Weakly-Supervised Vessel Detection in Ultra-Widefield Fundus Photography Via Iterative Multi-Modal Registration and Learning

Li Ding, *Student Member, IEEE*, Ajay E. Kuriyan, Rajeev S. Ramchandran, Charles C. Wykoff, and Gaurav Sharma, *Fellow, IEEE*

***Abstract*—We propose a deep-learning based annotation-efficient framework for vessel detection in ultra-widefield (UWF) fundus photography (FP) that does not require *de novo* labeled UWF FP vessel maps. Our approach utilizes concurrently captured UWF fluorescein angiography (FA) images, for which effective deep learning approaches have recently become available, and iterates between a multi-modal registration step and a weakly-supervised learning step. In the registration step, the UWF FA vessel maps detected with a pre-trained deep neural network (DNN) are registered with the UWF FP via parametric chamfer alignment. The warped vessel maps can be used as the tentative training data but inevitably contain incorrect (noisy) labels due to the differences between FA and FP modalities and the errors in the registration. In the learning step, a robust learning method is proposed to train DNNs with noisy labels. The detected FP vessel maps are used for the registration in the following iteration. The registration and the vessel detection benefit from each other and are progressively improved. Once trained, the UWF FP vessel detection DNN from the proposed approach allows FP vessel detection without requiring concurrently captured UWF FA images. We validate the proposed framework on a new UWF FP dataset, PRIME-FP20, and on existing narrow-field FP datasets. Experimental evaluation, using both pixel-wise metrics and the CAL metrics designed to provide better agreement with human assessment, shows that the proposed approach provides accurate vessel detection, without requiring manually labeled UWF FP training data.**

***Index Terms*—Retinal vessel detection, multi-modal registration, ultra-widefield fundus photography, noisy labels**

I. Introduction

Ophthalmologists recognize features of retinal vasculature as important biomarkers associated with multiple diseases. For example, diabetic retinopathy and retinal vein occlusion are characterized by increase in retinal vasculature tortuosity, vessel caliber expansion, and retinal non-perfusion [1]. Therefore, detecting vessels is a fundamental problem in retinal image analysis that has been extensively researched. Existing approaches classify into two main categories, supervised and unsupervised, depending on whether they do or do not use labeled training data [2]. Traditionally, the focus was on unsupervised methods that addressed the problem from a variety of perspectives, incuding hand-crafted match filtering [3], [4], morphological processing [5]–[7], multiscale approaches [8], [9], and matting-based techniques [9]. Recently, supervised learning appproaches, specifically deep neural networks (DNNs), have led to significant improvements in retinal vessel detection. A variety of DNN architectures have been proposed for retinal vessel detection, including per-pixel classifier [10], fully convolutional network [11], [12], U-Net [13]–[16], graph neural network [17], context encoder network [18], and generative adversarial networks [19]. Additionally, several works exploit novel loss functions [20]–[22] and training strategies [23]. These DNN based methods have primarily focused on narrow field (NF) fundus photography (FP), both because NF FP is the predominant format and modality of capture in the clinical setting and because recent efforts have created reasonable sized labeled ground truth datasets for DNN training [4], [24]–[29].

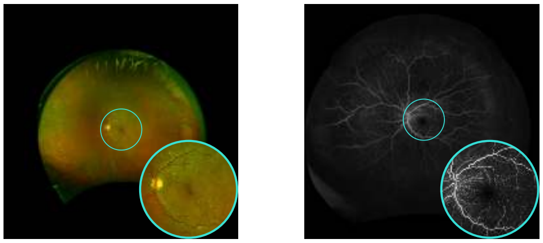

(a) UWF FP (b) UWF FA

Fig. 1: Concurrently captured ultra-widefield (UWF) fundus photography (FP) and UWF fluorescein angiography (FA) image pair. Cyan circles depict the approximate field-of-view for narrow-field FP.

Due to the additional diagnostic information they can offer, vessel detection is also of interest in formats and modalities other than NF FP [30]. Specifically, in this paper, we focus on ultra-wide field (UWF) FP [31] leveraging concurrently

The work was supported in part by a University of Rochester Research Award, by a distinguished researcher award from the New York state funded Rochester Center of Excellence in Data Science (CoE #3B C160189) at the University of Rochester, by an unrestricted grant to the Department of Ophthalmology from Research to Prevent Blindness, and grant P30EY001319-35 from the National Institutes of Health.

L. Ding and G. Sharma are with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627, USA (e-mail: {l.ding, gaurav.sharma}@rochester.edu).

A. E. Kuriyan is with Retina Service, Wills Eye Hospital, Philadelphia, PA 19107 & the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642, USA (e-mail: ajay.kuriyan@gmail.com).

R. S. Ramchandran is with the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642, USA (e-mail: rajeev_ramchandran@urmc.rochester.edu).

C. C. Wykoff is with Retina Consultants of Houston and Blanton Eye Institute, Houston Methodist Hospital & Weill Cornell Medical College, Houston, TX 77030, USA (e-mail: ccwmd@houstonretina.com).

captured UWF fluorescein angiography (FA) images. Like NF FP, UWF FP is noninvasive and only involves capture of the retinal images under low-power illumination; even pupil dilation is not required [31]. As shown in Fig. 1a, UWF FP images provide a wide 200° field-of-view (FOV) in a single high-resolution image, as opposed to the much narrower 30°–50° FOV for NF FP. Manual examination of the UWF images in diagnosis achieves reliable performance comparable to direct clinical examination using an opthalmoscope with pupil dilation. At the same time, UWF FP also reveals additional peripheral retinal vasculature structure that is of diagnostic importance when compared to NF FP [32], [33]. UWF FA, which is shown in Fig. 1b, represents an alternative modality that also offers a wide FOV and additional diagnostic utility, but has the limitation that it is more invasive, requiring intravenous injection of fluorescein sodium dye.

While DNNs trained on NF FP can be applied to UWF FP, the performance is relatively poor in the peripheral region (as demonstrated in Section III-F). The development of DNNs specifically for detecting vessels in UWF FP has been stymied by the paucity of labeled ground truth data. Manually annotating the binary vessel maps for UWF FP is particularly time-consuming and requires clinical-expertise. High-resolution UWF FP exhibits non-uniform illumination and contrast between vessels and background, which makes it challenging and time-intensive to accurately annotate both major and minor vessels across the large FOV; estimates indicate that approximately 18 hours are required for *de novo* manual annotation for one UWF FP image [34]. Prior work on UWF FP vessel detection [34] therefore proposed the use of pixel-wise hand-crafted features with a shallow, two-layer, multi-layer perceptron that were trained on a limited number of small labeled patches. The approach, however, does not take full advantage of deep learning advances that employ end-to-end training and also learn features in a data-driven fashion.

In this paper, we focus on innovative methodologies that train DNNs for UWF FP vessel detection in an annotation-efficient fashion and eliminate the requirement of manually labeled datasets for supervised learning. To this end, we make the following contributions:

- We present a novel iterative framework for vessel detection in UWF FP using DNNs that does not require *de novo* labeled UWF FP vessel maps. Instead, we rely on datasets that also include concurrently captured UWF FA images, for which effective deep learning approaches for vessel detection have recently become available allowing for accurate vessel detection. The proposed framework then jointly addresses precise registration between the vessel images for the modalities and vessel detection in UWF FP, where the two tasks synergistically benefit each other as iterations progress despite the differences in geometry and modality.
- We construct a new ground truth labeled dataset, PRIME-FP20, to evaluate retinal vessel detection in UWF FP and to facilitate further work on this problem.
- The proposed framework provides a method for accurate vessel detection in UWF FP imagery, a modality that has received limited attention in prior works. The proposed approach significantly outperforms existing methods on the PRIME-FP20 dataset and, on NF FP datasets, achieves performance comparable with state-of-the-art methods designed specifically for NF FP.

We note that an alternative framework for joint vessel detection and registration on paired NF FP and NF FA images has also been proposed in [35]. The framework in [35] formulates vessel detection as a style transfer task (from retinal images to binary vessel maps) and uses one vessel map from the existing dataset as the style target for all training images. In this setting, the supervision signal, which is based on perceptual loss, is relatively weak and sensitive to the selection of the style target. In contrast, the proposed framework directly transfers vessel maps from UWF FA to UWF FP providing pixel-wise supervision, which is more effective.

The rest of the paper is organized as follows. Section II describes the proposed iterative registration and learning framework. In Section III, we perform the detailed analysis of the proposed framework and present the experimental results of vessel detection. Section IV concludes the paper.

II. Iterative Registration and Learning approach

As already mentioned, instead of labeled data, training in the proposed approach is accomplished by using a set of concurrently captured UWF FP and UWF FA images, which we denote as $\{(\boldsymbol{X}_c^i, \boldsymbol{X}_a^i)\}_{i=1}^M$, where $(\boldsymbol{X}_c^i, \boldsymbol{X}_a^i)$ denotes a simultaneously captured UWF FP and UWF FA image pair (in that order) and M is the number of image pairs. Importantly, we note that while the image pairs for the two modalities are captured during the same clinical visit, they are not aligned and have significant differences in geometry in addition to fundamental differences in the information they contain arising from the differences in the modalities. In the ensuing discussion, we illustrate and describe the processing for one pair $(\boldsymbol{X}_c^i, \boldsymbol{X}_a^i)$, the i^{th} pair, for situations where the same processing flow applies to all pairs.

For each UWF FA image $\boldsymbol{X}_a^i$, a corresponding vessel map $\boldsymbol{Y}_a^i$ is obtained using a pre-trained DNN for this modality (shown in green Fig. 2). Our implementation uses [36] though alternative approaches could also be utilized for this purpose. The training of the desired DNN for FP vessel detection is then accomplished as shown in Fig. 2 by iterating between two steps comprising (a) multi-modal registration between the estimated UWF FA vessel map and a current estimate for the UWF FP vessel map and (b) weakly-supervised learning from noisy labels. Specifically, at iteration t, using parametric chamfer alignment, the detected UWF FA vessel map $\boldsymbol{Y}_a^i$ is registered with the current estimate $\boldsymbol{Y}_c^{i,t}$ of the UWF FP vessel map. The UWF FA vessel map $\boldsymbol{Y}_a^i$ is warped using the estimated registration transformation to obtain tentative/noisy training labels $\boldsymbol{Y}_{a\to c}^{i,t}$ for pixels in the corresponding UWF FP image $\boldsymbol{X}_c^i$. The collective set of such pairs of images for the concurrent UWF FP and FA captured images form the (noisy-labeled) training data $\{(\boldsymbol{X}_c^i, \boldsymbol{Y}_{a\to c}^{i,t})\}_{i=1}^M$. The fundamental differences between FA and FP imaging modalities and invariable errors in the registration contribute to the noise in the labeling. In particular, FA imaging captures fine vessels

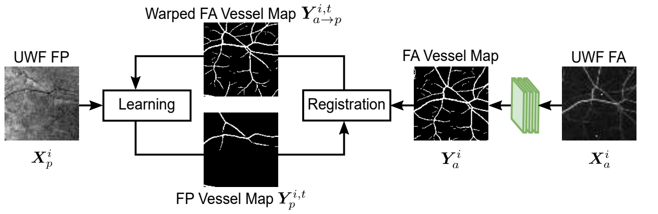


Fig. 2: Proposed iterative registration and learning approach for retinal vessel detection in UWF FP without requiring labeled FP vessel data. For clarity, the figure illustrates the processing flow for only the i^{th} pair $(\boldsymbol{X}_c^i, \boldsymbol{X}_a^i)$ of UWF FP and UWF FA images, from the complete set of M pairs $\{(\boldsymbol{X}_c^i, \boldsymbol{Y}_{a\to c}^{i,t})\}_{i=1}^{M}$ used for the training.

that are not visible in FP [31]. Consequently, the warped vessel maps $\boldsymbol{Y}_{a\to c}^{i,t}$ contain a large amount of "false positive" labels that are actually background in $\boldsymbol{X}_c^i$.

In the learning step, we propose a robust weakly-supervised learning approach to train DNN that identifies and corrects the noisy labels in the generated dataset. The detected UWF FP vessel map $\boldsymbol{Y}_c^{i,t+1}$, estimated by the trained DNN, is used for the registration step in the $(t+1)^{\text{th}}$ iteration.

The proposed framework iteratively addresses precise registration and vessel detection, where two tasks synergistically benefit each other as iterations progress. Precise alignment is important to obtain high-quality training labels. Even a small misalignment between the FA and FP images can significantly deteriorate the training data quality by assigning incorrect labels to the image pixels. On the other hand, accurate UWF FP vessel detection, estimated using the weakly-supervised learning approach, helps estimate the registration parameters because chamfer alignment uses detected UWF FP vessel maps for anchoring. Using concurrently captured UWF FP and FA images, the proposed framework accomplishes the training of a DNN for FP vessel detection *without requiring labeled UWF FP data*. Note that vessel detection in UWF FP images can be performed using the trained DNN without requiring concurrently captured UWF FA images.

Next we provide details for the registration and learning steps that constitute the two major steps in the proposed iterative framework.

A. Vessel Registration via Chamfer Alignment

Binary UWF FA vessel maps $\boldsymbol{Y}_a^i$ are transferred to the corresponding UWF FP images $\boldsymbol{X}_c^i$ by using a geometric transform that is estimated using the chamfer alignment technique from [36]. To make the presentation self-contained, we include a brief overview here that conveys the key intuition.

We denote the locations of estimated vessel pixels in the UWF FA vessel map $\boldsymbol{Y}_a^i$ by $\mathcal{Q}_a = \{\boldsymbol{q}_j^a\}_{j=1}^{N_a}$, where $\boldsymbol{q}_j^a$ are the 2D coordinates of vessel pixel j and N_a is the number of vessel pixels in $\boldsymbol{Y}_a^i$. Similarly, the locations of the N_c vessel pixels in the estimated UWF FP vessel map $\boldsymbol{Y}_c^{i,t}$ at iteration t are represented as $\mathcal{Q}_c^t = \{\boldsymbol{q}_k^c\}_{k=1}^{N_c}$. Chamfer alignment [37] estimates a parametric geometric transformation $\mathcal{T}_{\boldsymbol{\beta}}$ to register the points in $\mathcal{Q}_a$ to those in $\mathcal{Q}_c^t$ by minimizing the average squared Euclidean distance between the transformed locations $\mathcal{T}_{\boldsymbol{\beta}}\left(\boldsymbol{q}_j^a\right)$ and the closest point in $\mathcal{Q}_c$, where $\boldsymbol{\beta}$ denotes the vector of parameters for the geometric transform. Specifically, define the objective function

$$L\left(\boldsymbol{\beta}\right) = \frac{1}{N_a}\sum_{j=1}^{N_a} D_j(\boldsymbol{q}_k^c, \boldsymbol{q}_j^a), \tag{1}$$

with $D_j(\boldsymbol{q}_k^c, \boldsymbol{q}_j^a) = \min_k \|\boldsymbol{q}_k^c - \mathcal{T}_{\boldsymbol{\beta}}\left(\boldsymbol{q}_j^a\right)\|^2$. Then the estimated registration transform is obtained as $\mathcal{T}_{\boldsymbol{\beta}^*}$ where $\boldsymbol{\beta}^*$ minimizes $L\left(\boldsymbol{\beta}\right)$. We use a second order polynomial transformation for $\mathcal{T}_{\boldsymbol{\beta}}$, which is parameterized by a 12-dimensional parameter vector $\boldsymbol{\beta}$ and has been shown to be suitable for retinal vessel registration in prior work [36], [38].

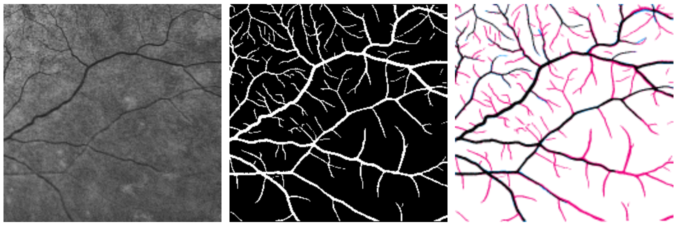
In practice, we use a refinement of the basic chamfer alignment approach outlined above that uses a latent-variable based probablistic formulation along with the expectation maximization (EM) algorithm [39] to provide robustness against outlier points that exist in $\mathcal{Q}_a$ but do not have correspondences in $\mathcal{Q}_c$. The robustness against such outliers is particularly crucial in this application setting because, as noted earlier, some fine vessels appear only in $\mathcal{Q}_a$ because the FA modality detects these much better than FP. We refer readers to [36] for detailed derivations of the parameter estimation with the EM approach. Here we only note that the key intuition can be understood from the fact that, in the EM approach, the arithmetic average in (1) is replaced by a weighted average where the weight for the squared error $D_j(\boldsymbol{q}_k^c, \boldsymbol{q}_j^a)$ corresponding to the j^{th} point in $\mathcal{Q}_a$ corresponds to the estimated posterior probability that it is not an outlier (and has a corresponding point in $\mathcal{Q}_c$). When these posterior probabilities are accurately estimated, the errors for the outlier points effectively drop out from the weighted average, as desired.

For the t^{th} iteration, once the registration transform parameters have been estimated, by applying the corresponding transformation $\mathcal{T}_{\boldsymbol{\beta}^*}$ to the UWF FA vessel maps $\boldsymbol{Y}_a^i$ we obtain the warped version $\boldsymbol{Y}_{a\to c}^{i,t}$ as the current estimate of the FA vessel map aligned with the FP imagery, which serves as "noisy labels" for the learning step.

B. Weakly-Supervised Learning with Noisy Labels

While the multi-modal registration provides tentative dataset $\{(\boldsymbol{X}_c^i, \boldsymbol{Y}_{a\to c}^{i,t})\}_{i=1}^{M}$ to train a DNN for detecting vessels in UWF FP, the labels in $\boldsymbol{Y}_{a\to c}^{i,t}$ inevitably contain noise (incorrect labels) due to the fundamental differences in FA and FP modalities. In this sub-section, we analyze the characteristic of the label noise and propose a weakly-supervised learning method to train DNN against label noise.

FA imaging is able to capture the fine retinal vessels better than FP [31]. Consequently, the warped FA vessel maps $\boldsymbol{Y}_{a\to c}^{i,t}$ contain a large number of vessel branches, especially fine vessels, that are not visible in FP modality. Figures 3a and 3b show a sample UWF FP patch selected from the peripheral region and the corresponding warped UWF FA vessel map, respectively. From these two figures, one can appreciate that the majority of fine vessels are not captured in UWF FP image. In Fig. 3c, we compare and visualize the differences between the warped vessel map and ground truth labels that are manually annotated from scratch by a human annotator. The



(a) (b) (c)

Fig. 3: (a) Sample UWF FP patch and (b) the corresponding labels from warped UWF FA vessel map. Red and blue pixels in (c) indicate incorrect vessel labels (“false positive”) and background labels (“false negative”), respectively, in the warped UWF FA vessel map.

red pixels in Fig. 3c depict a large proportion of vessel labels in the warped UWF FA vessel map $\boldsymbol{Y}^{i,t}_{a\to c}$ that are actually background in the UWF FP image. On the other hand, the FP vessel pixels that are not in the warped UWF FA vessel map $\boldsymbol{Y}^{i,t}_{a\to c}$, shown in blue in Fig. 3c, are a rather small fraction of the FP vessel pixels. Thus, treated as an estimate of the FP vessel map, the warped FA vessel map $\boldsymbol{Y}^{i,t}_{a\to c}$ has low precision but high recall. Therefore, the label noise in $\boldsymbol{Y}^{i,t}_{a\to c}$ is asymmetric: the background labels are largely accurate and the vessel labels potentially have errors.

We exploit the asymmetry of the label noise and propose a weakly-supervised learning approach to train a DNN using $\boldsymbol{Y}^{i,t}_{a\to c}$ as noisy labels. Formally, we divide pixels in $\boldsymbol{Y}^{i,t}_{a\to c}$ into two sets, $\mathcal{Y}^{i,t}_v$ and $\mathcal{Y}^{i,t}_b$, where pixels in $\mathcal{Y}^{i,t}_v$ are labeled as vessels (white pixels in Fig. 3b) and those in $\mathcal{Y}^{i,t}_b$ are labeled as background (black pixels in Fig. 3b). We further denote $\mathcal{Y}^t_v = \mathcal{Y}^{1,t}_v \bigcup \mathcal{Y}^{2,t}_v \bigcup \cdots \mathcal{Y}^{M,t}_v$ and $\mathcal{Y}^t_b = \mathcal{Y}^{1,t}_b \bigcup \mathcal{Y}^{2,t}_b \bigcup \cdots \mathcal{Y}^{M,t}_b$. Our goal is to train a DNN, modeled as a function f with learnable weights $\boldsymbol{W}$, that outputs a probabilistic vessel map $\boldsymbol{Y}_c = f(\boldsymbol{X}_c; \boldsymbol{W})$ in response to an input FP image $\boldsymbol{X}_c$. In the t^{th} iteration, weight parameters $\boldsymbol{W}^t$ for the DNN are estimated by minimizing the binary cross-entropy loss, viz.,

$$\mathcal{L}^t = \frac{1}{|\mathcal{Y}^t_v| + |\mathcal{Y}^t_b|}\left(\sum_{v\in\mathcal{Y}^t_v} l^t_v + \sum_{b\in\mathcal{Y}^t_b} l^t_b\right), \tag{2}$$

where $l^t_v = -\log(y^t_{c,v})$ and $l^t_b = -\log(1 - y^t_{c,b})$ are the binary cross-entropy loss computed from the predicted vessel probability $y^t_{c,v}$ and $y^t_{c,b}$ in $\mathcal{Y}^t_v$ and $\mathcal{Y}^t_b$, respectively, and $|\cdot|$ represents the cardinality. Our motivation is that while DNNs can be over-fitted on noisy labels with sufficient training epochs, in the early training epochs [40], DNNs tend to first learn on the correct labels. Thus the correct and the incorrect labels can be distinguished based on the loss values [41].

In Fig. 4 (left), we plot the training loss values l^t_v computed after each training epoch for both correct (green) and incorrect (red) labels in $\mathcal{Y}^t_v$. At the early stage of the training, pixels with incorrect labels have larger loss values than the correctly labeled pixels, allowing one to identify the noisy labels from the loss values. In Fig. 4 (right), we show the loss distribution after 20 training epochs for both correctly and incorrectly labeled pixels. We see that the distribution is bimodal and can be modeled as a two-component mixture model.

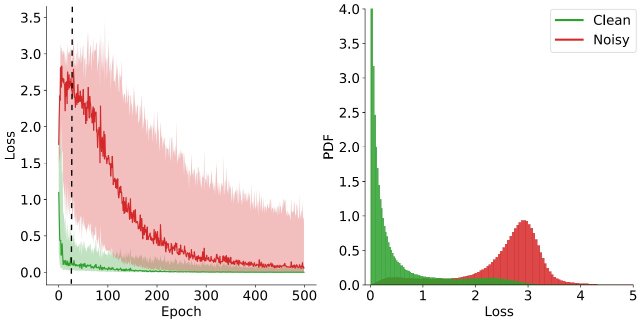


Fig. 4: Left: The binary cross-entropy loss for correct (green) and incorrect (red) vessel labels in the course of training. The curve shows the median loss value and the shaded region represents the range between the 15th and the 85th percentile of the loss values. Right: histogram of the training loss after 20 training epochs (indicated by the dash line in the left plot).

To estimate the distribution of l^t_v, we use the latent variable $Z^t_v \in \{0, 1\}$ to indicate if the pixel v in $\mathcal{Y}^t_v$ is mislabeled. Given that the label is correct ($Z^t_v = 1$), the conditional probability of l^t_v is modeled as an exponential distribution $\lambda\exp(-\lambda l^t_v)$ with parameter λ. See the green distribution in Fig. 4. And, given $Z^t_v = 0$, the conditional probability of l^t_v is modeled as a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ with mean μ and standard deviation σ. See the red distribution in Fig. 4. The distribution of the mixture model for l^t_v takes the form of

$$p\left(l^t_v\right) = \pi\lambda e^{-\lambda l^t_v} + (1-\pi)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(l^t_v-\mu)^2}{2\sigma^2}}, \tag{3}$$

where $\pi = p\left(Z^t_v = 1\right)$ is the mixing weight that represents the prior probability of latent variable Z^t_v. We adopt the EM algorithm [39] to fit the proposed mixture model. EM algorithm alternates between the E-step and the M-step. In the E-step, we compute the posterior probability $p^t_v = p\left(Z^t_v = 1 \mid l^t_v\right)$, which can be obtained using Bayes’ rule:

$$p^t_v = \frac{\pi\lambda e^{-\lambda l^t_v}}{\pi\lambda e^{-\lambda l^t_v} + (1-\pi)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(l^t_v-\mu)^2}{2\sigma^2}}}. \tag{4}$$

In the M-step, we update the parameters of the mixture model. Using the estimated posterior probability, we obtain

$$\begin{aligned} \pi^t &= \frac{\sum_{v\in\mathcal{Y}^t_v} p^t_v}{|\mathcal{Y}^t_v|}, \quad \mu^t = \frac{\sum_{v\in\mathcal{Y}^t_v}\left(1-p^t_v\right)l^t_v}{\sum_{v\in\mathcal{Y}^t_v}\left(1-p^t_v\right)}, \\ \lambda^t &= \frac{\sum_{v\in\mathcal{Y}^t_v} p^t_v}{\sum_{v\in\mathcal{Y}^t_v} p^t_v l^t_v}, \ \sigma^t = \sqrt{\frac{\sum_{v\in\mathcal{Y}^t_v}\left(1-p^t_v\right)\left(l^t_v-u\right)^2}{\sum_{v\in\mathcal{Y}^t_v}\left(1-p^t_v\right)}}. \end{aligned} \tag{5}$$

The process is repeated until parameters converge. The fitted mixture model provides a tool for analyzing the label noise in the warped vessel maps. The prior probability $p\left(Z^t_v = 1\right)$ is an estimate of the amount of correct labels in $\mathcal{Y}_v$. More importantly, the posterior probability $p\left(Z^t_v = 1 \mid l^t_v\right)$ indicates the probability of pixel being correctly labeled, which allows us to update labels in $\mathcal{Y}^t_v$. Specifically, the label in updated ground truth $\boldsymbol{Y}^{i,t}_u$ is computed as

$$y^t_{u,v} = \begin{cases} p^t_v y^t_{a\to c,v} + \left(1-p^t_v\right) y^t_{c,v}, & \text{if } v \in \mathcal{Y}^i_v, \qquad (6a)\\ 0, & \text{if } v \in \mathcal{Y}^i_b. \qquad (6b) \end{cases}$$

In (6a), the updated label for pixel in $\mathcal{Y}^t_v$ is a linear combination of the label in the warped vessel map $y^t_{a\to c,v}$ and the predicted probability vessel map $y^t_{c,v}$ where the coefficients are determined by the posterior probability p^t_v. Intuitively, if the posterior probability p^t_v is close to 1, we trust the label $y^t_{a\to c,v}$ in the warped vessel map because the corresponding pixel is correctly labeled. Otherwise, we reduce the weights of label $y^t_{a\to c,v}$ and rely more on the network-predicted probability vessel maps $y^t_{c,v}$. In (6b), we do not update background labels in $\mathcal{Y}^t_b$ because these labels are considered accurate.

The training process is divided into two stages. First, we train the DNN on the tentative noisy dataset $\{(\boldsymbol{X}^i_c, \boldsymbol{Y}^{i,t}_{a\to c})\}^M_{i=1}$ for E_0 epochs. Then we fit the proposed mixture model on the loss values l^t_v and obtain the updated labels $\boldsymbol{Y}^t_u$. In the second stage, we continue to train the DNN on $\{(\boldsymbol{X}^i_c, \boldsymbol{Y}^{i,t}_u)\}^M_{i=1}$ for another E_1 epochs. The overall algorithm for the proposed framework is summarized in Algorithm 1.

Note that both the vessel registration and the robust learning steps utilize the EM framework to estimate the posterior probabilities of a pixel being outlier/mislabeled. However, the objectives in these two steps are different and we can not use the posterior probabilities estimated in one step for the other. In the registration step, the EM framework mitigates the effects of outlier vessel points. The outliers are defined as the vessel points in $\boldsymbol{Y}^i_a$ that do not have correspondences in the current estimated vessel map $\boldsymbol{Y}^{i,t}_c$. As we show in Section III-D, some vessels are not properly detected in $\boldsymbol{Y}^{i,t}_c$ in the first few iterations. As a result, the outlier pixels in the registration step are not necessarily the same as the mislabeled pixels that need to be identified in the training step.

III. Experiments

In this section, we first introduce a new dataset, PRIME-FP20, that is used for implementing the proposed iterative framework and for evaluating the vessel detection performance. Next, we summarize the evaluation metrics in Section III-B, and describe the implementation details and alternative methods used as baselines in Section III-C. The experimental results are structured as follows. We provide detailed analysis to demonstrate the effectiveness of the proposed iterative framework and the weakly-supervised learning method in Section III-D and Section III-D, respectively. We then compare the proposed framework with alternative methods on the PRIME-FP20 dataset in Section III-F. Finally, we show the boarder utility of the proposed framework for detecting vessels in NF FP in Section III-G.

A. PRIME-FP20 Dataset

We construct a new dataset, PRIME-FP20, for evaluating the performance of vessel detection in UWF FP. The PRIME-FP20 dataset consists of 15 pairs of concurrently captured UWF FP and UWF FA images that are selected from baseline images of patients enrolled in the PRIME study[1]. The images are captured using Optos California and 200Tx cameras (Optos plc, Dunfermline, United Kingdom) [42]. The system uses a scanning ophthamoscope with a low power laser to capture dual red and green channel UWF FP images and a single channel FA image. All images have the same resolution of 4000 × 4000 pixels and are stored as 8-bit TIFF format with lossless LZW compression. The green channel UWF FP image is used as the input $\boldsymbol{X}_c$ for our vessel detection because it captures information for layers with the retinal vasculature, whereas the red channel captures information from other layers (from the retinal pigment epithileum to the choroid) [42]. For evaluation, ground truth vessel maps for the UWF FP modality are manually labeled by a human annotator using the ImageJ software [43] with the segmentation editor plugins.

Algorithm 1: Iterative training of DNN for vessel detection in UWF FP without *de novo* labeled data

```
Given : DNN architecture that outputs a probabilistic
        FP vessel map $f(\boldsymbol{X}_c; \boldsymbol{W})$, where $\boldsymbol{X}_c$ is an
        input FP image and $\boldsymbol{W}$ are the weights for
        the network
Input : UWF FP image $\boldsymbol{X}^i_c$, UWF FA vessel map
        $\boldsymbol{Y}^i_a$, number of iterations $T$, training epochs
        $E_0$ and $E_1$
Output: Trained DNN weights $\boldsymbol{W}^*$
Initialization:
$t = 0$ ;
Detect preliminary FP vessel map $\boldsymbol{Y}^{i,0}_c$;
Extract vessel pixel coordinates $\mathcal{Q}^i_a$ from $\boldsymbol{Y}^i_a$ ;
repeat /*registration and learning iterations*/
    Vessel registration and Warping
    for $i = 1 : M$ do
        Extract vessel coordinates $\mathcal{Q}^{i,t}_c$ from $\boldsymbol{Y}^{i,t}_c$;
        Estimate second-order transformation $\mathcal{T}^i_{\boldsymbol{\beta}^*}$ from
          $\mathcal{Q}^i_a$ to $\mathcal{Q}^{i,t}_c$. See Sect. II-A and [36] for details;
        Warp FA vessel map to FP: $\boldsymbol{Y}^{i,t}_{a\to c} \leftarrow \mathcal{T}^i_{\boldsymbol{\beta}^*}(\boldsymbol{Y}^i_a)$;
    end
    Learn using $\boldsymbol{Y}^{i,t}_{a\to c}$ as noisy labels
    Obtain weights $\boldsymbol{W}^t$ by training DNN $f(\cdot;\cdot)$ for $E_0$
      epochs on $\{(\boldsymbol{X}^i_c, \boldsymbol{Y}^{i,t}_{a\to c})\}^M_{i=1}$;
    Compute loss values $l^t_v$ for pixels in $\mathcal{Y}^t_v$ using (2);
    repeat /*EM for label noise mixture model*/
        Compute posterior probabilities $p^t_v$ using (4);
        Update parameters $\pi^t$, $\lambda^t$, $\mu^t$, and $\sigma^t$ using (5);
    until Parameter converge;
    Update labels $\boldsymbol{Y}^{i,t}_u$ using (6a) and (6b);
    Fine-tune weights $\boldsymbol{W}^t$ by training DNN $f(\cdot;\cdot)$ for
      $E_1$ epochs on $\{(\boldsymbol{X}^i_c, \boldsymbol{Y}^{i,t}_u)\}^M_{i=1}$ ;
    Update
    $t \leftarrow t + 1$; $\boldsymbol{Y}^{i,t}_c \leftarrow f(\boldsymbol{X}^i_c; \boldsymbol{W}^t)$;
until $(t = T)$;
$\boldsymbol{W}^* \leftarrow \boldsymbol{W}^T$
```

[1] The study (ClinicalTrials.gov Identifier: NCT03531294) evaluates the impact of intravitreal aflibercept in diabetic retinopathy patients with a baseline diabetic retinopathy severity score level of 47A to 71A inclusive.

The available selection tools in ImageJ, such as brush tool and free-hand selection tool were used to mark the vessel pixels in the UWF FP. The annotator repeatedly adjusted image brightness and contrast to precisely label both major and minor vessel branches in different regions. For each UWF FP, we also provide a binary mask for the FOV of the image. To obtain the mask, we simply binarize the green channel of the UWF FP because the pixels intensities out of FOV are close to zero.

B. Evaluation Metrics

For quantitative evaluation, we report the area under the Precision-Recall curve (AUC PR)[2], the Dice coefficient (DC), and the CAL metric [44]. The computation of these metrics is summarized in Section S.III of the Supplementary Material.

The AUC PR and the Dice coefficient, although widely used in prior literature, are based on the pixel-wise comparison of the ground truth and the estimated vessel map. However, the pixel-wise comparison does not consider the structure of retinal vasculature and is sensitive to the label ambiguities, particularly for peripheral pixels that only partially belong to vessels. The CAL metric [44] is designed to be less sensitive to label uncertainties and provides better agreement with human assessment of higher level structure. CAL evaluates the consistency between the binary ground truth and the binary predicted vessel map by calculating three individual factors that quantify the consistency with respect to the connectivity (C), the area (A), and the corresponding length of skeletons (L). Each factor ranges between 0 and 1 where 1 indicates perfect consistency to the ground truth. The product of three factors is defined as the overall CAL metrics. The computation of the CAL metrics requires a binary vessel map, to obtain which, we binarize the predicted probabilistic vessel map $\boldsymbol{Y}_p$ with a threshold $\tau = 0.5$.

For the experiments on the PRIME-FP20 dataset, we perform the K-fold cross-validation [45] to evaluate the performance of vessel detection, where K is set to 5, and report the statistics of the five evaluation metrics. We only consider pixels within the FOV mask when computing the metrics.

C. Implementation Details and Alternative Methods

To detect UWF FA vessels $\boldsymbol{X}_a$, we train the U-Net [13] model on the RECOVERY-FA19 dataset [36] that provides eight high-resolution (3900 × 3072 pixels) UWF FA images and the ground truth vessel maps. We use the U-Net model because of its superior performance in medical image segmentation [46]. Detailed training protocol is included in the Supplementary Material (Section S.II-B). We apply the trained model to the UWF FA images $\boldsymbol{X}_a^i$ and binarize the estimated vessel map $\boldsymbol{Y}_a^i$ with a threshold $\tau = 0.5$.

For the proposed iterative framework, we use the pairs of UWF FP and UWF FA images in the PRIME-FP20 dataset. Note that the proposed framework does not require the ground truth vessel maps for UWF FP in the PRIME-FP20 dataset. *These manually labeled ground truth are only used for evaluation in our experiments*. We implement the chamfer alignment and the weakly-supervised learning using MATLAB$^{\text{TM}}$ and PyTorch [47], respectively. We perform three iterations between registration and learning ($T = 3$) and provide an empirical evaluation of different number of iterations in Section III-D. In the first iteration, we use a preliminary UWF FP vessel map $\boldsymbol{Y}_c^{i,0}$ for chamfer alignment, which is obtained from a DNN pre-trained on existing NF FP dataset. For the weakly-supervised learning step, we use the U-Net [13] model. We set the training epochs $E_0 = 25$ and $E_1 = 30$. Detailed network architectures and training protocol are included in Section S.II of the Supplementary Material.

We consider existing learning-based vessel detection methods for as baselines for comparison. These methods include HED [48], U-Net [13], DRIU [11], CRF [49],NestUNet [14], M2U-Net [50], CE-Net [18], CS-Net [51], RU-Net [15], and IterNet [16]. We train all methods on the IOSTAR [26] dataset where the images are captured with the scanning laser ophthalmoscopy (SLO) technique that is also used in the PRIME-FP20 dataset. Our experiments show that these baseline methods trained on the IOSTAR achieve the best generalization performance on the PRIME-FP20 dataset.

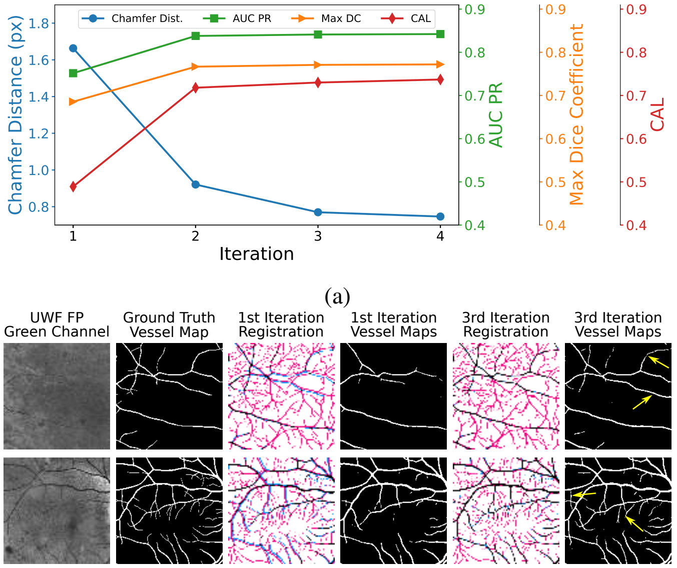
D. Iterative Registration and Learning Framework

We demonstrate the effectiveness of the proposed iterative framework by showing that both registration and learning benefit from each other and improve progressively.

To quantify registration accuracy, we compute the chamfer distance as the average Euclidean distance between each point in the ground truth binary vessel maps and its closest point in the transformed vessel maps detected in UWF FA. The average chamfer distance under the second-order transformation can be treated as a proxy for the registration error. The blue line with circle markers in Fig. 5a shows the average chamfer distance over 4 iterations. In the first iteration, the chamfer distance is on average 1.66 pixels. While the misalignment is slight, it can significantly deteriorate the quality of the training data. The generated tentative ground truth in the first iteration only has a recall of 0.63, which means that 37.0% of true vessels are labeled as background (false negative labels). The third column in Fig. 5b shows sample results of the generated ground truth in the first iteration, where the blue pixels highlight the false negative labels. In the third iteration, the chamfer distance drops to 0.77 pixels, yielding accurate training data with a recall increased to 0.83. The fifth column in Fig. 5b shows training data obtained from the third iteration.

The improved ground truth dataset in turn benefits network training for vessels detection in UWF FP. The fourth and the last columns in Fig. 5b show the predicted vessel maps in the first and the third iteration, respectively. The yellow arrows highlight the improved vessel detections that are not correctly identified in the first iteration. We quantify and visualize the performance of vessel detection obtained over 4 iterations in Fig. 5a. The axes on the right side correspond to the three metrics used for evaluation. It is clear that, as the registration and training proceed, the performance of vessel detection is improved progressively. Additionally, we see that the DNN

[2] We do not choose the Receiver Operating Characteristic (ROC) curve as the evaluation metric because the ground truth label is highly skewed. We provide additional discussion in Supplementary Material (Section S.III).



(a)

(b)

Fig. 5: Registration and vessel detection performance as a function of iteration count. (a) the residual chamfer distance, which serve as a good proxy for the registration error, is labeled on the left axis and shown in a corresponding plot. The three axes labeled on the right side and corresponding plots show the metrics used for evaluating the vessel detection performance. (b) Sample vessel maps obtained in the first and the third iterations. Red and blue pixels indicate incorrect vessel labels ("false positive") and background labels ("false negative"), respectively, in the warped UWF FA vessel map.

performance becomes stable after three iterations and going to the fourth iteration offers limited improvement. Thus, we set the total number of iterations T to 3 in our experiments.

E. Robust Learning with Noisy Labels

We conduct detailed analysis for a better understanding of the proposed method for robust learning from noisy labels. Because we focus on the robust learning method, all experimental results reported in this section are performed on the noisy training data that is generated from the last iteration in the proposed framework.

To justify the effectiveness of the proposed robust learning method, we compare the performance of vessel detection with the following alternative training strategies: (1) the standard training approach that directly trains a DNN without any techniques particularly attuned to noisy labels, (2) the re-labeling method that dynamically updates the labels in the training dataset [52], and (3) a re-weighting method that reduces the weight for the noisy labels in the loss function. The re-labeling method seeks to obtain a clean dataset by dynamically updating the training labels using the probabilistic vessel maps predicted from the DNN. The training process is formulated as a joint framework that alternatively optimizes the DNN parameters and the training labels. For the re-weighting method, the idea is to adaptively assign small weights to the potential noisy pixels and to emphasize the clean pixels in the loss function. Specifically, we assign the posterior probability p_v as the weighting factor to each pixel in $\mathcal{Y}_v$ and set the weights to 1 for all pixels in $\mathcal{Y}_b$.

Methods	AUC PR	Max DC	CAL (C, A, L)
Direct Training	0.802	0.745	0.628 (0.998, 0.777, 0.809)
Re-labeling [52]	0.837	0.769	0.586 (0.999, 0.729, 0.805)
Re-weighting	0.842	0.768	0.713 (0.999, 0.833, 0.856)
Proposed	**0.842**	**0.772**	**0.730** (**0.999**, **0.849**, **0.860**)

TABLE I: Accuracy metrics for vessel detection results obtained with alternative training strategies. All DNNs are trained on the dataset obtained from the third iteration in the proposed framework. The best result is shown in bold.

The quantitative results obtained from different training methods are listed in Table I. Directly training on the incorrect labels adversely impacts the performance of vessel detection, even though we apply early stopping to prevent the DNN from over-fitting the noisy labels. In addition, it is difficult to determine the stopping criterion because no validation dataset is available in this settings. The re-weighting and the proposed approaches, both of which utilize the posterior probabilities p_v to train DNNs, show significant improvement over the direct training and the re-labeling methods. This also demonstrate the effectiveness of the proposed mixture-model-based noisy label identification. Unlike the re-weighting method, which uses p_v to reduce the effects of incorrect labels, the proposed robust training approach updates the noisy labels and therefore explicitly forces DNN to learn on the correct prediction.

Next, we assess the effects of different mixture models on fitting the loss distribution and estimating the posterior probabilities p_v. Specifically, we compare the proposed mixture model with a two-component Gaussian mixture model (GMM) and a two-component beta mixture model (BMM) [41]. A proper mixture model, which provide a good approximation to the loss distribution, should lead to an accurate estimation of the posterior probability p_v and an accurate update on the training labels $\boldsymbol{Y}_u$. Thus, we compare the quality of the updated labels with respect to the manually labeled ground truth. To do so, we fit the mixture models on the same loss distribution and update the labels using (6a) and (6b). Figure 6 plots the AUC PR obtained after each training epoch for different mixture models. We have several observations from this figure. First, the GMM is not a good approximation for the loss distribution and the accuracy of noisy label correction decreases as the training proceeds and is significantly worse than other two mixture models. Second, compared to the BMM, the proposed mixture model provides the more accurate results and the performance is largely stable in the first 70 training epochs. In Fig. 7, we show the sample results of updated labels, the corresponding noisy labels from the warped vessel maps, and the manually labeled ground truth. The "false positive" labels are removed from the warped vessel maps, highlighted by the yellow arrows in Fig. 7, yielding to updated labels that is similar to the ground truth labels.

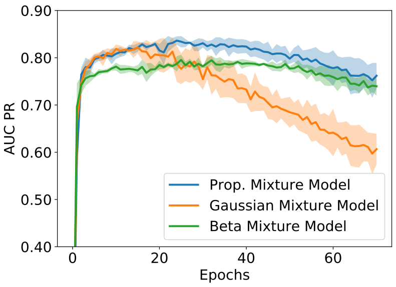


Fig. 6: AUC PR obtained with alternative mixture models for modeling the loss distribution as a function of training epochs. The curves show the average AUC PR values over 5-fold cross-validation, and the shaded region represents the one standard deviation from the mean AUC PR values.

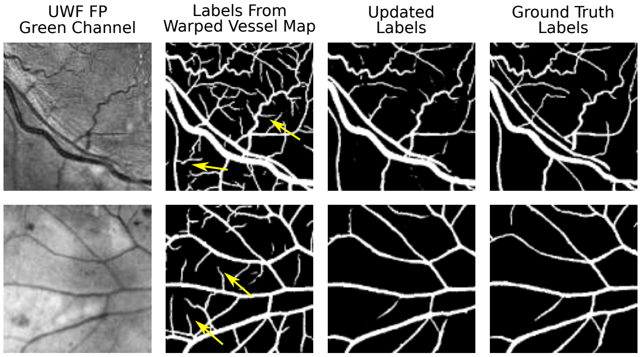


Fig. 7: Sample images and vessel maps illustrating noisy label correction in the proposed framework.

F. Evaluation on the PRIME-FP20 Dataset

As mentioned in Section III-B, we perform 5-fold cross-validation to assess the results of vessel detection on the PRIME-FP20 dataset. Table II lists the quantitative results obtained from the proposed iterative framework and the existing methods. The proposed iterative framework performs remarkably well and significantly outperforms other methods with respect to all evaluation metrics, achieving an AUC PR of 0.845, the maximum Dice coefficient of 0.776, and an overall CAL of 0.730. Notably, the performance metrics for the proposed framework are quite close to the annotation-intensive approach, where a U-Net model is trained on *manually labeled clean dataset* with the same 5-fold cross-validation (The row labeled U-Net* in Table II). We show sample results of the detected vessel maps obtained from different methods in Fig. 8 and provide more visual results in the Section S.IV of the Supplementary Material. In Fig. 8, we see that the existing DNNs trained on NF fundus images perform poorly in the peripheral region. We attribute this poor performance to the fact that the peripheral region contains artifacts that are not visible in the NF dataset. Such artifacts normally have dark and curvilinear structures that can be misinterpreted as vessels in the image. For example, the yellow arrows in the enlarged view of region III highlight the "false positive" detection region that is not a vessel but an eyelash shadow appearing in the periphery. Compared to the DNNs trained on NF images, the proposed iterative framework accurately detects vessel maps from different regions in UWF FP. See the enlarged view of regions III and IV for the result patches selected from the periphery and the central retina, respectively.

Methods	Year	AUC PR	Max DC	CAL (C, A, L)
U-Net* [13]	-	0.869	0.796	0.755 (0.999, 0.869, 0.870)
HED [48]	2015	0.723	0.683	0.451 (0.997, 0.640, 0.700)
U-Net [13]	2015	0.746	0.704	0.547 (0.998, 0.727, 0.752)
DRIU [11]	2016	0.728	0.691	0.495 (0.999, 0.698, 0.705)
CRF [49]	2017	0.563	0.550	0.341 (0.994, 0.577, 0.590)
NestUNet [14]	2018	0.754	0.716	0.567 (0.999, 0.747, 0.757)
M2U-Net [50]	2019	0.727	0.694	0.534 (0.998, 0.720, 0.742)
CE-Net [18]	2019	0.757	0.718	0.574 (0.999, 0.752, 0.762)
CS-Net [51]	2019	0.772	0.721	0.565 (0.998, 0.746, 0.755)
RU-Net [15]	2019	0.757	0.707	0.559 (0.999, 0.737, 0.758)
IterNet [16]	2020	0.746	0.717	0.553 (0.999, 0.732, 0.753)
Proposed	2020	**0.841**	**0.771**	**0.730** (**0.999**, **0.849**, **0.860**)

TABLE II: Quantitative metrics assessing vessel detection accuracy for different methods on the PRIME-FP20 dataset. The row U-Net* lists the results from a U-Net trained on a manually labeled dataset. The best result is shown in bold.

We also notice that, under the precise registration, the warped vessel maps with noisy labels are still valuable for training DNNs. Comparing the results listed in Tables I and II, the direct training approach already has a better performance than the existing methods trained on NF fundus images. These results further reinforce the benefits of the transfer approach for generating training data for UWF FP modality.

G. Evaluation on Narrow-Field Fundus Photography

Fundus photography shares common characteristic between the ultra-widefield and the narrow-field modalities. In this section, we demonstrate that the DNN trained only on ultra-widefield images using the proposed framework is capable of detecting vessels in NF FP. To this end, we test the performance of the trained DNN on two public datasets, DRIVE [24] and STARE [4], and compare with the existing learning-based methods for vessel detection. Note that we train the DNN on ultra-widefield images using the proposed weakly-supervised learning approach and evaluate the performance on the NF images. We refer to this experiment as the cross-training evaluation [12], [53] where the training and the test data come from two independent sources. For existing learning-based methods, the models are trained on the DRIVE [24] and evaluated on the STARE [4], and vice versa. These two datasets provide two independent ground truth vessel maps manually labeled by two human annotators. We choose the vessel maps from the first annotator as the ground truth also report the human performance by evaluating the vessel maps made by the second annotator, which is commonly accepted approach in the literature.

Complete results are listed in Table S.1 in the Supplementary Material. On the DRIVE dataset, the proposed framework achieves the best performance with the AUC PR of 0.886, the maximum DC of 0.803, and the overall CAL metric of 0.827. Note that the CAL metric is significantly better than those obtained from prior alternatives by large margins and is close to human performance (0.839). The second-best performing method, HED [48], achieves an overall CAL of 0.743. The

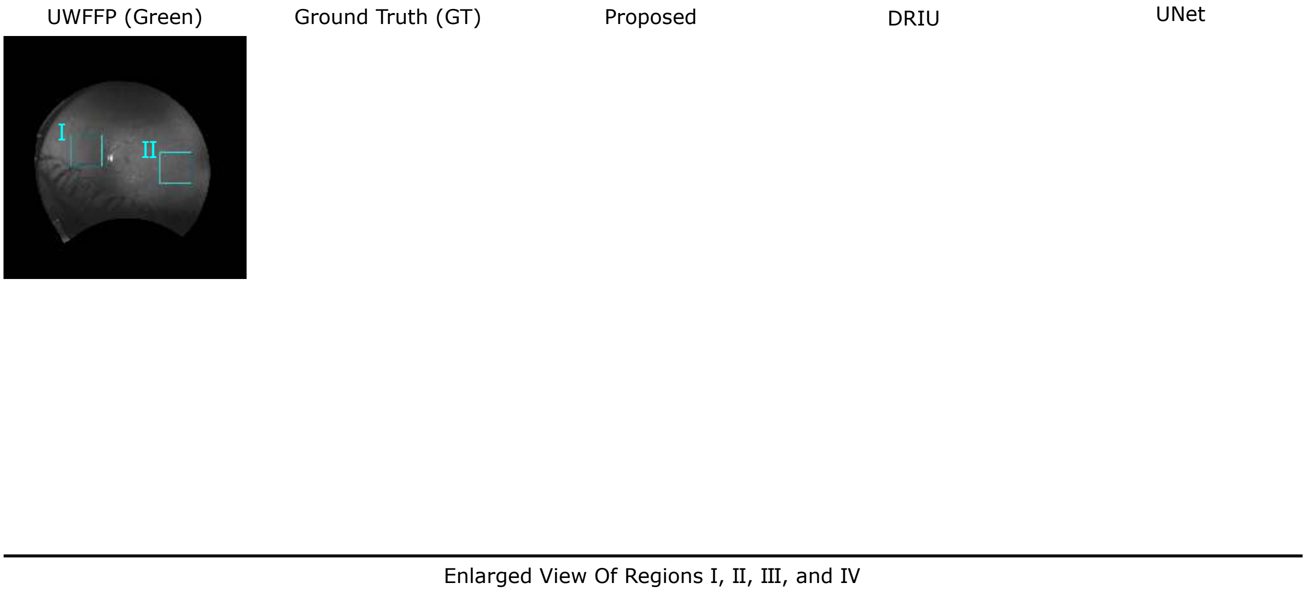


Fig. 8: Sample images and detected vessel maps for the proposed approach and alternatives from the PRIME-FP20 dataset. The contrast-enhanced enlarged views I-IV, marked by the cyan rectangles in the full image, are included. Additional visual results are provided in Section S.IV of the Supplementary Material.

performance on the STARE dataset, while slightly worse than the best performing method, is comparable to other methods. Specifically, the results obtained from the proposed framework has the AUC PR of 0.884, the maximum DC of 0.795, and the overall CAL metric of 0.756. The results on both datasets reinforce the robustness and the accuracy of the proposed iterative framework. We provide visual results of detected vessel maps in Section S.V of the Supplementary Material.

IV. Conclusion

The iterative registration and deep-learning framework proposed in this paper provides an effective and annotation-efficient approach for detecting retinal blood vessels in UWF FP imagery without requiring manually labeled UWF FP vessel maps. Experimental evaluations demonstrate that the proposed approach significantly outperforms the existing methods on a new UWF FP dataset, PRIME-FP20, and achieves comparable performance with the state-of-the-arts on existing NF FP datasets. The PRIME-FP20 is made publicly available [54][3]. to facilitate further work on retinal image analysis.

Acknowledgments

We thank the Center for Integrated Research Computing, University of Rochester, for providing access to computational resources, and Eric Kegley and Cary Stoever for acquisition of PRIME images. The PRIME study is an investigator-initiated trial supported by Regeneron (Tarrytown, New York) who had no role in the funding, design, or conduct of the development of the methodology presented in this paper.

References

[1] S. Rogers *et al.*, “The prevalence of retinal vein occlusion: pooled data from population studies from the United States, Europe, Asia, and Australia,” *Ophthalmology*, vol. 117, no. 2, pp. 313–319, 2010.

[2] C. L. Srinidhi, P. Aparna, and J. Rajan, “Recent advancements in retinal vessel segmentation,” *J. Med. Syst.*, vol. 41, p. 70, 2017.

[3] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, “Detection of blood vessels in retinal images using two-dimensional matched filters,” *IEEE Trans. Med. Imaging*, vol. 8, no. 3, pp. 263–269, Sep 1989.

[4] A. Hoover, V. Kouznetsova, and M. Goldbaum, “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response,” *IEEE Trans. Med. Imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[5] F. Zana and J. C. Klein, “Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation,” *IEEE Trans. Image Proc.*, vol. 10, no. 7, pp. 1010–1019, Jul 2001.

[6] L. Ding, A. Kuriyan, R. Ramchandran, and G. Sharma, “Multi-scale morphological analysis for retinal vessel detection in wide-field fluorescein angiography,” in *Proc. IEEE Western NY Image and Signal Proc. Wksp. (WNYISPW)*, Rochester, NY, Nov. 2017, pp. 1–5.

[3] A sample low resolution annotated image is currently provided and the full set of 15 high resolution images will be made available with the publication of the paper.

[7] A. M. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE Trans. Med. Imaging*, vol. 25, no. 9, pp. 1200–1213, Sept 2006.

[8] H. Yu, S. Barriga, C. Agurto, G. Zamora, W. Bauman, and P. Soliz, "Fast vessel segmentation in retinal images using multi-scale enhancement and second-order local entropy," in *SPIE Medical Imaging*, vol. 8315, 2012, p. 83151B.

[9] Z. Fan, J. Lu, C. Wei, H. Huang, X. Cai, and X. Chen, "A hierarchical image matting model for blood vessel segmentation in fundus images," *IEEE Trans. Image Proc.*, vol. 28, no. 5, pp. 2367–2377, May 2019.

[10] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imaging*, vol. 35, no. 11, pp. 2369–2380, Nov 2016.

[11] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *Intl. Conf. Med. Image Computing and Computer-Assisted Intervention*, 2016, pp. 140–148.

[12] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 109–118, 2016.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Intl. Conf. Med. Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[14] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Med. Image Analysis*, 2018, pp. 3–11.

[15] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imaging*, vol. 6, no. 1, pp. 1 – 16, 2019.

[16] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *IEEE Winter Conf. Applications of Comp. Vision*, 2020, to appear.

[17] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep vessel segmentation by learning graphical connectivity," *Med. Image Analysis*, vol. 58, p. 101556, 2019.

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

[19] J. Son, S. J. Park, and K.-H. Jung, "Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks," *J. Digital Imaging*, vol. 32, no. 3, pp. 499–512, Jun 2019.

[20] V. Cherukuri, V. Kumar B G, R. Bala, and V. Monga, "Deep retinal image segmentation with regularization under geometric priors," *IEEE Trans. Image Proc.*, vol. 29, pp. 2552–2567, 2020.

[21] A. Mosinska, P. Márquez-Neila, M. Koziński, and P. Fua, "Beyond the pixel-wise loss for topology-aware delineation," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2018, pp. 3136–3145.

[22] 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*, 2019, to appear.

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

[24] J. Staal, M. Abràmoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

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

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

[27] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *Intl. J. of Biomed. imaging*, vol. 2013, 2013.

[28] P. Bankhead, C. N. Scholfield, J. G. McGeown, and T. M. Curtis, "Fast retinal vessel detection and measurement using wavelets and edge location refinement," *PLOS ONE*, vol. 7, no. 3, pp. 1–12, 03 2012.

[29] E. Decencière *et al.*, "Feedback on a publicly distributed image database: the Messidor database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014.

[30] M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010.

[31] A. Nagiel, R. A. Lalane, S. R. Sadda, and S. D. Schwartz, "Ultra-widefield fundus imaging: A review of clinical applications and future trends," *RETINA*, vol. 36, no. 4, 2016.

[32] P. S. Silva, J. D. Cavallerano, J. K. Sun, J. Noble, L. M. Aiello, and L. P. Aiello, "Nonmydriatic ultrawide field retinal imaging compared with dilated standard 7-field 35-mm photography and retinal specialist examination for evaluation of diabetic retinopathy," *Amer. J. Ophthalmology*, vol. 154, no. 3, pp. 549 – 559.e2, 2012.

[33] P. S. Silva, J. D. Cavallerano, J. K. Sun, A. Z. Soliman, L. M. Aiello, and L. P. Aiello, "Peripheral lesions identified by mydriatic ultrawide field imaging: Distribution and potential impact on diabetic retinopathy severity," *Ophthalmology*, vol. 120, no. 12, pp. 2587 – 2595, 2013.

[34] E. Pellegrini *et al.*, "Blood vessel segmentation and width estimation in ultra-wide field scanning laser ophthalmoscopy," *Biomed. Opt. Express*, vol. 5, no. 12, pp. 4329–4337, Dec 2014.

[35] J. Zhang *et al.*, "Joint vessel segmentation and deformable registration on multi-modal retinal images based on style transfer," in *IEEE Intl. Conf. Image Proc.*, Sep. 2019, pp. 839–843.

[36] 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 Trans. Image Proc.*, vol. 29, no. 1, 2020, accepted for publication, to appear.

[37] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proc. Int. Joint Conf. Artificial Intell.*, 1977, pp. 659–663.

[38] Y. Gavet, M. Fernandes, and J.-C. Pinoli, "Quantitative evaluation of image registration techniques in the case of retinal images," *J. Electronic Imaging*, vol. 21, no. 2, pp. 1 – 8, 2012.

[39] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, pp. 1–38, 1977.

[40] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," in *Intl. Conf. Learning Representations*, 2017.

[41] E. Arazo, D. Ortego, P. Albert, N. O'Connor, and K. McGuinness, "Unsupervised label noise modeling and loss correction," in *Intl. Conf. on Mach. Learning*, vol. 97, 2019, pp. 312–321.

[42] *Optos California Tech Sheet*, Optos, 2015. [Online]. Available: https://www.optos.com/globalassets/www.optos.com/products/california/california-brochure.pdf

[43] J. Schindelin *et al.*, "Fiji: an open-source platform for biological-image analysis," *Nature methods*, vol. 9, no. 7, p. 676, 2012.

[44] M. E. Gegundez-Arias, A. Aquino, J. M. Bravo, and D. Marin, "A function for quality evaluation of retinal vessel segmentations," *IEEE Trans. Med. Imaging*, vol. 31, no. 2, pp. 231–239, Feb 2012.

[45] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY: Springer-Verlag, 2009.

[46] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Analysis*, vol. 42, pp. 60–88, 2017.

[47] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Adv. in Neural Info. Proc. Sys.*, 2019, pp. 8024–8035.

[48] S. Xie and Z. Tu, "Holistically-nested edge detection," in *IEEE Intl. Conf. Comp. Vision.*, Dec. 2015, pp. 1395–1403.

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

[50] T. Laibacher, T. Weyde, and S. Jalali, "M2U-Net: Effective and efficient retinal vessel segmentation for real-world applications," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog. Wksp.*, June 2019.

[51] L. Mou *et al.*, "CS-Net: Channel and spatial attention network for curvilinear structure segmentation," in *Intl. Conf. Med. Image Computing and Computer-Assisted Intervention*, 2019, pp. 721–730.

[52] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa, "Joint optimization framework for learning with noisy labels," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, 2018, pp. 5552–5560.

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

[54] L. Ding, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "PRIME-FP20: Ultra-widefield fundus photography vessel segmentation dataset," IEEE Dataport, 2020. [Online]. Available: https://doi.org/10.21227/ctgj-1367

Supplementary Material for "Weakly-Supervised Vessel Detection in Ultra-Widefield Fundus Photography Via Iterative Multi-Modal Registration and Learning"

Li Ding, *Student Member, IEEE*, Ajay E. Kuriyan, Rajeev S. Ramchandran, Charles C. Wykoff, and Gaurav Sharma, *Fellow, IEEE*

S.I. Overview

This document provides Supplementary Material for the paper [1]. Section S.II provides implementation details, including network architectures and training protocol. Section S.III provides a summary of the evaluation metrics. In Section S.IV, we show additional visual results of vessel detection on the PRIME-FP20 dataset. Finally, we include the complete results on the narrow-field fundus photography.

S.II. Implementation Details

A. Network Architectures

In the proposed framework, we adopt the U-Net [2] model that is an encoder-decoder architecture with skip connection. The encoder architecture is

$$C_1^e(64) \text{ - } C_2^e(128) \text{ - } C_3^e(256) \text{ - } C_4^e(512) \text{ - } C_5^e(512),$$

where $C_i^e(n)$ denotes the i-th layer in the encoder, which consists two consecutive convolutional layers followed by a max-pooling layer. The decoder architecture is

$$C_4^d(256) \text{ - } C_3^d(128) \text{ - } C_2^d(64) \text{ - } C_1^d(64) \text{ -} C_{out}(1),$$

where $C_i^d(n)$ denotes the i-th layer in the decoder that has the skip connection to the layer C_i^e in the encoder, and $C_{out}(1)$ is the output convolutional layer that returns the probabilistic vessel maps. The convolutional layers $C_i(n)$ have 3×3 kernel size, n output channels, and ReLU activation. The output layer $C_{out}(1)$ uses a 1×1 kernel and sigmoid activation.

B. Training Protocol

The input to the U-Net are 256×256 patches extracted from the training images with a stride of 128. Patches that are not completely in the FOV masks are not included. Data augmentation techniques are applied to enlarge the size of training data. To do so, we randomly apply a sequence of transformations to image patches, including (1) rotation with an angle randomly selected between $-90°$ and $90°$, (2) horizontal and vertical flip, (3) blurring with Gaussian filter, and (4) contrast and brightness adjustment. We use Adam optimizer [3] with a fixed learning rate of 0.0001. The parameters that are used for calculating the gradient averages and its square are set to 0.9 and 0.999, respectively. We shuffle the training dataset in each epoch and set the batch to 16. The network is trained on a NVidia Tesla V100 GPU.

S.III. Description of Evaluation Metrics

We report three metrics to quantify the performance of vessel detection, i.e., the area under the Precision-Recall curve (AUC PR), the Dice coefficient (DC), and the CAL metric [4]. The PR curve is plotted as the precision versus the recall obtained by binarizing the predicted vessel map with thresholds τ ranging 0 to 1. Precision, recall, and DC are computed as

$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN}, \quad DC = \frac{2TP}{2TP+FP+FN},$$

where TP, FP, and FN are true positive, false positive, and false negative, respectively.

L. Ding and G. Sharma are with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627-0231, USA (e-mail: {l.ding, gaurav.sharma}@rochester.edu).

A. E. Kuriyan is with Retina Service, Wills Eye Hospital, Philadelphia, PA 19107-5109 & the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642-0001, USA (e-mail: ajay.kuriyan@gmail.com).

R. S. Ramchandran is with the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642-0001, USA (e-mail: rajeev_ramchandran@urmc.rochester.edu).

C. C. Wykoff is with Retina Consultants of Houston and Blanton Eye Institute, Houston Methodist Hospital & Weill Cornell Medical College, Houston, TX 77030-2700, USA (e-mail: ccwmd@houstonretina.com).

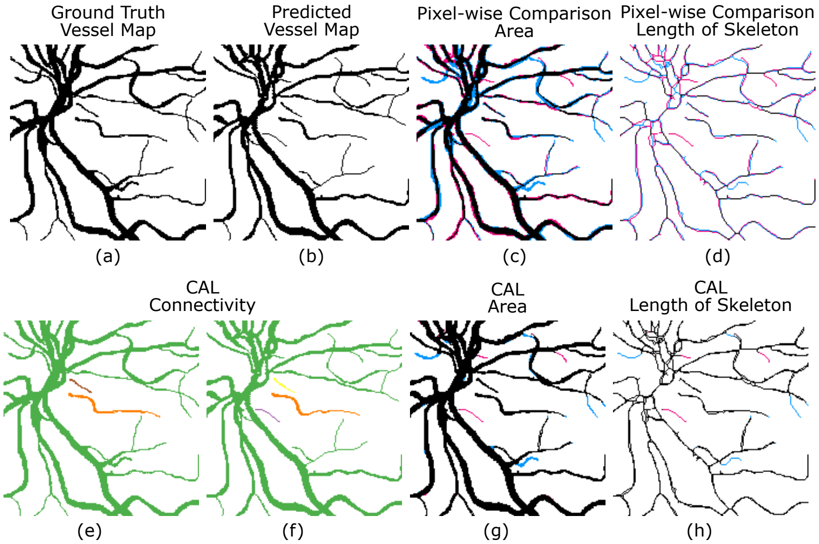


Fig. S.1: Schematic illustration of the pixel-wise based metrics and the CAL metric [4]. (a) and (b) show sample patches of the ground truth label and the binary predicted vessel map, respectively. (c) shows the pixel-wise comparison of the overlapping area, where true positive, false positive, and false negative are highlighted in black, red, and blue, respectively. (d) shows the pixel-wise comparison of the length of skeleton, which are obtained from the corresponding binary vessel maps. Figures in (e) - (h) illustrate the CAL metric that consists of three individual factors: the connectivity, the area, and the length of skeleton. (e) and (f) visualize the connected vessel segments in (a) and (b), respectively. (g) and (h) show the comparisons used in the CAL area factor and CAL length factor computations, respectively, demonstrating the resilience of these factors to differences in labeling of ambiguous pixels on vessel peripheries and slight displacements between vessel skeletons.

We do not choose the Receiver Operating Characteristic (ROC) curve, which is a plot of the true positive rate against the false positive rate, as the evaluation metric. For assessing the performance of vessel detection, the ROC curve is not informative because the ground truth labels are highly skewed where the majority is the negative labels (background pixels in the UWF FP). In this setting, the false positive rate, computed as the ratio between the number of false positive detection to the total number of negative labels, is dominated by the negative labels. As noted in [5], the PR curve is more preferable than the ROC curve when the dataset contains highly imbalanced labels.

Although the AUC PR and the DC are commonly reported in prior works, these metrics are based on pixel-wise comparison of the labeled ground truth and the predicted vessel map. However, as shown in Fig. S.1(c), the pixel-wise comparison is sensitive to the label ambiguities, particularly for pixels on vessel peripheries that can be partially belong to the vessel. In addition, the pixel-wise comparison does not reflect the performance with regard to the higher level structure of the vasculature, which is also of clinical interest. To overcome these concerns, we use the CAL metric [4] that provides resilience to labeling of ambiguous pixels on vessel peripheries and better agreement with human assessment (of higher level structure). The CAL metric assesses the consistency of the binary ground truth and the binary predicted vessel map using three individual factors, the connectivity (C), the area (A), and the length of skeleton (L). The connectivity factor C compares the number of connected vessel segments between the ground truth and the predicted vessel maps, as shown in Figs. S.1(e) and (f). The area factor A assesses the relative overlapping area between the ground truth vessel map and the predicted vessel map while disregard the labeling uncertainty in pixels on vessel peripheries using morphological dilation on binary vessel maps. It can be seen in Fig. S.1(g) that the area factor is more robust against label uncertainties than pixel-wise comparison. The length factor L assesses the consistency of the vessel skeleton obtained from the ground truth and the predicted vessel map. Similar to the area factor, the morphological dilation operation is performed to overcome the issue that vessel skeleton may be slightly displaced in one image relative to the other. Figures S.1(d) and (h) show the evaluation of vessel skeleton obtained from the pixel-wise comparison and the CAL metric, respectively. The overall CAL metric is defined as the product of individual C, A, and L factors. We refer the readers to the original paper [4] for detailed computation of the CAL metric.

S.IV. Additional Visual Results on PRIME-FP20 Dataset

We provide additional visual comparison of vessel detection on PRIME-FP20 dataset in Fig. S.2. The results reinforce the findings in the main manuscript that the proposed iterative framework offers significant improvement over existing methods.

[redacted]

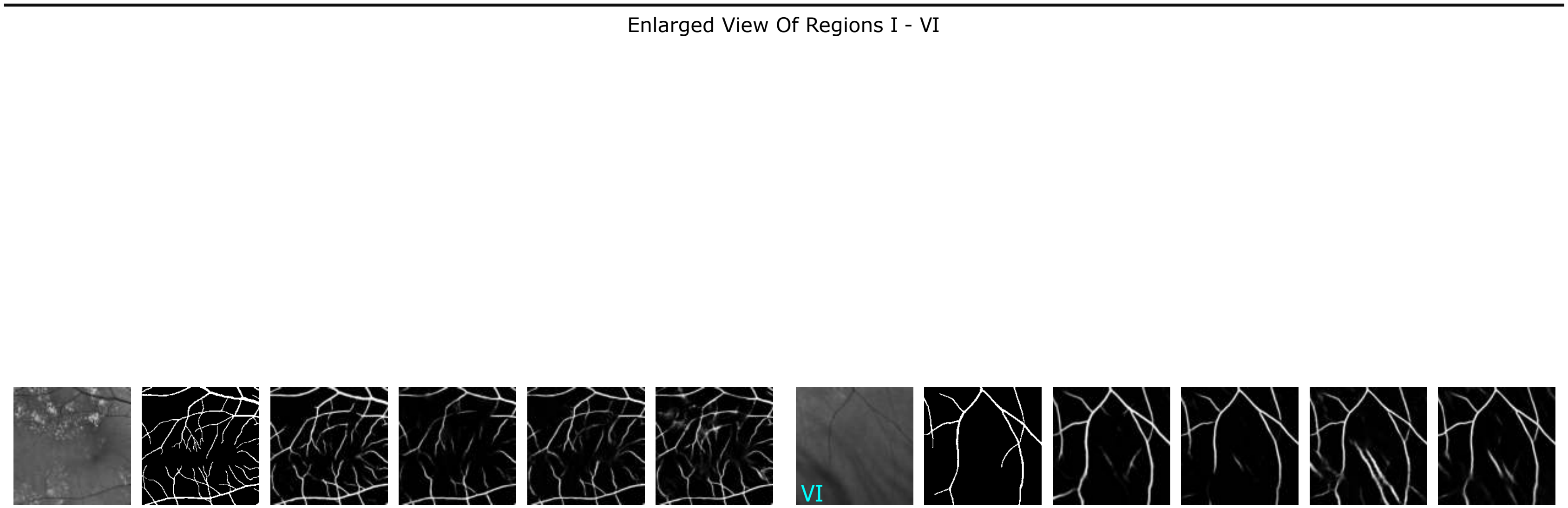


Fig. S.2: Additional sample images and detected vessel maps for the proposed approach and alternatives from the PRIME-FP20 dataset. Six contrast-enhanced enlarged views I-VI, marked by the cyan rectangles in the full image, are included.

S.V. Results on Narrow-Field Fundus Photography

In this section, we provide complete results of cross-training evaluation where the training and the test data are from two independent sources. Table S.1 lists the quantitative results. On the DRIVE dataset, the proposed framework has the best performance and significantly outperforms the existing alternatives, achieving an AUC PR of 0.886, the max DC of 0.803, and the overall CAL of 0.827. Although the proposed method is not the best performing method on the STARE dataset, the performance is only slightly worse than the best performing method (DRIU [6]). In Fig. S.3, we show sample results of the detected vessel maps on the DRIVE and the STARE datasets.

Method	Year	DRIVE (Trained On STARE)			STARE (Trained On DRIVE)		
		AUC PR	Max DC	CAL (C, A, L)	AUC PR	Max DC	CAL (C, A, L)
2nd Annotator	-	-	0.789	0.839 (1.000, 0.940, 0.892)	-	0.742	0.640 (1.000, 0.848, 0.753)
HED [7]	2015	0.879	0.797	0.743 (0.996, 0.900, 0.828)	0.838	0.748	0.574 (0.995, 0.773, 0.740)
U-Net [2]	2015	0.886	0.803	0.713 (0.997, 0.890, 0.803)	0.852	0.782	0.730 (0.996, 0.859, 0.842)
DRIU [6]	2016	0.877	0.793	0.629 (0.996, 0.847, 0.744)	**0.898**	**0.812**	**0.806 (0.996, 0.912, 0.886)**
NestUNet [8]	2018	0.877	0.795	0.688 (0.996, 0.876, 0.787)	0.892	0.805	0.786 (0.997, 0.895, 0.879)
M2U-Net [9]	2019	0.859	0.784	0.649 (0.995, 0.856, 0.760)	0.817	0.749	0.635 (0.995, 0.800, 0.785)
CE-Net [10]	2019	0.876	0.792	0.694 (0.997, 0.880, 0.790)	0.871	0.785	0.750 (0.997, 0.875, 0.855)
CS-Net [11]	2019	0.883	0.801	0.703 (0.996, 0.883, 0.798)	0.854	0.775	0.701 (0.996, 0.840, 0.821)
RU-Net [12]	2019	0.884	0.800	0.659 (0.996, 0.859, 0.769)	0.891	0.815	0.780 (0.996, 0.899, 0.869)
IterNet [13]	2020	0.845	0.795	0.698 (0.998, 0.882, 0.792)	0.815	0.794	0.727 (0.999, 0.861, 0.839)
Proposed	2020	**0.886**	**0.803**	**0.827 (0.998, 0.938, 0.883)**	0.884	0.795	0.756 (0.999, 0.880, 0.857)

TABLE S.1: Quantitative results of vessel detection obtained from different methods on the DRIVE and the STARE datasets. The best result is shown in bold.

References

[1] 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," submitted for review.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Intl. Conf. Med. Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[3] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Intl. Conf. Learning Representations*, 2015. [Online]. Available: https://arxiv.org/abs/1412.6980

[4] M. E. Gegundez-Arias, A. Aquino, J. M. Bravo, and D. Marin, "A function for quality evaluation of retinal vessel segmentations," *IEEE Trans. Med. Imaging*, vol. 31, no. 2, pp. 231–239, Feb 2012.

[5] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Intl. Conf. on Mach. Learning*, 2006, pp. 233–240.

[6] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *Intl. Conf. Med. Image Computing and Computer-Assisted Intervention*, 2016, pp. 140–148.

[7] S. Xie and Z. Tu, "Holistically-nested edge detection," in *IEEE Intl. Conf. Comp. Vision.*, Dec. 2015, pp. 1395–1403.

[8] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Med. Image Analysis*, 2018, pp. 3–11.

[9] T. Laibacher, T. Weyde, and S. Jalali, "M2U-Net: Effective and efficient retinal vessel segmentation for real-world applications," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog. Wksp.*, June 2019.

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

[11] L. Mou *et al.*, "CS-Net: Channel and spatial attention network for curvilinear structure segmentation," in *Intl. Conf. Med. Image Computing and Computer-Assisted Intervention*, 2019, pp. 721–730.

[12] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imaging*, vol. 6, no. 1, pp. 1 – 16, 2019.

[13] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *IEEE Winter Conf. Applications of Comp. Vision*, 2020, to appear.

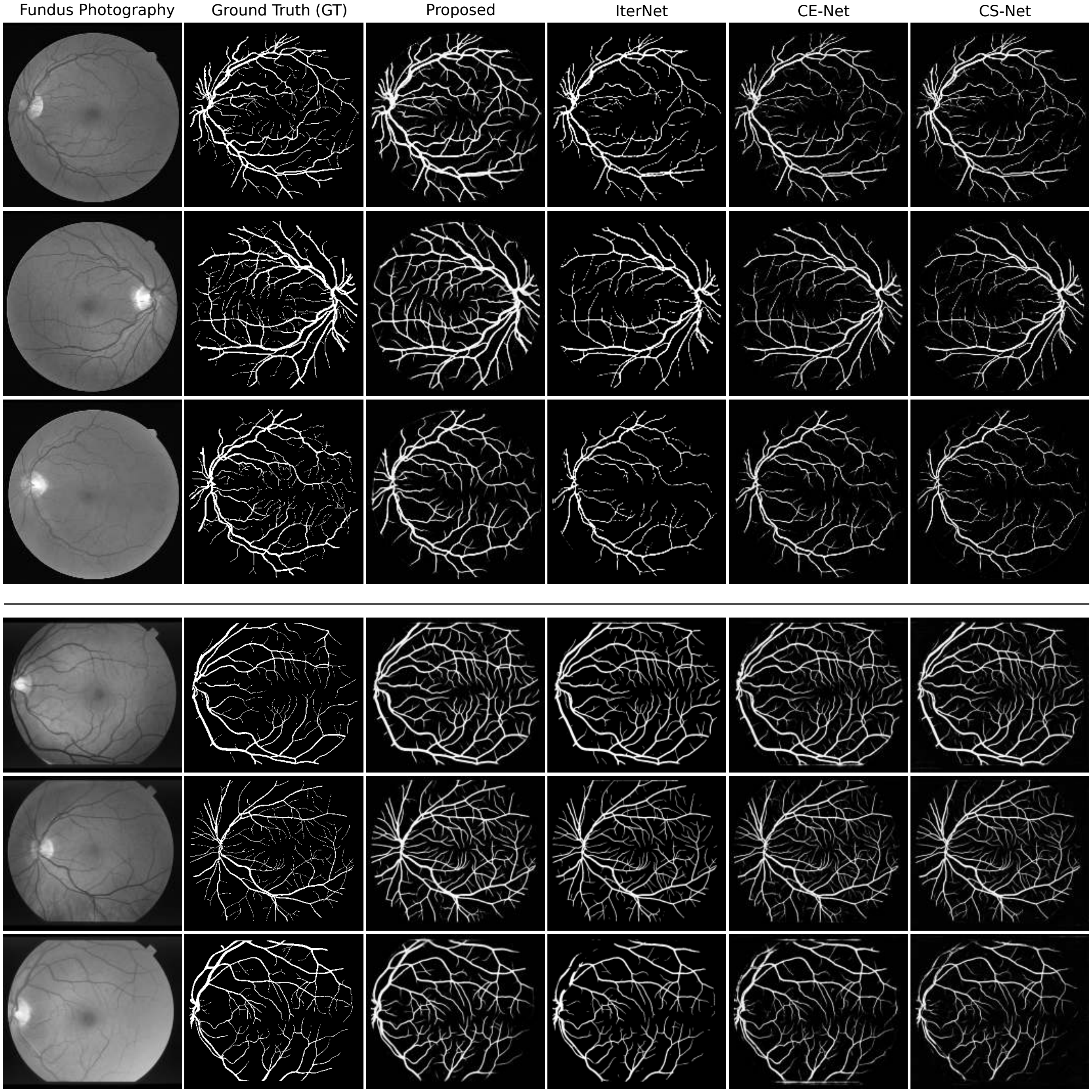


Fig. S.3: Sample images and detected vessel maps for the proposed approach and alternatives for cross-training evaluations on the DRIVE (Rows 1-3) and the STARE (Rows 4-6) datasets.