed by the resultant widening and accentuation of the vascular light reflex (“copper wiring” or “silver wiring”). By contrast, physiologic changes in venous vasculature include dilated and tortuous venules resulting from venous compression at arteriovenous crossing points subsequent to thickening of the adjacent arterial walls. Additionally, these physiological changes progress in a well-understood and sequential manner that is correlated to the duration and severity of disease. For this reason, clinically observable changes in retinal vasculature are among the primary metrics used in the diagnosis, progression, and response to treatment.

The development of an automated, reliable tool that reproducibly quantifies changes in the retinal vasculature in response to disease and intervention promises to augment or even disrupt current evaluation and treatment paradigms. By enabling physicians to detect disease, predict outcomes, and assess interventions at an earlier stage, such a tool can serve in the management of systemic conditions such as hypertension and diabetes with the potential to dramatically improve patient outcomes and public health at large.

A critical step in tracking such important structural changes of the retinal vasculature is segmentation of the retinal vessels in ophthalmologic images, as it enables localization of the veins and arteries and the extraction of relevant information such as the change in diameters of the vessels. The automated segmentation and serial analysis of both arterial and venous retinal vasculature is therefore a potentially valuable capability. Since the advent of “deep learning” in the field of computer vision, Convolutional Neural Networks (CNNs) have been widely employed for the automated segmentation of medical images, including retinal images. Using stochastic gradient descent optimization approaches, CNNs can be trained end-to-end to automatically learn and extract important features. However, training requires annotated datasets with detailed, pixel-wise ground truth labels, and such annotation processes are often notoriously time-consuming due to the complex structure of retinal vessels.¹ Consequently, publicly available annotated datasets for retinal vessel segmentation tend to be

¹Self-supervised and unsupervised deep learning approaches do not require annotations, but supervised approaches usually perform significantly better.

relatively small in size, as is often the case in medical imaging, a domain that is infamously deficient in adequate datasets.

In this paper, we introduce a dataset for training deep learning CNN models to perform detailed, pixel-resolution “semantic segmentation” of Retinal Arteries and Veins in IR imaging, dubbed RAVIR. In conjunction with our new RAVIR dataset, we propose a novel deep learning-based methodology, denoted SegRAVIR, which is tailored to the semantic segmentation of retinal arteries and veins. Specifically, SegRAVIR features a two-stream encoder-decoder CNN architecture consisting of main and auxiliary streams with shared encoders to simultaneously segment the vessels and reconstruct the input images. The role of the auxiliary stream is mainly to regularize the main stream and help the shared encoder learn effective representations of the retinal vessels.

Since the RAVIR dataset provides precise pixel-scale ground-truth information, we can leverage the output segmentation masks to measure the diameters of retinal arteries and veins. This can enable monitoring the morphological changes of retinal vasculature in a quantitative manner. Our extensive experiments validate the effectiveness of SegRAVIR, as it proves to outperform existing methodologies by a considerable margin. Furthermore, when trained from scratch on another color retinal artery and vein segmentation dataset called RITE, SegRAVIR again achieves state-of-the-art performance.

Additionally, we propose a novel knowledge distillation framework for the domain adaptation from IR to color images of models pretrained on the RAVIR dataset. We demonstrate the effectiveness of our domain adaptation approach by fine-tuning and testing SegRAVIR on three popular color retinal vessel segmentation datasets, DRIVE, STARE, and CHASE_DB1.

To summarize, our specific contributions are as follows:

- 1) We introduce RAVIR, a novel dataset for semantic segmentation and quantitative analysis of retinal arteries and veins in the IR imaging modality.
- 2) We propose a novel framework with tailored loss functions for the segmentation task. To this end, we propose the proxy task of reconstructing input images by using an autoencoder to regularize the segmentation output and improve its fine-grained details.
- 3) We provide extensive quantitative benchmarks for performance comparison of our RAVIR framework against state-of-the-art methodologies for retinal vessel segmentation.
- 4) We propose a novel knowledge distillation framework for the domain adaptation to color images of models pretrained on RAVIR, and our results establish new state-of-the-art benchmarks on the publicly available DRIVE, STARE, and CHASE_DB1 datasets by a substantial margin.

II. RELATED WORK

In this section, we review and discuss the publicly available datasets for training models to perform retinal artery and vein segmentation in different imaging modalities (Table I), as well as the notable deep-learning-based methodologies underlying such segmentation models.

TABLE I
COMPARISON BETWEEN THE RAVIR DATASET AND OTHER PUBLICLY AVAILABLE

| Dataset | retinal vessel segmentation datasets | | | | | | |
|--------------------|--------------------------------------|-----|------------|-----------|----------|--------------------------|----------|
| | Resolution | FOV | Train Size | Test Size | AV Label | Camera | Modality |
| RAVIR | 768 × 768 | 30 | 26 | 20 | Yes | Heidelberg Spectralis | IR |
| STARE [9] | 605 × 700 | 35 | 20 | - | No | Topcon TRV-50 | Color |
| DRIVE [10] | 565 × 584 | 45 | 20 | 20 | Yes | Canon CR-5 non-mydriatic | Color |
| RITE [11] | 565 × 584 | 45 | 20 | 20 | Yes | Canon CR-5 non-mydriatic | Color |
| CHASE_DB1 [12] | 999 × 960 | 30 | 20 | 8 | No | Nidek NM-200-D | Color |
| IOSTAR [13] | 1024 × 1024 | 45 | 30 | - | Yes | EasyScan | Color |
| LES-AV [14] | 1620 × 1444 | 30 | 22 | - | Yes | Canon CR-1 | Color |
| HRF [15] | 3504 × 2336 | 45 | 45 | - | Yes | Canon CR-1 | Color |
| ROSE [16] | 304 × 304 | - | 90 | 27 | No | RTVue-XR | OCTA |
| RECOVERY-FA19 [17] | 3900 × 3072 | 200 | 8 | - | No | Optos 200Tx | FA |
| PRIME-FP20 [18] | 4000 × 4000 | 200 | 15 | - | No | Optos 200Tx | FA |

A. Retinal Artery and Vein Segmentation Datasets

Traditional color photographs have long been utilized in the screening evaluation of retinal vascular diseases.

One such early effort is the STructured Analysis of the REtina (STARE) dataset [9], which was created at the University of California, San Diego, with the aim of developing a system capable of the automated diagnosis of various diseases in the human eye. This dataset comprises 20 color retinal images, acquired by a Topcon TRV-50 camera at a 35° field of view (FOV), and sized at 605 × 700 pixels with 8 bits per RGB color channel. Half of the images in this dataset depict pathology that manifests an abnormal appearance of retinal blood vessels and obscures them.

Another important effort is the Digital Retinal Images for Vessel Extraction (DRIVE) dataset [10], which was created as part of a diabetic retinopathy screening program in The Netherlands. It consists of 40 color retinal images with corresponding binary segmentation masks, which are equally divided into train and test sets. In the DRIVE dataset, 7 cases show signs of mild early diabetic retinopathy or conditions such as pigment epithelium changes, pigmentary epithelial atrophy, butterfly maculopathy with pigmented scar in fovea, choridiopathy, atrophy around the optic disc, background diabetic retinopathy, and other vascular abnormalities. The images were acquired using a Canon CR5 non-mydriatic 3CCD camera with a 45° FOV and are sized at 768 × 584 pixels with 8 bits per RGB color channel.

Using the original images in the DRIVE dataset, the Retinal Images vessel Tree Extraction (RITE) dataset [11] was developed to support the segmentation and classification of retinal arteries and veins. Qureshi *et al.* [19] also presented a manually labeled dataset with pixel-wise annotations for retinal arteries and veins for images in the DRIVE dataset.

The Child Heart and Health Study (CHASE) in England yielded the CHASE_DB1 dataset [12], including 28 color retinal images from both eyes of 14 child subjects. The images were captured by a hand-held Nidek NM-200-D fundus camera with a resolution of 1280 × 960 pixels at a 30° FOV. The vessels in these images are poorly contrasted against the background. Additionally, the artery cross-sectional intensity depicts a central vessel reflex. Eight images are used as the

training set and the remainder as a test set.

The IOSTAR [13] and LES-AV [14] datasets contain color retinal images that include artery and vein pixel-wise annotations. The IOSTAR dataset consists of 30 images of size 1024 × 1024 pixels acquired through a Scanning Laser Ophthalmoscopy (SLO) methodology from patients suffering from diabetic retinopathy, whereas the LES-AV dataset has 22 images of glaucomatous patients.

Another commonly used dataset is the High Resolution Fundus (HRF) dataset [15], which contains 45 color retinal images of size 3504 × 2336 pixels captured by a Canon CR-1 fundus camera with a 45° FOV. The dataset includes images of 15 subjects with diabetic retinopathy, 15 subjects with glaucoma, and 15 healthy subjects.

Aside from IR imaging and RAVIR, and in addition to the aforementioned color image datasets, retinal vessel segmentation datasets have recently been curated in other imaging modalities. The ROSE dataset [16] provides 117 Optical Coherence Tomography Angiography (OCTA) images of 304 × 304 pixel resolution from 39 subjects suffering from Alzheimer's disease and, with regard to the Fluorescein Angiography (FA) modality, the RECOVERY-FA19 dataset [17] contains 8 ultra-wide-field images of size 3900 × 3072 pixels from patients participating in the Intravitreal Aflibercept for Retinal Non-Perfusion in Proliferative Diabetic Retinopathy trial [20]. Another FA dataset is the PRIME-FP20 dataset [18], comprising 15 images with a resolution of 4000 × 4000 pixels acquired from patients with diabetic retinopathy.

B. Retinal Vessel Segmentation Methodologies

Subsequent to the success in medical and biomedical image segmentation of the seminal U-Net [21] architecture, many efforts have been proposed to leverage its encoder-decoder paradigm in the context of retinal vessel segmentation. Zhuang *et al.* [22] proposed an architecture based on U-Net that utilizes multiple-path networks to leverage a path-wise formulation in segmenting retinal vessels. Conversely, Wang *et al.* [23] proposed a Dense U-Net architecture with a patch-based training strategy and extensive data augmentation, which outperforms the baseline U-Net by a small margin. Jiang *et al.* [24] developed a multi-path encoder-decoder architecture

that fused the outputs of the network at different scales to obtain the final predictions.

Furthermore, Alom *et al.* [25] introduced a recurrent residual U-Net architecture in which residual convolutional layers are used for optimal feature representation learning. Jin *et al.* [26] proposed a deformable U-Net architecture that captures various scales and shapes of the vessels by adjusting the adaptive receptive field of the deformable convolutional layers. Wu *et al.* [27] used a U-like encoder-decoder architecture that leverages Inception residual blocks for improved feature representation and is supervised in a multi-scale manner to preserve features of different sizes. Zhang *et al.* [28] proposed an attention-guided network that effectively captures vessel structural information. Wang *et al.* [29] utilized two encoders, in a U-Net like architecture, to preserve the spatial information and semantic representations separately. Gu *et al.* [30] proposed a context encoder network, CE-Net, to effectively capture high-level semantic and spatial content for use in segmenting vessels. Additionally, Zhang *et al.* [31] proposed an edge-attention guidance network to improve the accuracies of the boundaries of segmented vessels.

Recently gaining traction is the application of cascaded architectures where several encoder-decoders are chained together to refine the segmentation outputs. Although these approaches typically result in more refined boundaries, they are memory-intensive and the inference time is substantially increased due to the complexity of the model. Yan *et al.* [32] introduced a methodology in which a three-step CNN was employed to separately segment thick and thin vessels and learn how to effectively fuse the results and refine the predictions. Francia *et al.* [33] proposed to use an architecture consisting of two residual U-Nets in which the second refines the feature maps of the first and resolves the ambiguities. Li *et al.* [34] used a cascade of U-Net architectures that share their weights and are connected by mid-level skip connections, thus yielding a deeper CNN that can effectively disambiguate the uncertain regions in the output segmented vessels. Guo *et al.* [35] introduced a bidirectional symmetric cascade with adaptive supervision of each layer according to the diameter scale of each vessel.

In addition to graph-based methodologies [36]–[38], several efforts have applied deep learning-based methods to the semantic segmentation of arteries and veins as distinct vessels in color retinal images. Welikala *et al.* [39] used an architecture consisting of three convolutional layers and three fully-connected layers to segment vessels in small 25×25 image patches. Hemelings *et al.* [40] proposed another patch-based approach to vessel segmentation by leveraging a regular U-Net with categorical cross-entropy as the loss function. Girard *et al.* [41] introduced a whole-image approach with a U-Net-like encoder-decoder in which the output predictions of the CNN are propagated through a graph representation of the retinal vasculature based on a likelihood score for detecting arteries, but the model requires training on small patches that are centered on retinal vessels, which is a limitation relative to the method by Hemelings *et al.* [40]; however, both methods require ad-hoc image pre-processing techniques to achieve optimal performance. Xu *et al.* [42] proposed a related whole-image approach requiring extensive data augmentation, and

the final segmentation maps are obtained via a heuristic for assigning pixel-wise outputs. Kang *et al.* [43] introduced an architecture for segmenting retinal arteries and veins consisting of a GoogLeNet [44] encoder, a decoder, and a weighted-attention fusion module that predicts the final segmentation outputs by fusing and refining segmentation probabilities of each semantic class; however, the absence of high-resolution skip connections between the encoder and decoder limits the ability to exploit multi-scale representations. Ma *et al.* [45] introduced a model consisting of a pretrained U-Net incorporating a ResNet [46] encoder with two parallel branches for extracting common retinal artery and vein representations, focusing on discriminative features whose segmentation output is generated by concatenating the outputs of each branch and using a customized activation block that emphasizes the importance of capillary vessels. Chen *et al.* [47] proposed a Generative Adversarial Network (GAN) approach to improving the connectivity of segmented retinal arteries and veins. Morano *et al.* [48] described a U-like architecture to independently segment retinal arteries, veins, and vessels using Binary Cross Entropy (BCE) as the loss function. Despite these advances, a primary shortcoming of the aforementioned methodologies is frequent classification errors in the detected vessels.

III. THE RAVIR DATASET

Recently there has been a substantial increase in the utilization of single wavelength confocal imaging that only captures reflected light passing through a pinhole, which will allow capturing high quality, high contrast fundus images. The images in our RAVIR dataset were captured using infrared (815nm) Scanning Laser Ophthalmoscopy (SLO), which in addition to having higher quality and contrast, is more convenient for the patient and is less affected by opacities in optical media and pupil size. This imaging modality is also commonly used in ophthalmology clinics. The RAVIR dataset consists of 46 IR retinal images from the UCLA Stein Eye Institute imaging database, divided into train and test sets of 26 and 20 images, respectively. We used 4 images from the 26 test images as our validation set. The images were captured using a Heidelberg Spectralis camera with a 30° FOV. They are sized at 768×768 pixels and compressed in the Portable Network Graphics (PNG) format. Each pixel in the images has a reference length of 12.5 microns. Table I compares our RAVIR dataset against other retinal vessel segmentation datasets.

Our study was carried out with the approval of the Institutional Review Board at UCLA and adhered to the tenets of the Declaration of Helsinki. Table II reports the demographics of the study population. The average age of the subjects was 55 years, with a range from 19 to 88. There were 29 male and 17 female subjects. One eye from each patient was imaged for the dataset, and there were 24 right eyes and 22 left eyes. Twenty-three images were normal, while 23 images depicted retinal pathology. The train and test sets of the RAVIR dataset have an equal number of healthy and pathological cases.

As described in Table III, a wide variety of retinal pathologies are covered by the dataset. These conditions may all affect the retinal vascular pattern and infrared retinal background.

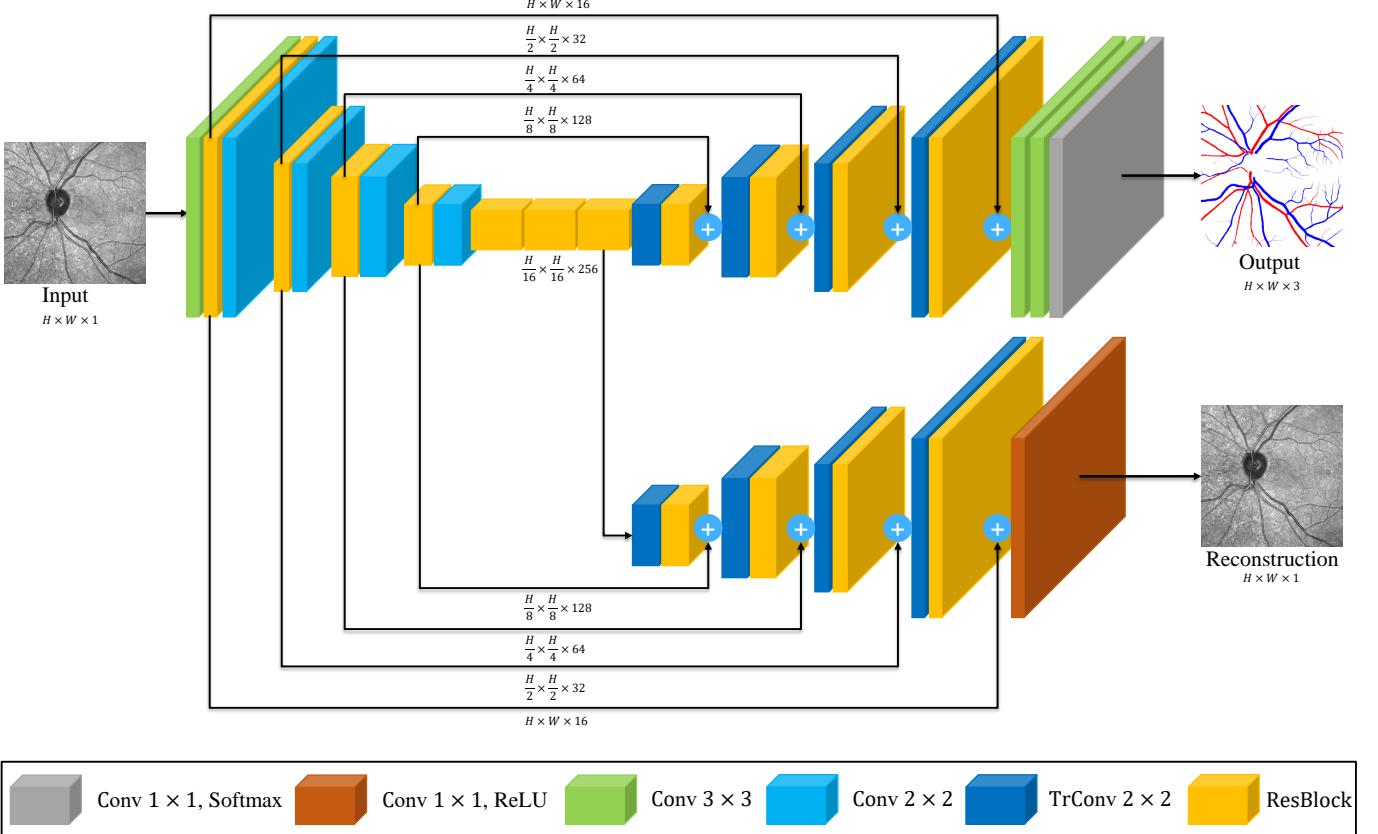


Fig. 2. SegRAVIR consists of a main stream for segmenting the arteries and veins as well as an autoencoder that learns to reconstruct the input and acts as a regularizer. The architectural building blocks are depicted at bottom —Conv, TrConv, and ResBlock denote convolution, transposed convolution, and residual blocks, respectively.

TABLE II
RAVIR DATASET DEMOGRAPHICS

| Statistics | Quantity |
|----------------------|----------|
| Average Age (years) | 55 ± 18 |
| Age Range (years) | 19 – 88 |
| Number of Males | 29 |
| Number of Females | 17 |
| Number of Right Eyes | 24 |
| Number of Left Eyes | 22 |

TABLE III
IR IMAGING MANIFESTATIONS OF STUDY POPULATION IN RAVIR

| Clinical Finding | #Eyes | Imaging Manifestation |
|----------------------------|-------|---|
| Retinal vein occlusion | 10 | Venous tortuosity, retinal hemorrhage |
| Hypertensive retinopathy | 10 | Vessel narrowing, arteriovenous nicking |
| Peripapillary Atrophy | 8 | Increased background reflectivity |
| Diabetic Retinopathy | 3 | Intraretinal hemorrhage, vessel sclerosis |
| Isolated vessel tortuosity | 3 | Abnormal vessel tortuosity |
| High Myopia | 2 | Transmission of choroidal vessels |
| Media opacities | 1 | Retinal haziness |

For example, diabetic retinopathy and vein occlusions may cause vascular caliber changes and signal blocking from retinal hemorrhage or edema. Peripapillary atrophy causes increased infrared reflectance around the Optic Nerve Head (ONH). Hypertensive retinopathy can manifest as vessel narrowing

and arteriovenous nicking, which is the segmental narrowing of veins at arterial crossing points. Severe myopia can cause increased signal transmission of choroidal vessels.

Manual pixel-wise annotations were performed and verified by our experienced retinal image analysis specialist. Per-pixel labels were applied to all vessel regions that could accurately be identified as artery or vein. Vessels were labeled over the ONH in those images where the arteries and veins could be resolved. However, for images in which the veins and arteries over the ONH were indistinguishable, the ONH region was blocked for masking.

IV. SEGRAVIR METHODOLOGY

This section develops our SegRAVIR methodology, including the network architecture and loss function, as well as our vessel width estimation technique and domain adaptation strategy.

A. The SegRAVIR Architecture

To analyze retinal arteries and veins, we devise an encoder-decoder architecture dubbed SegRAVIR (Fig. 2), which consists of a main stream for segmenting and an auxiliary stream for reconstructing the input images. The main stream leverages residual blocks and skip connections that connect the encoder and decoder. The auxiliary stream benefits from a decoder that learns how to reconstruct the input images and acts as a

regularizer. SegRAVIR’s building blocks are residual blocks consisting of two 3×3 convolutional layers and an identity skip connection that adds the input of the block to the convolutional outputs. Each convolutional layer with kernel W is followed by a Rectified Linear Unit (ReLU) and a batch normalization $BN_{\gamma,\beta}$ with learnable parameters γ and β :

$$C(i, j) = BN_{\gamma,\beta} \left(\text{Re} \left(\sum_{t=-1}^{+1} \sum_{l=-1}^{+1} K[t, l] W[i-t, j-l] \right) \right), \quad (1)$$

where C and K denote the input and output of the convolution layer, respectively, and i, j, t, l are pixel indices. Additionally, the output of each convolution is fed into a dropout layer for regularization. SegRAVIR’s encoder spans across 4 different resolutions, each of which consists of a residual block followed by a convolutional layer with a stride of 2 and appropriate padding to reduce the resolution of learned representations by a factor of 2. At its lowest resolution, the output of the convolution layer is fed into 3 consecutive residual blocks and then into the decoder.

SegRAVIR’s decoder has 4 corresponding resolutions consisting of transposed convolutional layers with a stride of 2 and appropriate padding to increase the resolution of learned representations by a factor of 2, followed by a residual block. At each resolution, the output of the residual blocks in the encoder is added to the output of the corresponding residual block in the decoder via a skip connection. Finally, the output of the decoder is fed into two consecutive 3×3 convolutional layers with ReLU activation and batch normalization, followed by a 1×1 convolutional layer with softmax activation and three output channels that correspond to background, vein, and artery classes.

Furthermore, we employ an auxiliary stream that learns to reconstruct the input images and acts as a regularizer. It is an auto-encoder that regularizes the shared encoder by learning an additional proxy task of image reconstruction in order to avoid overfitting. At each resolution, the outputs of residual blocks in the main-stream encoder are fed into a decoder that shares the same architecture as the main-stream decoder, except for the last three layers which use a 1×1 convolutional layer followed by a ReLU activation function. The output of the auxiliary stream has the same size as the input image; hence, a proper image-based reconstruction loss can be employed at the original resolution.

Unlike a typical U-Net, the SegRAVIR model leverages residual blocks in both the encoder and decoder of the main stream and benefits from three residual blocks in the bottleneck of the architecture, thus enabling the capture of fine-grained details of the retinal vasculature. Moreover, its auxiliary stream regularizes the segmentation outputs, further improving performance. Fig. 2 provides additional SegRAVIR architectural details.

B. SegRAVIR Loss Function

To supervise the outputs of the main stream as well as the autoencoder of the auxiliary stream, we employ a hybrid loss

function consisting of the following terms:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{Dice}} + \lambda_2 \mathcal{L}_{\text{CE}} + \lambda_3 \mathcal{L}_{\text{L2}}, \quad (2)$$

where λ_i are weighting parameters associated with the loss terms. Letting the number of pixels and number of semantic classes be V and K , respectively, the Dice loss function is defined as

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2}{K} \sum_{k=1}^K \frac{\sum_{v=1}^V G_{v,k} P_{v,k}}{\sum_{v=1}^V G_{v,k} + \sum_{v=1}^V P_{v,k}}, \quad (3)$$

where $P_{v,k}$ and $G_{v,k}$ denote the output probability prediction and corresponding one-hot encoded ground truth for class k at pixel v . The Cross Entropy (CE) loss is defined as

$$\mathcal{L}_{\text{CE}} = -\frac{1}{V} \sum_{v=1}^V \sum_{k=1}^K G_{v,k} \log P_{v,k}, \quad (4)$$

The L2 reconstruction loss of the input image is defined as

$$\mathcal{L}_{\text{L2}} = \|X_i - X_r\|_2^2. \quad (5)$$

where X_i and X_r denote the input and reconstructed images, respectively. While \mathcal{L}_{L2} supervises the output of the autoencoder, a combination of $\mathcal{L}_{\text{Dice}}$ and \mathcal{L}_{CE} serves to supervise the output segmentation of the main stream. The hybrid loss function enables the learning of different semantics as the \mathcal{L}_{CE} and $\mathcal{L}_{\text{Dice}}$ loss terms capture pixel-wise and region-based information, respectively.

C. Vessel Width Estimation

From the segmentation predicted by the SegRAVIR network, our model measures the diameter of the arteries and veins. To this end, the segmentation probability map is first thresholded, using a constant value of 0.5, to obtain the medial curves of the vessels by iteratively identifying and removing border pixels while maintaining vessel connectivity, in an approach similar to the thinning algorithm presented by Zhang *et al.* [49]. Then, the distance transform of the medial curve mask is multiplied with the segmentation mask in a pixel-wise manner. The result is the diameter distance map with respect to the medial curves of the segmented vessels.

D. Knowledge Adaptation via Distillation

Knowledge distillation [50] [51], which was originally proposed as a model compression methodology, enables the training of a smaller student model under the supervision of a larger pretrained teacher network. We leverage knowledge distillation as a feature alignment mechanism for the domain adaptation of RAVIR-pretrained networks to color retinal images. Specifically, we first use transfer learning to create a teacher network in the new modality (color) from a pretrained SegRAVIR model. For this purpose, we use the main stream (encoder and segmentation decoder) of a SegRAVIR network and initialize the weights from a previous pretrained checkpoint. However, we replace the last layer of the network with a 1×1 convolutional layer whose weights are randomly initialized, followed by a softmax activation layer to account for the single output channel of the target segmentation task (i.e., binary

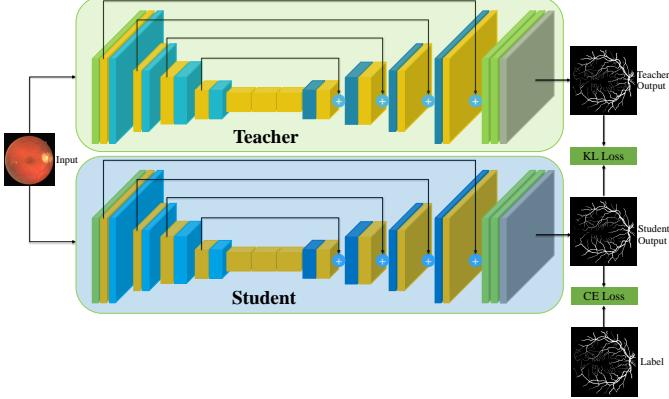


Fig. 3. Knowledge distillation strategy. A SegRAVIR network fine-tuned on color images supervises the output of a SegRAVIR student network by minimizing the KL divergence of prediction probability distributions.

vessel segmentation). We fine-tune this teacher network on the target image modality.

Furthermore, we create a student network with the same architecture and supervise its segmentation outputs according to the soft labels generated by the teacher network predictions. Fig. 3 overviews the proposed strategy.

Specifically, let $Z_T \in \mathbb{R}^{H \times W \times 2}$ denote the pixel-wise output logits of the teacher network. We employ temperature τ scaling to generate soft labels with smoother probability distributions, hence enabling the identification of inter-class relationships, according to

$$Y_T = \frac{\exp(Z_T/\tau)}{\sum \exp(Z_T/\tau)}. \quad (6)$$

Similarly, we generate soft labels Y_S by applying temperature scaling to the output logits Z_S of the student network. Consequently, the distillation loss is calculated as

$$\mathcal{L}_{\text{Distill}} = \lambda_D \tau^2 \mathcal{L}_{\text{KL}}(Y_S, Y_T) + (1 - \lambda_D) \mathcal{L}_{\text{CE}}(Y_S, G), \quad (7)$$

where λ_D is a hyper-parameter, G denotes the segmentation ground truth, and \mathcal{L}_{KL} denotes the KL divergence loss.

As is shown in Section VII, our adaptation via distillation strategy is more effective than naive transfer learning, since it addresses the domain gap between the RAVIR dataset and color fundus image datasets by feature alignment in the target space through KL divergence minimization.

V. EXPERIMENTS

A. Implementation Details

All the models were implemented in TensorFlow 2.0 [52]. Experiments were performed on an Nvidia Titan RTX GPU, and an Intel® Core™ i7-7700K CPU @ 4.20 GHz.

All the networks employ a sliding window approach and were trained on randomly extracted patches of size 256×256 . Additionally, we used common data augmentation techniques such as random rotations, flipping, and change of contrast in order to increase the available training samples. We trained all the networks for 600 epochs and used the Adam optimization algorithm [53] with minibatches of size 6 and an initial learning rate of 0.001 that is reduced by a factor of 2 every 50 epochs.

In all experiments, the network is evaluated on a validation set consisting of 4 images from the original 26 images, and the instance with the best average F1 score on the artery and vein classes is used for testing.

We used a grid search approach and found the following optimal values for the hyper-parameters in (2): $\lambda_1 = \lambda_2 = 1.0$ and $\lambda_3 = 0.001$. These hyper-parameters enable the balancing of their respective loss terms so as to facilitate fast convergence to optimal solutions. In addition, for the hyper-parameters in (7), we use $\tau = 3$ and $\lambda_D = 0.1$. All the baseline methods were implemented by adopting their official released repositories. For all the baseline experiments, we used a combination of Dice and CE as the training loss function.

B. Evaluation Metrics

1) *Segmentation*: To assess the accuracy of model predictions, we used various evaluation metrics, including Sensitivity (SE), Specificity (SP), Accuracy (Acc), Precision (PR), Recall (RE), Area Under the Receiver Operating Characteristics (ROC) Curve (AUC), and the *F1* score, which is the same as the Dice coefficient in the context of image segmentation.

2) *Width Estimation*: We refer to the diameter of the vessel as twice the maximum distance to its medial axis. To assess the accuracy of diameter estimations, we employ the Mean Absolute Percentage Error (MAPE) defined as

$$MAPE = \frac{1}{N} \sum_{m=1}^N \left| \frac{y_m - \hat{y}_m}{y_m} \right|, \quad (8)$$

where N , \hat{y}_m , and y_m denote the total number of cases, predictions, and ground truth values, respectively.

VI. RESULTS

In this section, we first present quantitative and qualitative comparisons between the SegRAVIR model and competing methodologies on the RAVIR dataset as well as on the RITE dataset [11]. Then we report the pixel-wise diameter profiles of SegRAVIR-segmented arteries and veins with quantitative error analysis. Finally, we present the benchmarks of pretrained RAVIR networks on three publicly available color datasets, the DRIVE [10], STARE [9], and CHASE_DB1 [12] datasets, and demonstrate the effectiveness of our proposed distillation methodology.

A. Retinal Artery and Vein Segmentation

As reported in Table IV, we compared the SegRAVIR model against competing deep learning-based segmentation approaches on the RAVIR dataset. Evidently, SegRAVIR outperforms these methods as judged by all metrics for artery and vein classes with a healthy margin. In terms of Dice score, SegRAVIR outperforms CE-NET, IterNet and AG-Net by 16.28%, 26.71% and 27.48% for artery segmentation and by 16.28%, 26.71% and 27.48% for vein segmentation, respectively. Fig. 4 presents a qualitative comparison of the semantic segmentation outputs of SegRAVIR, CE-Net, and U-Net. Specifically, SegRAVIR yields more accurate vessel

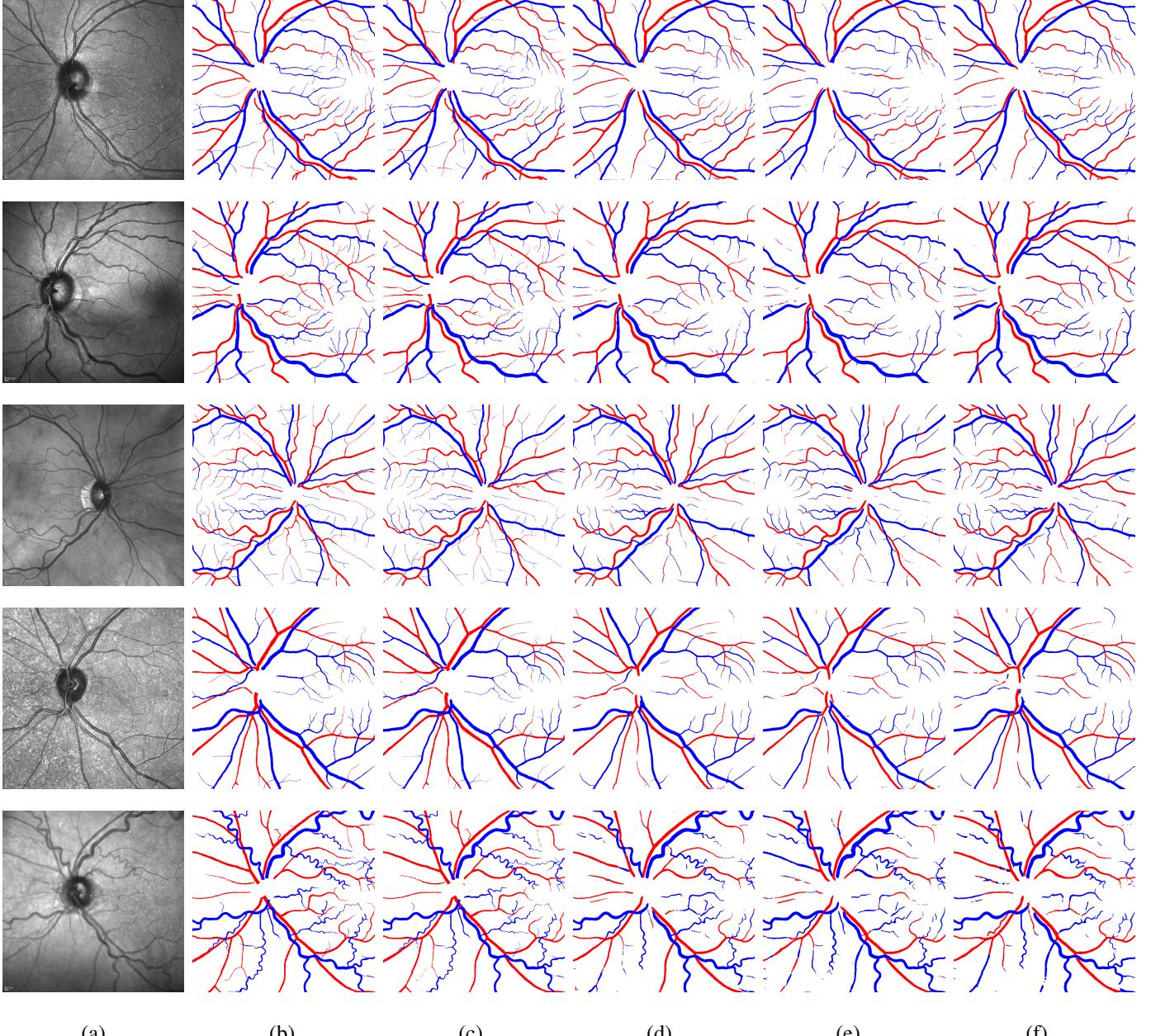


Fig. 4. Qualitative comparison of segmentation outputs on the RAVIR test set. (a) Input image. (b) Semantic Labels. (c) SegRAVIR. (d) CE-Net. (e) R2U-Net. (f) U-Net. The artery and vein classes are indicated by red and blue, respectively.

TABLE IV
QUANTITATIVE COMPARISON OF SEMANTIC SEGMENTATION PERFORMANCE ON THE RAVIR DATASET

| Method | Artery | | | | | Vein | | | | | Average |
|----------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | SE | SP | Acc | AUC | Dice | SE | SP | Acc | AUC | Dice | |
| U-Net [21] | 0.7275 | 0.9788 | 0.9670 | 0.9665 | 0.7510 | 0.7389 | 0.9795 | 0.9685 | 0.9713 | 0.7674 | 0.7592 |
| Dense U-Net [54] | 0.7296 | 0.9795 | 0.9681 | 0.9673 | 0.7584 | 0.7410 | 0.9808 | 0.9691 | 0.9736 | 0.7704 | 0.7644 |
| Residual U-Net [25] | 0.7375 | 0.9828 | 0.9697 | 0.9689 | 0.7602 | 0.7455 | 0.9839 | 0.9701 | 0.9773 | 0.7768 | 0.7685 |
| R2U-Net [25] | 0.7408 | 0.9810 | 0.9706 | 0.9727 | 0.7621 | 0.7492 | 0.9842 | 0.9728 | 0.9802 | 0.7826 | 0.7723 |
| Recurrent U-Net [25] | 0.7389 | 0.9833 | 0.9718 | 0.9749 | 0.7690 | 0.7478 | 0.9849 | 0.9735 | 0.9810 | 0.7898 | 0.7794 |
| U-Net++ [55] | 0.7406 | 0.9827 | 0.9729 | 0.9772 | 0.7781 | 0.7527 | 0.9837 | 0.9743 | 0.9821 | 0.7909 | 0.7845 |
| DU-Net [26] | 0.7432 | 0.9854 | 0.9750 | 0.9785 | 0.7833 | 0.7569 | 0.9870 | 0.9762 | 0.9840 | 0.7972 | 0.7902 |
| AG-Net [28] | 0.7426 | 0.9860 | 0.9751 | 0.9792 | 0.7854 | 0.7562 | 0.9879 | 0.9768 | 0.9844 | 0.7983 | 0.7918 |
| IterNet [34] | 0.7431 | 0.9863 | 0.9765 | 0.9820 | 0.7898 | 0.7586 | 0.9895 | 0.9779 | 0.9852 | 0.8009 | 0.7953 |
| CE-Net [30] | 0.7501 | 0.9881 | 0.9771 | 0.9834 | 0.7928 | 0.7682 | 0.9908 | 0.9786 | 0.9871 | 0.8053 | 0.7990 |
| SegRAVIR | 0.7772 | 0.9925 | 0.9815 | 0.9902 | 0.8287 | 0.8086 | 0.9951 | 0.9829 | 0.9916 | 0.8301 | 0.8294 |

TABLE V
QUANTITATIVE COMPARISON OF SEMANTIC SEGMENTATION PERFORMANCE ON THE RITE DATASET

| Method | SE | SP | Acc |
|--------------------------------|---------------|---------------|---------------|
| Dashtbozorg <i>et al.</i> [36] | 0.9000 | 0.8400 | 0.8740 |
| Estrada <i>et al.</i> [37] | 0.9300 | 0.9410 | 0.9350 |
| Girard <i>et al.</i> [41] | 0.8630 | 0.8660 | 0.8650 |
| Hemelings <i>et al.</i> [40] | - | - | 0.9487 |
| Ma <i>et al.</i> [45] | 0.9220 | 0.9300 | 0.9260 |
| Kang <i>et al.</i> [43] | 0.8863 | 0.9272 | 0.9081 |
| Morano <i>et al.</i> [48] | 0.8747 | 0.9089 | 0.8924 |
| SegRAVIR | 0.9310 | 0.9431 | 0.9513 |

TABLE VI
MEASURED DIAMETERS (IN MICRONS) OF SEGMENTED ARTERIES AND VEINS IN THE RAVIR TEST SET

| Method | Artery | | Vein | |
|----------------------|-------------------|---------------|---------------------|---------------|
| | Measured | MAPE | Measured | MAPE |
| Dense U-Net [54] | 99.98±8.12 | 0.0618 | 112.54±9.89 | 0.0607 |
| U-Net [21] | 99.93±7.31 | 0.0612 | 112.42±9.76 | 0.0595 |
| Residual U-Net [25] | 99.88±6.18 | 0.0596 | 112.30±10.14 | 0.0586 |
| Recurrent U-Net [25] | 99.76±6.53 | 0.0583 | 112.57±8.59 | 0.0535 |
| R2U-Net [25] | 96.94±6.14 | 0.0578 | 113.86±9.85 | 0.0526 |
| U-Net++ [55] | 93.96±5.64 | 0.0559 | 111.63±10.83 | 0.0520 |
| AG-Net [28] | 94.34±6.76 | 0.0548 | 110.84±9.42 | 0.0514 |
| DU-Net [26] | 97.81±5.54 | 0.0539 | 110.57±9.48 | 0.0501 |
| IterNet [34] | 97.54±6.10 | 0.0502 | 111.55±9.77 | 0.0498 |
| CE-Net [30] | 97.33±5.75 | 0.0463 | 113.73±11.23 | 0.0457 |
| SegRAVIR | 98.29±5.48 | 0.0409 | 112.34±10.02 | 0.0393 |

MAPE: Mean Absolute Percentage Error

topology (i.e., thickness and orientation) segmentation with higher pixel-wise classification accuracy.

Table V presents quantitative performance benchmarks of SegRAVIR and other competing approaches for retinal artery and vein classification on the RITE dataset [11]. SegRAVIR outperforms previous state-of-the-art approaches in terms of accuracy, sensitivity, and specificity. Fig. 5 provides a qualitative comparison between segmentation outputs of SegRAVIR and the method of Hemelings *et al.* [40] on the RITE test set.

B. Vessel Width Estimation

Table VI presents a quantitative comparison of the measured diameters using the segmentation outputs of SegRAVIR and competing approaches. Using the pixel-wise annotated masks, the reference average diameter of the arteries and veins in the test set of the RAVIR dataset were measured as 97.93 ± 8.62 and 113.98 ± 13.93 , respectively. According to our analysis, SegRAVIR can accurately measure the diameter of the vessels and it achieves the smallest MAPE among the competing approaches. Specifically, in comparison to CE-Net, Iter-Net, and DU-Net, respectively, SegRAVIR is on average 13.20%, 22.73% and 31.78% more accurate in terms of MAPE for the measured diameter of arteries and 16.28%, 26.71% and 27.48% in terms of MAPE for the measured diameter of veins. Fig. 6 presents qualitative comparisons of reference and SegRAVIR estimated diameter maps.

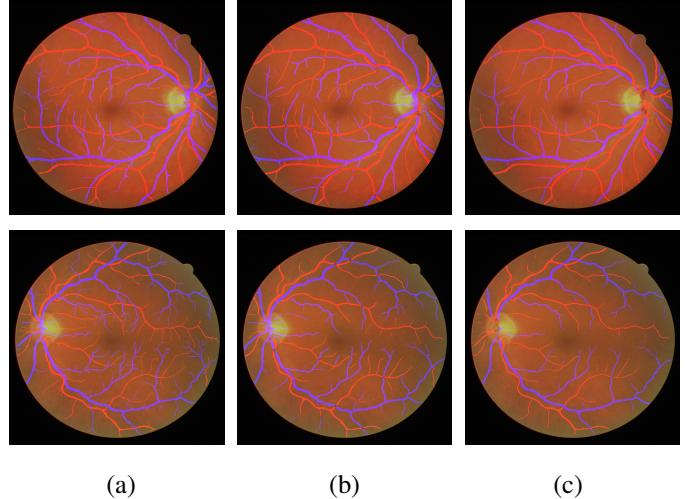


Fig. 5. Qualitative comparison of segmentation outputs on the RITE test set. (a) Reference labels. (b) SegRAVIR. (c) Hemelings *et al.* [40]. The artery and vein classes are indicated by red and blue, respectively.

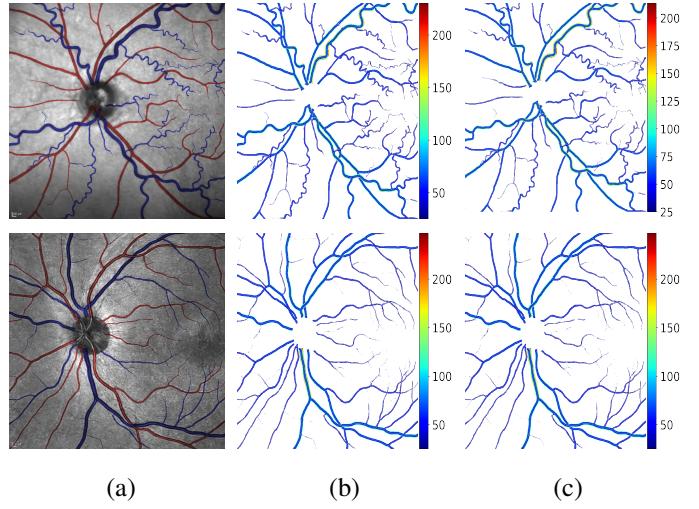


Fig. 6. Qualitative comparisons of pixel-wise width maps (measured in microns), with the red and blue masks denoting artery and vein classes, respectively. (a) SegRAVIR segmentation output. (b) Reference width maps. (c) Computed width maps.

C. Knowledge Distillation

Using the final pretrained student SegRAVIR model in our knowledge distillation framework described in Section IV-D, we fine-tuned and tested the network on three publicly available datasets of color images, DRIVE [10], STARE [9], and CHASE_DB1 [12]. Fig. 7 shows example of SegRAVIR segmentation outputs on DRIVE and CHASE_DB1 datasets. Table VII presents a quantitative comparison between the segmentation performance of our SegRAVIR model and state-of-the-art models: R2U-Net [25], DU-Net [26], and IterNet [34]. By all evaluation metrics, SegRAVIR has achieved new state-of-the-art results on the DRIVE, STARE [9], and CHASE_DB1 datasets.

VII. ABLATION STUDY

In Table VIII, we evaluate the effectiveness of each term of SegRAVIR’s loss function (2). Our proposed hybrid loss

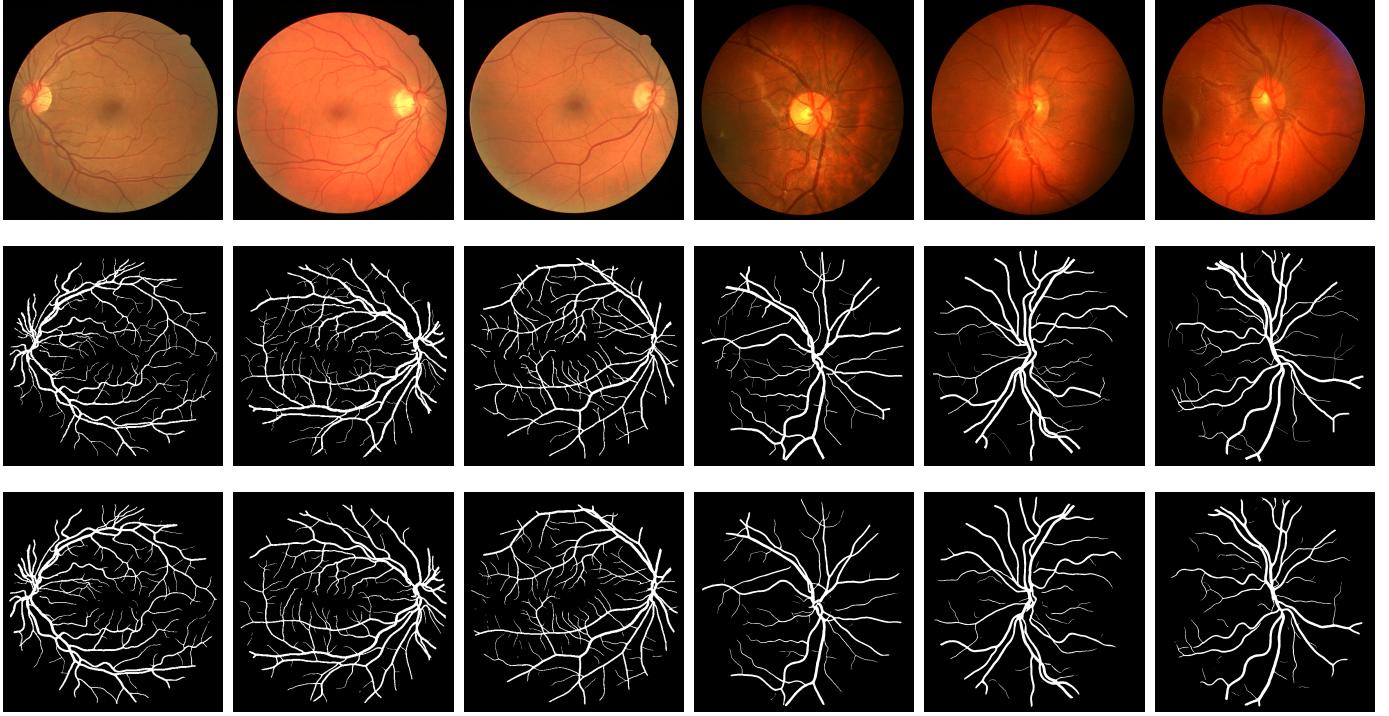


Fig. 7. Example SegRAVIR segmentation outputs for images in the test sets of the DRIVE (first three columns) and CHASE_DB1 datasets. From top, the rows show input images, groundtruth labels, and segmentations.

TABLE VII
QUANTITATIVE COMPARISON OF SEGMENTATION PERFORMANCE ON THE DRIVE, CHASE_DB1, AND STARE DATASETS.

| Method | DRIVE | | | | | CHASE_DB1 | | | | | STARE | | | | |
|----------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | SE | SP | Acc | AUC | Dice | SE | SP | Acc | AUC | Dice | SE | SP | Acc | AUC | Dice |
| U-Net [21] | 0.7822 | 0.9808 | 0.9555 | 0.9752 | 0.8174 | 0.7840 | 0.9880 | 0.9752 | 0.9870 | 0.7993 | 0.6681 | 0.9915 | 0.9639 | 0.9710 | 0.7595 |
| Dense U-Net [54] | 0.7928 | 0.9776 | 0.9541 | 0.9756 | 0.8146 | 0.7877 | 0.9848 | 0.9743 | 0.9880 | 0.8005 | 0.6807 | 0.9916 | 0.9651 | 0.9755 | 0.7691 |
| Residual U-Net [25] | 0.7726 | 0.9820 | 0.9553 | 0.9779 | 0.8149 | 0.7726 | 0.9820 | 0.9553 | 0.9779 | 0.7800 | 0.8203 | 0.9856 | 0.9700 | 0.9904 | 0.8388 |
| Recurrent U-Net [25] | 0.7751 | 0.9816 | 0.9556 | 0.9782 | 0.8155 | 0.7459 | 0.9836 | 0.9622 | 0.9803 | 0.7810 | 0.8108 | 0.9871 | 0.9706 | 0.9909 | 0.8396 |
| R2U-Net [25] | 0.7792 | 0.9813 | 0.9556 | 0.9784 | 0.8171 | 0.7756 | 0.9820 | 0.9634 | 0.9815 | 0.7928 | 0.8298 | 0.9862 | 0.9712 | 0.9914 | 0.8475 |
| DU-Net [26] | 0.7863 | 0.9805 | 0.9558 | 0.9778 | 0.8190 | 0.7858 | 0.9880 | 0.9752 | 0.9887 | 0.8000 | 0.6810 | 0.9903 | 0.9639 | 0.9758 | 0.7629 |
| IterNet [34] | 0.7735 | 0.9838 | 0.9573 | 0.9816 | 0.8205 | 0.7969 | 0.9881 | 0.9760 | 0.9899 | 0.8072 | 0.7715 | 0.9886 | 0.9701 | 0.9881 | 0.8146 |
| DAVS-Net [56] | 0.8286 | 0.9824 | 0.9689 | 0.9825 | - | 0.8144 | 0.9843 | 0.9726 | 0.9855 | - | 0.8238 | 0.9866 | 0.9744 | 0.9877 | - |
| SegRAVIR | 0.8245 | 0.9903 | 0.9698 | 0.9889 | 0.8324 | 0.8172 | 0.9924 | 0.9788 | 0.9935 | 0.8535 | 0.8314 | 0.9942 | 0.9751 | 0.9921 | 0.8489 |

TABLE VIII
EFFECT OF VARIOUS LOSS FUNCTIONS ON THE SEGMENTATION PERFORMANCE OF SEGRAVIR IN TERMS OF DICE SCORE

| Loss function | Artery | Vein | Average |
|---------------|---------------|---------------|---------------|
| Dice | 0.8033 | 0.8095 | 0.8064 |
| CE | 0.8078 | 0.8138 | 0.8108 |
| CE+Dice | 0.8193 | 0.8225 | 0.8209 |
| CE+Dice+L2 | 0.8287 | 0.8301 | 0.8294 |

function, which incorporates Dice and CE losses for the main stream and L2 loss for the auxiliary stream, achieves consistently better segmentation performance for both the artery and vein classes. On average, it outperforms SegRAVIR networks trained with Dice, CE, and combined Dice and CE losses by 2.77%, 2.24%, and 1.02%, respectively. We hypothesize that the dominance of Dice and CE losses is due to combined improvements in artery and vein pixel-wise classification and mask prediction by the CE and Dice losses,

TABLE IX
QUANTITATIVE COMPARISON OF DIFFERENT PRETRAINING STRATEGIES ON THE DRIVE, CHASE_DB1 AND STARE DATASETS.

| Method | DRIVE | | CHASE_DB1 | | STARE | |
|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | AUC | F1 | AUC | F1 | AUC | F1 |
| Random | 0.9773 | 0.8151 | 0.9817 | 0.7998 | 0.9879 | 0.8119 |
| Pretrained | 0.9822 | 0.8216 | 0.9903 | 0.8297 | 0.9902 | 0.8275 |
| Proposed | 0.9889 | 0.8324 | 0.9935 | 0.8535 | 0.9921 | 0.8489 |

respectively.

In Table IX, we present the segmentation performance of SegRAVIR, with different pretraining strategies, on the DRIVE, STARE [9], and CHASE_DB1 datasets. To validate the effectiveness of our proposed knowledge distillation approach, we compared against the RAVIR network pretrained on the RAVIR dataset as well as the network with random weight initialization. On the DRIVE and STARE datasets, our approach yields

2.12% and 4.55% improvements in Dice score, respectively, relative to random weight initialization and 1.31% and 2.58% relative to pretrained weight initialization. Similarly, on the CHASE_DB1 dataset, our approach outperforms random and pretrained weight initialization by 6.71% and 2.86% in terms of the Dice score, respectively.

VIII. DISCUSSION

We have tackled the problem of semantic segmentation of retinal arteries and veins in IR imagery. According to our extensive experiments, the SegRAVIR model outperforms competing approaches, evidently excelling in accurately delineating both arteries and veins. Since the latter usually manifest as larger vessels, all methods showed a better performance in segmenting veins compared to arteries. Our study reveals that the two important issues in segmenting retinal IR images are maintaining structural consistency across the vessels and correctly distinguishing between arteries and veins. SegRAVIR addresses these challenges by introducing an additional auxiliary stream that learns to reconstruct the input image via an encoder shared with the main stream and by enforcing a combined Dice and CE loss function that improves both pixel-level accuracy and contour delineation. Although, some of the competing models leverage attention modules to further localize desired retinal structures, our experiments did not reveal significant performance gains, if any.

Our experiments show that SegRAVIR can successfully segment major arteries and veins while properly capturing the fine-grained structural details of smaller vessels. Furthermore, we have quantitatively studied the problem of vessel diameter estimation using the segmentation results of our model. Results show that the segmentation outputs of SegRAVIR achieve the smallest MAPE error for both arterioles and venules and can be leveraged for precise vascular diameter analysis. In particular, such a measurement tool enables the study of morphological changes of retinal arterioles and venules and opens the door to creating early predictive models for the diagnosis of various anomalies such as diabetic and hypertensive retinopathy.

We have explored the segmentation of retinal vessels in other imaging modalities. SegRAVIR outperforms the previous state-of-the-art approaches to retinal artery and vein segmentation based on the RITE dataset [11], hence validating the generalizability of our approach to other datasets. In addition, we proposed a knowledge distillation approach to domain adaptation of RAVIR-pretrained models. Although naive pretraining on the RAVIR dataset improves the segmentation performance on color datasets, our approach significantly improves this performance baseline. As a result, our model establishes new state-of-the-art performance benchmarks on the DRIVE [10], STARE [9], and CHASE_DB1 [12] datasets. This indicates that our RAVIR dataset can be leveraged to address the annotated data scarcity problem in this domain, and our SegRAVIR model can be effectively used for multi-modal segmentation across different domains.

IX. CONCLUSIONS

In summary, RAVIR is a new retinal IR image dataset that has engendered SegRAVIR, a novel, deep learning model for

the fully-automated semantic segmentation and quantitative analysis of retinal vasculature. SegRAVIR can accurately segment both arteries and veins without additional preprocessing or post-processing, and its output enables precise vessel diameter mensuration critical to the quantitative assessment of retinal vasculature. Furthermore, via a knowledge distillation approach, our dataset and model can be effectively applied to retinal vasculature segmentation based on other imaging modalities. This enables new opportunities in retinal image analysis while helping alleviate the problem of annotated data scarcity in this domain.

REFERENCES

- [1] O. Tekin, S. Çukur, C. Uraldi, B. Işık, A. Özkar, H. Kurtaran, F. Çatal, and I. Üstün, “Relationship between retinopathy and cognitive impairment among hypertensive subjects,” *European Neurology*, vol. 52, no. 3, pp. 156–161, 2004.
- [2] Y.-T. Ong, T. Y. Wong, R. Klein, B. E. Klein, P. Mitchell, A. R. Sharrett, D. J. Couper, and M. K. Ikram, “Hypertensive retinopathy and risk of stroke,” *Hypertension*, vol. 62, no. 4, pp. 706–711, 2013.
- [3] A. Omotoso, P. Kolo, T. Olanrewaju, J. Owoeye, S. Biliaminu, and V. Olatunji, “Relationship between retinopathy and renal abnormalities in black hypertensive patients,” *Clinical Hypertension*, vol. 22, no. 1, p. 19, 2016.
- [4] L. E. Downie, L. A. Hodgson, C. DSylva, R. L. McIntosh, S. L. Rogers, P. Connell, and T. Y. Wong, “Hypertensive retinopathy: Comparing the Keith-Wagener-Barker to a simplified classification,” *Journal of Hypertension*, vol. 31, no. 5, pp. 960–965, 2013.
- [5] C. Cuspidi, C. Sala, and G. Grassi, “Updated classification of hypertensive retinopathy: Which role for cardiovascular risk stratification?” *Journal of Hypertension*, vol. 33, no. 11, pp. 2204–2206, 2015.
- [6] S. Guo, S. Yin, G. Tse, G. Li, L. Su, and T. Liu, “Association between caliber of retinal vessels and cardiovascular disease: A systematic review and meta-analysis,” *Current Atherosclerosis Reports*, vol. 22, no. 4, pp. 16–16, 2020.
- [7] N. Patton, T. Aslam, T. MacGillivray, A. Pattie, I. J. Deary, and B. Dhillon, “Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: A rationale based on homology between cerebral and retinal microvasculatures,” *Journal of Anatomy*, vol. 206, no. 4, pp. 319–348, 2005.
- [8] A. Ajaz, B. Aliahamad, M. Sarossy, and D. K. Kumar, “Relation between retinal vasculature and retinal thickness in macular edema,” in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 5593–5596.
- [9] A. Hoover, V. Kouznetsova, and M. Goldbaum, “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response,” *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, 2000.
- [10] 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 Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
- [11] Q. Hu, M. D. Abràmoff, and M. K. Garvin, “Automated separation of binary overlapping trees in low-contrast color retinal images,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 436–443.
- [12] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, “An ensemble classification-based approach applied to retinal blood vessel segmentation,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012.
- [13] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. Pluim, R. Duits, and B. M. ter Haar Romeny, “Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 12, pp. 2631–2644, 2016.
- [14] J. I. Orlando, J. B. Breda, K. Van Keer, M. B. Blaschko, P. J. Blanco, and C. A. Bulant, “Towards a glaucoma risk index based on simulated hemodynamics from fundus images,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 65–73.
- [15] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, “Robust vessel segmentation in fundus images,” *International Journal of Biomedical Imaging*, vol. 2013, 2013.

- [16] Y. Ma, H. Hao, J. Xie, H. Fu, J. Zhang, J. Yang, Z. Wang, J. Liu, Y. Zheng, and Y. Zhao, "ROSE: A retinal OCT-angiography vessel segmentation dataset and new model," *IEEE Transactions on Medical Imaging*, 2020.
- [17] L. Ding, M. H. Bawany, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "A novel deep learning pipeline for retinal vessel detection in fluorescein angiography," *IEEE Transactions on Image Processing*, vol. 29, pp. 6561–6573, 2020.
- [18] L. Ding, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "Weakly-supervised vessel detection in ultra-widefield fundus photography via iterative multi-modal registration and learning," *IEEE Transactions on Medical Imaging*, 2020.
- [19] T. A. Qureshi, M. Habib, A. Hunter, and B. Al-Diri, "A manually-labeled, artery/vein classified benchmark for the drive dataset," in *IEEE International Symposium on Computer-Based Medical Systems*, 2013, pp. 485–488.
- [20] C. C. Wykoff, M. G. Nittala, B. Zhou, W. Fan, S. B. Velaga, S. I. Lampen, A. M. Rusakevich, J. P. Ehlers, A. Babiuch, D. M. Brown *et al.*, "Intravitreal afibercept for retinal nonperfusion in proliferative diabetic retinopathy: Outcomes from the randomized recovery trial," *Ophthalmology Retina*, vol. 3, no. 12, pp. 1076–1086, 2019.
- [21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.
- [22] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," *arXiv preprint arXiv:1810.07810*, 2018.
- [23] C. Wang, Z. Zhao, Q. Ren, Y. Xu, and Y. Yu, "Dense U-net based on patch-based learning for retinal vessel segmentation," *Entropy*, vol. 21, no. 2, p. 168, 2019.
- [24] Y. Jiang, H. Zhang, N. Tan, and L. Chen, "Automatic retinal blood vessel segmentation based on fully convolutional neural networks," *Symmetry*, vol. 11, no. 9, p. 1112, 2019.
- [25] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 1, p. 014006, 2019.
- [26] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "Dunet: A deformable network for retinal vessel segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149–162, 2019.
- [27] Y. Wu, Y. Xia, Y. Song, D. Zhang, D. Liu, C. Zhang, and W. Cai, "Vessel-Net: Retinal vessel segmentation under multi-path supervision," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 264–272.
- [28] S. Zhang, H. Fu, Y. Yan, Y. Zhang, Q. Wu, M. Yang, M. Tan, and Y. Xu, "Attention guided network for retinal image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 797–805.
- [29] B. Wang, S. Qiu, and H. He, "Dual encoding U-Net for retinal vessel segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 84–92.
- [30] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.
- [31] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, "ET-Net: A generic edge-attention guidance network for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 442–450.
- [32] Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1427–1436, 2018.
- [33] G. A. Francia, C. Pedraza, M. Aceves, and S. Tovar-Arriaga, "Chaining a U-Net with a residual U-Net for retinal blood vessels segmentation," *IEEE Access*, vol. 8, pp. 38493–38500, 2020.
- [34] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2020.
- [35] Y. Guo and Y. Peng, "BSCN: Bidirectional symmetric cascade network for retinal vessel segmentation," *BMC Medical Imaging*, vol. 20, no. 1, pp. 1–22, 2020.
- [36] B. Dashtbozorg, A. M. Mendonça, and A. Campilho, "An automatic graph-based approach for artery/vein classification in retinal images," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1073–1083, 2013.
- [37] R. Estrada, M. J. Allingham, P. S. Mettu, S. W. Cousins, C. Tomasi, and S. Farsiu, "Retinal artery-vein classification via topology estimation," *IEEE Transactions on Medical Imaging*, vol. 34, no. 12, pp. 2518–2534, 2015.
- [38] Q. Hu, M. D. Abràmoff, and M. K. Garvin, "Automated construction of arterial and venous trees in retinal images," *Journal of Medical Imaging*, vol. 2, no. 4, p. 044001, 2015.
- [39] R. Welikala, P. Foster, P. Whincup, A. R. Rudnicka, C. G. Owen, D. Strachan, S. Barman *et al.*, "Automated arteriole and venule classification using deep learning for retinal images from the UK Biobank cohort," *Computers in Biology and Medicine*, vol. 90, pp. 23–32, 2017.
- [40] R. Hemelings, B. Elen, I. Stalmans, K. Van Keer, P. De Boever, and M. B. Blaschko, "Artery–vein segmentation in fundus images using a fully convolutional network," *Computerized Medical Imaging and Graphics*, vol. 76, p. 101636, 2019.
- [41] F. Girard, C. Kavalec, and F. Cheriet, "Joint segmentation and classification of retinal arteries/veins from fundus images," *Artificial Intelligence in Medicine*, vol. 94, pp. 96–109, 2019.
- [42] X. Xu, R. Wang, P. Lv, B. Gao, C. Li, Z. Tian, T. Tan, and F. Xu, "Simultaneous arteriole and venule segmentation with domain-specific loss function on a new public database," *Biomedical Optics Express*, vol. 9, no. 7, pp. 3153–3166, 2018.
- [43] H. Kang, Y. Gao, S. Guo, X. Xu, T. Li, and K. Wang, "Avnet: A retinal artery/vein classification network with category-attention weighted fusion," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105629, 2020.
- [44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [45] W. Ma, S. Yu, K. Ma, J. Wang, X. Ding, and Y. Zheng, "Multi-task neural networks with spatial activation for retinal vessel segmentation and artery/vein classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 769–778.
- [46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [47] W. Chen, S. Yu, J. Wu, K. Ma, C. Bian, C. Chu, L. Shen, and Y. Zheng, "TR-GAN: Topology ranking gan with triplet loss for retinal artery/vein classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 616–625.
- [48] J. Morano, Á. S. Herrvela, J. Novo, and J. Rouco, "Simultaneous segmentation and classification of the retinal arteries and veins from color fundus images," *Artificial Intelligence in Medicine*, p. 102116, 2021.
- [49] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, Mar. 1984.
- [50] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, pp. 1–9, 2015.
- [51] R. Adriana, B. Nicolas, K. S. Ebrahimi, C. Antoine, G. Carlo, and B. Yoshua, "Fitnets: Hints for thin deep nets," *Proc. ICLR*, 2015.
- [52] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *OSDI*, vol. 16, 2016, pp. 265–283.
- [53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [54] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 568–576, 2019.
- [55] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, 2019.
- [56] M. Raza, K. Naveed, A. Akram, N. Salem, A. Afafq, H. A. Madni *et al.*, "DAVS-NET: Dense aggregation vessel segmentation network for retinal vasculature detection in fundus images," *PLOS One*, vol. 16, no. 12, p. e0261698, 2021.