Research Article | Vol. 11, No. 7 / 1 July 2020 / Biomedical Optics Express 3843
Biomedical Optics EXPRESS

Semi-supervised deep learning based 3D analysis of the peripapillary region

MORGAN HEISLER,[1] MAHADEV BHALLA,[2] JULIAN LO,[1] ZAID MAMMO,[3] SIEUN LEE,[1] MYEONG JIN JU,[3,4] MIRZA FAISAL BEG,[1] AND MARINKO V. SARUNIC[1,*]

[1]*Simon Fraser University, Department of Engineering Science, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada*
[2]*University of British Columbia, Faculty of Medicine, 317-2194 Health Sciences Mall, Vancouver, BC, V6 T 1Z3, Canada*
[3]*University of British Columbia, Department of Ophthalmology and Vision Science, 2550 Willow Street, Vancouver, BC, V5Z 3N9, Canada*
[4]*University of British Columbia, School of Biomedical Engineering, 251-2222 Health Sciences Mall, Vancouver, BC, V6 T 1Z3, Canada*
[*]*msarunic@sfu.ca*

Abstract: Optical coherence tomography (OCT) has become an essential tool in the evaluation of glaucoma, typically through analyzing retinal nerve fiber layer changes in circumpapillary scans. Three-dimensional OCT volumes enable a much more thorough analysis of the optic nerve head (ONH) region, which may be the site of initial glaucomatous optic nerve damage. Automated analysis of this region is of great interest, though large anatomical variations and the termination of layers make the requisite peripapillary layer and Bruch's membrane opening (BMO) segmentation a challenging task. Several machine learning-based segmentation methods have been proposed for retinal layer segmentation, and a few for the ONH region, but they typically depend on either heavily averaged or pre-processed B-scans or a large amount of annotated data, which is a tedious task and resource-intensive. We evaluated a semi-supervised adversarial deep learning method for segmenting peripapillary retinal layers in OCT B-scans to take advantage of unlabeled data. We show that the use of a generative adversarial network and unlabeled data can improve the performance of segmentation. Additionally, we use a Faster R-CNN architecture to automatically segment the BMO. The proposed methods are then used for the 3D morphometric analysis of both control and glaucomatous ONH volumes to demonstrate the potential for clinical utility.

1. Introduction

Glaucoma is a leading cause of irreversible blindness worldwide [1] and the second most common cause of blindness in the developed world [2]. It is characterized by the degeneration of retinal ganglion cells and the loss of their axons [3], which manifests as narrowing of the neuroretinal rim and structural remodeling of the optic nerve head (ONH) [4]. Current clinical indices for evaluating the progression of glaucomatous optic neuropathy include assessment of the cup-to disc ratio and neuroretinal rim area in fundus photographs, visual field testing, and retinal nerve fiber layer (RNFL) analysis using optical coherence tomography (OCT) images.

Although the conventional RNFL analysis is limited to sectoral averaging of RNFL thickness and circumpapillary thickness profile, recent advances in OCT imaging have enabled the acquisition of high-resolution 3D images from which morphometric measurements and new biomarkers based on anatomical landmarks can be derived. One such landmark, the Bruch's Membrane Opening (BMO), has been shown to be more reliable than disc margin-based rim

#392648 https://doi.org/10.1364/BOE.392648
Journal © 2020 Received 18 Mar 2020; revised 30 Apr 2020; accepted 1 May 2020; published 19 Jun 2020

Biomedical Optics EXPRESS

evaluation of fundus photography [4–6]. As such, its segmentation has become a crucial step to enable the detection of ONH shape parameters.

Several traditional processing tools have been employed to automatically segment the ONH layers and extract ONH morphological parameters [7–11]. The high degree of variability in the appearance of this region, such as the large range of peripapillary tissue thickness and surface regularity due to the region's dense vasculature, age, and disease severity [12], makes accurate modeling particularly challenging. Additionally, the ONH cup can be shallow or deep, greatly changing the shape of the prelaminar neural tissue [13,14]. These factors, in addition to complications arising from typical OCT imaging conditions, such as different fields of view or shadowing artifacts, can yield traditional algorithms clinically unviable without the dependence on several ad hoc parameters to account for different imaging cases. In our previous works using volumetric images of the ONH and graph-cut segmentation [12,15–18], extensive manual corrections of the automated layer segmentations were sometimes necessary before quantification, especially in high myope, pathological, or poor quality images.

Recently, deep learning has emerged as a potential solution to previously inflexible retinal segmentation methods. Several studies have shown the successful segmentation of retinal layers from macular OCT images [19–24] as well as ONH OCT images [25–28]. Of note, Fang et al. [19] used a patch-based convolutional neural network to provide an initial segmentation which is refined using a graph based approach. More recently, Roy et al. [20] designed a fully convolutional network, ReLayNet, to segment retinal layers and macular edema in retinal OCT images. Both methods were applied and tested only on macula scans. Devalla et al. recently improved upon their previous patch-based ONH segmentation method [26] and 2D U-Net method [25] by using a 2-part method with an image enhancer network and 3D segmentation network [20], which presented good results on a limited dataset. Additionally, Zang et al [27] presented a method by which a U-Net based neural network and graph search were combined to delineate the boundary of the optic disc as well as the retinal boundaries. This method was tested on both control and glaucomatous OCT volumes and achieved good results when trained on 10,000+ scans.

In this study, we present a Generative Adversarial Network (GAN) based method for segmenting the peripapillary retinal layers. Although amassing a large amount of OCT images is typically not difficult in a clinical setting, attaining the corresponding manual annotations can be tedious and resource intensive, and having a large number of scans without annotations is not helpful in fully supervised methods. Through the use of semi-supervised (SS) learning, we demonstrate methods to train a network with a small amount of annotated training data while taking advantage of unlabeled scans. Additionally, as the BMO must be segmented in order to extract quantitative parameters, we present results using a Faster R-CNN for automatically segmenting the BMO.

2. Materials and methods

The use of the data for this study was approved from the ethics review boards of Simon Fraser University (SFU) and from the University of British Columbia (UBC). The study was conducted in accordance with the guidelines of the Declaration of Helsinki, and informed consent form was obtained from each participant.

2.1. Database demographics

All volumes in this study were previously acquired for our group's published studies [12,16,18,29,30]. A total of 63 subjects (122 eyes) were used for this study: twenty-one healthy controls (42 eyes, mean age ± standard deviation = 33.95 ± 11.37 years), and forty-two patients with glaucoma (80 eyes, mean age = 65.59 ± 8.91 years). A total of 326 OCT volumes were used in this study. A diagnosis of open-angle glaucoma was made clinically by a fellowship-trained glaucoma specialist based on conventional examination including dilated stereoscopic examination of the

Biomedical Optics EXPRESS

optic nerve, glaucomatous visual field abnormalities, and peripapillary SD-OCT analysis. In addition, severity of glaucomatous visual field loss was quantified by visual field mean deviation (MD) values.

As the purpose and parameters were different for each of the previous studies from which this data set was generated, the availability of manually corrected segmentations was different for each volume set and ranged from no segmentations, only layer segmentations, only BMO segmentations, and both layer and BMO segmentations. A breakdown of the participants' demographics for each dataset is shown in Table 1.

Table 1. Dataset Demographics for the varying levels of segmentation

	Control	Glaucomatous
Peripapillary Layer Segmentation Dataset		
Number of subjects (eyes)	13 (25)	16 (30)
Number of volumes	128	69
Number of B-scans	46,287	26,354
Age (years ± std)	36.80 ± 6.42	68.10 ± 8.44
Axial Length (mm ± std)	24.02 ± 1.08	25.70 ± 2.03
VF MD (dB ± std)	-0.92 ± 0.48	-9.95 ± 8.62
Male: Female	6:7	9:7
BMO Segmentation Dataset		
Number of subjects (eyes)	16 (30)	34 (65)
Number of volumes	145	136
Number of Radial frames	11,600	10,880
Age (years ± std)	32.30 ± 9.12	63.69 ± 13.71
Axial Length (mm ± std)	24.87 ± 1.24	25.58 ± 1.65
VF MD (dB ± std)	-1.00 ± 0.77	-9.68 ± 8.97
Male: Female	8:8	19:15
Unlabeled Dataset		
Number of subjects (eyes)	4 (8)	13 (22)
Number of volumes	8	22
Number of B-scans	2,662	6,796
Age (years ± std)	44.75 ± 16.31	58.41 ± 13.82
Axial Length (mm ± std)	25.02 ± 1.34	24.91 ± 1.36
VF MD (dB ± std)	-1.28 ± 0.65	-12.30 ± 10.16
Male: Female	1:3	7:6

2.2. OCT volume acquisition and processing

Details on the OCT system used in this study has been previously published [29]. In brief, a custom-built swept-source OCT system with a center wavelength of 1.06 μm and 100 kHz sweeping frequency was used to image the ONH region. The acquired three-dimensional (3D) images consisted of 400 B-scans, each with 400 A-scans, and 1024 pixels per A-scan. The imaged region in physical space spanned an axial depth of 2.8 mm and a square area of 5 × 5 to 8 × 8 mm^2 dependent on the axial length of the eye and scan angle. The resulting voxel dimensions were 2.7 μm in the axial direction and ranged from 12.5 to 20 μm in the lateral direction.

Axial displacement caused by involuntary axial eye motion during image acquisition was corrected using cross-correlation between adjacent frames, which were subsequently cropped to

Biomedical Optics EXPRESS

be 640 pixels in the A-scan direction. Each B-scan used for input to the network was averaged with the preceding and following 5 B-scans. Additionally, three-dimensional bounded variation smoothing was applied to the volumes used for BMO segmentation in order to reduce the effect of speckle while preserving and enhancing edges.

2.3. Ground truth labels

Automated layer segmentations of the inner limiting membrane (ILM), the posterior boundary of the RNFL, Bruch's membrane (BM), Bruch's membrane opening, and the choroid–sclera boundary (CS boundary) were generated in 3D using a graph-cut algorithm [15,31]. The area of the automated segmentation result outside of the optic cup was examined and corrected by trained research engineers in Amira (version 5.2; Visage Imaging, San Diego, CA, USA) or ITK-SNAP (version 3.2).

The BMO, defined as the termination point of the high-reflectance BM/retinal pigment epithelium (RPE) complex, was segmented on 80 radial slices extracted from the smoothed volume, intersecting at the approximate center of the BMO and spaced at a constant angle of 2.25°. The ONH is a relatively radially symmetric structure, and radial slices provide a more consistent cross-sectional view of the BMO compared to the raster scan pattern in which the volumes were acquired. For input into the BMO Segmentation Network, the BMO was considered to be the 50×50 pixel box centered on the manual BMO segmentation points. Example segmentations of the 4 layer boundaries and the Bruch's membrane opening are shown in Fig. 1.

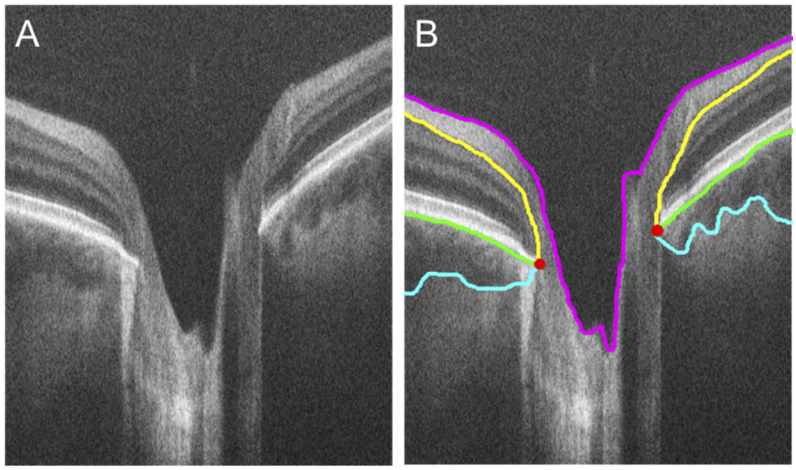

Fig. 1. Representative B scan (A) and its corresponding segmentations (B). The inner limiting membrane (magenta), posterior of the nerve fiber layer (yellow), Bruch's membrane (green), choroid-sclera boundary (cyan) and BMO points (red dots) are shown.

2.4. Layer segmentation network

2.4.1. Architecture

A Pix2Pix GAN [32] based approach was used with the layer segmentation. For the generator network, we used a U-Net based segmentation network [33] which is comprised of a contracting and expanding path connected by skip connections as denoted in Fig. 2. The contracting path contains 4 blocks consisting of a 2D convolutional layer (3×3), dropout layer, 2D convolutional layer (3×3) and max pooling layer. Similarly, the expanding path also contained 4 blocks consisting of a 2D transpose convolutional layer (2×2, stride 2×2) concatenated with the corresponding layer from the contracting path, a 2D convolutional layer (3×3), dropout layer,

Biomedical Optics EXPRESS

and 2D convolutional layer (3×3). To compare the effect of semi-supervised learning on a standard U-Net and on a GAN, the output layer was slightly different for each network. For the U-Net implementation, the output layer was implemented as a 1×1 convolution layer with 5 filters corresponding to the number of regions segmented and a sigmoid activation. Pixel-wise classifications were made based on the highest probability. For the GANs implementation, to make the output layer differentiable it was implemented as a 1×1 convolution layer with 1 filter, where the output is integer encoded to the number of regions segmented and a linear activation.

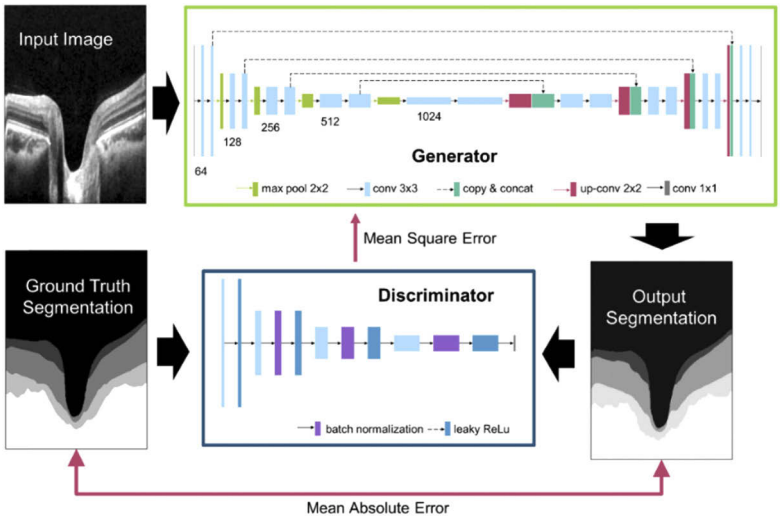

Fig. 2. Network architecture for adversarial layer segmentation of ONH peripapillary layers.

The discriminator model is implemented as a PatchGAN [32]. The output of the network is a single feature map of manual/automated predictions corresponding to a patch size of 16×16 on the original image that is then averaged to give a single score.

2.4.2. Training

Data was augmented during training with random horizontal flips, cropping from 0 to 10%, linear contrast stretching from 0.75 to 1, and rotations from -20 to +20 degrees. The Adam optimizer with a learning rate of 1e-4 was used and the loss function was categorical crossentropy. Additionally, a batch size of 1 was used and the maximum epochs was set to 50 with a callback set to stop training if the validation loss had not improved in 5 epochs. To test the effects of semi-supervised learning on the layer segmentation, the GAN was first trained in a fully supervised fashion on the available labeled dataset. Subsequently, the network was fine-tuned using a ratio of 10 supervised images to 1 pseudo-labelled image. When training the GAN using the unlabeled data, only the generator weights were updated to preserve the integrity of the discriminator, and the predicted output of the generator network was rounded to the nearest integer before being used as the pseudo-label. Training time for the U-Net architectures were an average of 4 hours, for the GAN an average of 20 hours and the pseudo-labeled GAN an average of 28 hours.

2.5. Bruch's membrane opening segmentation network

2.5.1. Architecture

The architecture for this network was based on the Faster R-CNN [34] architecture. A ResNet 50 backbone was used, and the ImageNet pretrained weights were loaded. The output feature map

Biomedical Optics EXPRESS

was then used as an input for a region proposal network (RPN). Each point on the feature map is considered an anchor, and as the OCT radial frames were approximately the same size and all ground truth inputs were 50×50 pixel squares (as mentioned in Section 2.3), we only used a 1:1 ratio anchor of size 50×50. The RPN consisted of a 3×3 convolutional layer connected to two 1×1 convolutional layer output channels for classification and box-regression.

After regions have been proposed, region of interest (ROI) pooling is performed. The output is then fed through three 3×3 convolutional layers, averaged with a 7×7 filter and flattened. The final step is a softmax function for classification and linear regression to fix the boxes' location. A high-level overview of this network is shown in Fig. 3.

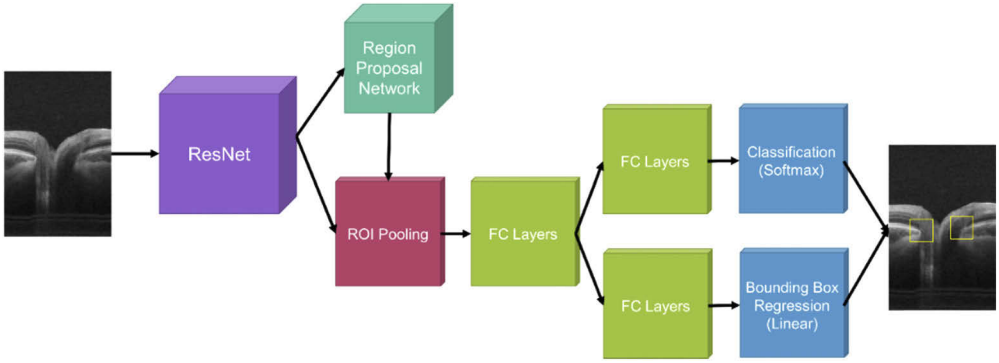

Fig. 3. High level overview of Faster-RCNN architecture used for BMO detection. The yellow boxes on the rightmost image represent the bounding boxes detected for the BMO, where the center of the box corresponds to the BMO.

2.5.2. Training

For training, all anchors are separated into BMO and non-BMO patches based on the Intersection over Union (IoU). Anchors that overlap a manual segmentation box with an IoU larger than 0.5 are considered BMO and anchors with an IoU less than 0.1 are considered non-BMO. A mini batch was chosen to be 256 of these anchors, equally split between BMO and non-BMO classes. Non-maximum suppression is applied to ensure there is no overlapping for the proposed regions.

The Adam optimizer with a learning rate of 1e-5 was used to train the RPN and classifier layers (labelled FC Layers in Fig. 3). The loss function was defined as the addition of the losses for the classification and bounding box regression. The classification loss was the log loss over the BMO and non-BMO classes. The box regression loss was the smooth L_1 loss for the coordinates. Three-fold cross-validation was performed on this dataset. Training time was 3 hours on average. Effort was made to keep the three different training sets equal in glaucoma/control ratio while maintaining separation of subjects in the training set to ensure that they were not in the test set.

2.6. Clinical parameters

Four boundaries were extracted using the layer segmentation network: the inner limiting membrane, posterior boundary of the RNFL, posterior boundary of the BM and Choroid-Sclera boundary. To extract these boundaries, the segmentations for each B-scan in a volume were grouped into a volume. The largest 3D connected component corresponding to each tissue was assigned to that tissue and holes within the connected component were filled by the corresponding tissue. The boundaries were taken to be the first pixel in the axial direction for each tissue. Three shape characteristics were measured using these extracted layers and BMO points as previously described [18]: RNFL thickness, choroidal thickness, and BMO area.

Biomedical Optics EXPRESS

Nerve fiber layer thickness was measured at each pixel of the posterior RNFL surface as the closest distance to the ILM surface. Similarly, choroidal thickness was measured at each pixel of the posterior CS boundary as the closest distance to the BM surface. For statistical analysis, the thickness measurements were averaged over an elliptical annulus, inwardly bounded at 0.75 mm from BMO and outwardly bounded at 1.75 mm from BMO. This provided a level of anatomical consistency in averaging measurements over multiple eyes with different image and BMO sizes.

To quantify the BMO shape, segmented points on the radial frames were first transformed back to the volume scans. Erroneous points were then eliminated by removing any segmentations more than one standard deviation away from the mean axial position of the segmented BMO points. An ellipse was then fitted to the segmented BMO points by first finding the best-fit plane using principal component analysis (PCA) and fitting an ellipse to the projection of the BMO points on the plane by least-squares criterion. Bruch's membrane opening area was calculated from the fitted ellipse.

2.7. Analysis

The Dice similarity coefficient was used to measure the spatial overlap between the manual and automated layer segmentation. It is defined between 0 and 1, where 0 represents no overlap and 1 represents complete overlap. The Dice similarity coefficient was calculated for each tissue as follows:

$$Dice(X) = \frac{2|X \cap Y|}{|X| + |Y|}\ , \tag{1}$$

where X denotes the set of pixels corresponding to the tissue in the manual segmentation, while Y denotes the set of pixels corresponding to the tissue in the automatically segmented image.

Mean Average Precision (mAP) was used to measure the accuracy of the BMO detection network. Average precision was calculated as

$$AP = \sum_{n}(R_n - R_{n-1})P_n\ , \tag{2}$$

where P_n and R_n are the precision and recall at the n^{th} threshold. A prediction was considered positive if the IoU ≥ 0.5.

Clinical parameters for the volumes which had both BMO and layer manual segmentations were presented in a table including the mean and standard deviation for both manual and automated measurements. Paired, two-tailed Student's t-Tests were run to compare the means of the parameters. Bland-Altman plots were also used to evaluate the agreement between manual and automated methods. The differences between manual and automated measures were plotted against the average of both measures. The mean and standard deviation (SD) of the differences, mean of the absolute differences and 95% confidence intervals (± 1.96 SD) were calculated. Statistical significance was set at $P<0.05$ for all the tests performed.

3. Results

3.1. Layer segmentation

To test the effects of pseudo labelling and adversarial training, the networks were trained using two different amounts of B-scans: 1000 B-scans (800 training, 200 validation) and 10,000 B-scans (8,000 training, 2,000 validation). The training set included data from 22 volumes (7 subjects, 9 eyes). The validation set included data from 4 volumes (2 subjects, 2 eyes). The training and validation sets had an equal split between control and glaucomatous B-scans. For the semi-supervised approaches, all unlabeled B-scans were available to be randomly selected for training. A total of 171 volumes (20 subjects, 40 eyes) were used for testing purposes. Subjects were divided between the training, validation, and testing sets such that they did not appear in

Biomedical Optics EXPRESS

more than one set. Longitudinal data (different time points) from the same subject were contained in the same set.

In order to compare the effects of the different architecture and training schemes, no post-processing was done on the network outputs for the mean DICE coefficients reported in Table 2. As shown the semi-supervised (SS) Pix2Pix GAN has a higher Dice value for the regions inside the retina and choroid. The Dice values were slightly worse for the vitreous and scleral regions in this scheme, but this was mainly due to fluctuations in noise in these regions and were simple to remove in the post-processing steps described in Section 2.6. Therefore, the SS Pix2Pix GAN was chosen for parameterization in Section 3.3.

Table 2. Mean DICE Coefficient for 57,319 B-scans before post-processing. The SS Pix2Pix GAN refers to the semi-supervised approach. Bolded values represent the best Dice Coefficient for that region out of all methods.

# of Training B-Scans	Start-ILM	ILM-RNFL	RNFL-BM	BM-CS Boundary	CS Boundary-End
U-Net					
1,000 B-scans	0.9291	0.7049	0.7903	0.6181	0.9359
10,000 B-scans	0.9694	0.8799	0.9319	0.8828	**0.9644**
Pix2Pix GAN					
1,000 B-scans	0.9696	0.8816	0.9296	0.8629	0.9614
10,000 B-scans	**0.9730**	0.8918	0.9338	0.8799	0.9641
SS Pix2Pix GAN					
1,000 B-scans	0.9621	0.8865	0.9308	0.8711	0.9594
10,000 B-scans	0.9701	**0.8958**	**0.9394**	**0.8840**	0.9626

Figure 4 shows representative B-scans from the test set and their segmentation results using the different architectures, which were trained on 10,000 labeled B-scans. The top two rows are from glaucomatous volumes, whereas the bottom two rows are from healthy volumes. The U-net examples show consistent spikes in the segmentations, particularly in the glaucomatous volumes, which are not present in the ground truth segmentations. As neither of the GAN based segmentations have this artifact, the adversarial learning technique provides segmentations that qualitatively look more similar to the ground truth. The qualitative differences between the semi-supervised and non-semi-supervised GAN are small, but can be seen in some instances such as the Bruch's membrane segmentation of the first row.

3.2. BMO segmentation

The mAP for the BMO segmentation network was 0.8547 for glaucomatous (n=21,596 manual BMO points) and 0.9567 for control subjects (n=23,052 manual BMO points). Qualitatively, the network performed quite well. Example B-scans showing BMO segmentations of both control and glaucomatous subjects are shown in Fig. 5.

3.3. Parameters

The clinical parameters extracted from the datasets are shown in Table 3. There was no statistical significance between the manual and automated measurements for the BMO area, though both thickness measurements showed significantly thinner measurements. However, the coefficient of determination was above 0.97 showing excellent correlation.

Example BMO segmentations with corresponding ground truth segmentations are overlaid on the sum-voxel, *en face* view of the OCT volumes in Fig. 6. Manual segmentations (purple dots), the fit ellipse (green circle) and fit ellipse center (green star) as well as the automated

Biomedical Optics EXPRESS

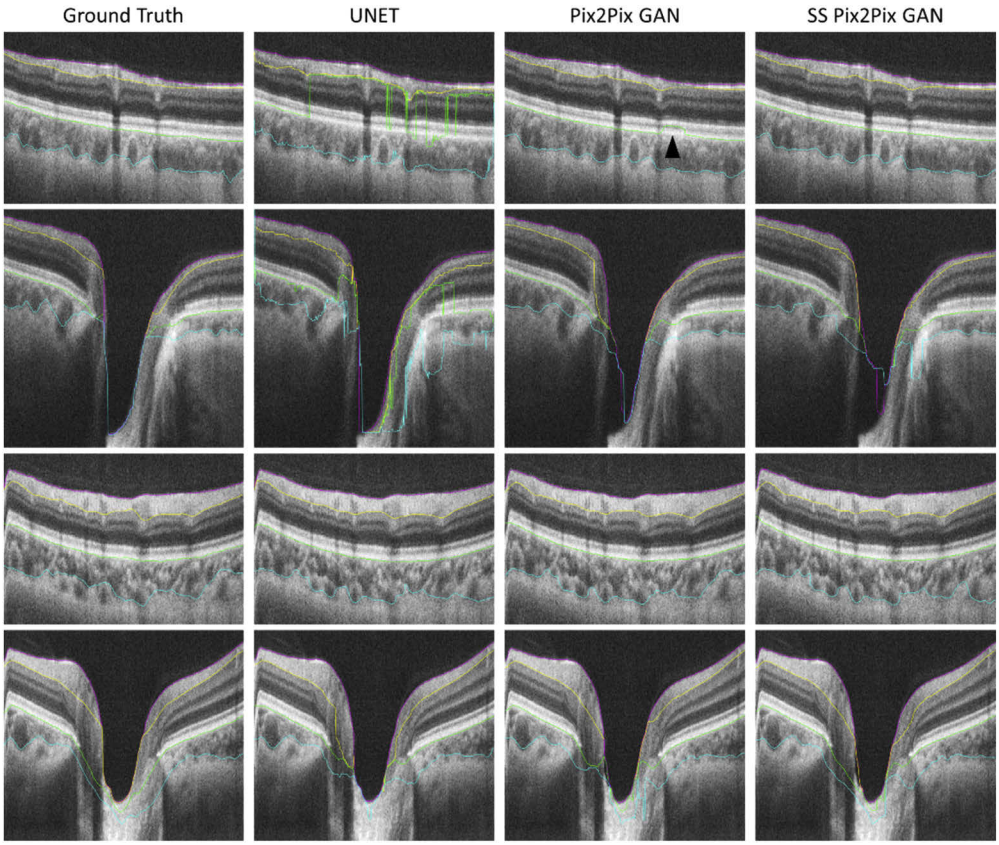

Fig. 4. Examples of layer segmentation results using the different methods. The ILM (magenta), posterior border of the NFL (yellow), Bruch's membrane (green) and CS boundary (cyan) are shown for examples both in the peripapillary region and through the optic disc. The top two rows are from glaucomatous subject volumes and the bottom two rows are from healthy subject volumes. The black arrowhead in the Pix2Pix GAN image in the first row shows a segmentation error that does not occur in the corresponding semi-supervised image.

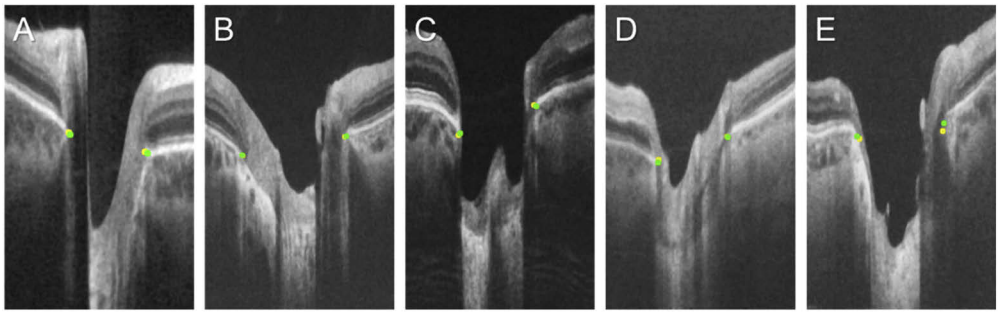

Fig. 5. Examples of BMO segmentations on control (A,B) and glaucomatous subjects (C,D,E). Ground truth labels are shown in green, with automated segmentations in yellow.

Biomedical Optics EXPRESS

Table 3. Mean values for the clinical parameters using both manual and automated methods.

	Mean Manual (±std)	Mean Automated (±std)	P-value	R^2
RNFL Thickness (n=102)	82.7 ± 16.1 μm	80.3 ± 14.9 μm	< 0.05	0.971
Choroid Thickness (n=102)	147.3 ± 57.4 μm	140.7 ± 55.8 μm	< 0.05	0.980
BMO Area (n=281)	2.6 ± 1.3 mm^2	2.6 ± 1.2 mm^2	0.14	0.980

segmentations (yellow dots), the fit ellipse (blue circle) and fit ellipse center (blue star) are shown for a control (A), high-myope control (B) and glaucomatous (C) subject.

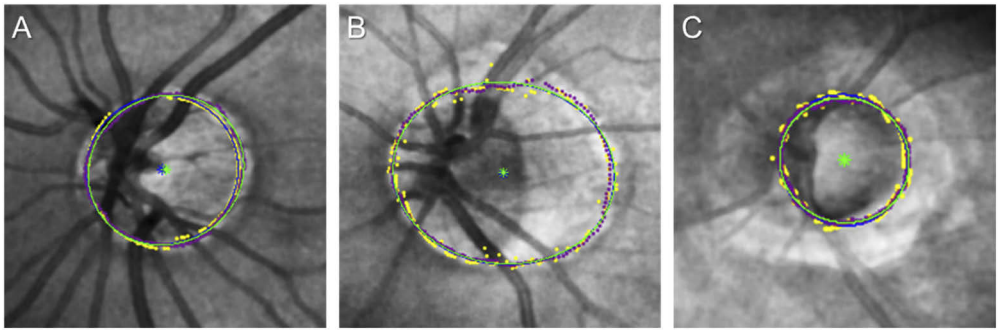

Fig. 6. Examples of manual and automated BMO segmentations. Manual segmentations (purple dots), the fit ellipse (green circle) and fit ellipse center (green star) as well as the automated segmentations (yellow dots), the fit ellipse (blue circle) and fit ellipse center (blue star) are shown for a control (A), high myope control (B) and glaucomatous (C) subject.

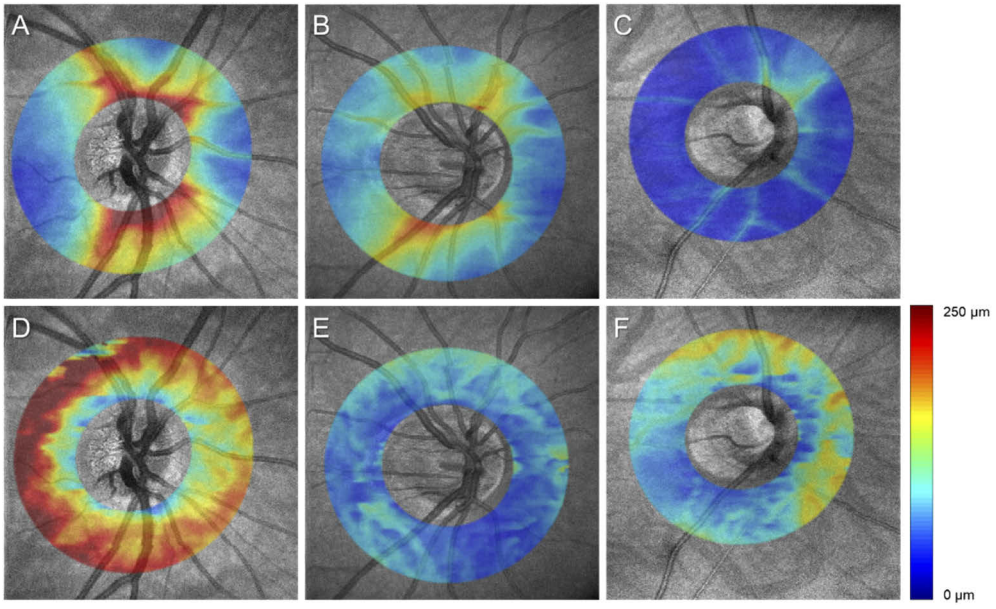

Fig. 7. Example images of the RNFL (A-C) and corresponding choroidal thickness (D-F) for a young control (A,D), myopic control (B,E) and glaucomatous (C,F) eye. Thickness measurements are inwardly bounded 0.25 mm from the best fit BMO ellipse and outwardly bounded at 1.25 mm from the best fit BMO ellipse.

Biomedical Optics EXPRESS

Example automated RNFL and choroidal thickness measurements from the semi-supervised pix-to-pix GAN method are overlaid on the sum-voxel, *en face* view of the OCT volumes in Fig. 7. Qualitatively, the RNFL thickness maps follow the characteristic pattern expected in the control eyes, whereas the glaucomatous eye (Fig. 7(C)) exhibits a much thinner RNFL, which is congruent with the pathophysiology of glaucoma. Additionally, the choroid is a highly vascular layer and large differences in thickness can be seen in all three eyes with no apparent pattern.

Scatter plots showing the Pearson's correlation coefficient for glaucomatous and control data points are shown in Fig. 8. All clinical parameters showed good correlation between those extracted from automated and manual segmentations.

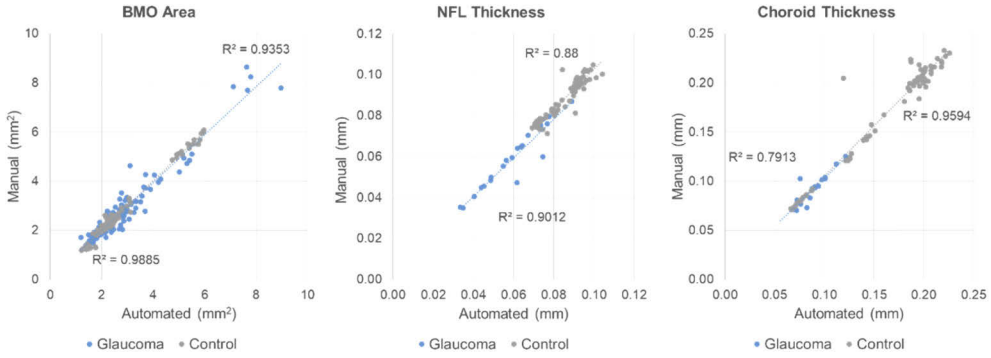

Fig. 8. Scatter plots for the BMO area (left), mean RNFL thickness (middle) and mean choroid thickness (right).

Bland-Altman plots for the clinical parameters further confirm the high reliability of the automated measurements and are shown in Fig. 9.

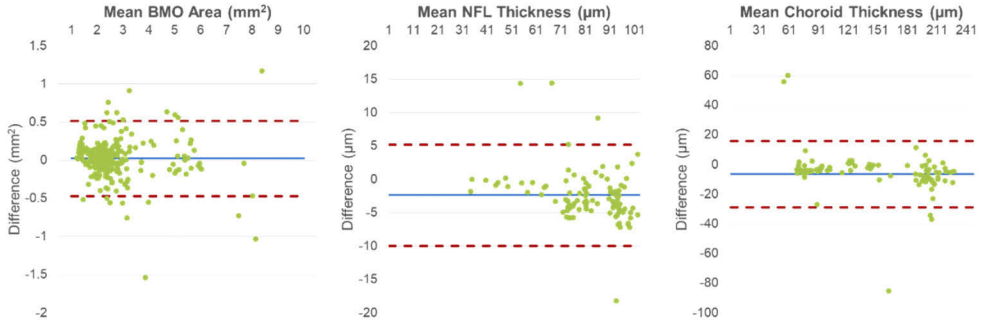

Fig. 9. Bland-Altman Plots for the BMO area (left), mean RNFL thickness (middle) and mean choroid thickness (right).

4. Discussion

Training neural networks to segment diseased data generally requires a large amount of manually annotated ground truth images for fully supervised learning. Although a general U-Net architecture did show good performance when trained on 1000 B-scans (the equivalent of 2.5 OCT volumes), significantly better performance was shown for both adversarial training and fine-tuning using pseudo-labels when using a small dataset. This is particularly useful for medical imaging modalities, as curating expertly segmented scans is a finite and highly limited resource. Importantly, the results demonstrate that the benefits to the semi-supervised GAN approaches with 10x more data are only an improvement of 1-2% in performance.

Biomedical Optics EXPRESS

Fine-tuning the network using pseudo-labelling generally improved the Dice scores for the regions of interest; however, it should be noted that the regions above the ILM and below the choroid-scleral boundary did get slightly worse after pseudo-labelling as indicated in Table 2. This may be due to the reinforcement of poor segmentations during the pseudo-labelling training, particularly in regions of noise such as above the ILM and below the choroid-sclera boundary. As these errors generally present themselves as smaller pockets fully encased by the correctly segmented region, they are easily filtered out during the post-processing steps described in Section 2.6.

It is important to address that although the Dice similarity coefficient reported above for the peripapillary layers was satisfactory, it did not reach as high as other papers [20,27] have reported for similar layer segmentation. This may be due to several factors. First, the variation in the anatomical shape and retinal layers in the region near the ONH is much higher than in the macula. Additionally, the manual segmentations provided for layer segmentation were completed for studies that did not consider the region inside the BMO. Therefore, manual raters were told not to correct inside the optic cup as it was not used for parameterization, leading to Dice scores inside the optic cup to be significantly lower. As such, the clinical measurements provide a better idea of the accuracy of the networks.

The clinical thickness parameters extracted from the automatically segmented volumes were shown to be slightly thinner than the manually segmented volumes. There was however, excellent correlation with the ground truth parameters, suggesting this was a small systematic difference. Additionally, through the Bland-Altman analysis, we see that the majority of datapoints fall within the limits of agreement. Although the thickness values were averaged to calculate a single score for ease of comparison, further 3D analysis could be done using these automated segmentations.

We were also able to show how a Faster R-CNN could be used to detect the BMO in radial OCT scans. From both the mAP and coefficient of determination values, BMO segmentation was shown to be better on control data than on glaucomatous data. The reason for this may be due to the larger number of control radial frames for training or the poorer quality of the glaucomatous dataset. However, most of the erroneous BMO segmentations were easily eliminated during post-processing (Section 2.6) leading to no significant difference in the BMO area parameter when comparing manual and automated methods as seen in Table 3, with good correlation for both glaucoma ($R^2 = 0.9353$) and control eyes ($R^2 = 0.9885$) as shown in Fig. 8.

Future works using this pipeline may also include looking at the 3D BMO minimum rim width (BMO-MRW), a parameter that has been shown to be useful in discriminating preperimetric and perimetric glaucoma [35]. Moreover, the datasets used contained a large amount of longitudinal data which increased the number of B-scans, but not subjects. Therefore, future studies with a larger number of subjects are warranted. Additionally, the methods described in this paper are readily translatable to more of the retinal layers which may also be of interest to clinicians, such as the ganglion cell layer.

5. Conclusion

In this study, we presented a Generative Adversarial Network based method for segmenting the peripapillary ONH layers which outperformed the vanilla U-Net. Through the use of pseudo-labelling, B-scans that did not have a corresponding manual segmentation were still able to be used and provided a further increase in performance. A Faster R-CNN was also used to segment the BMO from the volumes, allowing for comparison of volumetric parameters. The BMO area was shown to have no statistically significant difference, while the thickness parameters were slightly under segmented but highly correlated.

Biomedical Optics EXPRESS

Funding

Fondation Brain Canada; Natural Sciences and Engineering Research Council of Canada; Canadian Institutes of Health Research; Alzheimer Society; Michael Smith Foundation for Health Research; Genome British Columbia.

Disclosures

MVS: Seymour Vision, Inc. (I).

References

1. Y. C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C. Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis," Ophthalmology **121**(11), 2081–2090 (2014).
2. H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," Br. J. Ophthalmol. **90**(3), 262–267 (2006).
3. Y. H. H. Kwon, J. H. H. Fingert, M. H. H. Kuehn, and W. L. M. L. M. Alward, "Primary open-angle glaucoma," N. Engl. J. Med. **360**(11), 1113–1124 (2009).
4. B. C. Chauhan and C. F. Burgoyne, "From clinical examination of the optic disc to clinical assessment of the optic nerve head: A paradigm change," Am. J. Ophthalmol. **156**(2), 218–227 (2013).
5. A. S. C. Reis, N. O'Leary, H. Yang, G. P. Sharpe, M. T. Nicolela, C. F. Burgoyne, and B. C. Chauhan, "Influence of clinically invisible, but optical coherence tomography detected, optic disc margin anatomy on neuroretinal rim evaluation," Invest. Ophthalmol. Visual Sci. **53**(4), 1852–1860 (2012).
6. M. Young, S. Lee, M. Rateb, M. F. Beg, M. V. Sarunic, and P. J. Mackenzie, "Comparison of the clinical disc margin seen in stereo disc photographs with neural canal opening seen in optical coherence tomography images," J. Glaucoma **23**(6), 360–367 (2014).
7. P. Zang, S. S. Gao, T. S. Hwang, C. J. Flaxel, D. J. Wilson, J. C. Morrison, D. Huang, D. Li, and Y. Jia, "Automated boundary detection of the optic disc and layer segmentation of the peripapillary retina in volumetric structural and angiographic optical coherence tomography," Biomed. Opt. Express **8**(3), 1306 (2017).
8. M. K. Garvin, M. D. Abràmoff, R. Kardon, S. R. Russell, X. Wu, and M. Sonka, "Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search," IEEE Trans. Med. Imaging **27**(10), 1495–1505 (2008).
9. S. Li, D. Cunefare, C. Wang, R. H. Guymer, L. Fang, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," Biomed. Opt. Express **8**(5), 2732–2744(2017).
10. S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," Opt. Express **18**(18), 19413 (2010).
11. A. Lang, A. Carass, M. Hauser, E. S. Sotirchos, P. A. Calabresi, H. S. Ying, and J. L. Prince, "Retinal layer segmentation of macular OCT images using boundary classification," Biomed. Opt. Express **4**(7), 1133 (2013).
12. S. Lee, M. L. Heisler, K. Popuri, N. Charon, B. Charlier, A. Trouvé, P. J. Mackenzie, M. V. Sarunic, and M. F. Beg, "Age and glaucoma-related characteristics in retinal nerve fiber layer and choroid: Localized morphometrics and visualization using functional shapes registration," Front. Neurosci. **11**, 381 (2017).
13. C. Burgoyne, "The morphological difference between glaucoma and other optic neuropathies," J. Neuro-Ophthalmology **35**, S8–S21 (2015).
14. D. S. Greenfield and R. N. Weinreb, "Role of optic nerve imaging in glaucoma clinical practice and clinical trials," Am. J. Ophthalmol. **145**(4), 598–603 (2008).
15. S. Lee, N. Fallah, F. Forooghian, A. Ko, K. Pakzad-Vaezi, A. B. Merkur, A. W. Kirker, D. a. Albiani, M. Young, M. V. Sarunic, and M. F. Beg, "Comparative analysis of repeatability of manual and automated choroidal thickness measurements in nonneovascular age-related macular degeneration," Invest. Ophthalmol. Visual Sci. **54**(4), 2864 (2013).
16. S. Lee, M. Heisler, P. J. Mackenzie, M. V. Sarunic, and M. Faisal, "Quantifying variability in longitudinal peripapillary RNFL and choroidal layer thickness using surface based registration of OCT Images," Transl. Vis. Sci. Technol. **6**(1), 11–20 (2017).
17. S. Lee, N. Charon, B. Charlier, K. Popuri, E. Lebed, M. V. Sarunic, A. Trouvé, and M. F. Beg, "Atlas-based shape analysis and classification of retinal optical coherence tomography images using the functional shape (fshape) framework," Med. Image Anal. **35**, 570–581 (2017).
18. S. Lee, S. X. Han, M. Young, M. F. Beg, M. V. Sarunic, and P. J. Mackenzie, "Optic nerve head and peripapillary morphometrics in myopic glaucoma," Invest. Ophthalmol. Visual Sci. **55**(7), 4378–4393 (2014).
19. L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," Biomed. Opt. Express **8**(5), 2732 (2017).

Biomedical Optics EXPRESS

20. A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," Biomed. Opt. Express **8**(8), 3627–3642 (2017).
21. Y. He, A. Carass, Y. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Deep learning based topology guaranteed surface and MME segmentation of multiple sclerosis subjects from retinal OCT," Biomed. Opt. Express **10**(10), 5042 (2019).
22. S. Apostolopoulos, S. De Zanet, C. Ciller, S. Wolf, and R. Sznitman, "Pathological OCT retinal layer segmentation using branch residual U-shape networks," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Springer Verlag, 2017), Vol. 10435 LNCS, pp. 294–301.
23. M. Pekala, N. Joshi, T. Y. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina, "Deep learning based retinal OCT segmentation," Comput. Biol. Med. **114**, 103445 (2019).
24. S. Sedai, B. Antony, D. Mahapatra, and R. Garnavi, "Joint Segmentation and Uncertainty Visualization of Retinal Layers in Optical Coherence Tomography Images Using Bayesian Deep Learning," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Springer Verlag, 2018), Vol. 11039 LNCS, pp. 219–227.
25. S. K. Devalla, P. K. Renukanand, B. K. Sreedhar, G. Subramanian, L. Zhang, S. Perera, J.-M. Mari, K. S. Chin, T. A. Tun, N. G. Strouthidis, T. Aung, A. H. Thiéry, and M. J. A. Girard, "DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images," Biomed. Opt. Express **9**(7), 3244–3265 (2018).
26. S. K. Devalla, K. S. Chin, J. M. Mari, T. A. Tun, N. Strouthidis, T. Aung, A. H. Thiéry, and M. J. A. Girard, "A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head," Invest. Ophthalmol. Visual Sci. **59**, 63–74 (2018).
27. P. Zang, J. Wang, T. T. Hormel, L. Liu, D. Huang, and Y. Jia, "Automated segmentation of peripapillary retinal boundaries in OCT combining a convolutional neural network and a multi-weights graph search," Biomed. Opt. Express **10**(8), 4340 (2019).
28. S. K. Devalla, D. Pham, S. Kumar Panda, L. Zhang, G. Subramanian, A. Swaminathan, C. Z. Yun, M. Rajan, S. Mohan, R. Krishnadas, V. Senthil, J. Mark, S. De Leon, T. A. Tun, C.-Y. Cheng, L. Schmetterer, S. Perera, T. Aung, A. H. Thiéry, and M. J. A. Girard, "Towards label-free 3D segmentation of optical coherence tomography images of the optic nerve head using deep learning," (2017).
29. M. Young, S. Lee, M. F. Beg, P. J. Mackenzie, and M. V. Sarunic, "High speed morphometric imaging of the optic nerve head with 1 µm OCT," Invest. Ophthalmol. Visual Sci. **52**(9), 6720 (2011).
30. M. Bhalla, M. Heisler, S. X. Han, M. V. Sarunic, M. F. Beg, P. J. Mackenzie, and S. Lee, "Longitudinal analysis of Bruch membrane opening morphometry in myopic glaucoma," *J. Glaucoma* 28(10), 889–895 (2019).
31. S. Lee, M. F. Beg, and M. V. Sarunic, "Segmentation of the macular choroid in OCT images acquired at 830 nm and 1060 nm," in *Optical Coherence Tomography and Coherence Techniques VI* (2013), Vol. 8802, p. 88020J.
32. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, 5967–5976 (2016).
33. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," CoRR abs/1505.0 (2015).
34. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017).
35. J. M. D. Gmeiner, W. A. Schrems, C. Y. Mardin, R. Laemmer, F. E. Kruse, and L. M. Schrems-Hoesl, "Comparison of Bruch's membrane opening minimum rim width and peripapillary retinal nerve fiber layer thickness in early glaucoma assessment," Invest. Ophthalmol. Visual Sci. **57**(9), OCT575 (2016).