

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Automated macular OCT retinal surface segmentation in cases of severe glaucoma using deep learning

Hui Xie, Jui-Kai Wang, Randy Kardon, Mona Garvin, Xiaodong Wu

Hui Xie, Jui-Kai Wang, Randy H. Kardon, Mona K. Garvin, Xiaodong Wu, "Automated macular OCT retinal surface segmentation in cases of severe glaucoma using deep learning," Proc. SPIE 12032, Medical Imaging 2022: Image Processing, 120320W (4 April 2022); doi: 10.1117/12.2611859

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

Automated Macular OCT Retinal Surface Segmentation in Cases of Severe Glaucoma Using Deep Learning

Hui Xie^{a1}, Jui-Kai Wang^{2,b1,a1}, Randy H. Kardon^{2,b1}, Mona K. Garvin^{2,a1}, and Xiaodong Wu^{a1}

^aDepartment of Electrical and Computer Engineering,

^bDepartment of Ophthalmology and Visual Sciences,

¹The University of Iowa, Iowa City, IA, USA

²The Iowa City VA Center for the Prevention and Treatment of Visual Loss, IA, USA

ABSTRACT

Glaucoma is one of the leading causes of permanent blindness due to optic nerve damage. Optical coherence tomography (OCT) has become an important clinical tool for assessing structural damage from the loss of neurons. Traditional 2D and 3D methods have been successfully applied to quantify inner retinal layer thickness. However, these methods show less reliable segmentation in severe glaucoma when the retinal layers have become thin and violate algorithm assumptions. Deep learning (DL) is an alternative image analysis approach due to its powerful ability to extract features directly from data. State-of-the-art DL segmentation approaches can achieve sub-pixel accuracy at multiple retinal surfaces in OCT scans from normal eyes. However, limitations, such as spike-like segmentation errors (showing as high Hausdorff distances) and lack of contextual information from the input image, still need to be improved. To address these limitations, three novel solutions were proposed in this study. First, for data augmentation, we reconstructed more B-scans by reassembling A-scans at the vertical and jittered planes to expose DL to a greater variety of features encountered in OCT. Second, smoothed and contrast-enhanced images of each three adjacent B-scans were concatenated to provide a six-channel input image stack to the neural network with contextual information. Finally, we merged the predicted surfaces from both horizontal and vertical B-scans while maintaining retinal topological order. In our independently tested dataset, which included eyes with severe glaucoma, the proposed approach outperformed the state-of-the-art methods in mean absolute surface distances, Dice coefficients, and Hausdorff distance at multiple surfaces.

Keywords: OCT, glaucoma, deep learning, surface segmentation, data augmentation

1. INTRODUCTION

Glaucoma is one of the leading causes of permanent blindness due to optic nerve damage.¹ In addition to functional tests (e.g., standard automated perimetry, SAP) and intraocular pressure exams, optical coherence tomography (OCT) is often used for assessing retinal structural damage. Studies have suggested that progressive macular ganglion cell plus inner plexiform layer (GCIPL) thinning can be identified before corresponding retinal nerve fiber layer thinning at the optic disc.² To quantify retinal structural change, traditionally, 2D models and graph-search-based approaches^{3,4} have been widely used to segment the retinal layers in OCT. However, these methods have less reliable segmentation results in severe glaucoma cases because very thin retinas often violate underlying algorithm assumptions.

Recently, deep learning (DL) has become popularized as an alternative approach for OCT layer segmentation due to its powerful ability to directly extract latent features from the image.⁵ However, because of limitations imposed by the computational ability of standard computers and the relatively small training/testing datasets (compared to non-medical images), most DL methods applied to OCT-segmentation rely on the use of single 2D B-scans as input images. As an exploration of trying to use 3D information, Liu et al.⁶ suggested using a hybrid 2D-3D network to capture displacements of B-scans to offset the 2D segmentation accuracy. At present, the fully convolutional regression network (FCRN)^{7,8} and the internal-point-method (IPM) for surface segmentation⁹ are two examples of the most advanced OCT segmentation approaches. Although both approaches can achieve

Send correspondence to Jui-kai Wang/Xiaodong Wu: E-mail: {jui-kai-wang/xiaodong-wu}@uiowa.edu

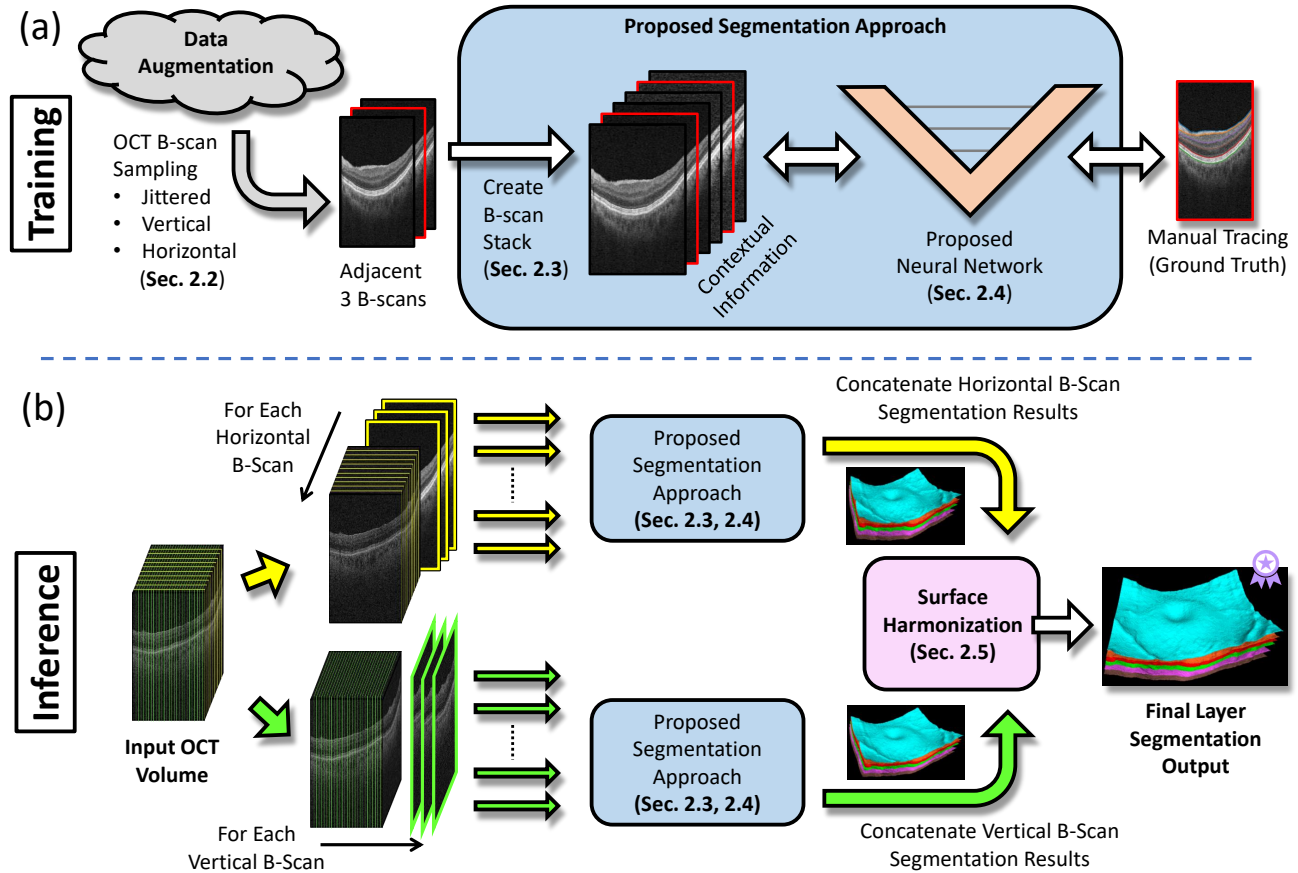


Figure 1. Flowcharts of the proposed method in training and inference.

sub-pixel accuracy at multiple retinal surfaces, spike-like segmentation errors (i.e., high Hausdorff distances) and lack of 3D contextual information are significant limitations that need to be addressed.

In this study, we proposed a novel DL framework to address the limitations of current state-of-the-art DL approaches and also to address the need for more robust OCT surface segmentation in cases of severe glaucoma with very thin layers. To this end, an innovative data augmentation method was introduced to reconstruct OCT B-scans by reassembling A-scans horizontally, vertically, and in a jittered plane to enrich input image patterns. In addition, local 3D contextual information was added to the input of the neural network. Within the novel DL framework, segmented surfaces were guaranteed correct topological orders without crossing each other in the 3D surface context. To evaluate the performance of the novel DL framework, we computed mean absolute surface distance (MASD), surface Hausdorff distance, and the Dice coefficient of GCIPL among the proposed method, graph-search-based method,⁴ and the FCRN method.^{7,8} Comparisons to manual tracing at six surfaces were also made, which included the ILM (internal limiting membrane), RNFL-GCL (the surface between retinal nerve fiber layer and ganglion cell layer), IPL-INL (the surface between inner plexiform layer and inner nuclear layer), OPL-HFL (the surface between outer plexiform and Henle's fiber layers), BMEIS (the boundary of myoid and ellipsoid of inner segments), and OB-RPE (the outer boundary of the retinal pigment epithelium).

2. METHODS

2.1 Overview

Fig. 1 illustrates overall flowcharts describing the proposed method in training and inference for robust surface segmentation in macular OCT scans (either $200 \times 200 \times 1024$ or $512 \times 128 \times 1024$ volume scans obtained by Zeiss

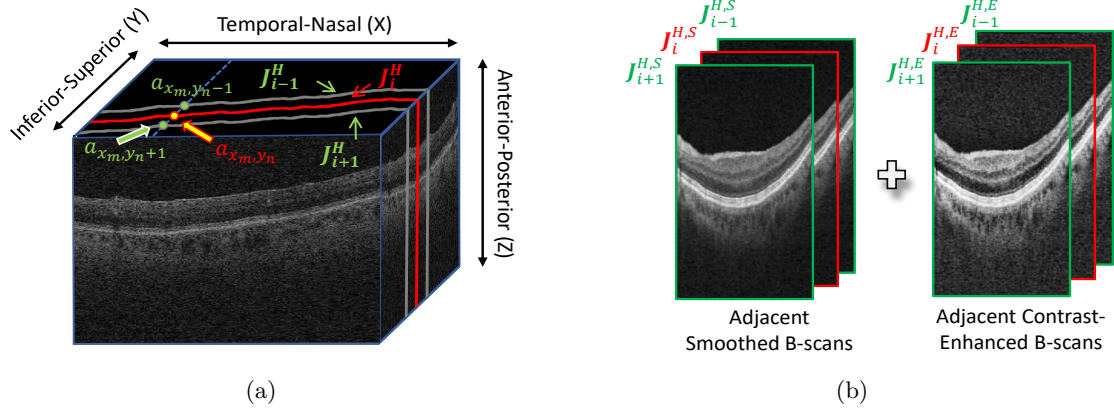


Figure 2. Data augmentation and input B-scan concatenation. (a) An example of a horizontal jittered B-scan (J_i^H) at location i , shown in the red jittery line; the two gray lines are the adjacent B-scans (J_{i-1}^H and J_{i+1}^H). (b) Based on the setting of (a), the input image stack is created by concatenating the adjacent (along the inferior-superior direction) smoothed B-scans, $J_{i-1}^{H,S}$, $J_i^{H,S}$, $J_{i+1}^{H,S}$, and the adjacent contrast limited adaptive histogram equalization (CLAHE) B-scans, $J_{i-1}^{H,E}$, $J_i^{H,E}$, $J_{i+1}^{H,E}$.

Meditec Cirrus OCT) with various levels of retinal thinning (including normal and glaucoma cases). For data augmentation (Sec. 2.2), horizontal, vertical, and jittered B-scan planes were introduced by reconstructing A-scans to enrich the input image patterns. In addition, every three adjacent smoothed B-scans and their contrast-enhanced images are concatenated to form a six-channel input image stack (Sec. 2.3). This design enabled capturing of local 3D contextual information to help achieve a smaller Hausdorff distance. The proposed network (Sec. 2.4) used a modified six-layer U-Net¹⁰ followed by a simple segmentation head to identify initial surface locations.^{8,9}

After the training process, to segment an input OCT volume, the horizontal and vertical B-scan stacks were first separately resampled (Fig. 1b). Each B-scan in the horizontal and vertical stacks was segmented using the proposed segmentation approach. The segmented retinal layer boundaries from the input B-scans were then concatenated together to form 3D retinal surfaces that are labeled as the horizontal or vertical group. Finally, the segmented retinal surfaces from both the horizontal and vertical groups were harmonized together according to a smoothness assumption (Sec. 2.5). A topology guarantee module was also applied to maintain the surface ordering in 3D.

2.2 Data Augmentation

Assume that an input macular OCT volumetric scan has image dimensions of $X \times Y \times Z$ voxels along the temporal-nasal, superior-inferior, and anterior-posterior direction, respectively (Fig. 2a). Currently, most of the common deep learning methods^{7-9,11-15} segment retinal layers in each individual input horizontal B-scan, which can be described as $\mathbf{H}_y = (\mathbf{a}_{x,y})_{x=0}^{X-1} = \{\mathbf{a}_{0,y}, \mathbf{a}_{1,y}, \dots, \mathbf{a}_{X-1,y}\}$, in which $(\bullet)_u^v$ represents an ordered sequence from u to v , and $\mathbf{a}_{x,y}$ is the A-scan at location (x,y) . In other words, $\mathbf{H}_y$ represents the horizontal B-scan at location y along the superior-inferior axis. Because horizontal scanning is the default OCT device imaging protocol, horizontal B-scans often have less motion artifacts (caused by small horizontal eye movements during the scan acquisition) and are commonly used as the input images for image-analysis automated approaches.

In this study, on top of using the common horizontal B-scans as the input of the proposed neural network (Sec. 2.4), we proposed another two novel schemes to increase the variety of the input images. First, we reconstructed A-scans along the vertical direction to simulate the B-scans that are generated from the OCT device using the vertical scanning protocol [i.e., along the superior-inferior direction scanning, $\mathbf{V}_x = (\mathbf{a}_{x,y})_{y=0}^{Y-1}$]. Second, based on the idea of reassembling A-scans, we constructed simulated jittered B-scans consisting of randomly selected adjacent A-scans along either the temporal-nasal (i.e., the X -axis) or superior-inferior (i.e., the Y -axis) direction. Meanwhile, adjacency constraints were added to ensure the retinal surfaces were still smooth in the

simulated B-scans. [Note: Only one-voxel-away A-scans could be randomly selected in this study.] The horizontal and vertical jittered B-scans are described below, and an example of a horizontal jittered B-scan is illustrated in Fig. 2a.

$$\begin{aligned}
 \text{Horizontal direction:} \quad \mathbf{J}^H &= (\mathbf{a}_{x,y_x})_{x=0}^{X-1}; \begin{cases} y_0: \text{randomly selected from } [0, Y-1] \\ y_{x+1} - y_x: \text{randomly selected from } [-1, 0, 1]; \text{ adjacency constraint} \\ \forall y, \{0 \leq y \leq Y-1\}; \end{cases} \\
 \text{Vertical direction:} \quad \mathbf{J}^V &= (\mathbf{a}_{x_y,y})_{y=0}^{Y-1}; \begin{cases} x_0: \text{randomly selected from } [0, X-1] \\ x_{y+1} - x_y: \text{randomly selected from } [-1, 0, 1]; \text{ adjacency constraint} \\ \forall x, \{0 \leq x \leq X-1\}; \end{cases}
 \end{aligned} \tag{1}$$

2.3 B-Scan Concatenation for Local Contextual Information

To enable local 3D contextual information for the neural network, instead of using a single B-scan as the input, we concatenated three adjacent smoothed B-scans and their contrast-enhanced version to create a B-scan stack, which contains a total of six B-scans, as the input image stack. For each horizontal, vertical, and jittered B-scan, we first generated a smoothed B-scan by averaging with its two adjacent B-scans to reduce image artifacts and noise. Then, we applied the contrast-limited-adaptive-histogram-equalization (CLAHE¹⁶) method to enhance the local contrast of the image, thus increasing the visibility of layer boundaries. Next, we repeated the same smoothing-and-enhancing process for both adjacent B-scans of the original B-scan to create a B-scan stack consisting of three adjacent smoothed and three adjacent contrast-enhanced images. Fig 3 shows an example of a horizontal jittered B-scan stack.

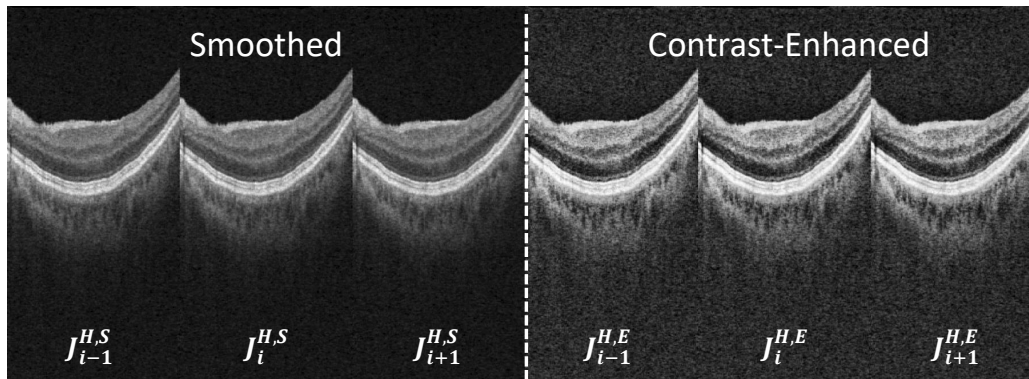


Figure 3. An example B-scan stack that consists of a horizontal jittered B-scan ($\mathbf{J}_i^H$) and its two adjacent B-scans ($\mathbf{J}_{i-1}^H$ and $\mathbf{J}_{i+1}^H$). The left and right panel shows the smoothed ($\mathbf{J}^{H,S}$) and contrast-enhanced ($\mathbf{J}^{H,E}$) version, respectively.

2.4 Network

We proposed a modified U-Net with a contracting (down-sampling) path and expansive (up-sampling) path in six levels with increasing channels along the network depth; for each level, the down-sampling module and up-sampling module were also connected (Fig. 4). The neural network was designed as a feature-extracting module before a surface segmentation head. Along the down-sampling path, each down-sampling module consisted of a 2×2 max-pooling, a 3×3 channel-change Conv2d module, and a cascade of three 3×3 channel-keep Conv2d modules with a residual connection. The down-sampling and up-sampling paths were symmetrical, except for each of the max-pooling layer in the down-sampling path was replaced by a bilinear up-sampling layer in the up-sampling path.

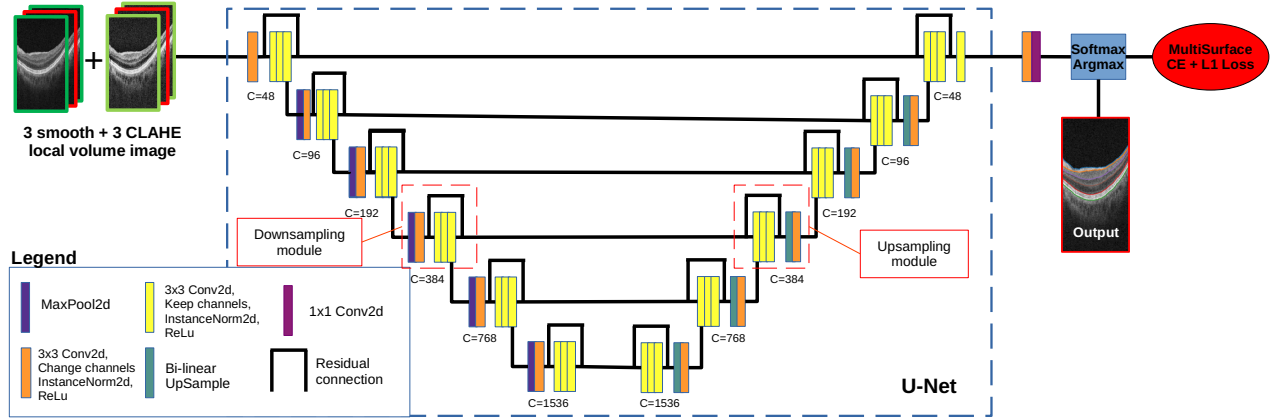


Figure 4. The architecture of the proposed neural network. The number of channels at each level is listed under the corresponding down-sampling/up-sampling module. The total loss function was the summation of the cross-entropy loss plus the L1 loss for all the desired surfaces.

The segmentation head was composed of a 128-channel 3×3 Conv2d module and a 1×1 Conv2d layer. The output logits had dimensions of $X \times Z \times K$, in which X and Z represent the width and height of the input OCT B-scan, and K is the total number of segmented retinal surfaces, in which element $p_{x,z,k} \in [0, 1]$ indicated the estimated probability of surface k at pixel location (x, z) . Assume that $g_{x,z,k} \in [0, 1]$ was the probability of manual tracing at the same location for surface k . A multi-surface cross-entropy loss was computed by

$$L_{CE} = - \frac{\sum_{k=0}^{K-1} \sum_{z=0}^{Z-1} \sum_{x=0}^{X-1} (g_{x,z,k} \ln(p_{x,z,k}) + (1 - g_{x,z,k}) \ln(1 - p_{x,z,k}))}{XZK}. \quad (2)$$

To estimate segmented surface locations, we utilized a softmax argmax to interpret the output logits from the segmentation head. For the same input B-scan, the location of segmented surface k at column x can be described as follows.

$$\mu_{x,k} = \sum_{z=0}^{Z-1} z \cdot p_{x,z,k} \quad (3)$$

In addition, mean absolute error (i.e., the L_1 loss) was also introduced to penalize the difference between the estimated surface locations and manual tracing. Given that $s_{x,k}$ was the corresponding pixel location of the ground truth, the L_1 loss can be described as

$$L_1 = \frac{\sum_{k=0}^{K-1} \sum_{x=0}^{X-1} \|\mu_{x,k} - s_{x,k}\|}{XK}. \quad (4)$$

The total loss for the network training was $L_{Total} = L_{CE} + L_1$.

2.5 Surface Harmonization and Surface Topology Constraint

In the reference stage, as shown in Fig. 1b, each input OCT volume was resampled and formed as a stack of horizontal B-scans and a stack of vertical B-scans. Based on the segmentation method described above, the desired retinal boundaries were automatically labeled in each input B-scan and followed by concatenating the B-scan outputs to form 3D surfaces. The reason for using both of the horizontal and vertical B-scan groups to segment the retinal surfaces was because image artifacts and noise can be sometimes more suppressed in one resampling direction over the other. So, the segmented surface can be locally smoother (or less noisy) in one direction than the other. However, the output surfaces from the horizontal and vertical B-scans should be similar in most cases.

The segmented 3D retinal surfaces from the horizontal and vertical B-scan group were harmonized using the methods below according to a local surface smoothness constraint in OCT images. For an estimated surface k at A-scan location x from the horizontal (H) and vertical (V) B-scan group, if $|\mu_{x,k}^H - \mu_{x,k}^V| < 3$, the harmonized surface at this location was the mean value, $\mu_{x,k}^M = 0.5 * (\mu_{x,k}^H + \mu_{x,k}^V)$. When a disagreement happened (i.e., $|\mu_{x,k}^H - \mu_{x,k}^V| \geq 3$), the harmonized surface location $\mu_{x,k}^M$ was chosen from either $\mu_{x,k}^H$ or $\mu_{x,k}^V$, according to whose local 5×5 neighborhood has a smaller standard deviation.

Finally, we applied a surface-topological-order constraint at the 3D harmonized surfaces. One of the state-of-the-art methods, FCRN,^{7,8} used a “hard” ReLU to limit the surface topological orders at each single A-scan context by fixing the upper surface location and then adjusting the lower surface locations. However, when a spike error occurred at the upper surface, the assumption of using the upper surface as the reference surface would fail and further affect the already-correctly segmented lower surfaces. To address this issue, another state-of-the-art method IPM network⁹ used $\mathbf{A}\hat{\boldsymbol{\mu}} \leq 0$, in which $\mathbf{A}$ is a matrix expressing the relations of adjacent surfaces, and $\hat{\boldsymbol{\mu}}$ is the K -surface location optimization vector along each A-scan. However, this method requires heavy computations. In this study, we proposed a light-computation topological-order constraint by dynamically assigning the reference location of the surface that had a smaller local 5×5 neighborhood standard deviation. Once the surface topological order violation occurred, the violated surfaces would be forced to be overlaid with the reference surface.

3. RESULTS

3.1 Dataset

In this study, a total of 41 OCT macular volumetric scans (Zeiss Cirrus, Carl Zeiss Meditec, Inc., Dublin CA) from 41 different subjects were randomly selected (one eye per subject) from an existing glaucoma study at the University of Iowa.¹⁷ Experiments used the same network for both normal and glaucoma cases. The training set included a total of 20 (7 normal + 13 glaucoma) scans, the validation set included 3 glaucoma scans, and the test set included 18 (9 normal + 9 glaucoma) scans. The training/validation and test datasets were strictly separate. The manual tracing for the ground truth was done by modifying the graph-search segmentation results⁴ by a human expert (J.-K. W.) using Iowa OCTExplorer.¹⁸ Based on the ground truth, the mean thicknesses ($\pm$ standard deviations) of the GCIPL and total retina in the normal group and glaucoma group for all the 41 OCT scans are shown in Table 1.

Table 1. Mean layer thickness ($\pm$ standard deviation) based on ground truth for the whole dataset

Group	Total Scans	Total Retinal Thickness (μm)	GC IPL Thickness (μm)
Normal	16	288.41 ± 12.19	60.86 ± 4.95
Glaucoma	25	261.31 ± 21.37	45.99 ± 7.54

3.2 Data Preprocessing

The randomly selected 41 OCT macula-centered volumes included two different protocols of $512 \times 128 \times 1024$ or $200 \times 200 \times 1024$ voxels covering $6 \times 6 \times 6 \text{ mm}^3$ along the temporal-nasal, superior-inferior, and anterior-posterior direction, respectively. To be consistent, all the OCT volumes were resized to $200 \times 200 \times 512$. Next, the data augmentation (Sec. 2.2) and B-scan concatenation (Sec. 2.3) were applied.

For the manual tracing, since the human expert traced the retinal surfaces in each individual horizontal B-scan, motion artifact could be introduced along the superior-inferior direction. To reduce this bias, a thin-plate-spline algorithm was applied to create 3D smooth manually-traced surfaces.

3.3 Experimental results

The proposed segmentation network was trained using the Adam optimizer with an initial learning rate of 0.1 without weight decay. The batch size was set to two. Data augmentation on the fly was used by adding Gaussian noise and salt/pepper noise as well as randomly flipping the whole input B-scan stack. In experiments, for all

Table 2. Measurement results on the test set among different methods. Bold numbers indicate the best results in corresponding column and case group.

Mean absolute surface distance [†] (MASD) in $\mu\text{m} \pm$ standard deviation								
Cases	Method	Overall	ILM	RNFL-GCL	IPL-INL	OPL-HFL	BMEIS	OB-RPE
Normal [‡]	Standard FCRN	2.45 $\pm$ 0.97	1.66 $\pm$ 0.42	2.88 $\pm$ 0.52	3.50 $\pm$ 0.92	3.16 $\pm$ 0.67	1.50 $\pm$ 0.45	2.03 $\pm$ 0.44
	FCRN + proposed augmentation	2.23 $\pm$ 0.97	1.46 $\pm$ 0.21	2.71$\pm$0.64	2.94 $\pm$ 1.07	2.98$\pm$0.75	1.15$\pm$0.28	2.11 $\pm$ 0.59
	Our proposed method	2.18$\pm$0.91	1.35$\pm$0.16	2.89 $\pm$ 0.70	2.70$\pm$0.62	3.06 $\pm$ 0.75	1.24 $\pm$ 0.35	1.82$\pm$0.36
Glaucoma	Graph search	3.70 $\pm$ 4.04	2.09 $\pm$ 0.33	6.02 $\pm$ 4.39	7.47 $\pm$ 6.52	3.15$\pm$2.39	1.69 $\pm$ 0.23	1.78$\pm$0.22
	Standard FCRN	3.58 $\pm$ 1.99	2.27 $\pm$ 0.41	5.32 $\pm$ 2.05	4.92 $\pm$ 2.38	4.34 $\pm$ 1.30	2.08 $\pm$ 0.60	2.53 $\pm$ 1.10
	FCRN + proposed augmentation	3.36 $\pm$ 2.08	2.09 $\pm$ 0.44	4.96 $\pm$ 2.28	4.83 $\pm$ 2.61	4.20 $\pm$ 1.46	1.74 $\pm$ 0.61	2.31 $\pm$ 0.75
	Our proposed method	3.02$\pm$1.85	1.87$\pm$0.33	4.93$\pm$2.80	4.03$\pm$1.30	3.71 $\pm$ 1.00	1.67$\pm$0.55	1.90 $\pm$ 0.24

GCIPL Dice coefficients and Hausdorff distances [‡] in pixels								
Cases	Method	GCIPL Dice	ILM	RNFL-GCL	IPL-INL	OPL-HFL	BMEIS	OB-RPE
Normal [‡]	Standard FCRN	0.945 $\pm$ 0.018	29.6	13.0	10.5	16.0	10.2	15.2
	FCRN + proposed augmentation	0.952 $\pm$ 0.019	8.2	19.3	15.3	15.2	6.5	8.0
	Our proposed method	0.952$\pm$0.015	6.4	12.1	14.5	24.2	11.8	6.5
Glaucoma	Graph search	0.844 $\pm$ 0.133	25.5	31.3	35.3	42.2	9.3	9.4
	Standard FCRN	0.889 $\pm$ 0.024	48.4	330.9	321.3	309.8	295.8	344.8
	FCRN + proposed augmentation	0.896 $\pm$ 0.030	141.1	273.9	320.5	307.7	294.5	279.9
	Our proposed method	0.899$\pm$0.027	11.9	28.9	34.8	39.5	12.1	15.6

[†] For the MASD and Hausdorff distance, the lower the better; For the Dice coefficients, the higher the better.

[‡] All the surface segmentation results of normal cases from the graph-search method were manually examined and directly considered as ground truth.

* Voxel size along the A-scan direction is 3.91 μm /pixel.

the test scans, we compared the proposed method to the graph-search method,⁴ the standard FCRN,^{7,8} and the FCRN with our proposed vertical and jittered B-scan augmentation planes. In Table 2, compared to the ground truth, the proposed method overall achieved mean absolute surface distances (MASD) of $2.18 \pm 0.91 \mu\text{m}$ (mean $\pm$ standard deviation) in the normal cases and $3.02 \pm 1.85 \mu\text{m}$ in the glaucoma cases; significantly less than the other methods with p -values < 0.001 . Our proposed network and data augmentation methods improved GCIPL Dice coefficients in both normal and glaucoma cases and substantially reduced Hausdorff distances in the glaucoma cases compared to other DL methods. Superposing our proposed data augmentation methods on the standard FCRN method also explicitly improves the MASD and Dice coefficient. Fig. 5 shows two examples where the proposed method overcame two typical segmentation errors, including the spike-like error (top row) and the GCIPL-shifting error (bottom row).

To achieve the best segmentation results, we proposed the jittered/vertical B-scan data augmentation, local

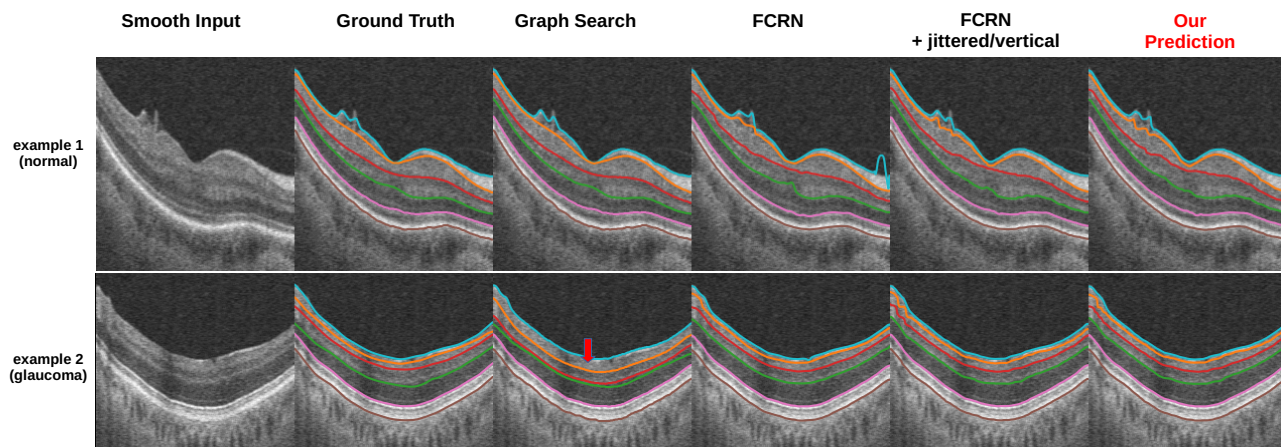


Figure 5. Examples of how the proposed method performed well for two common types of segmentation errors. Top row: spike-like error occurred at the ILM (i.e., the cyan surface) Bottom row: the GCIPL (between the orange and red surfaces) segmentation is completely shifting down in the graph-search results (pointed by the red arrow) because of the extremely thin adjacent RNFL layer (between the cyan and orange surfaces), which violated the graph-search underlying assumptions.

smoothed and CLAHE-enhanced B-scan concatenation, and 3D surface harmonization and topological constraints in this study. We then investigated the effect of each single step on the segmentation performance. In the following ablation experiments, the network architecture and configuration were fixed except for the ablation part. The ablation mean absolute surface distance (MASD) errors were computed in micrometers ($\pm$ standard deviation) for the normal and glaucoma cases (Table 3). The combination of all the proposed methods still showed the best overall results for all the desired surfaces in both normal and glaucoma cases in the test set. Using only one-smoothed and one-CLAHE B-scans affected the segmentation accuracy the most in the normal cases (i.e., increasing the overall MASD errors the most); the removal of the 3D topological ordering constraints affected the overall MASD the most in the glaucoma cases.

Table 3. Measurement results on the test set in the ablation experiments. Bold numbers indicate the best results in the corresponding column and case group.

Mean absolute surface distance [†] (MASD) in $\mu\text{m} \pm$ standard deviation								
Cases	Method	Overall	ILM	RNFL-GCL	IPL-INL	OPL-HFL	BMEIS	OB-RPE
Normal	Our proposed baseline method	2.18$\pm$0.91	1.35$\pm$0.16	2.89$\pm$0.70	2.70$\pm$0.62	3.06$\pm$0.75	1.24$\pm$0.35	1.82$\pm$0.36
	No 3D topological order guaranteeing	2.56 $\pm$ 0.99	1.89 $\pm$ 0.28	3.02 $\pm$ 0.70	3.12 $\pm$ 1.07	3.43 $\pm$ 1.13	1.66 $\pm$ 0.39	2.24 $\pm$ 0.37
	1 smoothed B-scan only input	2.58 $\pm$ 0.84	2.05 $\pm$ 0.40	3.23 $\pm$ 0.66	2.90 $\pm$ 0.61	3.25 $\pm$ 0.87	1.67 $\pm$ 0.41	2.37 $\pm$ 0.41
	3 smoothed B-scans input	2.50 $\pm$ 0.85	1.99 $\pm$ 0.35	2.97 $\pm$ 0.68	2.85 $\pm$ 0.68	3.27 $\pm$ 1.05	1.78 $\pm$ 0.46	2.16 $\pm$ 0.42
	1 smoothed and 1 CLAHE B-scan input	2.73 $\pm$ 1.06	2.22 $\pm$ 0.89	3.32 $\pm$ 1.12	3.32 $\pm$ 1.06	3.44 $\pm$ 0.82	1.75 $\pm$ 0.36	2.36 $\pm$ 0.43
	No jittered/vertical augmentation	2.62 $\pm$ 0.88	2.11 $\pm$ 0.29	3.38 $\pm$ 0.99	3.01 $\pm$ 0.62	3.23 $\pm$ 0.71	1.68 $\pm$ 0.37	2.28 $\pm$ 0.39
Glaucoma	Our proposed baseline method	3.02$\pm$1.85	1.87$\pm$0.33	4.93 $\pm$ 2.80	4.03$\pm$1.30	3.71$\pm$1.00	1.67$\pm$0.55	1.90$\pm$0.24
	No 3D topological order guaranteeing	3.33 $\pm$ 2.44	2.13 $\pm$ 0.63	5.43 $\pm$ 3.94	4.57 $\pm$ 2.41	3.96 $\pm$ 1.34	1.80 $\pm$ 0.55	2.08 $\pm$ 0.32
	1 smoothed B-scan only input	3.24 $\pm$ 2.16	2.05 $\pm$ 0.48	5.11 $\pm$ 3.01	4.58 $\pm$ 2.57	3.82 $\pm$ 1.09	1.76 $\pm$ 0.47	2.14 $\pm$ 0.37
	3 smoothed B-scans input	3.19 $\pm$ 1.75	2.02 $\pm$ 0.48	4.71$\pm$1.65	4.50 $\pm$ 1.99	4.11 $\pm$ 1.24	1.82 $\pm$ 0.57	2.00 $\pm$ 0.31
	1 smoothed and 1 CLAHE B-scan input	3.26 $\pm$ 2.00	2.11 $\pm$ 0.49	5.20 $\pm$ 3.00	4.31 $\pm$ 1.75	4.00 $\pm$ 1.17	1.76 $\pm$ 0.33	2.20 $\pm$ 0.38
	No jittered/vertical augmentation	3.21 $\pm$ 1.82	2.09 $\pm$ 0.56	4.82 $\pm$ 2.45	4.50 $\pm$ 1.73	3.89 $\pm$ 1.13	1.74 $\pm$ 0.37	2.20 $\pm$ 0.31

[†]The lower the better.

★ Voxel size along the A-scan direction is 3.91 $\mu\text{m}/\text{pixel}$.

4. CONCLUSIONS

In this study, we proposed novel approaches to enrich the variety of retinal layer patterns in OCT used for deep learning along all three dimensions by reconstructing A-scans at the horizontal, vertical, and jittered planes. Concatenating the adjacent and contrast-enhanced B-scans to form a six-channel image-stack as the input image further enriched the learning of the segmentation network using 3D local contextual information, leading to improved segmentation accuracy and robustness. In addition, the 3D surface harmonization and topological constraints took advantage of surface local regional information to overcome image artifacts and ensure the segmented surfaces followed a correct order rather than only considering a single A-scan context.⁷⁻⁹ Our experiments showed that the proposed method outperformed the state-of-the-art methods with fewer spike errors and lower mean absolute surface distances from ground truth (manually segmented scans). The proposed method offers a significant improvement that can be applied to other terrain-like surface segmentation needs. A potential limitation of this study is that the manual tracing was obtained by modifying the graph-search segmentation results as a starting point, so bias to the surfaces might be introduced. Also, the proposed data augmentation methods would be violated if the input OCT scan has an ultra-high image resolution in one direction on the retina but is sparsely scanned along the other direction (in other words, if there is an unbalance between the total number of A-scans in each B-scan and the total number of B-scans in the input OCT volume).

ACKNOWLEDGMENTS

This study was also supported, in part, by the Iowa City VA Center for the Prevention and Treatment for Visual Loss, funded by the VA Rehabilitation Research and Development (RR&D) Division I50RX003002, a National Institutes of Health (NIH) grant R01EY031544, and a National Science Foundation (NSF) grant #1733742.

REFERENCES

- [1] Tham, Y. C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., and Cheng, C. Y., “Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis,” *Ophthalmology* **121**(11), 2081–2090 (2014).

- [2] Shin, J. W., Sung, K. R., and Song, M. K., “Ganglion cell-inner plexiform layer and retinal nerve fiber layer changes in glaucoma suspects enable prediction of glaucoma development,” *American Journal of Ophthalmology* **210**, 26–34 (2020).
- [3] Li, K., Wu, X., Chen, D., and Sonka, M., “Optimal surface segmentation in volumetric images - a graph-theoretic approach,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(1), 119–134 (2006).
- [4] Garvin, M. K., Abramoff, M. D., Wu, X., Russell, S. R., Burns, T. L., and Sonka, M., “Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images,” *IEEE Transactions on Medical Imaging* **28**(9), 1436–1447 (2009).
- [5] Pekala, M., Joshi, N., Liu, T. Y., Bressler, N. M., Cabrera DeBuc, D., and Burlina, P., “OCT segmentation via deep learning: A review of recent work,” in [*Asian Conference on Computer Vision - ACCV 2019*], **11367**, 316–322 (2019).
- [6] Liu, H., Wei, D., Lu, D., Li, Y., Ma, K., Wang, L., and Zheng, Y., “Simultaneous alignment and surface regression using hybrid 2d-3d networks for 3D coherent layer segmentation of retina OCT images,” in [*Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*], 108–118, Springer International Publishing (2021).
- [7] He, Y., Carass, A., Liu, Y., Jedynek, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A., and Prince, J. L., “Fully convolutional boundary regression for retina OCT segmentation,” in [*Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*], **11764**, 120–128 (2019).
- [8] He, Y., Carass, A., Liu, Y., Jedynek, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A., and Prince, J. L., “Structured layer surface segmentation for retina OCT using fully convolutional regression networks,” *Medical Image Analysis* **68** (2021).
- [9] Xie, H., Pan, Z., Zhou, L., Zaman, F. A., Chen, D., Jonas, J. B., Wang, Y., and Wu, X., “Globally optimal segmentation of mutually interacting surfaces using deep learning,” *arXiv pre-print, cs.CV* **2007.01259** (2020).
- [10] Ronneberger, O., Fischer, P., and Brox, T., “U-Net: Convolutional networks for biomedical image segmentation,” in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*], 234–241 (2015).
- [11] Shah, A., Abramoff, M. D., and Wu, X., “Simultaneous multiple surface segmentation using deep learning,” in [*Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*], Springer International Publishing (2017).
- [12] Shah, A., Zhou, L., Abramoff, M. D., and Wu, X., “Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in OCT images,” *Biomed. Opt. Express* **9**, 4509–4526 (Sep 2018).
- [13] Lee, C. S., Tying, A. J., Deruyter, N. P., Wu, Y., Rokem, A., and Lee, A. Y., “Deep-learning based, automated segmentation of macular edema in optical coherence tomography,” *Biomed. Opt. Express* **8**, 3440–3448 (Jul 2017).
- [14] Schlegl, T., Waldstein, S. M., Bogunovic, H., Endstraßer, F., Sadeghipour, A., Philip, A.-M., Podkowinski, D., Gerendas, B. S., Langs, G., and Schmidt-Erfurth, U., “Fully automated detection and quantification of macular fluid in OCT using deep learning,” *Ophthalmology* **125**(4), 549–558 (2018).
- [15] Pekala, M., Joshi, N., Liu, T. A., Bressler, N., DeBuc, D. C., and Burlina, P., “Deep learning based retinal OCT segmentation,” *Computers in Biology and Medicine* **114**, 103445 (2019).
- [16] Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T., “Scikit-image: image processing in python,” *PeerJ* **2**, e453 (2014).
- [17] Wall, M., Zamba, G. K., and Artes, P. H., “The effective dynamic ranges for glaucomatous visual field progression with standard automated perimetry and stimulus sizes III and V,” *Investigative Ophthalmology and Visual Science* **59**(1), 439–445 (2018).
- [18] Lee, K., Abramoff, M. D., Garvin, M., Sonka, M., and Wu, X., “Iowa reference algorithm – OCTExplorer,” (March 2020). <https://www.iibi.uiowa.edu/oct-reference>.