tvst

Special Issue

A Deep-Learning Approach for Automated OCT En-Face Retinal Vessel Segmentation in Cases of Optic Disc Swelling Using Multiple En-Face Images as Input

Mohammad Shafkat Islam[1], Jui-Kai Wang[1,2], Samuel S. Johnson[1], Matthew J. Thurtell[3], Randy H. Kardon[2,3], and Mona K. Garvin[1,2]

[1] Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA, USA
[2] Iowa City VA Health Care System and Iowa City VA Center for the Prevention and Treatment of Visual Loss, Iowa City, IA, USA
[3] Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA, USA

Correspondence: Mona K. Garvin, 4318 Seamans Center for the Engineering Arts and Sciences, University of Iowa, Iowa City, IA 52242, USA. e-mail: mona-garvin@uiowa.edu

Received: August 19, 2019
Accepted: December 29, 2019
Published: March 24, 2020

Keywords: optic disc swelling; vessel segmentation; optical coherence tomography; multiple en-face images; deep learning; U-Net; papilledema; retinal blood vessels

Citation: Islam MS, Wang J-K, Johnson SS, Thurtell MJ, Kardon RH, Garvin MK. A deep-learning approach for automated OCT en-face retinal vessel segmentation in cases of optic disc swelling using multiple en-face images as input. Trans Vis Sci Tech. 2020;9(2):17, https://doi.org/10.1167/tvst.9.2.17

Purpose: In cases of optic disc swelling, segmentation of projected retinal blood vessels from optical coherence tomography (OCT) volumes is challenging due to swelling-based shadowing artifacts. Based on our hypothesis that simultaneously considering vessel information from multiple projected retinal layers can substantially increase vessel visibility, in this work, we propose a deep-learning-based approach to segment vessels involving the simultaneous use of three OCT en-face images as input.

Methods: A human expert vessel tracing combining information from OCT en-face images of the retinal pigment epithelium (RPE), inner retina, and total retina as well as a registered fundus image served as the reference standard. The deep neural network was trained from the imaging data from 18 patients with optic disc swelling to output a vessel probability map from three OCT en-face input images. The vessels from the OCT en-face images were also manually traced in three separate stages to compare with the performance of the proposed approach.

Results: On an independent volume-matched test set of 18 patients, the proposed deep-learning-based approach outperformed the three OCT-based manual tracing stages. The manual tracing based on three OCT en-face images also outperformed the manual tracing using only the traditional RPE en-face image.

Conclusions: In cases of optic disc swelling, use of multiple en-face images enables better vessel segmentation when compared with the traditional use of a single en-face image.

Translational Relevance: Improved vessel segmentation approaches in cases of optic disc swelling can be used as features for an improved assessment of the severity and cause of the swelling.

Introduction

Retinal blood vessel attributes, such as the width, location, obscuration, integrity, and tortuosity, are commonly considered important features in assessments of optic disc swelling.[1] For example, in the modified Frisén grading system, the number of obscured vessel segments leaving the optic disc, is considered one of the key features to distinguish papilledema from mild to severe levels using fundus photographs.[2–5] Echegaray et al.[6] have also shown that the measurement of vessel discontinuity can be helpful for a machine-learned Frisén grading system to achieve a substantial agreement between its output and the human expert's decision. Figure 1a shows three example fundus photographs with mild, moderate, and severe optic disc swelling (from the top to the bottom row in the figure). As indicated with yellow arrows, the vessel attributes change substantially on the swollen optic disc among these cases.

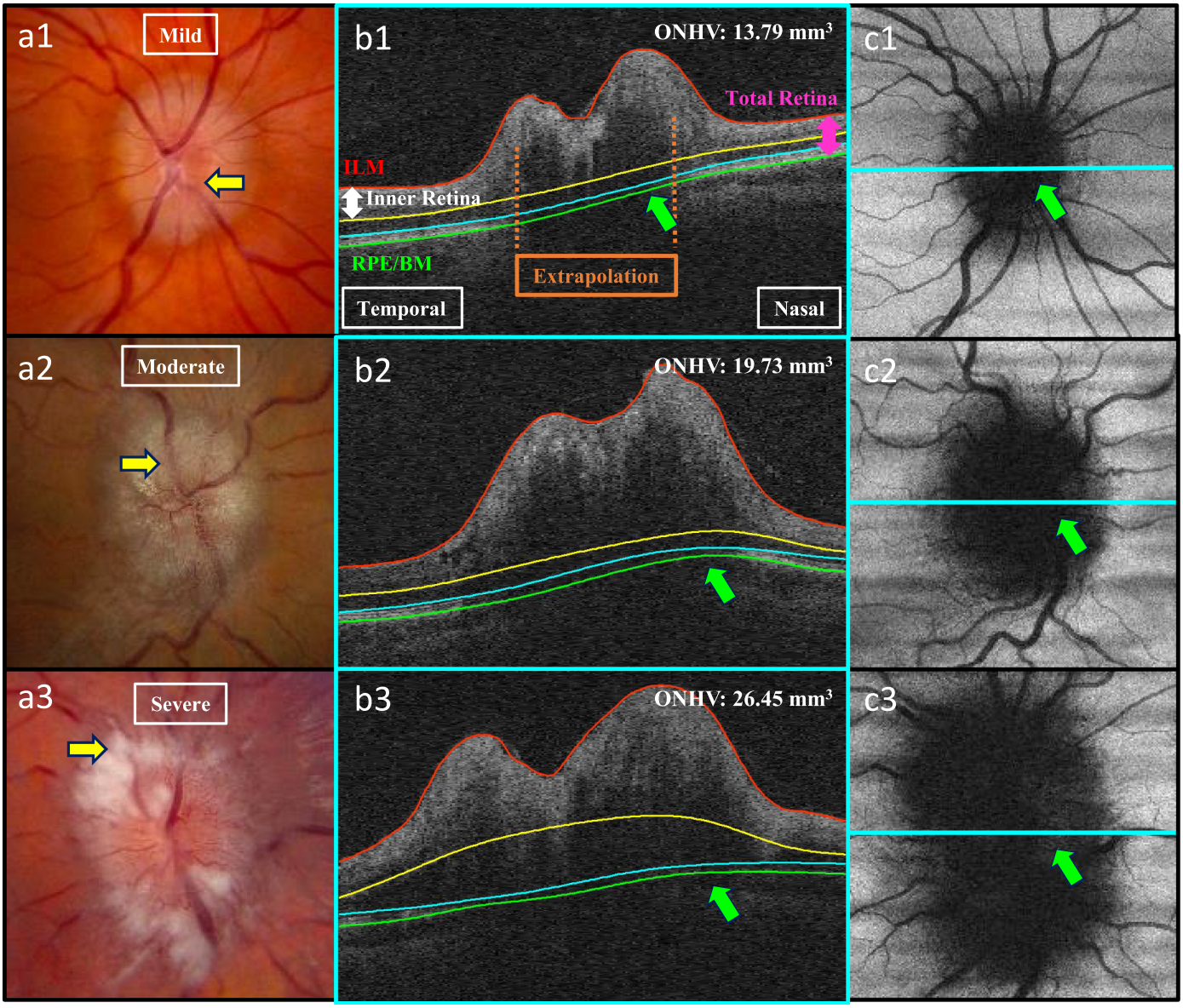


Figure 1. Comparisons of the fundus photograph and OCT pairs with mild optic disc swelling (*top row*: a1, b1, c1), moderate swelling (*middle row*: a2, b2, c2), and severe swelling (*bottom row*: a3, b3, c3). *Left column* (a1, a2, a3) shows fundus photographs. *Middle column* (b1, b2, b3) shows the OCT central B-scans with automated layer segmentation. *Right column* (c1, c2, c3) shows the OCT RPE en-face images. Note: In case of swelling, the *yellow arrows* indicate vessel attributes changes in (a), the *cyan lines* in (c) represent the location of the central B-scans, and the *green arrows* in (b) and (c) indicate the matched shadow regions.

Spectral-domain optical coherence tomography[7–10] (OCT) is another imaging modality that is regularly used for assessing optic disc swelling. To date, most OCT-based measurements that have been used in the clinic and in research studies in cases of optic disc swelling (such as the retinal nerve fiber layer as well as the total retinal layer thicknesses, the optic nerve head [ONH] volume, and Bruch's membrane shapes)[5,8,11–15] do not incorporate vessel information. However, especially for purposes of developing automated systems for assessing the severity and causes of optic disc swelling, having robust automated approaches for the OCT-based segmentation of retinal vessels is needed not only for computation of vessel-based features but also as one of the preprocessing steps for computation of other features. For example, removing retinal vessels is often involved as a part of preprocessing for further retinal texture analyses, such as retinal fold analysis.[16,17] Having an accurate vessel tree location map can substantially reduce the

false-positive rate for an automated method to detect retinal folds in OCT.[17] Furthermore, vessels are often an important structure used for the alignment of images (e.g., color fundus to OCT or OCT images over time), which can be used for multimodal analyses and regional longitudinal analyses. Thus, the motivation for an OCT-based vessel segmentation in cases of optic disc swelling includes the need for the direct computation of vessel-related features in OCT (especially in cases where fundus photography is not available) for direct measures of severity or for differentiation, the need for additional contextual information in the development of techniques for the automated segmentation and analysis of other structures (e.g., for fold/wrinkle detection) that may help in differentiation, and the need for an alignment technique for region-based longitudinal analyses.

Although OCT has been widely used for capturing cross-sectional information of the retina, observing the vessels in the common B-scan orientation is not straightforward (as shown in Fig. 1b). A common method to display the vessels in OCT is to create an en-face view by projecting the pixel intensity values within the retinal pigment epithelium (RPE) complex along each A-scan.[18–20] In cases without optic disc swelling, projection at the level of the RPE works well given the high contrast between the bright RPE and inner-retinal vessel shadows. However, in cases of optic disc swelling, the presence of swelling can cause image shadows, making the task of vessel segmentation much more challenging. Figure 1c continues to show the RPE en-face images from the same three patients; it is noticeable that the challenge of delineating the vessels for both manual and automated approaches increases when the image shadow (from the swollen disc) grows.

However, based on our prior preliminary experience with the segmentation of vessels in the OCT scans of mice whereby using multiple en-face images was advantageous over a single projection image[21] as well as our observation that the vessels could sometimes be seen more prominently in layers other than the RPE layer in cases of optic disc swelling in humans, we hypothesized that simultaneously considering vessel information from various projected retinal layers in cases of optic disc swelling would substantially increase the vessel visibility and enable a better segmentation. Thus, instead of relying on a single projection image at the level of the RPE, we have developed a deep-learning approach (using a modification of a U-Net[22] architecture) to simultaneously input three OCT en-face images from the RPE complex, inner retina, and total retina and to output an OCT vessel probability map (Fig. 2). Although current deep neural networks have shown prominent performance among the frequent implementations of vessel segmentation algorithms,[23–28] there is no study specifically focusing on the OCT cases with optic disc swelling. Both quantitative and qualitative comparisons between manual tracings in en-face images from various retinal layers and the automated segmentation results are performed.

Methods

Training/Testing Data

From 122 patients with various causes of optic disc swelling who had been recruited for research use of their clinical imaging data from the Neuro-Ophthalmology Clinic at the University of Iowa, we had previously analyzed the volumetric OCT imaging data of 22 of these patients for a preliminary fold detection image analysis approach.[17] To best ensure a true separation of training and testing sets (whereby evaluation on the testing set is limited to untouched data), our training set was selected from the 22 previously analyzed images. More specifically, of the 22 patients previously analyzed, we selected the 18 patients who had (1) both volumetric ONH-centered OCT scans (Zeiss Cirrus, Carl Zeiss Meditec, Inc., Dublin, CA, USA) and the corresponding fundus photographs (Topcon Medical Systems, Inc.) available at the same visit and (2) an intact retinal structure in the OCT scans to allow the automated retinal layer segmentation[12] to process correctly. The training data set was used for the purposes of designing the neural network architecture, deciding the hyperparameters, and training the neuron weights. For the independent testing data set, an additional 18 pairs of OCT scans and fundus photographs of 18 patients with optic disc swelling collected from the same set of 122 patients having optic disc swelling were included by matching the total retinal volume distribution with the training data set. The reason for this volume-matching process was to maintain a similar vessel visibility (which can be substantially affected by the degree of optic disc swelling, as shown in Fig. 1) in the OCT en-face images within different patients between the training and testing data sets.

Figure 3 shows the data distributions (by ONH volume) of the training and testing data sets. For the causes of optic disc swelling in the training data set, among the 18 training patients, 13 had papilledema, 1 had nonarteritic anterior ischemic optic neuropathy (NAION), and 4 had other causes of optic disc swelling. For the 18 testing participants, 15 had papilledema, 2 had NAION, and 1 had another cause of optic disc swelling. The training data set consisted

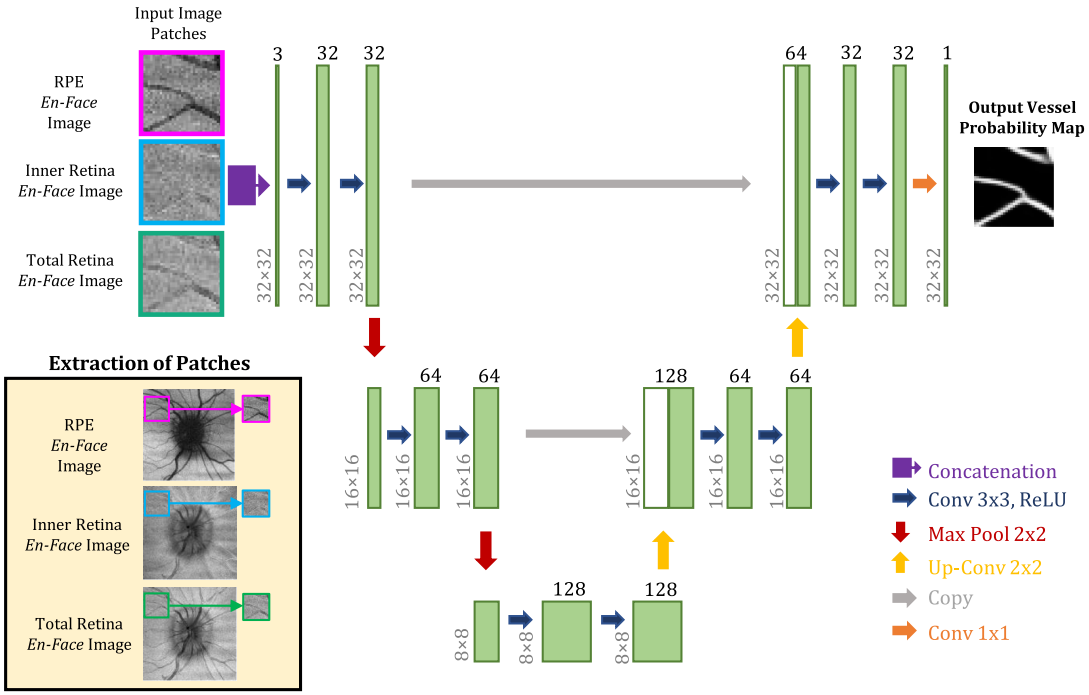


Figure 2. Architecture of the proposed deep-learning approach. Three image patches (32 × 32 pixels) are separately extracted from the OCT en-face images of the RPE complex, the inner retina, and the total retina. Next, these three patches are concatenated to each other at the first layer in the network. The numbers in *black* and in *gray* at each block represent the number of channels and dimensions at the current network layer, and the colors of the *arrows* represent different network operations.

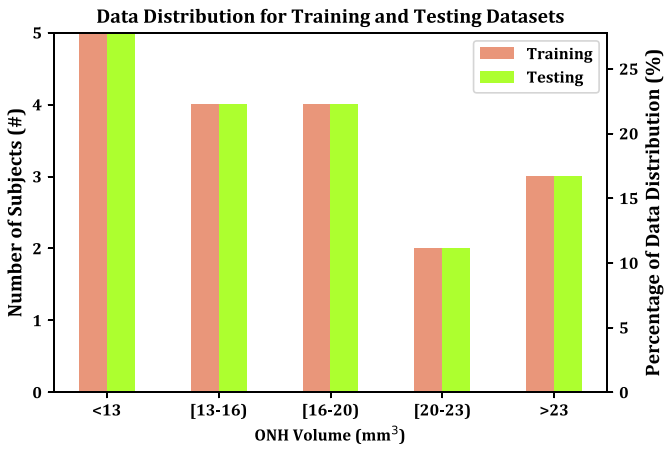


Figure 3. Data distribution (by ONH volume) of the training and testing data sets (shown in *pink* and *green bars*, respectively). There are 36 patients in total: 18 in the training data set and the other 18 in the testing data set. The severity of the disc swelling has been matched between both data sets based on the ONH volumes.

of 16 women and 2 men with a mean ± standard deviation (SD) age of 36.4 ± 11.4 years while the volume-matched testing data set consisted of 17 women and 1 man with a mean ± SD age of 31.6 ± 12.9 years. In total, the 36 patients had a mean ± SD age of 34 ± 12.4 years.

All the fundus photographs were obtained using a retinal camera (TRC-50DX; Topcon Medical Systems, Inc.) with 2392 × 2048 pixels. Each OCT scan (Cirrus; Carl Zeiss Meditec, Dublin, CA, USA) was centered at the ONH and had 200 × 200 × 1024 voxels covering (approximately) 6 × 6 × 2 mm^3. Note that the fundus photographs in this study were only used for helping to create ground truth images (as discussed in the section on the manual tracing process) and were registered and cropped with respect to the OCT en-face images. The study protocol was approved by the University of Iowa's Institutional Review Board and adhered to the tenets of the Declaration of Helsinki.

En-Face Images from Multiple Retinal Layers (Inputs to Deep-Learning Approach)

A customized three-dimensional graph-based algorithm[12] was utilized to segment the swollen retinal layers (Fig. 1b) as a part of the preprocessing. Then, based on the segmentation results, en-face images were generated of the RPE complex (between cyan and green surfaces in Fig. 1b), the inner retina (between red and yellow surfaces in Fig. 1b), and the total retina (between red and green surfaces in Fig. 1b) by

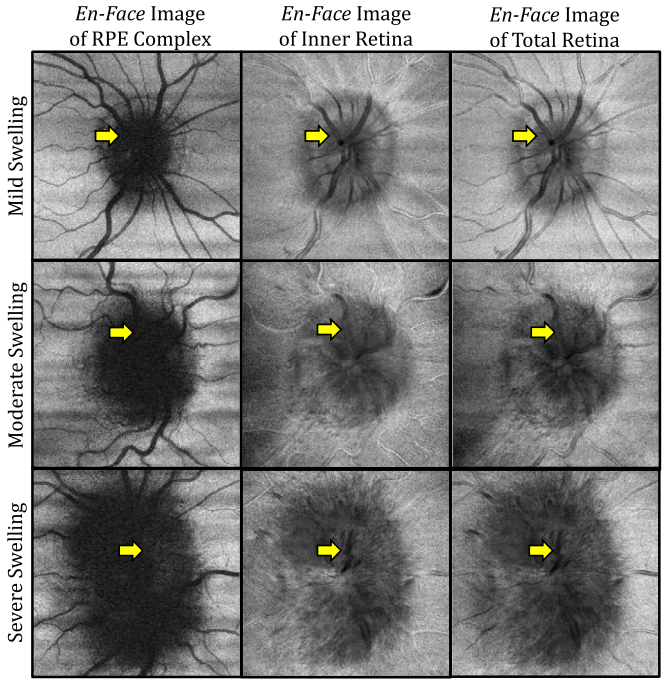


Figure 4. Demonstrations of vessel visibility of en-face images from different retinal layers in mild (*top row*), moderate (*middle row*), and severe (*bottom row*) optic disc swelling (continued from Fig. 1): *left column*, the RPE en-face image; *middle column*, the inner retina en-face image; and *right column*, the total retina en-face image. The *yellow arrows* indicate the vessel visibility changes.

averaging the pixel intensities within the interested layers along each A-scan in the OCT.

Figure 4 demonstrates the vessel visibility of the en-face images from these three layers with mild, moderate, and severe optic disc swelling. As shown in Figure 4, retinal vessels away from the swollen optic disc appear more distinct in the RPE en-face image given the high contrast between the RPE and the vessel shadow, whereas the retinal vessels in the swollen regions appear more distinct in the en-face images of the inner retina and total retina. All three of these images were simultaneously used as inputs to the deep-learning network, as discussed in the next section. Furthermore, all three of these images (as well as the registered fundus photograph) were used to create the reference "ground truth" image used for training and evaluation.

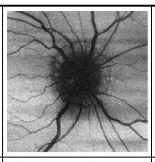
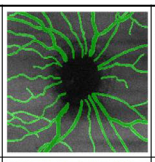
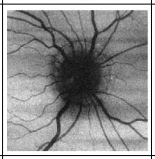
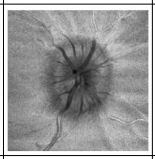
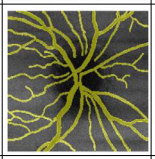
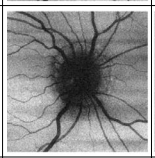
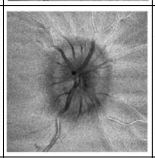
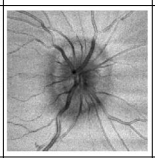
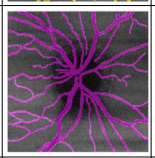
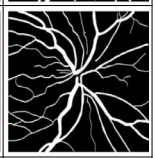
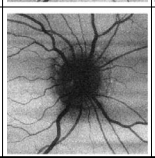
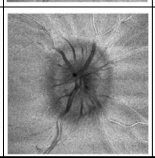
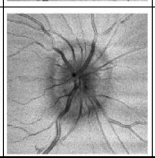
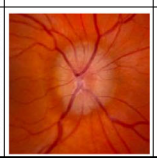
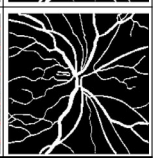
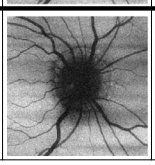
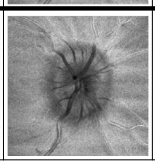
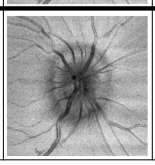
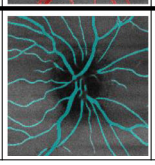
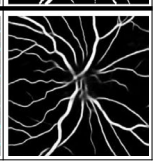
Architecture of the Proposed Deep-Learning Neural Network

As was shown in Figure 2, our proposed deep-learning network is designed to take (patches of) the three en-face images described above as input and output a pixel-based vessel probability map (with values close to 1 indicating a high vesselness probability and values close to 0 indicating a low vesselness probability). The high-level architecture of our proposed deep-learning network is based on a well-known U-shaped deep neural network (U-Net),[22] with modifications to allow for three inputs rather than one and modifications to the number of layers. More specifically, the architecture of the proposed deep-learning approach (Fig. 2) contains a total of 16 neural layers, including 1 concatenation layer, 13 convolution layers, and 2 max-pooling layers. The proposed approach is designed to obtain image features in different resolutions by passing the concatenated input image patches through a contracting path (i.e., the first half; repeatedly uses a combination of convolutional layers, rectified linear units [ReLU], and a max-pooling layer] and then an "up-sampling" path (i.e., the second half; repeatedly uses a combination of convolutional layers, ReLU, and "up-convolutional" layers). Moreover, the corresponding feature maps between both paths are also concatenated in different resolutions. For the input of the proposed approach, location-matched image patches (size: 32 × 32 pixels) were first extracted from the three input en-face images to reduce the computational time as well as computer memory. At the end of the proposed approach, a soft-max layer, which is a 1 × 1 convolution layer, was applied to compute the probability value of the retinal vessel at each corresponding pixel location in the input image patch coordinates. For each patient, these location-matched image patches slid (one pixel each time) through the entire en-face image dimensions, and the outputted small vessel probability maps from all the image patches were stitched together (by averaging all the overlapping regions) to form a complete vessel probability map with the same coordinates as the input en-face images. More details about the deep neural network and its hyperparameters are described in the Appendix.

Manual Tracings of Retinal Blood Vessels and Ground Truth Images

In order to compare how a human expert would segment the vessels with access to various combinations of the input en-face images with the results from the proposed deep-learning-based approach, all the en-face images in both training and testing data sets were independently traced in three separate stages (by J-KW): stage I, referring only to the RPE en-face image (Table 1, column 1); stage II, referring to the combination of the RPE and inner-retina en-face images (Table 1, columns 1–2); and stage III, referring to the combination of the RPE, inner-retinal, and

Table 1. Inputs and Outputs for Manual Tracing (MT) Stages I, II, and III and the Proposed Deep-Learning Approach

	RPE *En-Face* Image	Inner Retina *En-Face* Image	Total Retina *En-Face* Image	ONH-Registered Fundus Photograph	Output Vessel Tracing on the *En-Face* Image	Binary Vessel Map
Manual Tracing (MT) Stage I — RPE *En-Face* Image Only		.....	.....	.....		
Manual Tracing (MT) Stage II — RPE *En-Face* Image + Inner Retinal *En-Face* Image			.....	.....		
Manual Tracing (MT) Stage III — RPE *En-Face* Image + Inner Retinal *En-Face* Image + Outer Retinal *En-Face* Image				.....		
Ground Truth — RPE *En-Face* Image + Inner Retinal *En-Face* Image + Outer Retinal *En-Face* Image + Fundus Photograph						
Proposed Deep Learning Approach — RPE *En-Face* Image + Inner Retinal *En-Face* Image + Outer Retinal *En-Face* Image				.....		

The *green*, *yellow*, *pink*, *red*, and *cyan* vessels overlaid images show the outputs of MT stages I, II, and III; the ground truth; and the proposed approach. The *black* and *white* binary images are also shown at the next column to help visualization.

total-retinal en-face images (Table 1, columns 1–3). Furthermore, to serve as the overall reference standard (also known as the "ground truth") for training and evaluation purposes, the images of the retinal vessels were again separately created by the same expert not only using all the RPE + inner-retinal + total-retinal en-face images but also referring to the registered ONH-centered fundus photographs to obtain the most vessel information (Table 1, columns 1–4). Also note that the proposed deep-learning approach output is also shown (Table 1, row 5) for a comparison. More specific details regarding this manual-tracing process are provided in the Appendix.

Overview of Evaluation Approach

In the training process, leave-one-subject-out cross-validation (for a total of 18 training participants, 17 participants were used for training, 1 patient for validation, and then rotating the selected patient until all the patients had been tested) was used to help design and decide the architecture (using area under the receiver operating characteristic curve, AUC, as the evaluation metric) and hyperparameters in the proposed method. Next, based on the decided network, all the 18 patients were trained together to result in the final proposed deep neural network. The separate testing data set of 18 patients with matched ONH volumes was used with the quantitative measurements (described below) to compare the performances among the manual tracing stages I, II, and III and the proposed deep-learning approach.

Quantitative Evaluation Measurements

Pixel-based evaluation metrics were used to compare each ground truth binary image (0 = background and 1 = vessel object) to the binary results from each of the three OCT-based manual tracing stages, to the output probability map from the proposed deep-learning method, and to a binarized version (using Otsu's thresholding algorithm[29]) of

Table 2. Qualitative Measurements among the Manual Tracing Stages and Proposed Deep Neural Network in the Testing Data Set (Including the Mean Processing Time per Patient)

Methods		Testing Data Set (18 Patients) Mean AUC	Average Precision (AP)	Mean Square Error (MSE)	Mean R^2 Score	Mean Accuracy (ACC)	Mean Processing Time*
Manual tracing, stage I		0.79 ± 0.04	0.73 ± 0.04	0.071 ± 0.01	0.38 ± 0.08	0.93 ± 0.01	11 min 19 s
Manual tracing, stage II		0.83 ± 0.03	0.77 ± 0.03	0.061 ± 0.01	0.46 ± 0.06	0.94 ± 0.01	13 min 51 s
Manual tracing, stage III		0.85 ± 0.03	0.78 ± 0.04	0.061 ± 0.01	0.47 ± 0.08	0.94 ± 0.01	14 min 7 s
Proposed deep-learning approach	Probability map†	0.96 ± 0.02	0.84 ± 0.07	0.047 ± 0.01	0.59 ± 0.09	N/A	1 min 5 s
	Binarymap†	0.83 ± 0.05	0.77 ± 0.05	0.061 ± 0.01	0.46 ± 0.10	0.94 ± 0.01	

*Hardware details: A Linux machine is used with single GPU (NVIDIA GeForce GTX 1080 Ti) and 128 GB RAM. The time to train the proposed neural network, including all the 18 patients in the training data set, was 2 hours 2 minutes 12 seconds.

†The probability map is the direct output from the proposed deep neural network; the binary map is automatically obtained using the Otsu thresholding algorithm. (Note: The outputs from manual tracings are inherently binary maps.)

the proposed deep-learning method. More specifically, given a ground truth image and a corresponding binary map from another approach (e.g., an OCT-based manual tracing stage or a thresholded version of the output probability map), the true positive (TP), false positive (FP), true negative (TN), false negative (FN), and the total number of pixels ($K = TP + FP + TN + FN$) can be computed. We correspondingly computed the area under the receiver operating characteristic curves (AUC), for all approaches (manual tracings from the three stages, the original probability map, and the Otsu-thresholded binary map), to measure the TP rate against the FP rate across all possible threshold values. We also computed the average precision (AP) (for all approaches) to quantify the relationship between the precision (P) and recall (R) across all possible threshold values; $AP = \sum_n (R_n - R_{n-1})P_n$, where $R = TP/(TP + FN)$, $P = TP/(TP + FP)$, and n is the nth threshold value. For the binary-only results (manual tracings and Otsu-thresholded probability map but not the original probability map), we also computed the accuracy (ACC) to compute the ratio of correctly classified pixels to the total number of pixels in the OCT en-face image; $ACC = (TP + TN)/K$. In addition, for all approaches, we computed the mean squared error (MSE) as a measure of the label distance between the approaches and the ground truth; $MSE = \frac{1}{K}\sum_{i=0}^{K-1}(\hat{y}_i - y_i)^2$, where $\hat{y}_i$ and y_i represent the predicted label value at pixel location i from the approach and the truth label, respectively; and the coefficient of determination (R^2 score) to estimate how well the approach is with respect to the ground truth in the sense of regression; $R^2 = 1 - \frac{\sum_{i=0}^{K-1}(\hat{y}_i - y_i)^2}{\sum_{i=0}^{K-1}(\bar{y} - y_i)^2}$, where $\bar{y} = \frac{1}{K}(\sum_{i=0}^{K-1} y_i)$.

Results

The mean area under the ROC curves (AUC) of probability maps from a leave-one-subject-out cross-validation over the 18 patients in the training data set was 0.93. Using the independent imaging data from the test set, the mean AUCs for the manual tracing stages I, II, and III were 0.79, 0.83, and 0.85, respectively; for the proposed deep-learning approach, the mean AUC was 0.96 for the direct output of the vessel probability map, but the mean AUC was 0.83 when the probability maps were converted to binary maps using the Otsu algorithm.[29] For the AP, the results of the manual tracing stages I, II, and III were 0.73, 0.77, and 0.78, respectively; the results of the proposed method were 0.84 and 0.77 for the probability map and binary map, respectively. Other results among the manual tracing stages I, II, and III and the probability map as well as the binary map from the proposed method were as follows: MSE, 0.071, 0.061, 0.061, 0.047, and 0.061; mean coefficient of determination (R^2), 0.38, 0.46, 0.47, 0.59, and 0.46; and mean accuracy (ACC), 0.93, 0.94, 0.94, N/A (Not applicable), and 0.94, respectively. Table 2 shows the summary of the quantitative results.

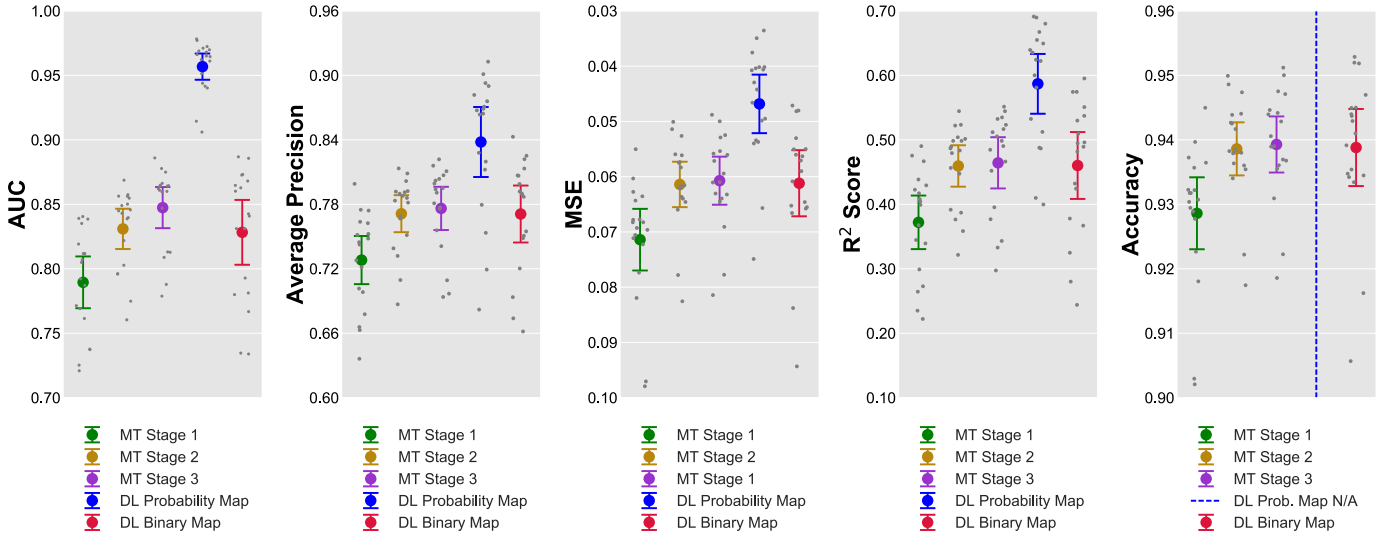


Figure 5. Data dot plots for the measurements of area under ROC curve (AUC), average precision (AP), mean square error (MSE), mean coefficient of determination (R^2), and mean accuracy (ACC) with 95% confidence intervals: *green*, manual tracing (MT) stage I; *dark yellow*, MT stage II; *purple*, MT stage III; *blue*, the probability map from the proposed deep learning (DL) approach; and *red*, the binary map, which is obtained using Ostu algorithm) from the DL approach.

Figure 5 displays all the data points from the 18 testing participants with a 95% confidence interval of the mean for the quantitative results in the testing data set. The direct output (i.e., the vessel probability maps) of the proposed deep-learning approach shows the best performance in all five quantitative measurements. Furthermore, for all the testing participants, the thresholded binary map of the proposed deep-learning approach provides better vessel segmentation results than manual tracing stage I (tracing on only the RPE en-face image) for all five measurements of all 18 testing participants (additional details in Appendix). The results from manual tracing stages II and III and that of the thresholded binary map of the proposed deep-learning approach were similar. More specific details regarding the subject-wise comparison between different approaches are provided in the Appendix.

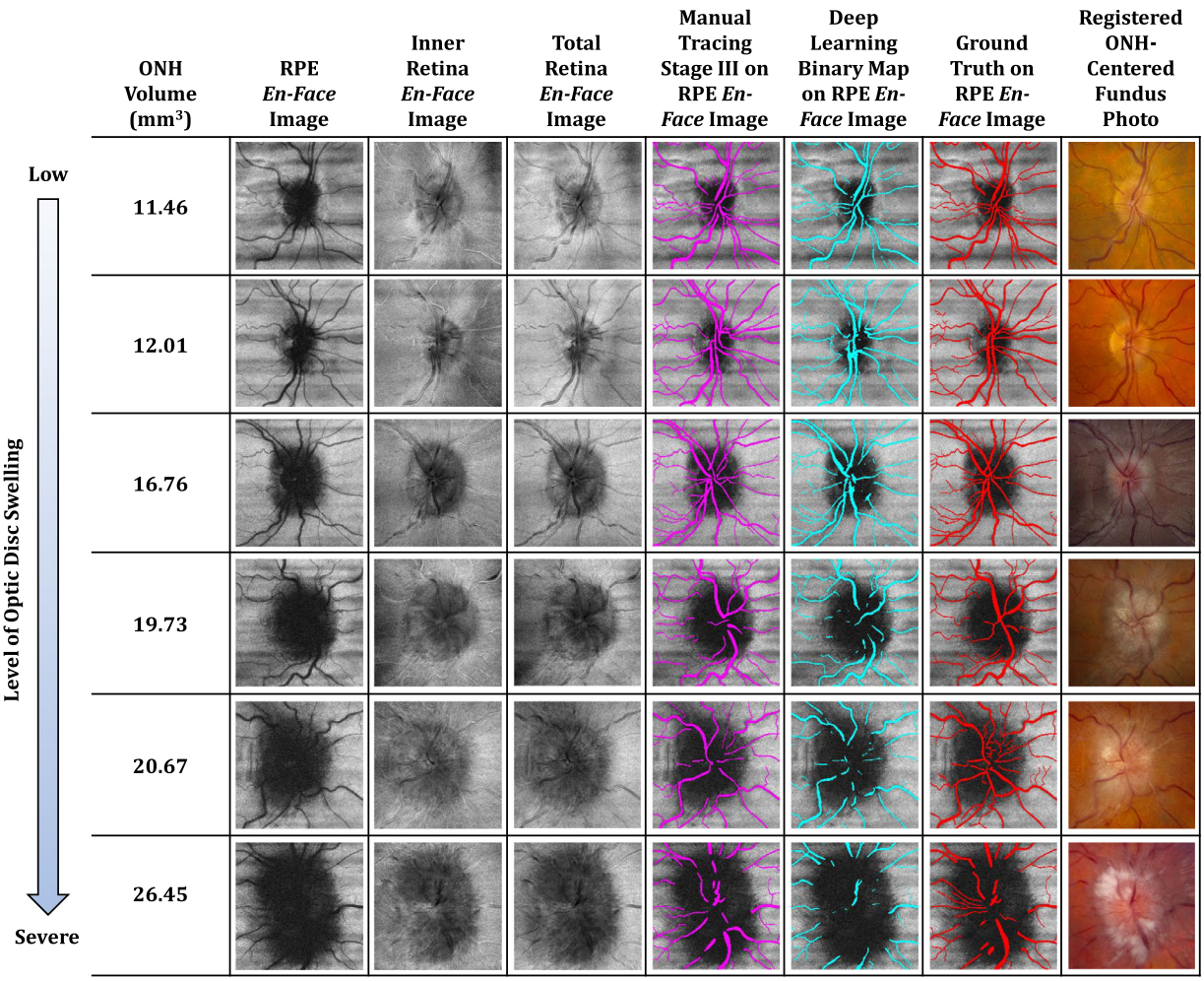
The mean processing times per patient for each method in the testing data set were also recorded (Table 2). The three manual tracing stages using Adobe Illustrator Draw (version 4.6.1; Adobe Systems, Inc., San Jose, CA, USA) on an iPad (Apple, Inc., Cupertino, CA, USA) required at least 11 minutes for each patient in the testing data set (precisely, 11 minutes 19 seconds, 13 minutes 51 seconds, and 14 minutes 7 seconds for stages I, II, and III, respectively). The computation for the proposed deep neural network was performed using a Linux machine with single GPU, NVIDIA GeForce GTX 1080 Ti (NVIDIA Corporation, Santa Clara, CA, USA) and 128 GB RAM. The mean processing time for each patient in the testing data set for the proposed deep neural network was 1 minute 5 seconds. (Note: The total time of training the proposed neural network, including all the 18 patients in the training data set, was 2 hours 2 minutes 12 seconds.)

Six patients with various levels of optic disc swelling (the ONH volume range is from 11.46 mm^3 [the top row] to 26.45 mm^3 [the bottom row]) are shown as qualitative results in Table 3. The en-face images of the RPE complex, the inner retina, and the total retina are listed in the table to show the shadow region growth in different degrees of swelling. The manual tracing stage III (highlighted in purple), the proposed deep-learning approach binary maps (highlighted in cyan), and the ground truth (highlighted in red) are displayed in the next three columns, respectively. The corresponding ONH-registered fundus photographs are also added at the last column for reference.

Discussion

Although a preliminary OCT-based vessel segmentation for swollen optic disc had been utilized in our previous studies as a part of the preprocessing,[17] this is the first time that we directly focus on OCT to reveal the obscured vessels from the image shadow by using a modified deep neural network that simultaneously considers multiple OCT en-face images from various retinal layers as its inputs. Furthermore, while use of an RPE en-face image would traditionally

Table 3. Examples of Vessel Segmentation in Six Patients with Various Levels of Optic Disc Swelling



The patients are arranged based on the ONH volumes from the top to the bottom rows.

be considered sufficient for visualization of projected retinal vessels in OCT (especially in nonswollen cases), our results also show that having access to multiple en-face views is important for even a human expert to properly visualize the vessels. Overall, our deep-learning approach (even when using a nonoptimized thresholding approach to generate the binary maps) performs at least as well as one would expect from a human expert having access to multiple en-face OCT views and better than what one would expect from a human expert having access to only the RPE en-face image (the traditional approach). In fact, when using the probability map itself (rather than the binary map), perhaps surprisingly, our results suggest that the deep-learning approach even outperforms the human expert in segmenting the retinal vessels with access to only OCT en-face images (discussed further below).

Effectively, having multiple simultaneous input images enabled the proposed deep-learning network to learn to extract the vessel information from the region at the retinal layer with better signal response to compensate the regions that are eclipsed. Figure 4 shows examples that the en-face images around the optic disc of the inner retina contain substantially more vessel information than the one from the RPE complex in cases with optic disc swelling; however, the vessel visibility at the peripheral region is still the clearest in the RPE en-face image compared with the others. Since the vessels shown in the en-face images are the mean intensity values of the shadow from the superficial vessels, it is not surprising that the vessel visibility in the RPE en-face image is considerably deteriorated around the optic disc in which the OCT signal is greatly weakened by passing through the swollen inner retinal layers. Overall, the total-retinal en-face image displays in-between vessel clarity, which is potentially helpful for the proposed deep-learning approach having an extra reference when the vessel

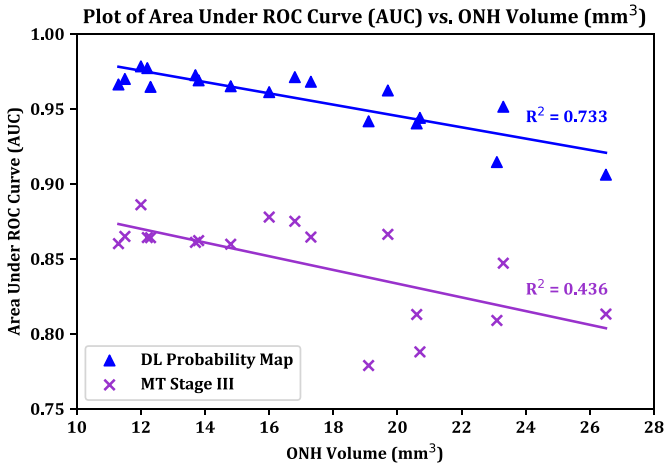


Figure 6. Scatterplots of 18 testing participants for displaying the relationships between the AUC and ONH volume from the manual tracing (MT) stage III (the best performance in all three manual tracing stages; shown as *magenta crosses*) and proposed deep-learning (DL) approach probability map (shown as *cyan triangles*).

patterns are inconsistent between the RPE and inner-retinal en-face images.

When the optic disc swelling is severe, in addition to the optic disc elevation, other pathogenic conditions, such as hemorrhage and nerve fiber layer infarcts (cotton-wool spots), may appear.[1,30] Meanwhile, the vessel appearance may also be affected: the retinal vessels can be seen extremely blurred, discontinuous, and/or covered by the cotton-wool spots. Under these circumstances, it is sometimes tricky to clearly define the boundary of the vessels in the OCT en-face images. The bottom row in Table 3 illustrates the difficulty of the process of tracing complete vessel trees even with the extra information from the fundus photograph (i.e., the ground truth). Figure 6 shows that the performances of both manual tracing and proposed method gradually decline when the ONH volume increases; however, the proposed method still seems visually more robust than the manual tracing stage III in the testing data set.

Because the degree of optic swelling influences the difficulty of the vessel segmentation, it is important to keep in mind our overall reported results are, in part, reflective of the distribution of swelling levels tested. Our training set (and, correspondingly, our test set because of the volume-matching process) likely reflected a higher proportion of cases with moderate-to-severe optic disc swelling than one might encounter in clinical practice. If a larger percentage of cases with milder swelling were evaluated, we would actually expect to obtain better overall performance numbers. However, while we evaluated the approach on a reasonably balanced data set of cases with optic disc swelling, one limitation of our work is that we have not quantitatively evaluated the result of the proposed approach on a separate normative data set. Nevertheless, based on visually assessing the results from separately applying our trained neural network on eyes with no apparent swelling (having approximate optic nerve head volume of 8–10 mm^3), we found that the proposed approach still successfully segments the major vessels in such eyes. Thus, while not quantitatively evaluated on eyes without optic disc swelling, we still expect the approach to be robust in these cases as well (where traditional approaches involving only use of an RPE en-face image may already work). Also note that in computing the ONH volumes used for estimating the degree of optic disc swelling, we were unable to correct for ocular magnification due to the lack of axial length information, and thus the reported volumetric measures are technically approximations. However, while we did use the estimated measures to provide a similar distribution of swelling severity in the training and testing data sets as well as to provide some insight into the dependence of the performance on degree of optic disc swelling, correcting for ocular magnification is not needed for appropriately training the algorithm.

As previously mentioned, the quantitative results in Table 2 and Figure 5 demonstrate that the probability map from the proposed deep neural network has the best performance compared with all the other methods. However, after thresholding the probability map (by the Otsu algorithm) into a binary map, the performance from the neural network declines to a similar level of manual tracing stages II and III. This is because of a part of vessel information gets lost in the process of thresholding. For example, in the probability map, some thinner vessels may have a smaller probability, which could be thresholded to become background in the binary map. The performance gap between the probability and binary maps can possibly be shortened by developing more sophisticated thresholding methods. Using regionally adaptive threshold values based on vessel continuity rather than one global threshold value is one of the options to improve.

Regarding the stages of manual tracing, it is worthy to note that we strictly followed the order that we described in the Appendix to avoid the extra information from the higher stages, especially for the ground truth images, to prejudice the tracing in the lower stages. Table 2 and Figure 5c have shown that, for the same human expert, only using the RPE en-face image to trace the retinal vessels (i.e., the traditional method) provides the worst performance among all three stages for all the measurements. After adding consideration of the inner-retinal and total-retinal en-face images, the performance in stages II and III noticeably increased. Also, our manual tracings are performed using Adobe Illustrator Draw (Adobe Systems, Inc., version 4.6.1)

on an iPad, which allows us to overlay all input en-face images on each other so that the vessel information can be intuitively accumulated from all the input retinal layers. The image-overlay method is potentially a more robust method than separately tracing these en-face images and then adding the results together.

Also note that the time-consuming multistage manual tracing stages were one factor that limited the size of our training and test sets (18 cases each). While 18 cases in a training set would likely be considered too small for an image-level classification task for a deep-learning approach (e.g., determining the cause of the swelling), since our work focused on a pixel-level task in determining the probability of each pixel being a vessel (in combination with our use of a U-Net-based architecture), we were able to have sufficient data to be able to train the approach to provide a good performance overall on the independent test set. However, confirming the performance on a larger data set (perhaps using less time-consuming multistage reference standard) would be useful in future work.

It is conceivable that our proposed deep-learning approach can be extended to detect other tubular objects in the OCT volumes. True three-dimensional vessel segmentation (instead of on the projected en-face planes) could be a subject of future study. Also, analyses of retinal folds[16,17,31] are believed to be one of the key features to scrutinize the mechanisms of stress/strain at the ONH region. Automatically detecting the retinal folds and further quantifying them can be another possible extension for our proposed neural network.

Acknowledgments

Supported in part by I01 RX001786, I50 RX003002, and R01 EY023279.

A preliminary version of this work was presented as an abstract at ARVO 2019 (Islam et al., ARVO Abstract #1510, 2019).

Disclosure: **M.S. Islam**, None; **J.-K. Wang**, None; **S.S. Johnson**, None; **M.J. Thurtell**, None; **R.H. Kardon**, Fight for Sight, Inc. (S), Department of Veterans Affairs Research Foundation, Iowa City, IA (S); **M.K. Garvin**, University of Iowa (P)

References

1. Van Stavern GP. Optic disc edema. *Semin Neurol*. 2007;27:233–243.
2. Frisén L. Swelling of the optic nerve head: a staging scheme. *J Neurol Neurosurg Psychiatry*. 1982;45:13–18.
3. Hayreh SS. Optic disc edema in raised intracranial pressure, V: pathogenesis. *Arch Ophthalmol*. 1977;95:1553–1565.
4. Trobe JD. Papilledema: the vexing issues. *J Neuroophthalmol*. 2011;31:175–186.
5. Scott CJ, Kardon RH, Lee AG, et al. Diagnosis and grading of papilledema in patients with raised intracranial pressure using optical coherence tomography vs clinical expert assessment using a clinical staging scale. *Arch Ophthalmol*. 2010;128:705–711.
6. Echegaray S, Zamora G, Yu H, Luo W, Soliz P, Kardon R. Automated analysis of optic nerve images for detection and staging of papilledema. *Invest Ophthalmol Vis Sci*. 2011;52:7470–7478.
7. Klein KA, Hedges TR, III. Optical coherence tomography and optic nerve edema. In: Grzybowski A, Barboni P, eds. *OCT in Central Nervous System Diseases: The Eye as a Window to the Brain*. Basel, Switzerland: Springer International Publishing; 2016:35–68.
8. Menke MN, Feke GT, Trempe CL. OCT measurements in patients with optic disc edema. *Invest Ophthalmol Vis Sci*. 2005;46:3807–3811.
9. Rebolleda G, Diez-Alvarez L, Casado A, et al. OCT: new perspectives in neuro-ophthalmology. *Saudi J Ophthalmol*. 2015;29:9–25.
10. Chen JJ, Costello F. The role of optical coherence tomography in neuro-ophthalmology. *Ann Eye Sci*. 2018;3:35–35.
11. Sibony PA, Kupersmith MJ, James Rohlf F. Shape analysis of the peripapillary RPE layer in papilledema and ischemic optic neuropathy. *Invest Ophthalmol Vis Sci*. 2011;52:7987–7995.
12. Wang J-K, Kardon RH, Kupersmith MJ, Garvin MK. Automated quantification of volumetric optic disc swelling in papilledema using spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2012;53:4069–4075.
13. Kaufhold F, Kadas EM, Schmidt C, et al. Optic nerve head quantification in idiopathic intracranial hypertension by spectral domain OCT. *PLoS One*. 2012;7:e36965.
14. OCT Sub-Study Committee for the NORDIC Idiopathic Intracranial Hypertension Study Group. Baseline OCT measurements in the idiopathic intracranial hypertension treatment trial, part I: quality control, comparisons, and variability. *Invest Ophthalmol Vis Sci*. 2014;55:8180–8188.
15. Wang J-K, Sibony PA, Kardon RH, Kupersmith MJ, Garvin MK. Semi-automated 2D Bruch's membrane shape analysis in papilledema using spectral-domain optical coherence tomography. In: Gimi B, Molthen RC, eds. *Medical Imaging 2015: Biomedical Applications in Molecular,*

Structural, and Functional Imaging. Bellingham, United States: SPIE - International Society for Optics and Photonics. 9417;2015:941721-1–941721-10.
16. Sibony PA, Kupersmith MJ, Feldon SE, et al. Retinal and choroidal folds in papilledema. *Invest Ophthalmol Vis Sci*. 2015;56:5670–5680.
17. Agne J, Wang J-K, Kardon RH, Garvin MK. Automatic detection of folds and wrinkles due to swelling of optic disc. In: Cardoso J, Arbel T, Melbourne A, et al., eds. *Fetal, Infant and Ophthalmic Medical Image Analysis. FIFI 2017, Ophthalmic Medical Image Analysis Fourth International Workshop OMIA 2017*. Vol 10554. Cham, Switzerland: Springer; 2017:235–242.
18. Niemeijer M, Garvin MK, van Ginneken B, Sonka M, Abràmoff MD. Vessel segmentation in 3D spectral OCT scans of the retina. In: Reinhardt JM, Pluim J, eds. *Medical Imaging 2008: Image Processing*. Bellingham, United States: SPIE - International Society for Optics and Photonics; 2008:69141R-1–69141R-8.
19. Hu Z, Abràmoff MD, Kwon YH, Lee K, Garvin MK. Automated segmentation of neural canal opening and optic cup in 3D spectral optical coherence tomography volumes of the optic nerve head. *Invest Ophthalmol Vis Sci*. 2010;51:5708–5717.
20. Li Y, Gregori G, Lam BL, Rosenfeld PJ. Automatic montage of SD-OCT data sets. *Opt Express*. 2011;19:26239–26248.
21. Deng W, Antony B, Sohn EH, Abramoff MD, Garvin MK. Segmentation of the retinal vasculature within spectral-domain optical coherence tomography volumes of mice. In: Liu J, Trucco E, Xu Y, Chen XGM, eds. *Proceedings of the Ophthalmic Medical Image Analysis Second International Workshop, OMIA 2015, Held in Conjunction with MICCAI 2015*. Cham, Switzerland: Springer; 2015:65–72.
22. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer; 2015:234–241.
23. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
24. Fraz MM, Remagnino P, Hoppe A, et al. Blood vessel segmentation methodologies in retinal images: a survey. *Comput Methods Programs Biomed*. 2012;108:407–433.
25. Melinscak M, Prentasic P, Loncaric S. Retinal vessel segmentation using deep neural networks. In: Braz J, Battiato S, Imai F, eds. *International Conference on Computer Vision Theory and Applications (VISAPP 2015)*. Setúbal, Portugal: SciTePress - Science and Technology Publications2015:577–582.
26. Fu H, Xu Y, Lin S, Wong DWK, Liu J. DeepVessel: retinal vessel segmentation via deep learning and conditional random field. In: Ourselin S, Joskowicz L, Sabuncu MR, Gozde U, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2016*. Vol 9901. Cham, Switzerland: Springer; 2016:132–139.
27. Jordan KC, Menolotto M, Bolster NM, Livingstone IA, Giardini ME. A review of feature-based retinal image analysis. *Expert Rev Ophthalmol*. 2017;12:207–220.
28. Moccia S, De Momi E, El Hadji S, Mattos LS. Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics. *Comput Methods Programs Biomed*. 2018;158:71–91.
29. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. 1979;9:62–66.
30. Sadun AA, Wang MY. *Abnormalities of the Optic Disc*. Vol 102. Amsterdam, Netherlands: Elsevier; 2011.
31. Kupersmith MJ, Sibony PA, Feldon SE, et al. The effect of treatment of idiopathic intracranial hypertension on prevalence of retinal and choroidal folds. *Am J Ophthalmol*. 2017;176:77–86.
32. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv Prepr* 2014. *arXiv1412.6980*.
33. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929–1958.
34. Jia Y, Shelhamer E, Donahue J, et al. Caffe: convolutional architecture for fast feature embedding. *arXiv Prepr* 2014. *arXiv1408.5093*.

Appendix

Additional Details Regarding the Modified U-Net Architecture

As displayed in Figure 2, the first layer of the neural network first concatenates the three input location-matched en-face image patches together to form an image blob (dimensions: 32 × 32 × 3). Next, for the contracting path, a combination of a convolutional layer (with the filter size of 3 × 3 pixels, 32 feature channels, and zero padding of 1 pixel for correcting the boundary effect) plus a rectified linear unit (ReLU) is used twice before a max-pooling layer (filter size: 2 ×

2 pixels, stride: 2 pixels). It is worthy of note that the feature map dimensions have been down-sampled to 16 × 16 pixels due to the max-pooling layer, but the depth has increased to 64 channels. A similar process is then repeated once, and the feature map dimensions achieve 8 × 8 pixels with 128 channels at the end of the contracting path.

In the up-sampling path, the "up-convolutional" layers are applied to double the dimensions of the feature map, and then the outputting feature maps are concatenated with the ones in the contracting path to "reconsider" past features (i.e., the gray arrows in the Fig. 2). Following by similar convolutional processes again but with different amounts of feature channels, the feature map dimensions finally recover back to 32 × 32 pixels and with 32 channels. The last layer of the neural network is a soft-max layer, which is a 1 × 1 × 32 convolutional operator, to estimate the probability value (range from 0 to 1) of the retinal vessel at each pixel location.

As mentioned in the Methods, the architecture of the proposed neural network is designed using a leave-one-subject-out cross-validation in the training data set (18 patients), and the hyperparameters include a first-order gradient-based optimization algorithm (Adam[32]) with momentum 1 (β_1) of 0.9, momentum 2 (β_2) of 0.999, delta (ε) of 10^{-8}, learning rate (α) of 0.01, and gamma (γ) of 0.9. In addition, a simple technique of dropping out half of the neurons after convolution layers is adopted to prevent data overfitting.[33] The proposed deep neural network is implemented using a Caffe library (1.0.0)[34] on a Linux machine with single GPU (NVIDIA GeForce GTX 1080 Ti), Intel Core i7-6800K Broadwell-E 6-Core 3.4G, and 128 GB RAM.

After the architecture of the proposed neural network has been decided and all the hyperparameters are fixed from the cross-validation step in the training data set, the proposed neural network is retrained by using all the 18 patients from the training data set and officially tested in the volume-matched testing data set. The step of completely separating the training and testing data sets helps reduce the overall bias of the proposed method.

Additional Details Regarding the Manual Tracings of Retinal Blood Vessels

Tools and more precise steps of manually tracing the retinal blood vessels are described as follows. Three stages of human tracing were designed to compete with the proposed deep-learning approach. In stage I, for each OCT image in the training set, its RPE en-face image was first loaded into an iPad Pro tablet (Apple, Inc.) using Adobe Illustrator Draw (Adobe Systems, Inc., version 4.6.1) and saved as an independent profile. Next, the visible vessels in this en-face image were manually highlighted by the expert drawing on the tablet using an Apple Pencil (Apple, Inc.) (Table 1, row 1). Then, the same process was repeatedly applied to the rest of the OCT images in the training set. In stage II, to have an independent vessel tracing from the previous stage, for each OCT image in the training set, its RPE en-face image was reloaded and then saved as a clean, separate profile using Adobe Illustrator Draw. Then, the corresponding inner-retina en-face image was loaded as the second image layer so that the expert could trace the retinal vessels on the RPE en-face image meanwhile accessing the vessel information from the inner-retina en-face image (Table 1, row 2). Then, the same process was repeatedly applied to the rest of the OCT images in the training set. In stage III, similar processes from stage II were repeated again and adding the third image layer from the total-retina en-face image (Table 1, row 3). Finally, the ground truth image was created by simultaneously accessing four images (all the three en-face images + the registered fundus photographs) (Table 1, row 4). The same procedures were next utilized for the images in the testing data set. Note that because the amount of vessel information increases starting from stage I, stage II, and stage III and maximizing at the ground truth images, the order of our manual tracings helped ensure that the manual tracings in lower-level stages had less of a chance to be affected by the prior knowledge from the higher-level stages. Also note that the proposed deep-learning approach output is also shown (Table 1, row 5) for a comparison.

Additional Details Regarding the Quantitative Comparison between Different Approaches

Since the performances of manual tracing stages II and III are similar (as shown in Fig. 5), only the results from stage III are kept for further comparisons. Figures A1a–A1c show the comparison between quantitative values of different evaluation metrics for all the 18 testing participants when different approaches are used. The lines are found by joining the quantitative values of evaluation metrices from two different approaches (e.g., manual tracing stage I and deep learning binary map for Fig. A1a, manual tracing stage I and manual tracing stage III for Fig. A1b, manual tracing stage III versus deep-learning binary map for Fig. A1c). An upward slope indicates a better

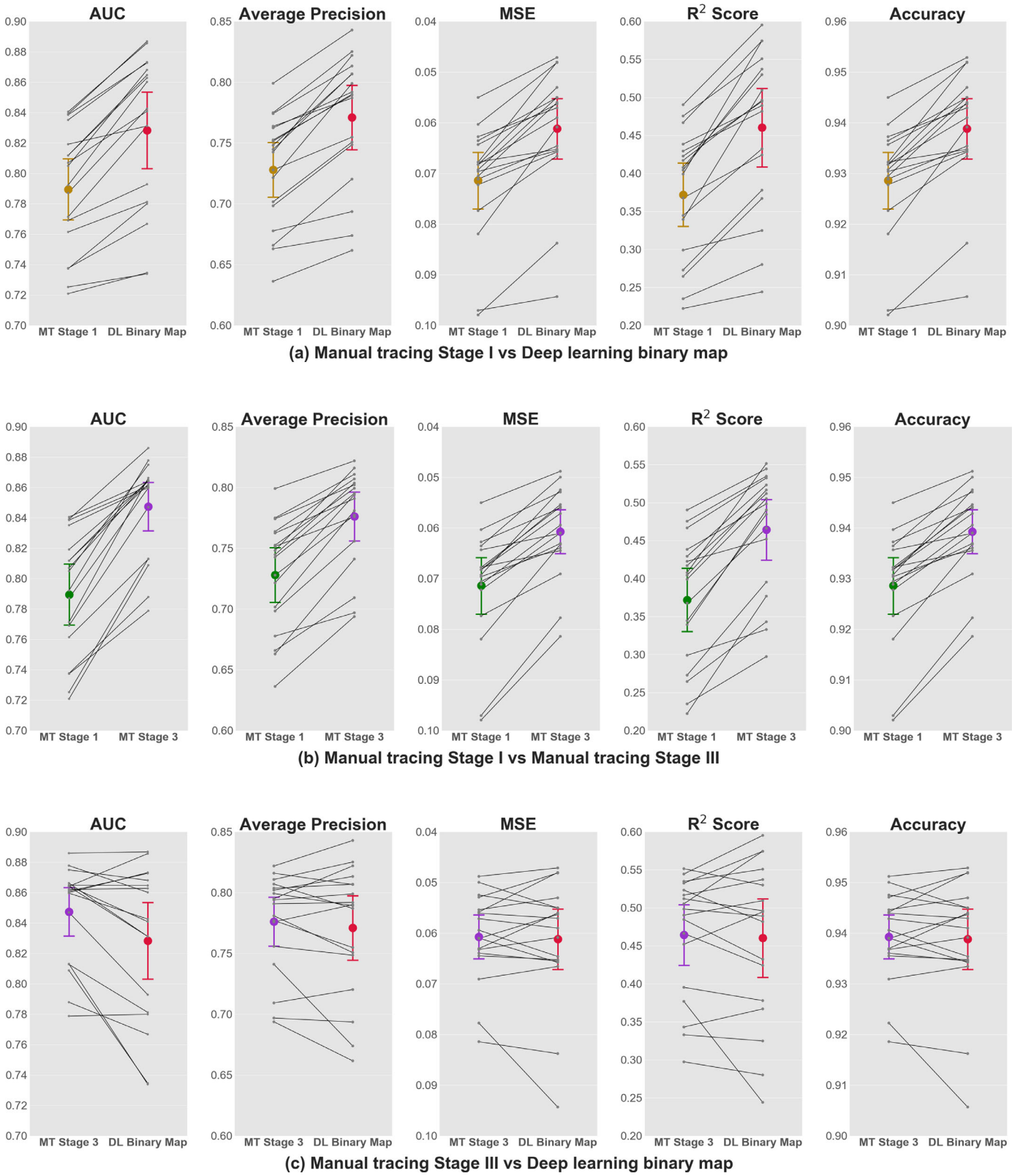


Figure A1. Trajectory connection dot plots for the measurements of area under ROC curve (AUC), average precision (AP), mean square error (MSE), mean coefficient of determination (R^2), and mean accuracy (ACC) with 95% confidence intervals: *Dark yellow and green*, manual tracing (MT) stage I (in Fig. A1(a) and (b) respectively); *purple*, MT stage III; and *red*, the binary map (which is obtained using Ostu algorithm) from the DL approach. (a) Trajectory connection between the MT stage I and DL binary map results. (b) Trajectory connection between the MT stage I and III results. (c) Trajectory connection between the MT stage III and DL binary map results.

performance using the second approach in the figure and a downward slope indicates the first approach provides better performance. Based on all the quantitative results, Figure A1a suggests that the proposed deep-learning approach binary map provides better vessel segmentation results than the manual tracing stage I by showing that the majority of the slopes are positive. Similarly, Figure A1b suggests that the manual tracing stage III has better performance than stage I does. Finally, Figure A1c suggests that the performances between the deep-learning binary map and the manual tracing stage III are similar.