

Supplementary Material for “A Novel Deep Learning Pipeline for Retinal Vessel Detection In Fluorescein Angiography”

Li Ding, *Student Member, IEEE*, Mohammad H. Bawany, 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]. In Section S.II, we provide detailed derivations for parameter updates in the M-step for the robust EM-based chamfer registration. In Section S.III, we compare chamfer alignment under alternative geometric transformations and validate that the second-order polynomial transformation that we utilize is appropriate in this setting. In Section S.IV, we provide implementation details, including details of the unsupervised preliminary vessel detection approach, additional network architectures, and the training protocol. In Section S.V, we report on additional validation performed for the proposed human-in-the-loop labeling. In Section S.VI and S.VII, we present additional quantitative and sample visual results for the VAMPIRE [2] and the DRIsfahanCFnFA [3] datasets, respectively. Section S.VIII includes a summary of the CAL metric [4]. Larger versions of Figs. 1 and 4 of the main manuscript are provided in Section S.IX to provide better visualization.

S.II. DETAILED DERIVATION FOR PARAMETER ESTIMATIONS IN THE M-STEP

The expectation of the complete-data log likelihood $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)})$ is

$$\begin{aligned} Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)}) &= \mathbb{E} \left[\sum_{j=1}^{N_Q} \log p(d_j, w_j | \boldsymbol{\theta}) \right] \\ &= \sum_{j=1}^{N_Q} \sum_{w_j \in \{0,1\}} p(w_j | d_j, \boldsymbol{\theta}) \log p(d_j, w_j | \boldsymbol{\theta}) \\ &= \sum_{j=1}^{N_Q} p_j^{(l)} [-\lambda d_j + \log(\pi) + \log(\lambda)] \\ &\quad + (1 - p_j^{(l)}) [\log(1 - \pi) - \log(D_{max})]. \end{aligned} \tag{S.1}$$

Setting the derivatives of $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(l)})$ in (S.1) with respect to π , λ , and $\boldsymbol{\beta}$ to zero, we obtain the expression for the optimal parameters in (7) and (8).

The optimization problem in (8) can be solved by Levenberg-Marquardt (LM) [5] algorithm that is an iterative method for solving non-linear least squares problems. The LM algorithm starts with an initial estimate $\hat{\boldsymbol{\beta}}^{(0)}$ and proceed to refine the parameters $\boldsymbol{\beta}$ at each iteration. Specifically, $\boldsymbol{\beta}$ is adjusted by a parameter increment $\boldsymbol{\Delta}$ to $\boldsymbol{\beta} + \boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is obtained by solving the linear system of equations

$$\left(\sum_{j=1}^{N_Q} \mathbf{J}_j^T \mathbf{J}_j + \sigma \mathbf{I} \right) \boldsymbol{\Delta} = 2 \sum_{j=1}^{N_Q} p_j \mathbf{J}_j \mathbf{r}_j, \tag{S.2}$$

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 & the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642, USA (e-mail: ajay_kuriyan@urmc.rochester.edu).

M. H. Bawany and R. S. Ramchandran are with the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642, USA (e-mail: {mohammad_bawany, 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).

where r_j is the residual vector for point r_j that can be efficiently calculated by using the distance transform [6], $\mathbf{I} \in \mathbb{R}^{12 \times 12}$ is the identity matrix, $\mathbf{J}_j \in \mathbb{R}^{2 \times 12}$ is the Jacobian matrix at each transformed target point $\mathcal{T}_\beta(\mathbf{q}_j)$, which is computed as

$$\frac{\partial \mathcal{T}_\beta(\mathbf{q}_j)}{\partial \beta} = \begin{bmatrix} 1 & u_j & v_j & u_j^2 & u_j v_j & v_j^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & u_j & v_j & u_j^2 & u_j v_j & v_j^2 \end{bmatrix}, \quad (\text{S.3})$$

and σ is a damping parameter varying from iteration to iteration. If the increment Δ leads to a reduction in the error, then σ is divided by a factor of 10 for the next iteration, whereas if Δ gives an increased error, σ is then multiplied by the same factor. This process is repeated until the convergence criterion is met.

S.III. EVALUATION OF ALTERNATIVE TRANSFORMATIONS FOR CHAMFER ALIGNMENT

In this section, we provide empirical justification for the suitability of the second-order polynomial transformation as an appropriate registration transformation for our application. To do so, we evaluate the registration results under alternative geometric transformations, including common non-elastic transformations (Euclidean, similarity, affine, and homography) and elastic transformations (second-order and third-order polynomial transformations). The bar plot in Fig. S.1(a) shows the average of the residual chamfer distance

$$\frac{1}{N_Q} \sum_{j=1}^{N_Q} \sqrt{d_j(\hat{\beta})} \quad (\text{S.4})$$

computed over all pairs of corresponding CF and FA images in the DRISfahanCFnFA dataset under the optimal registration transformations. This residual chamfer distance provides a good proxy for the registration error. From Fig. S.1(a), it is clear that elastic transformations offers significant improvements over the alternative non-elastic transformations. In addition, there appears to be very minor benefit from going to a polynomial transformation of high-order than a second-order (2.65 pixels for the third-order versus 2.74 pixels for the second-order). Therefore, we adopt a second-order polynomial transformation to align CF and FA vessel images. Figure S.1(b) shows a sample result of chamfer alignment for one image in the DRISfahanCFnFA dataset. From the figure, one can also appreciate that the second order polynomial transformation provides adequate accuracy and it is hard to see any visual improvement with the third-order polynomial transformation.

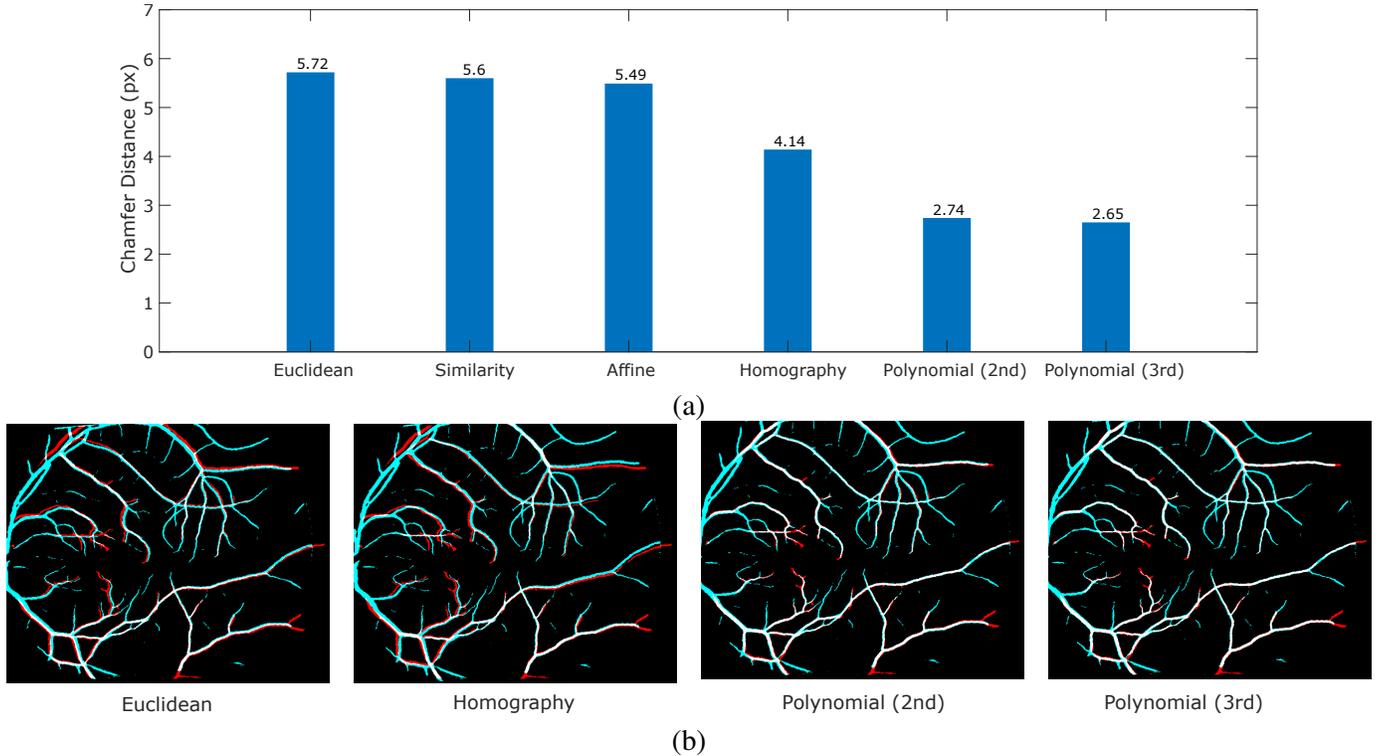


Fig. S.1: (a) Estimated registration errors for different parametric transformations. The errors are calculated as average residual chamfer distance over registered CF and FA images in the DRISfahanCFnFA dataset. (b) Sample visual results of registered CF and FA vessel maps with the alternative transformations.

S.IV. IMPLEMENTATION DETAILS

A. Preliminary Vessel Detection In FA Image

Figure S.2 provides a detailed overview of the unsupervised method used for preliminary FA vessel detection. The input FA image is decomposed into a two scale Gaussian image pyramid: one at the original image resolution (bottom branch in Fig. S.2) and the other downsampled by a factor of 2 (top branch in Fig. S.2). The images in each scale are processed independently. The resulting vessel maps in the lower resolution are upsampled to the original size using Gaussian pyramid expansion. Pixels where vessel are detected at any scale collectively comprise the estimated vessel map.

To extract bright and curvilinear vessel structures in each scale, we apply the modified top-hat operation (Equation (1) of the main manuscript) with nine line structuring elements, chosen to nominally be spaced 20 degree apart in angle, with lengths of 6/3 for the original/down-sampled scale (generated, for the original scale using the MATLAB function $strel('line', 6, \alpha)$, where α is the angle of linear structuring elements in degrees). Each top hat filtering yields a response image, where vessel pixel locations with a matching orientation are invariably high and other locations for background are usually low. The maximum value of 9 responses across different orientations at each pixel location is selected in the overall response map in which high and low values are likely for vessel and background pixels, respectively. The obtained soft vessel map is converted into a binary segmentation by locally adaptive thresholding. The threshold value for each pixel is based on the local mean intensity in the neighborhood of the pixel (estimated using the MATLAB function $adaptthresh(I)$, where I is the soft vessel map). As a post-processing step, we remove small incoherent random segments that have fewer than 100 pixels from the binary vessel maps (using the MATLAB function $bwareaopen(BW, 100)$, where BW is the binary vessel map from the adaptive thresholding).

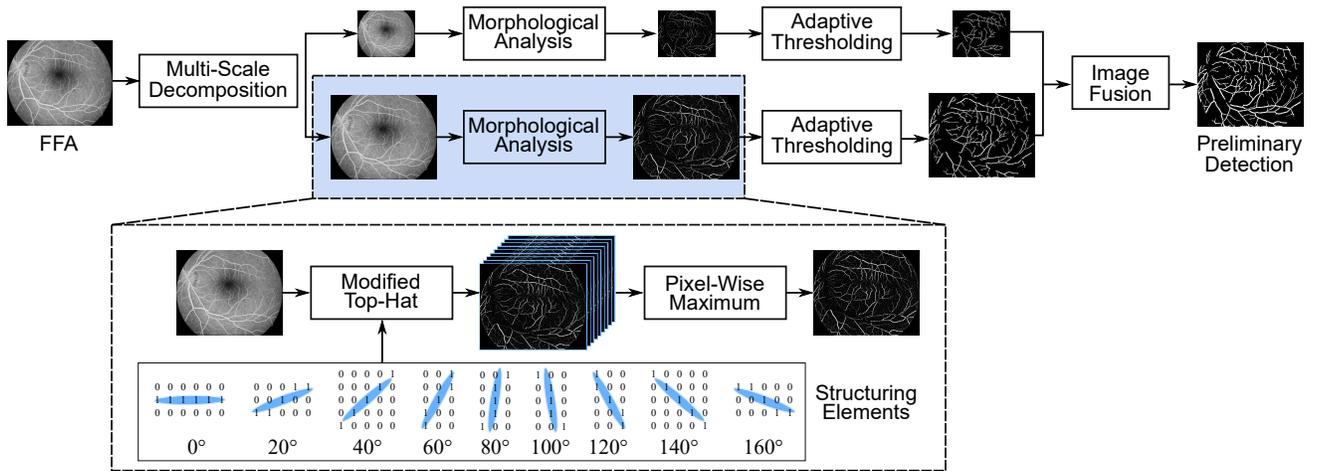


Fig. S.2: Overview of the unsupervised method for preliminary vessel detection in FFA.

B. Network Architectures and Loss Function

The Prop. + UNet uses the same network architecture as the generator in the Prop. + GAN, as shown in Fig. 6 (the generator block). For the Prop. + NestUNet, we adopt the NestUNet architecture proposed in [7]. All convolutional layers use 3×3 kernels with stride 1. We use MaxPooling layers with 2×2 kernels to reduce spatial resolutions. The objective function in (10) (in the main manuscript) is used for the Prop. + UNet and the Prop. + NestUNet.

C. Training Protocol

We feed the network 256×256 patches extracted from the FA training data with a fixed stride length 128. Patches that contains less than 1% vessel pixels are excluded. To prevent neural networks from over-fitting, we further enlarge the training set by performing on-the-fly data augmentation, i.e., randomly applying a list of transformations with different probabilities to each image before feeding into neural network as training data. Specifically, we consider following transformations: (1) rotating the image by an angle from -90° to 90° , (2) horizontally and vertically flipping the image, (3) scaling the image by a factor of 2, (4) blurring the image using Gaussian filter, and (5) adjusting the brightness and contrast of the image.

The network parameters are optimized using Adam optimizer [8] on a NVidia Tesla V100 GPU. The learning rate is fixed as 0.001. The coefficients used for computing running averages of gradient and its square are 0.9 and 0.999, respectively. The batch size is 16 and the training dataset is shuffled between epochs. We split the data into a training set (80%) and a validation set (20%) and use the model that has the best performance on the validation set. The lambda in (11) is set to 1.

S.V. ADDITIONAL VALIDATION OF HUMAN-IN-THE-LOOP LABELING

As additional validation, we also assessed the agreement between the ground truth vessel maps generated from the human-in-the-loop annotation and from *de novo* manual annotation. In order to do this, we chose two rectangular patches (512×512 pixels) from the image we used in the last (7th) human-in-the-loop iteration and manually annotated the binary vessel maps from scratch. The selected patches cover both central and peripheral retina and represent both major and minor vessel branches. We measure the intraclass correlation coefficient (ICC) [9] to quantify the agreement between the two labeling approaches. As a benchmark for comparison, we also quantified the level of agreement that may be expected in independent human annotation of labeled vessel maps. Because multiple independent annotations of the same images were not available for the FA modality, we assessed the level of agreement between two independent human annotations using the CF datasets DRIVE [10], STARE [11], and CHASE_DB1 [12] datasets, which provide two independent human annotations. The level of agreement was quantified using the ICC computed between the two alternative human annotations for these datasets. Table S.1 lists the ICC values computed from each dataset. The ICC is computed by a 2-way mixed-effects model, average measures, and absolute agreement [9]. The proposed human-in-the-loop annotation achieves an ICC of 0.870 that indicates good reliability and is comparable to the ICC values in the existing CF datasets.

Datasets	RECOVERY-FA19	(CF) DRIVE [10]	(CF) STARE [11]	(CF) CHASE_DB1 [12]
ICC	0.870	0.863	0.835	0.863

TABLE S.1: ICC values for assessing the level of agreement between alternative labels. For the RECOVERY-FA19 dataset the ICC quantifies the level of agreement between the labels obtained with the proposed human-in-the-loop approach and *de novo* labeling from scratch. For the (pre-existing) CF datasets the ICC quantifies the level of agreement between two independent human annotations.

S.VI. EVALUATIONS ON THE VAMPIRE DATASET

We provide quantitative metrics comparing the different methods on the VAMPIRE dataset [2]. Table S.2 lists the AUC for both ROC and PR curves and the maximum DC using the limited accuracy original (vessel) labeling. In addition, we also report in Table S.2 results evaluated on the “AMD2” and the “GER4” images using both the original labeling and the (refined) ground truth. Additionally, Table S.3 lists the CAL metrics evaluated on the refined ground truth. In Table S.2, it is interesting to see that the accuracy of the unsupervised methods, SFAT [2] and MSMA [13], is decreased when they are evaluated using the (refined) ground truth. This is because the (refined) ground truth data contains more vessel that are not detected by SFAT [2] and MSMA [13] (implying that these methods actually have more false negatives than would be indicated by the original, inaccurate, labeling). The performance gains of the DNN approaches using the proposed framework improve when the methods are evaluated on the more accurate (refined) ground truth compared with evaluations using the original (flawed) labeling. The CAL metrics in Table S.3 reinforce these findings: the proposed methods offer significant improvement over prior alternatives.

Methods	All Test Images: Original Lbl.			“AMD2” and “GER4”: Original Lbl.			“AMD2” and “GER4”: Refined GT		
	AUC ROC	AUC PR	Max DC	AUC ROC	AUC PR	Max DC	AUC ROC	AUC PR	Max DC
SFAT [2]	-	-	0.624	-	-	0.647	-	-	0.573
MSMA [13]	-	-	0.647	-	-	0.713	-	-	0.654
VDGAN [14]	0.957	0.702	0.686	0.971	0.738	0.680	0.965	0.783	0.707
Prop. + UNet	0.978	0.786	0.715	0.987	0.809	0.713	0.994	0.948	0.878
Prop. + NestUNet	0.979	0.779	0.715	0.987	0.801	0.727	0.995	0.953	0.883
Prop. + GAN	0.978	0.780	0.715	0.988	0.808	0.731	0.995	0.950	0.878

TABLE S.2: Quantitative results obtained from different methods on the VAMPIRE dataset. All deep neural networks are trained on the generated ground truth from the DRISfahanCFnFA dataset. The best result is shown in bold.

Methods	CAL	C	A	L
SFAT [2]	0.322	0.999	0.541	0.583
MSMA [13]	0.393	0.999	0.613	0.640
VDGAN [14]	0.465	0.998	0.651	0.717
Prop. + UNet	0.775	0.999	0.897	0.864
Prop. + NestUNet	0.786	0.999	0.904	0.870
Prop. + GAN	0.772	0.999	0.897	0.861

TABLE S.3: Quantitative CAL metrics for the different methods on the “AMD2” and “GER4” images from the VAMPIRE dataset evaluated using the refined labels. The best result is shown in bold.

Figure S.3 plots the ROC and the PR curves evaluated on the “AMD2” and the “GER4” images using the (refined) ground truth. The Prop. + GAN network achieves an AUC ROC of 0.995, an AUC PR of 0.950, the maximum DC of 0.878, and a CAL of 0.772. For reference, we also plot the accuracy of the original (vessel) labeling on the plots (shown by the points

marked by the blue cross \times), which achieves the maximum DC of 0.642. These plots highlight that the network-predicted vessel maps are significantly better than original vessel map labels, which further demonstrates the utility of the proposed pipeline.

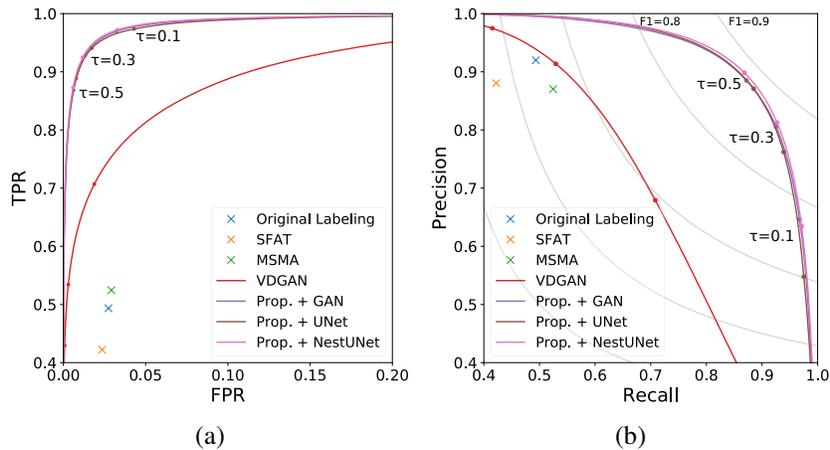


Fig. S.3: (a) ROC and (b) PR curves for different methods. The results are evaluated on the “AMD2” and the “GER4” images from the VAMPIRE [2] dataset with the refined ground truth. The gray curves in (b) represent the isolines of Dice coefficients. The small circular dots on the curves in (a) and (b) identify the corresponding values of the threshold τ .

In addition to the evaluation on *de novo* annotated image patches (Table III of the main manuscript), we also performed an additional assessment of the proposed approach and alternatives on labeled data from the VAMPIRE dataset that is obtained (largely) independently of the iterative human-in-the-loop labeling process. Specifically, we identified “trusted regions” of the VAMPIRE dataset, where the annotations were accurate and evaluated the methods on these “trusted regions” while excluding the non-trusted regions where the VAMPIRE dataset was missing vessel annotations (see Fig. 10 of the main manuscript). The approach relied on the observation that the VAMPIRE annotations were missing vessel annotations but did not have false positives. Figure S.4 shows an example image showing the identification of “trusted regions”, which were estimated to conservatively exclude regions that could include pixels that were clearly labeled incorrectly. These “trusted regions” were created using simple morphological operations and then visually validated. Note that the labels over the trusted regions are those provided directly in the VAMPIRE dataset and therefore not subject to potential confirmation bias in the labeling process. We report area under the ROC and Precision-recall curves and the maximum Dice coefficient for different methods in Table S.4¹. Just like the results presented earlier, these results highlight the fact that the proposed approach offers a much higher accuracy than the prior SFAT and MSMA approaches also improves significantly over the VDGAN precursor for the proposed approach.

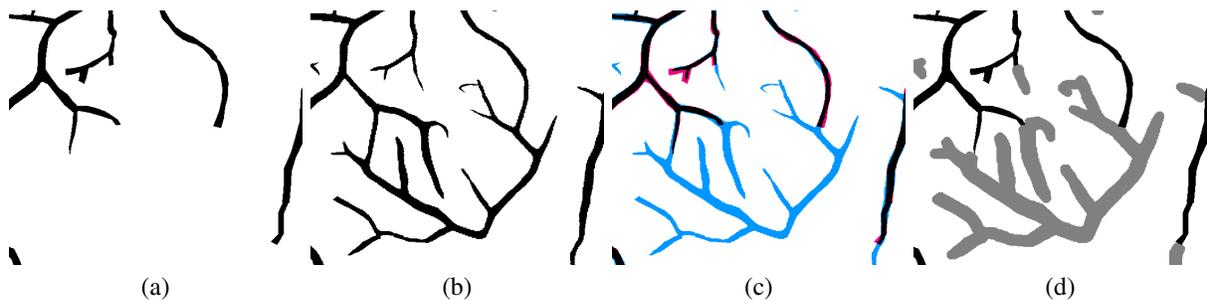


Fig. S.4: Identification of the “trusted regions” of the VAMPIRE dataset. (a) The original labels. (b) The refined labels. (c) Comparison between the original and the refined labels. The red and the blue regions show where the vessels are annotated only in the original and the refined labels, respectively. (d) Pixels annotated as vessels in the VAMPIRE dataset that were in close vicinity to pixels labeled as vessels via our proposed iterative human-in-the-loop approach were designated as “trusted” vessel pixels (black). Pixels that were labeled as background in the VAMPIRE dataset but were close to pixels labeled as vessels via our proposed iterative human-in-the-loop approach and not close to pixels labeled as vessels in the VAMPIRE annotation, were designated as “non trusted” background pixels (gray) and the remaining background pixels in the original annotation were considered “trusted” background pixels (white).

¹Note that the computation of the CAL metrics is not meaningful with the restriction of the evaluation to only the trusted regions and CAL metrics are therefore not reported.

Methods	AUC ROC	AUC PR	Max DC
SFAT [2]	-	-	0.670
MSMA [13]	-	-	0.728
VDGAN [14]	0.958	0.791	0.729
Prop. + UNet	0.989	0.892	0.831
Prop. + NestUNet	0.990	0.890	0.832
Prop. + GAN	0.989	0.893	0.833

TABLE S.4: Quantitative results obtained from different methods on the “trusted region” from the VAMPIRE dataset. The best result is shown in bold.

S.VII. EVALUATIONS ON THE DRISFAHANCfNFA DATASET

Figure S.5 show the ROC and the PR curves for different methods on the DRISfahanCFnFA dataset [3]. The best performing method (Prop. + GAN) achieves an AUC ROC of 0.974, an AUC PR 0.887, and the maximum DC of 0.808, outperforming other baseline methods. The unsupervised methods, SFAT [2] and MSMA [13], are developed for detecting vessels from UWFFA images rather than fundus FA and thus have relatively low DC (0.607 and 0.691, respectively). While deep neural networks are trained only using the UWFFA images from the RECOVERY-FA19 dataset, they have the generalization ability to detect vessels from fundus FA images and show significant improvement over the unsupervised methods. In Fig. S.6, we show visual results of the vessel detection obtained with the Prop. + GAN and the Prop. + NestUNet. While deep neural networks are trained only using the UWFFA images from the RECOVERY-FA19 dataset, it has the generalization ability to detect vessels from fundus FA images. Compared with the Prop. + NestUNet, the GAN loss term \mathcal{L}_{GAN} encourages consistency between the predicted vessel maps and the ground truth globally over the entire image patch in consideration (orange arrows in the enlarged views in Fig. S.6 illustrate this).

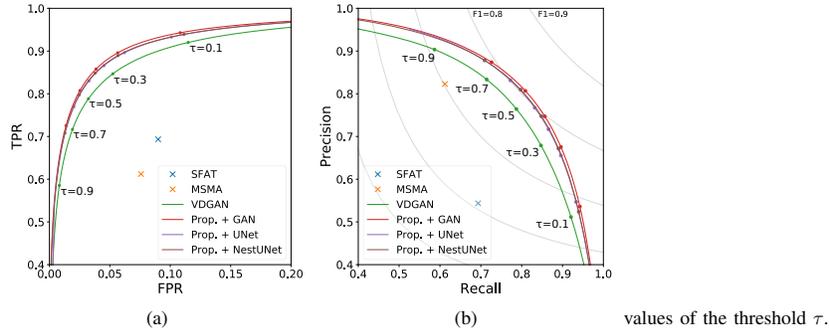


Fig. S.5: (a) ROC and (b) PR curves for different methods on the DRISfahanCFnFA dataset [3]. The gray curves in (b) represent the isolines of Dice coefficients. The small circular dots on the curves in (a) and (b) identify the corresponding values of the threshold τ .

S.VIII. DESCRIPTION OF CAL METRIC

CAL is a function for quantitatively evaluating the performance of retinal vessel detection [4] that takes into account the structure of retinal vasculature, unlike pixel-wise comparison metrics, such as, precision, recall, and Dice coefficient. Specifically, CAL assesses the overall consistency between a binary ground truth vessel map \mathbf{V}_G and a binary predicted vessel map \mathbf{V}_P by computing three factors between 0 and 1 that, respectively, assess the consistency with regard to connectivity (C), overlapping area (A), and the corresponding skeletons (L) (lengths). The connectivity C compares the number of the connected segments in \mathbf{V}_G and \mathbf{V}_P with respect to the number of vessels pixels in \mathbf{V}_G , and is computed as

$$C = 1 - \min\left(1, \frac{|n(\mathbf{V}_P) - n(\mathbf{V}_G)|}{|\mathbf{V}_G|}\right), \quad (\text{S.5})$$

where $n(\cdot)$ denotes the number of connected segments and $|\mathbf{W}|$ represents the cardinality, i.e., the number of elements, in the set \mathbf{W} . The area factor A is computed as

$$A = \frac{|(\mathbf{V}_P^{r_1} \cap \mathbf{V}_G) \cup (\mathbf{V}_P \cap \mathbf{V}_G^{r_1})|}{|\mathbf{V}_G \cup \mathbf{V}_P|}, \quad (\text{S.6})$$

where \mathbf{V}^r denotes the dilated version of the image \mathbf{V} using a disk-shaped structuring element, r pixels in radius. The motivation for the dilation comes from the fact that pixels adjacent to the periphery of labeled vessels may span both vessel and background regions and therefore their labeling may be inherently uncertain. The computation of the area factor A in (S.6) measures the relative overlap between two vessel maps while discounting for the ambiguity in pixels that are adjacent to the periphery of

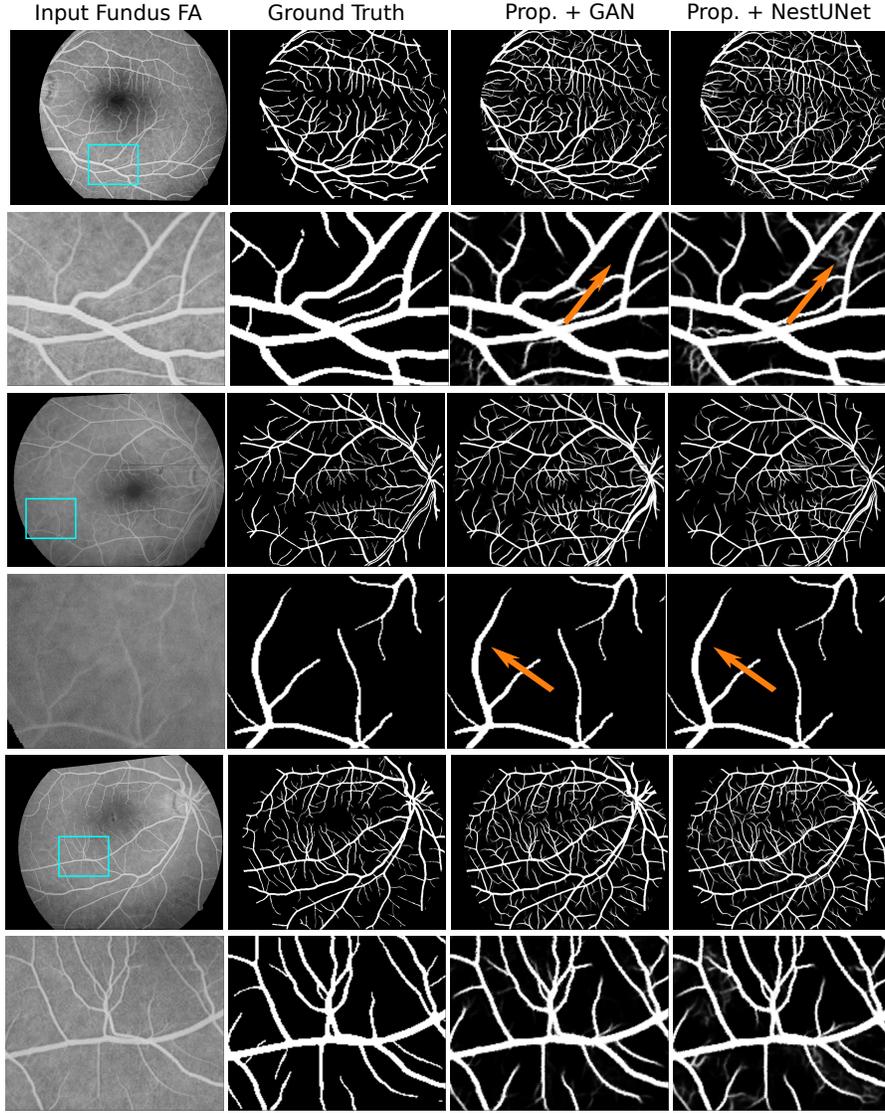


Fig. S.6: Qualitative comparison of results from the DRIsFahanCFnFA dataset [3]. The orange arrows highlight regions where the proposed adversarial network produces more accurate detections than NestUNet [7].

labeled vessels in either V_G and V_P . For this reason, compared with direct pixel-wise metrics (TPR (Recall)/FPR/Precision/Dice coefficient), the area metric provides robustness against variations in the labeling of these uncertain pixels. The length factor L ; computed as

$$L = \frac{|(S(V_P) \cap V_G^{r_2}) \cup (V_P^{r_2} \cap S(V_G))|}{|S(V_P) \cup S(V_G)|}, \quad (\text{S.7})$$

where $S(\cdot)$ denotes the morphological skeletonization operation; measures the consistency of the skeletons for the vessel maps V_G and V_P . The global score of CAL metric is defined as the product of C , A , and L and ranges in value between 0 and 1, where a high value indicates better match with the ground truth. Default values of $r_1 = r_2 = 2$ are used in our computation of the CAL metrics.

S.IX. HIGH RESOLUTION FIGURES

We provide larger views of Figs. 1 and 4 of the main manuscript as Figs. 1H, 4H, respectively. Although, the key ideas that these images illustrate can also be seen in the versions included in the main manuscript, the larger versions included here (and the corresponding images in the dataset accompanying the paper [15]) provide clearer visualizations.



Fig. 1H: Sample fluorescein angiography (FA) images. left: fundus FA. Middle: ultra-widefield FA. Right: enlarged view of the cyan rectangle (top and bottom: the original and the contrast-enhanced views, respectively).

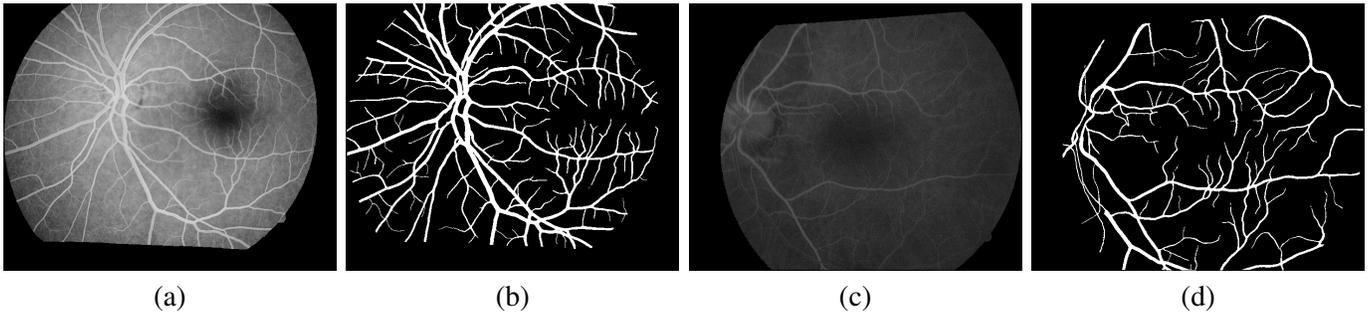


Fig. 4H: Sample results of generated training data for FA imagery in DRIsFahanCFnFA dataset. (a) and (c) show two FA images, and (b) and (d) are the corresponding vessel maps. Notice that the generated vessel maps are robust under different contrast conditions.

REFERENCES

- [1] 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. <https://doi.org/10.1109/TIP.2020.2991530> Code [16] and associated RECOVERY-FA19 dataset [15] available.
- [2] A. Perez-Rovira, K. Zutis, J. Hubschman, and E. Trucco, "Improving vessel segmentation in ultra-wide field-of-view retinal fluorescein angiograms," in *IEEE Intl. Conf. Eng. in Med. and Biol. Soc.*, Aug. 2011, pp. 2614–2617.
- [3] S. Hajeb-Mohammad-Alipour, H. Rabbani, and M. R. Akhlaghi, "Diabetic retinopathy grading by digital curvelet transform," *Computational and Mathematical Methods in Med.*, vol. 2012, 2012.
- [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. Nocedal and S. Wright, *Numerical optimization*. Springer, 2006.
- [6] G. Borgefors, "Distance transformations in digital images," *Comp. Vis., Graphics and Image Proc.*, vol. 34, no. 3, pp. 344–371, 1986.
- [7] 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.
- [8] 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>
- [9] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, 2016.
- [10] J. Staal, M. Abramoff, 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.
- [11] 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.
- [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 Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, 2012.
- [13] 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.
- [14] L. Ding, A. Kuriyan, R. Ramchandran, and G. Sharma, "Retinal vessel detection in wide-field fluorescein angiography with deep neural networks: A novel training data generation approach," in *IEEE Intl. Conf. Image Proc.*, Oct 2018, pp. 356–360.
- [15] L. Ding, M. H. Bawany, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "RECOVERY-FA19: Ultra-widefield fluorescein angiography vessel detection dataset," IEEE Dataport, 2019. [Online]. Available: <http://dx.doi.org/10.21227/m9yw-xs04>
- [16] L. Ding, M. H. Bawany, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "(Code Ocean capsule): Deep vessel segmentation for fluorescein angiography and evaluation," <https://doi.org/10.24433/CO.1133548.v1>, May 2020.