Towards label-free 3D segmentation of optical coherence tomography images of the optic nerve head using deep learning

SRIPAD KRISHNA DEVALLA,[1] TAN HUNG PHAM,[1,2] SATISH KUMAR PANDA,[1] LIANG ZHANG,[1] GIRIDHAR SUBRAMANIAN,[1] ANIRUDH SWAMINATHAN,[1] CHIN ZHI YUN,[1] MOHAN RAJAN,[3] SUJATHA MOHAN,[3] RAMASWAMI KRISHNADAS,[4] VIJAYALAKSHMI SENTHIL,[4] (ID) JOHN MARK S. DE LEON,[5] TIN A. TUN,[1,2] CHING-YU CHENG,[2,6] LEOPOLD SCHMETTERER,[2,7,9,10,11] SHAMIRA PERERA,[2,8] TIN AUNG,[2,8] ALEXANDRE H. THIÉRY,[12,14] AND MICHAËL J. A. GIRARD[13,15]

[1] *Ophthalmic Engineering & Innovation Laboratory, Department of Biomedical Engineering, Faculty of Engineering, National University of Singapore, Singapore*

[2] *Singapore Eye Research Institute, Singapore National Eye Centre, Singapore*

[3] *Rajan Eye Care Hospital, Chennai, India*

[4] *Glaucoma Services, Aravind Eye Care Systems, Madurai, India*

[5] *Department of Health Eye Center, East Avenue Medical Center, Quezon City, Philippines*

[6] *Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore*

[7] *Nanyang Technological University, Singapore*

[8] *Duke-NUS Graduate Medical School, 8 College Rd, Singapore 169857, Singapore*

[9] *Department of Clinical Pharmacology, Medical University of Vienna, Austria*

[10] *Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria*

[11] *Institute of Clinical and Molecular Ophthalmology, Basel, Switzerland*

[12] *Department of Statistics and Applied Probability, National University of Singapore, Singapore*

[13] *Ophthalmic Engineering and Innovation Laboratory (OEIL), Singapore Eye Research Institute, 20 College Road, Singapore 169856, Singapore*

[14] *a.h.thiery@nus.edu.sg*

[15] *mgirard@ophthalmic.engineering*

Abstract: Recently proposed deep learning (DL) algorithms for the segmentation of optical coherence tomography (OCT) images to quantify the morphological changes to the optic nerve head (ONH) tissues during glaucoma have limited clinical adoption due to their device specific nature and the difficulty in preparing manual segmentations (training data). We propose a DL-based 3D segmentation framework that is easily translatable across OCT devices in a label-free manner (i.e. without the need to manually re-segment data for each device). Specifically, we developed 2 sets of DL networks: the 'enhancer' (enhance OCT image quality and harmonize image characteristics from 3 devices) and the 'ONH-Net' (3D segmentation of 6 ONH tissues). We found that only when the 'enhancer' was used to preprocess the OCT images, the 'ONH-Net' trained on any of the 3 devices successfully segmented ONH tissues from the other two unseen devices with high performance (Dice coefficients > 0.92). We demonstrate that is possible to automatically segment OCT images from new devices without ever needing manual segmentation data from them.

1. Introduction

The complex 3D structural changes of the optic nerve head (ONH) tissues that manifest with the progression of glaucoma has been extensively studied and better understood owing to the advancements in optical coherence tomography (OCT) imaging [1]. These include changes such as the thinning of the retinal nerve fiber layer (RNFL) [2,3], changes in the choroidal thickness [4], minimum rim width [5], and lamina curvature and depth [6,7]. The automated segmentation and analysis of these parameters in 3D from OCT volumes could improve the current clinical management of glaucoma.

Robustly segmenting OCT volumes remains extremely challenging. While commercial OCTs have in-built proprietary segmentation software, they can segment some, but not all the ONH tissues [8–10]. To address this, several research groups have developed an overwhelming number of traditional image processing based 2D [8,11–15] and 3D [16–21] segmentation tools, however they are generally tissue-specific [11–13,15,16,21], computationally expensive [20,22], require manual input [17,19], and are often prone to errors in scans with pathology [20,23,24].

Recent deep learning (DL) based systems have however exploited a combination of low- (i.e. edge-information, contrast and intensity profile) and high-level features (i.e. speckle pattern, texture, noise) from OCT volumes to identify different tissues, yielding human-level [25–32] and pathology invariant [25,26,31] segmentations. Yet, given the variability in image characteristics (e.g. contrast or speckle noise) across devices as a result of proprietary processing software [33], a DL system designed for one device cannot be directly translated to others [34]. Since it is common for clinics to own different OCT devices, and for patients to be imaged by different OCT devices during their care, the device-specific nature of these DL algorithms considerably limit their clinical adoption.

While there currently exists only a few major commercial manufacturers of spectral-domain OCT (SD-OCT) such as Carl Zeiss Meditec (Dublin, CA, USA), Heidelberg Engineering (Heidelberg, Germany), Optovue Inc. (Fremont, CA, USA), Nidek (Aichi, Japan), Optopol Technology (Zawiercie, Poland), Canon Inc. (Tokyo, Japan), Lecia Microsystems (Wetzlar, Germany), etc., several others have already started to or will soon be releasing the next-generation OCT devices. This further increases the complexity in deploying DL algorithms clinically. Given that reliable segmentations [33] are an important step towards diagnosing glaucoma accurately, there is a need for a single DL segmentation framework that is not only translatable across devices, but also versatile to accept data from next-generation OCT devices.

In this study, we developed a DL-based 3D segmentation framework that is easily translatable across OCT devices in a label-free manner (without the need to manually re-segment data for each device). To achieve this, we first designed an 'enhancer': a DL network that can improve the quality of OCT B-scans and harmonize image characteristics across OCT devices. Because of such pre-processing, we demonstrate that a segmentation framework trained on one device can be used to segment volumes from other unseen devices.

2. Methods

2.1. Overview

The proposed study consisted of two parts: (1) image enhancement, and (2) 3D segmentation.

We first designed and validated a DL based image enhancement network to simultaneously de-noise (reduce speckle noise), compensate (improve tissue visibility and eliminate artefacts) [35], contrast enhance (better differentiate tissue boundaries) [35], and histogram equalize (reduce intensity inhomogeneity) OCT B-scans from three commercially available SD-OCT devices (Spectralis, Cirrus, RTVue). The network was trained and tested with images from all three devices.

A 3D DL-based segmentation framework was then designed and validated to isolate six ONH tissues from OCT volumes. The framework was trained and tested separately on OCT volumes from each of the three devices with and without image enhancement. The overall schematic of the study is shown in Supplement 1 (**Fig. S1**).

2.2. Patient recruitment

A total of 450 patients were recruited from four centers: the Singapore National Eye Center (Singapore), Rajan Eye Care Hospital (Chennai, India), Aravind Eye Hospital (Madurai, India), and the East Avenue Medical Center (Quezon City, Philippines) **(Table 1)**. All subjects gave written informed consent. The study adhered to the tenets of the Declaration of Helsinki and was approved by the institutional review board of the respective hospitals. The cohort comprised of 225 healthy and 225 glaucoma subjects. The inclusion criteria for healthy subjects were: an intraocular pressure (IOP) less than 21 mmHg, healthy optic discs with a vertical cup-disc ratio (VCDR) less than or equal to 0.5, and normal visual fields tests. Glaucoma was diagnosed with the presence of glaucomatous optic neuropathy (GON), VCDR > 0.7 and/or neuroretinal rim narrowing with repeatable glaucomatous visual field defects. We excluded subjects with corneal abnormalities that could preclude the quality of the scans.

Table 1. Patient Populations and Scanning Specifications

Device	Hospital	No of subjects		Scanning Specifications
		Normal	Glaucoma	
Spectralis	Singapore National Eye Center	57	11	97 horizontal B-scans (32μm distance between B-scans, 384 A-scans per B-scan); covering an area of 15 ° x 10 ° centered on the ONH; 20x signal averaging.
	Aravind Eye Hospital	18	64	
Cirrus	Rajan Eye Care Hospital	75	75	200 horizontal B-scans (30 μm,200 A-scans per B-scans); covering an area of 6mm x 6mm centered on the ONH.
RTVue	East Avenue Medical Center	75	75	101 horizontal B-scans (40 μm distance between B-scans; 101 A-scans per B-scan); covering an area of 20 ° x 20 ° centered on the ONH.

2.3. Optical coherence tomography imaging

All 450 subjects were seated and imaged using spectral-domain OCT under dark room conditions in the respective hospitals. 150 subjects (75 healthy + 75 glaucoma) had one of their ONHs imaged using Spectralis (Heidelberg Engineering, Heidelberg, Germany), 150 (75 healthy + 75 glaucoma) using Cirrus (model: HD 5000, Carl Zeiss Meditec, Dublin, CA, USA), and another 150 (75 healthy + 75 glaucoma) using RTVue (Optovue Inc., Fermont, CA, USA). For glaucoma subjects, the eye with GON was imaged, and if both eyes met the inclusion criteria, one eye was randomly selected. For healthy controls, the right ONH was imaged. The scanning specifications for each device can be found in **Table 1**.

From the dataset of 450 volumes, 390 (130 from each device) were used for training and testing the image enhancement network, while the remaining 60 (20 from each device) were used for training and testing the 3D segmentation framework.

2.4. Image enhancement

The enhancer network was trained to reproduce simple mathematical operations including spatial averaging, compensation, contrast enhancement, and histogram equalization. When using images from a single device, the use of a DL network to perform such operations would be seen as

unnecessary, as one could readily use the mathematical operators instead. However, when mixing images from multiple devices, besides performing such enhancement operations, the network also reduces the differences in the image characteristics across the devices, resulting in images that are 'harmonized' (i.e. less device specific) – a necessary step to perform robust device-independent 3D segmentation.

2.4.1.　Image enhancement–dataset preparation

The 390 volumes were first resized (in pixels) to 448 (height) x 352 (width) x 96 (number of B-scans), and a total of 37,440 baseline B-scans (12,480 per device) were obtained. Each B-scan (**Fig. 1, (A)** [1]) was then digitally enhanced (**Fig. 1, (A)** [4]) by performing spatial averaging (each pixel value was replaced by the mean of its 8 lateral neighbours; **Fig. 1, (A)** [2]) [36], compensation with contrast enhancement (contrast exponent = 2; **Fig. 1, (A)** [3]) [35], and histogram equalization (contrast limited adaptive histogram equalization [CLAHE], clip limit = 2; **Fig. 1, (A)** [4]) [37]. The compensated image with contrast enhancement I^{SC} was defined as:

$$M_{i,j} = \sum_{k=i}^{N} I_{k,j}^{n} \tag{1}$$

$$I_{i,j}^{SC} = \frac{I_{i,j}^{n}}{2M_{i,j}} \tag{2}$$

where I was the intensity map of the image ($i = 0$: top of the image; $i = N$:bottom of the image); $M_{i,j}$ was the compensation profile that enhanced the A-scan pixel intensity at depth i for a given A-scan j; and n was the exponent used to control the contrast profile (also known as contrast exponent; $n = 2$ was used based on the results from the earlier study [35]).

The detailed implementation for CLAHE can be found in [38]. The clip limit was a factor that limited the extent of intensity over-amplification during the process of histogram equalization. A clip limit of 2 [38] was empirically chosen to prevent the intensity over-amplification, especially for the already hyperreflective structures such as a the retinal pigment epithelium.

The image enhancement network was then trained with a training dataset of 36,000 pairs (12,000 per device) of baseline and digitally-enhanced B-scans, respectively. During the process of hyperparameter tuning, 80% (28,800 pairs) of the training dataset were used for the initial training, while the remaining 20% (7,200 pairs) were used for the subsequent validation. An independent test set of 1,440 pairs were used to truly evaluate the performance of the enhancer network. B-scans from a same patient were not shared between training and testing datasets.

2.4.2.　Image enhancement–network description

Briefly, as in our earlier DL based image enhancement study [39], the proposed enhancer exploited the inherent advantages of U-Net [40] and its skip connections [41], residual learning [42], dilated convolutions [43], and multi-scale hierarchical feature extraction [44]. We used the same network architecture, except that the output layer was now activated by the sigmoid activation function [45] (originally tanh). The design (**Fig. S1**; refer to Supplement 1), implementation, significance of each component, and data augmentation details can be referred to from our earlier study [39]. The loss function was a weighted combination of both the root mean square error (RMSE) and a multi-scale perceptual loss [46] function that was based on the VGG19 DL model [47].

Pixel-to-pixel loss functions (e.g., RMSE) compare only the low-level features (i.e., edge information) between the DL prediction and their corresponding ground-truth often leading to over-smoothened (blur) images [46], especially in image-to-image translation problems (e.g., de-noising). However, perceptual loss based functions exploit the high-level features (i.e., texture, abstract patterns) [46,48–50] in these images to assess their differences, enabling the DL network
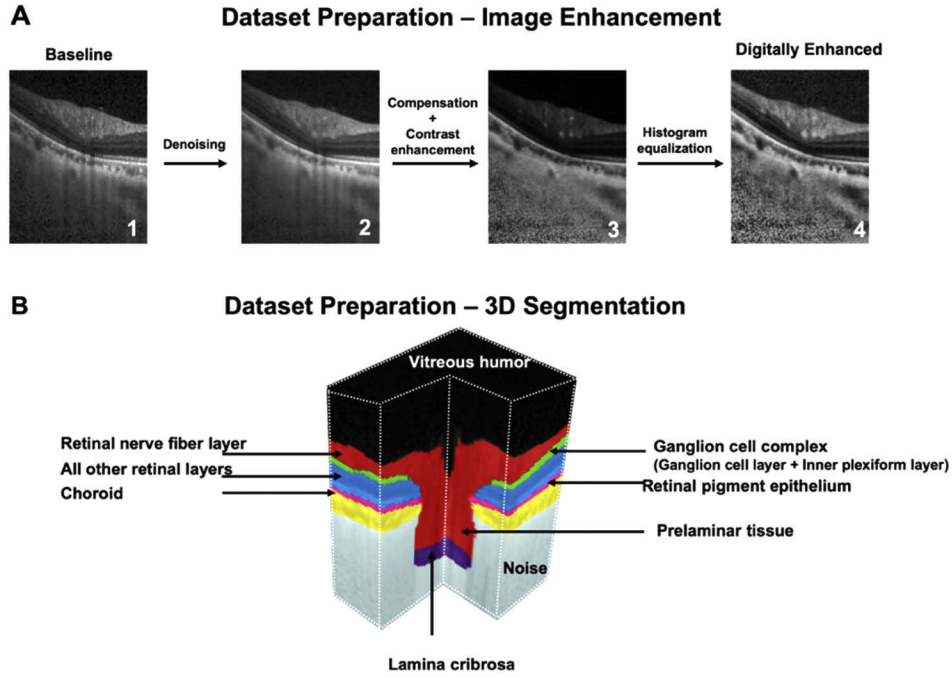
A   **Dataset Preparation – Image Enhancement**

B   **Dataset Preparation – 3D Segmentation**

Fig. 1. The dataset preparation for the image enhancement network is shown in (A). Each B-scan (A [1]) was digitally enhanced (4) by performing spatial averaging (each pixel value was replaced by the mean of its 8 lateral neighbors; A [2]) [36], compensation and contrast enhancement (contrast exponent = 2; A [3]) [35], and histogram equalization (contrast limited adaptive histogram equalization [CLAHE], clip limit = 2; A [4]) [37]. For training the 3D segmentation framework (B), the following tissues were manually segmented from OCT volumes: (1) the RNFL and prelamina (in red), (2) the ganglion cell complex (GCC; ganglion cell layer + inner plexiform layer; in cyan), (3) all other retinal layers (in blue); (4) the retinal pigment epithelium (RPE; in pink); (5) the choroid (in yellow); and (6) the lamina cribrosa (LC; in indigo). Noise (in grey) and vitreous humor (in black) were also isolated.

to achieve human-like visual understanding [19]. Thus, a weighted combination of both the loss functions allows the DL network to preserve the low- and high- level features in its predictions, limiting the effects of blurring.

To compute the perceptual loss, the output of the enhancer (referred to as 'DL-enhanced' B-scan) and its corresponding digitally-enhanced B-scan was separately passed through the VGG-19 [47] DL model that was pre-trained on the ImageNet dataset [51]. Feature maps at multiple scales (5 scales; outputs from the 2nd, 4th, 6th, 10th, and 14th convolutional layers) were extracted, and the perceptual loss was computed as the mean RMSE (average of all scales) between the extracted features from the 'DL-enhanced' and its corresponding 'digitally-enhanced' B-scan.

Experimentally, the RMSE and perceptual loss when combined (total loss) in a weighted-ratio of 1.0:0.01 offered the best performance (qualitative and quantitative; as described below). The individual and total loss functions were defined as:

$$\mathcal{L}_{RMSE}(I_{DL\ Enhanced}, I_{Digitally\ Enhanced}) = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(I(h,w)_{DL\ Enhanced} - I(h,w)_{Digitally\ Enhanced})^2} \quad (3)$$

$$\mathcal{L}_{Perceptual}(I_{DL\ Enhanced}, I_{Digitally\ Enhanced}) = \frac{\sqrt{\sum_{i=2,4,6,10,14}\frac{1}{C_iH_iW_i}||P_i(I_{DL\ Enhanced}) - P_i(I_{Digitally\ Enhanced})||^2}}{5} \quad (4)$$

$$\mathcal{L}_{Total} = \mathcal{L}_{RMSE} + 0.01 \times \mathcal{L}_{Perceptual} \tag{5}$$

where $I_{DL\ Enhanced}$ and $I_{Digitally\ Enhanced}$ are the intensity maps of the DL predicted and the ground-truth images; H and W are the height and width of the image; C_i, H_i, and W_i represent the channel depth, height, and width for the convolution layer i.

The enhancer comprised of a total of 900 K trainable parameters, and was trained end-to-end using the Adam optimizer [52], with a learning rate of 0.0001. We trained and tested on an NVIDIA GTX 1080 founders edition GPU with CUDA 10.1 and cuDNN v7.5 acceleration. Using the given hardware configuration, the DL network enhanced a single 'baseline' B-scan in under 25 ms.

2.4.3. Image enhancement–quality analysis

Upon training, the network was used to enhance the unseen baseline B-scans from all the three devices. The DL-enhanced B-scans were qualitatively assessed by two expert observers (S.K.D & T.P.H) for the following: (1) noise reduction, (2) deep tissue visibility and blood vessel shadows, (3) contrast enhancement and intensity inhomogeneity, and (4) DL induced artifacts.

2.4.4. Image enhancement–quantitative analysis

The following metrics were used to quantitatively assess the performance of the enhancer: (1) universal image quality index (UIQI) [53], and (2) structural similarity index (SSIM) [54]. We used the UIQI to assess the extent of image enhancement (baseline vs. DL-enhanced B-scans), while the MSSIM was used to assess the structural reliability of the DL-enhanced B-scans (digitally-enhanced vs. DL-enhanced).

Unlike the traditional error summation methods (e.g., RMSE etc.) that compared only the intensity differences, the UIQI jointly modeled the (1) loss of correlation (L_C) (2) luminance distortion (D_L), and (3) contrast distortion (D_C) to assess image quality [53]. It was defined as (x: baseline; y: DL-enhanced B-scan):

$$UIQI(x, y) = L_C \times D_L \times D_C \tag{6}$$

where,

$$L_C = \frac{\sigma_{xy}}{\sigma_{xy}\sigma_{xy}}; D_L = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2}; D_C = \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \tag{7}$$

L_C measured the degree of linear correlation between the baseline and DL-enhanced B-scans; D_L and D_C measured the distortion in luminance and contrast respectively; μ_x, σ_x, σ_x^2 denoted the mean, standard deviation, and variance of the intensity for B-scan x, while μ_y, σ_y, σ_y^2 denoted the same for the B-scan y; σ_{xy} was the cross-covariance between the two B-scans. The UIQI was defined between -1 (poor quality) and +1 (excellent quality).

As in our previous study [39], the SSIM (x: digitally-enhanced; y: DL-enhanced B-scan) was defined as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{8}$$

The constants C_1 and C_2 (to stabilize the division) were chosen as 6.50 and 58.52, as recommended in a previous study [54]. The SSIM was defined between -1 (no similarity) and +1 (perfect similarity).

2.5. 3D segmentation–dataset preparation

The 60 volumes used for training and testing the 3D segmentation framework (20 from each device, balanced with respect to pathology) were manually segmented (slice-wise) by an expert

observer (SD) using Amira (version 6, FEI, Hillsboro, OR). The following classes of tissues were segmented (**Fig. 1, (B)**): (1) the RNFL and prelamina (in red), (2) the ganglion cell complex (GCC; ganglion cell layer + inner plexiform layer; in cyan), (3) all other retinal layers (in blue); (4) the retinal pigment epithelium (RPE; in pink); (5) the choroid (in yellow); and (6) the lamina cribrosa (LC; in indigo). Noise (all regions below the choroid-sclera interface; in grey) and vitreous humor (black) were also isolated. We were unable to obtain a full-thickness segmentation of the LC due to limited visibility [55]. We also excluded the peripapillary sclera due to its poor visibility and the extreme subjectivity of its boundaries especially in Cirrus and RTVue volumes. To optimize computational speed, the volumes (baseline OCT + labels) for all three devices were resized (in voxels) to 112 (height) x 88 (width) x 48 (number of B-scans).

2.5.1. Deep learning based 3D segmentation of the ONH

Recent studies have demonstrated that 3D CNNs can improve the reliability of automated segmentation [56–63], and even out-perform their 2D variants [57]. This is because 3D CNNs not only harness the information from each image, but also effectively combine it with the depth-wise spatial information from adjacent images. Despite its tremendous potential, the applications of 3D CNNs in ophthalmology is still in its infancy [64–69], and has not yet been explored for the segmentation of the ONH tissues.

Further, there exist discrepancies in the delineation of ambiguous regions (e.g., choroid-sclera boundary, LC boundary) even among different well-trained DL model depending upon the type and complexity of architecture/feature extraction, learning method, etc., causing variability in the automated measurements. To address this, recent DL studies have explored ensemble learning [31,70–78], a meta-learning approach that synergizes (combine and fine-tune) [75] the predictions from multiple networks, to offer a single prediction that is closest to the ground-truth. Specifically, ensemble learning has shown to better generalize and increase the robustness of segmentations in OCT [31,71] and other medical imaging modalities [72–74,77].

In this study, we designed and validated 'ONH-Net', a 3D segmentation framework inspired by the popular 3D U-Net [58] to isolate six ONH tissues from OCT volumes. The ONH-Net consisted of three segmentation networks (3D CNNs) and one 3D CNN for ensemble learning (referred to as the 'ensembler'). Each of the three segmentation CNNs offered an equally plausible segmentation, which were then synergized by the 'ensembler' to yield the final 3D segmentation of the ONH tissues.

2.5.2. 3D segmentation–network description

The design of the three segmentation CNNs was based on the 3D U-Net [58] and its variants [65]. Briefly, each CNN (**Fig. 2, (A)**) comprised of four micro-U-Nets (**Fig. 2, (B)**; **μ-U-Nets**) and a latent space (**Fig. 2, (C); LS**). We used multi-scale hierarchical feature extraction [39,44] to obtain smoother tissue boundaries. The three CNNs differed from each other only in the design of the 'feature extraction' (FE) units (**Fig. 2, (D); Types 1-3**), thus resulting in three equally plausible segmentations.

The ensembler (**Fig. 2, (E)**) consisted of three sets of 3D convolutional layers, with each set separated by a dropout layer (50%) [79] to limit overfitting and improve generalizability.

Each of the three segmentation CNNs were first trained end-to-end with the same labeled-dataset. The ONH-Net was then assembled by using the three trained CNNs as parallel input pipelines to the ensembler network (**Fig. 2, (F)**). Finally, we trained the ONH-Net (ensembler weights: trainable; segmentation CNN weights: frozen) end-to-end using the same aforementioned labeled-dataset. During this process, each segmentation CNN provided equally plausible segmentation feature maps (obtained from the last 3D convolution layer), which were then concatenated and fed to the ensembler for fine-tuning. The ONH-Net was trained separately for each device. The design and implementation details can be found in the **Supplement 1**.
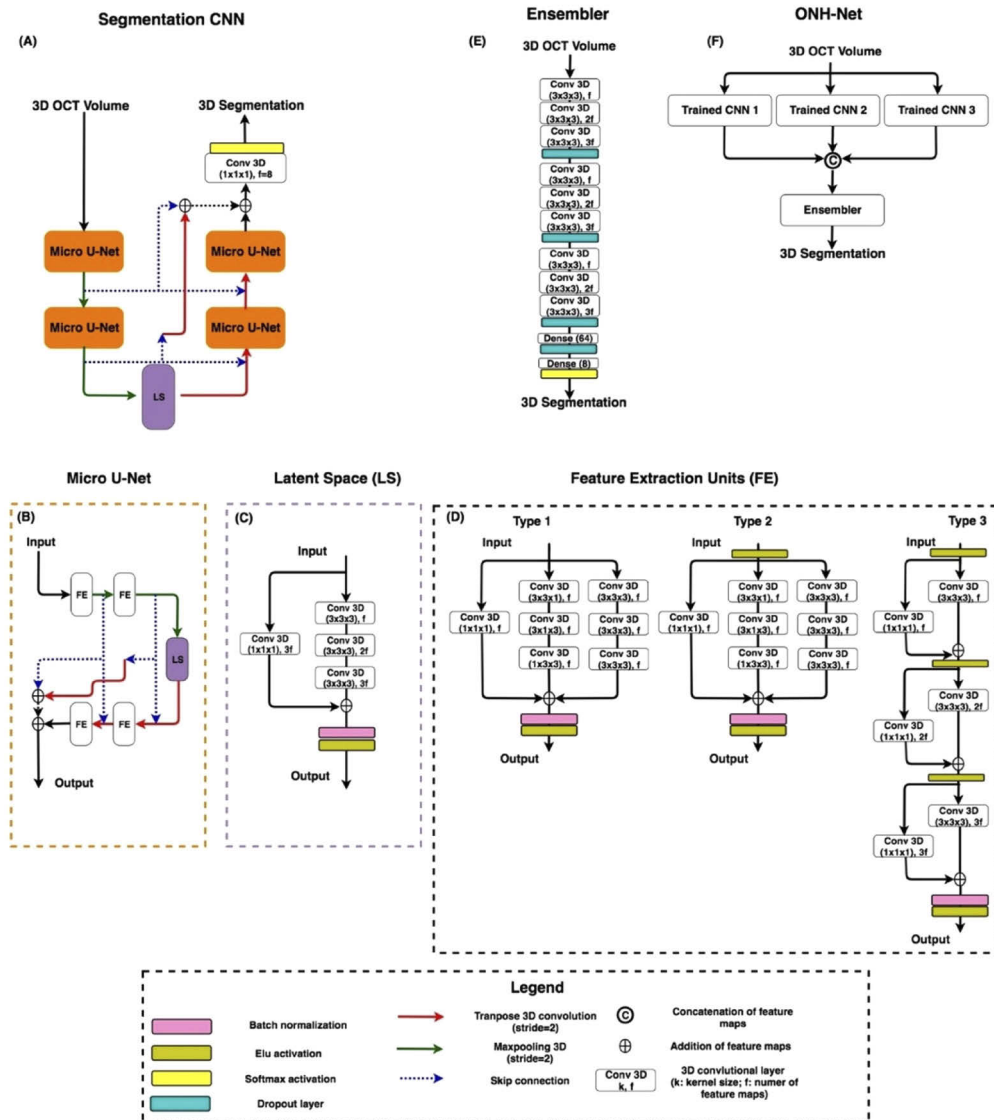
Fig. 2. The DL architecture of the proposed 3D segmentation framework (three segmentation CNNs + one ensembler network) is shown. Each CNN (A) comprised of four micro-U-Nets (μ-U-Nets; B) and a latent space (LS; C). The three CNNs differed from each other only in the design of the 'feature extraction' (FE) units (D; Types 1-3). The ensembler (E) consisted of three sets of 3D convolutional layers, with each set separated by a dropout layer. ONH-Net (F) was then assembled by using the three trained CNNs as parallel input pipelines to the ensembler network.

All the DL networks (segmentation CNNs, ONH-Net) were trained with the stochastic gradient descent (SGD; learning rate:0.01; Nesterov momentum:0.05 [80]) optimizer, and the Jaccard distance was used as the loss function [26]. We empirically observed that the use of SGD optimizer with Nesterov momentum offered a better generalizability and faster convergence compared to Adam optimizer [52] for OCT segmentation problems that typically use limited data, while Adam performed better for image-to-image translation problems (i.e., enhancement [39]) that use much larger datasets. However, we are unable to theoretically explain this yet for our case.

Given the limitations in hardware, all the DL networks were trained with a batch size of 1. To circumvent the scarcity in data, all the DL networks used custom data augmentation techniques (B-scans wise) as in our earlier study [26]. We ensured that the same data augmentation was used for each B-scan in a given volume.

The three CNNs consisted of 7.2 M (Type 1), 7.2 M (Type 2), and 12.4 M (Type 3) trainable parameters, while the ONH-Net consisted of 28.86 M parameters (2.06M trainable parameters [ensembler], 26.8M non-trainable parameters [trained CNNs with weights frozen]).

All the DL networks were trained and tested on NVIDIA GTX 1080 founders edition GPU with CUDA 10.1 and cuDNN v7.5 acceleration. Using the given hardware configuration, the ONH-Net was trained in 12 hours (10 hours for each CNN [trained in parallel; one per GPU]; 2 hours for fine-tuning with the ensembler). Once trained, each OCT volume was segmented in about 120 ms.

2.5.3. 3D segmentation–training and testing

We used a five-fold cross-validation approach (for each device) to train and test the performance of ONH-Net. In this process, the labeled-dataset (20 OCT volumes + manual segmentations) was split into five equal parts. One part ('left-out' set; 4 OCT volumes + manual segmentations) was used as the testing dataset, while the remaining four parts (16 OCT volumes + manual segmentations) were used as the training dataset. The entire process was repeated five times, each with a different 'left-out' testing dataset (and corresponding training dataset). Totally, for each device, the segmentation performance was assessed on 20 OCT volumes (4 per validation; 5-fold cross-validation).

2.5.4. 3D segmentation–qualitative analysis

The segmentations obtained from the trained ONH-Net on unseen data were manually reviewed by expert observers (S.D. & T.P.H) and compared against their corresponding manual segmentations.

2.5.5. 3D segmentation–quantitative analysis

We used the following metrics to quantitatively assess the segmentation performance: (1) Dice coefficient (*DC*); (2) specificity (*Sp*); and (3) sensitivity (*Sn*). The metrics were computed in 3D for the following tissues: (1) the RNFL and prelamina; (2) the GCC; (3) all other retinal layers; (4) the RPE; and (5) the choroid. Given the subjectivity in the visibility of the posterior LC boundary [55], we excluded LC from quantitative assessment. Noise and vitreous humor were also excluded from quantitative assessment.

The Dice coefficient (*DC*) was used to assess the spatial overlap between the manual and DL segmentations (between 0 [no overlap] and 1 [perfect overlap]). For each tissue, the *DC* was computed as:

$$DC = \frac{2 \times |D \cap M|}{|D| + |M|} \tag{9}$$

where D and M were the voxels that represented the chosen tissue in the DL segmented and the corresponding manually segmented volumes.

Specificity (*Sp*) was used to assess the true negative rate of the segmentation framework and was defined as:

$$Sp = \frac{|\overline{D} \cap \overline{M}|}{|M|} \tag{10}$$

where $\overline{D}$ represented the voxels that did not belong to the chosen tissue in the DL segmented volume, while $\overline{M}$ represented the same in the corresponding manually segmented volume.

Sensitivity (*Sn*) was used to assess the true positive rate and was defined as:

$$Sn = \frac{|D \cap M|}{|M|} \tag{11}$$

2.5.6. 3D segmentation–effect of image enhancement

To assess if image enhancement had an effect on segmentation performance, we trained and tested ONH-Net on the baseline and the DL-enhanced datasets. For both datasets, ONH-Net was trained on any one device (Spectralis/Cirrus/RTVue), but tested on all the three devices (Spectralis, Cirrus, and RTVue). Paired t-tests were used to compare the differences (means) in the segmentation performance (Dice coefficients, sensitivities, specificities; mean of all tissues) for both cases. For all experiments, the segmentation performance was compared between glaucoma and healthy subjects.

2.5.7. 3D segmentation–device independency

When tested on a given device (Spectralis/Cirrus/RTVue), paired t-tests were used to assess the differences (Spectralis vs. Cirrus; Cirrus vs. RTVue; RTVue vs. Spectralis) in the segmentation performance depending on the device used for training ONH-Net. The process was performed with both baseline and DL-enhanced datasets.

3. Results

3.1. Image enhancement–qualitative analysis

The enhancer was tested on a total of 1440 (480 from each device) unseen baseline B-scans. In the DL-enhanced B-scans from all the three devices (**Fig. 3, 3rd column**), the ONH-tissue boundaries appeared sharper with a uniformly enhanced intensity profile (compared to respective 'baseline' B-scans). The blood vessel shadows were also reduced with improved deep tissue (choroid-scleral interface, LC) visibility. In all cases, the DL-enhanced B-scans were consistently similar to their corresponding digitally-enhanced B-scans (**Fig. 3, 2nd column**), with no DL induced artifacts.

3.2. Image enhancement–quantitative analysis

The mean UIQI (mean ± SD) for the DL-enhanced B-scans (compared to baseline B-scans) were: 0.94 ± 0.02, 0.95 ± 0.03, and 0.97 ± 0.01 for Spectralis, Cirrus, and RTVue, respectively, indicating improved image quality.

In all cases, the mean SSIM (mean ± SD) for the DL-enhanced B-scans (compared to digitally-enhanced B-scans) were: 0.95 ± 0.02, 0.91 ± 0.02, and 0.93 ± 0.03, for Spectralis, Cirrus, and RTVue, respectively, indicating strong structural similarity.

3.3. 3D segmentation performance–qualitative analysis

When trained and tested on the baseline volumes from the same device (**Fig. 4, 5, and 6; 4th column**), ONH-Net successfully isolated all ONH layers. Further, the DL segmentations appeared consistent with their respective manual segmentations (**Fig. 4, 5, and 6; 3rd column; refer Fig. S3 in Supplement 1 for 3D visualization**), with no difference in the segmentation performance between glaucoma and healthy OCT volumes.
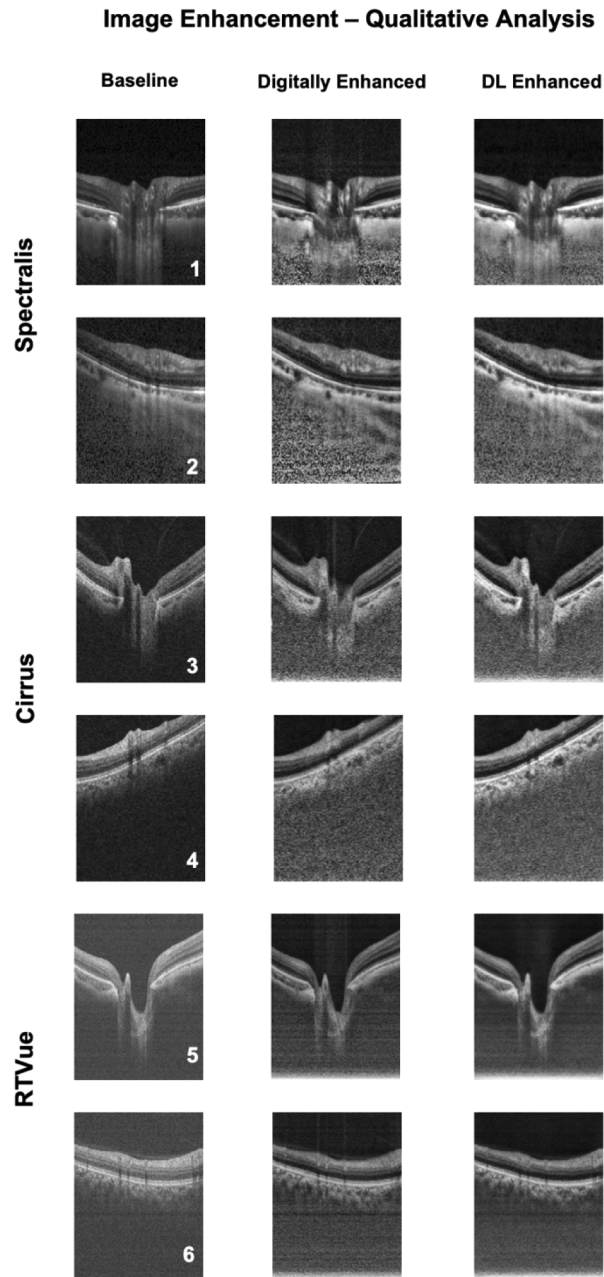
Fig. 3. The qualitative performance of the image enhancement network is shown for six randomly selected (1-6) subjects (2 per device). The 1ˢᵗ, 2ⁿᵈ and 3ʳᵈ columns represent the baseline, digitally-enhanced, and the corresponding DL-enhanced B-scans for patients imaged with Spectralis (1-2), Cirrus (3-4), and RTVue (5-6) devices, respectively.
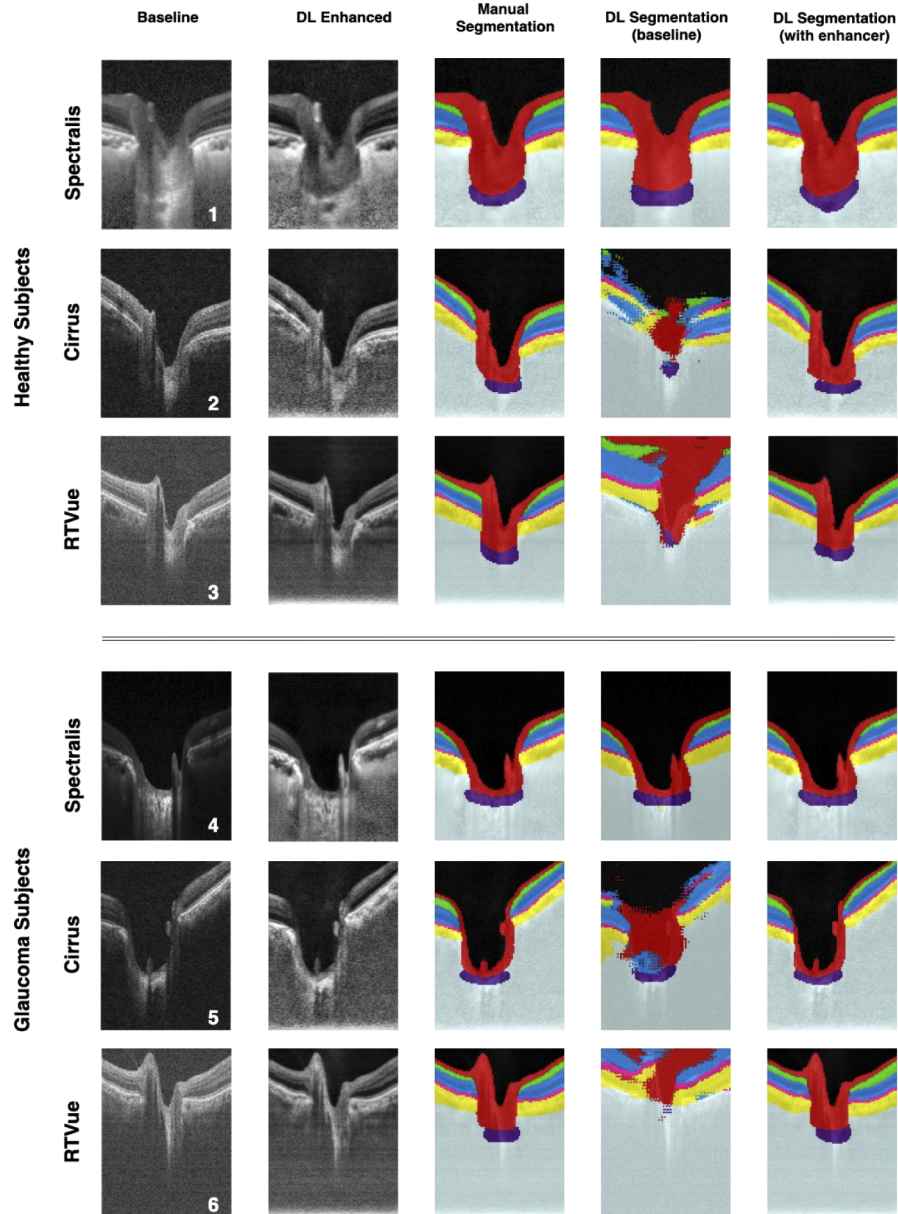
Fig. 4. The qualitative performance (one randomly chosen B-scan per volume) of the ONH-Net 3D segmentation framework for three healthy (1-3) and three glaucoma (4-6) subjects is shown. The framework was trained on volumes from Spectralis, and tested on Spectralis (1, 4), Cirrus (2,5), and RTVue (3,6) devices respectively. The 1,[st] 2,[nd] and 3rd columns represent the baseline, DL enhanced, and the corresponding manual segmentation for the chosen B-scan. The 4th and 5th columns represent the DL segmentations when ONH-Net was trained and tested using the baseline and DL enhanced volumes, respectively.
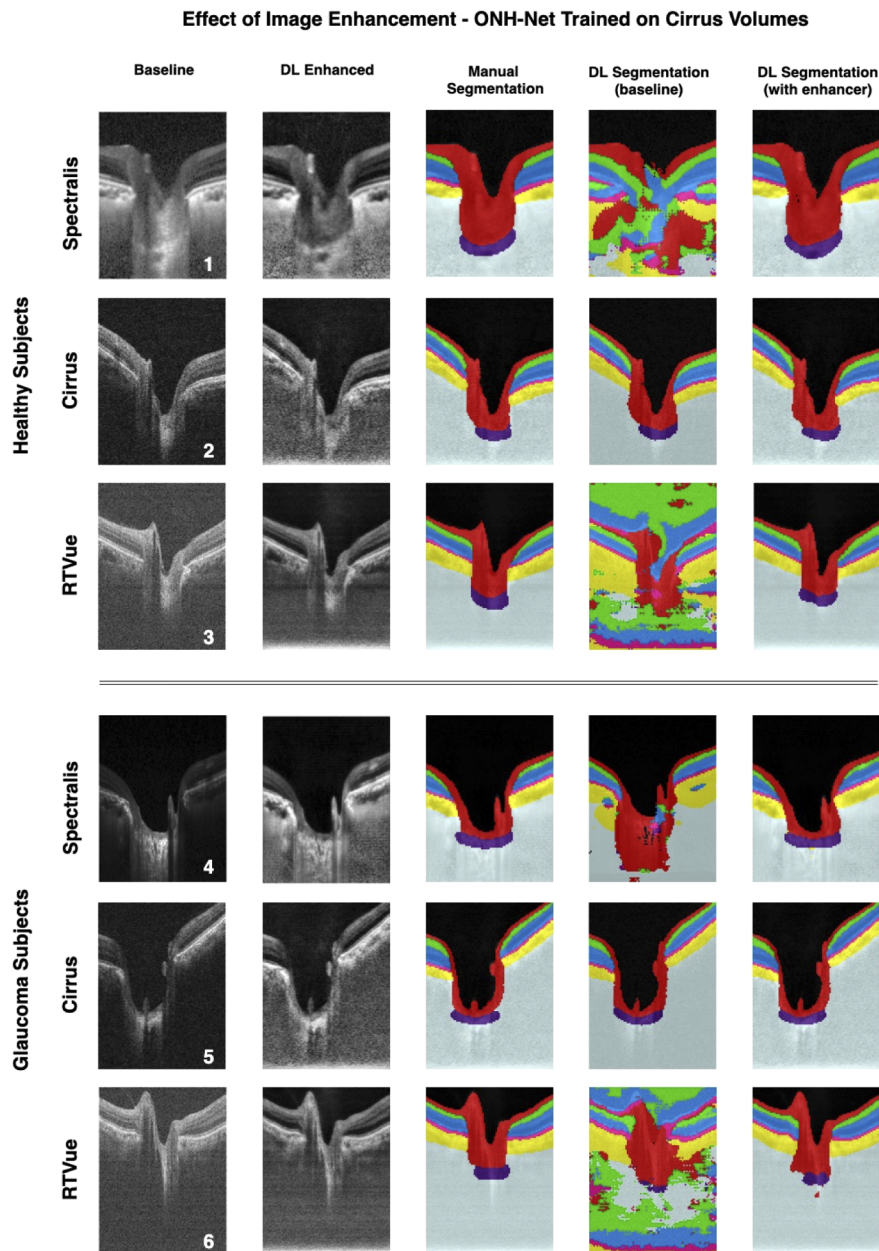
Fig. 5. The qualitative performance (one randomly chosen B-scan per volume) of the ONH-Net 3D segmentation framework for three healthy (1-3) and three glaucoma (4-6) subjects is shown. The framework was trained on volumes from Cirrus, and tested on Spectralis (1, 4), Cirrus (2,5), and RTVue (3,6) devices respectively. The 1,[st] 2,[nd] and 3rd columns represent the baseline, DL enhanced, and the corresponding manual segmentation for the chosen B-scan. The 4th and 5th columns represent the DL segmentations when the ONH-Net was trained and tested using the baseline and DL enhanced volumes respectively.

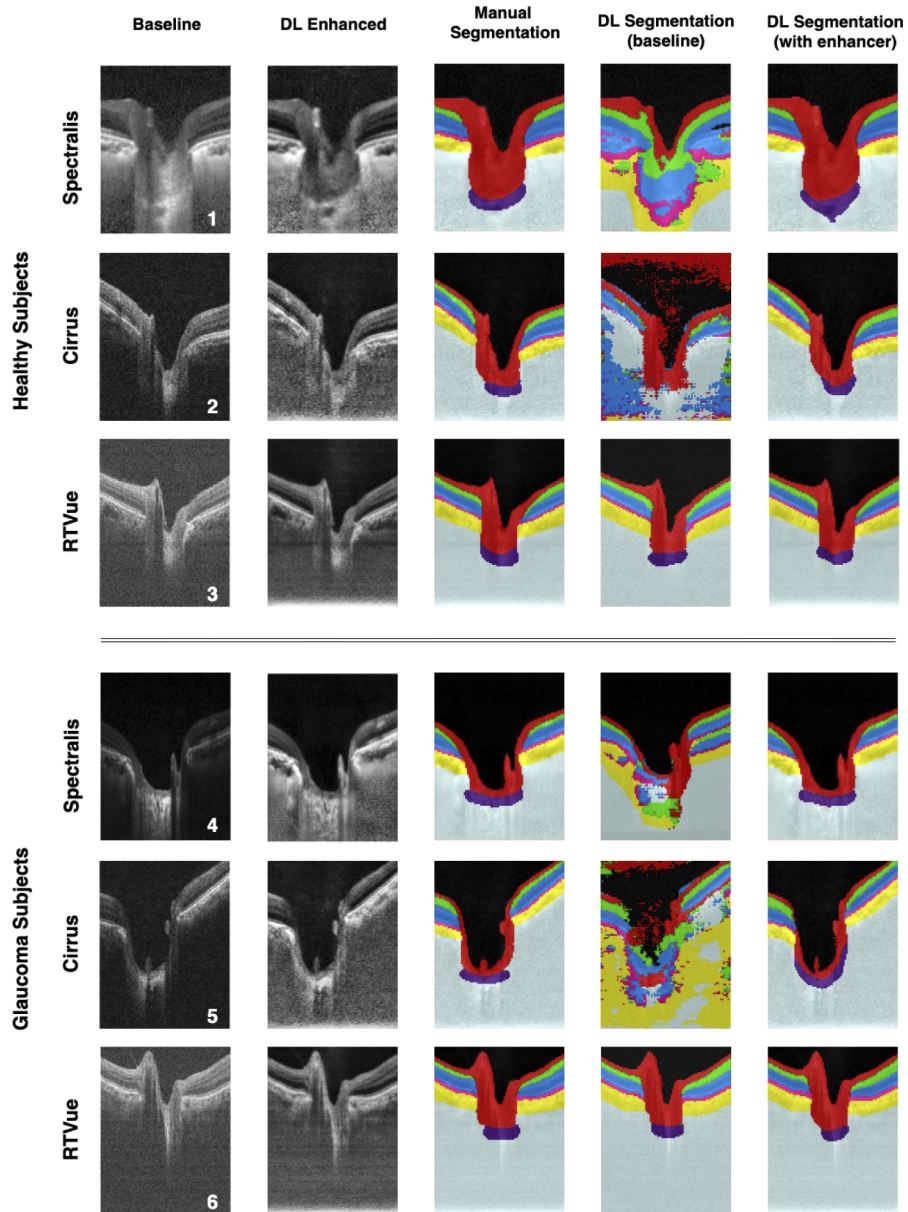**Effect of Image Enhancement - ONH-Net Trained on RTVue Volumes**

Fig. 6. The qualitative performance (one randomly chosen B-scan per volume) of the ONH-Net 3D segmentation framework from three healthy (1-3) and three glaucoma (4-6) subjects is shown. The framework was trained on volumes from Cirrus, and tested on Spectralis (1, 4), Cirrus (2,5), and RTVue (3,6) devices respectively. The 1,[st] 2,[nd] and 3rd columns represent the baseline, DL enhanced, and the corresponding manual segmentation for the chosen B-scan. The 4th and 5th columns represent the DL segmentations when the ONH-Net was trained and tested using the baseline and DL enhanced volumes respectively.

3.4. 3D segmentation performance–quantitative analysis

When trained and tested on the baseline volumes (same device), the mean Dice coefficients (mean of all tissues; mean ± SD) were: 0.93 ± 0.02, 0.93 ± 0.02, and 0.93 ± 0.02 for Spectralis, Cirrus, and RTVue, respectively. The mean sensitivities / specificities (mean of all tissues; mean ± SD) were: $0.94 \pm 0.02 / 0.99 \pm 0.00$, $0.93 \pm 0.02 / 0.99 \pm 0.00$, and $0.93 \pm 0.02 / 0.99 \pm 0.00$, respectively.

3.5. 3D segmentation performance–effect of image enhancement and device independency

Without image enhancement (baseline dataset), ONH-Net trained with one device was unable to segment even a single ONH tissue reliably on the other two devices (**Fig. 4; 2nd, 3rd, 5th, 6th rows; 4th column; similarly, for Fig. 5–6**). In all cases, dice coefficients were always lower than 0.65, sensitivities lower than 0.77, and specificities lower than 0.80.
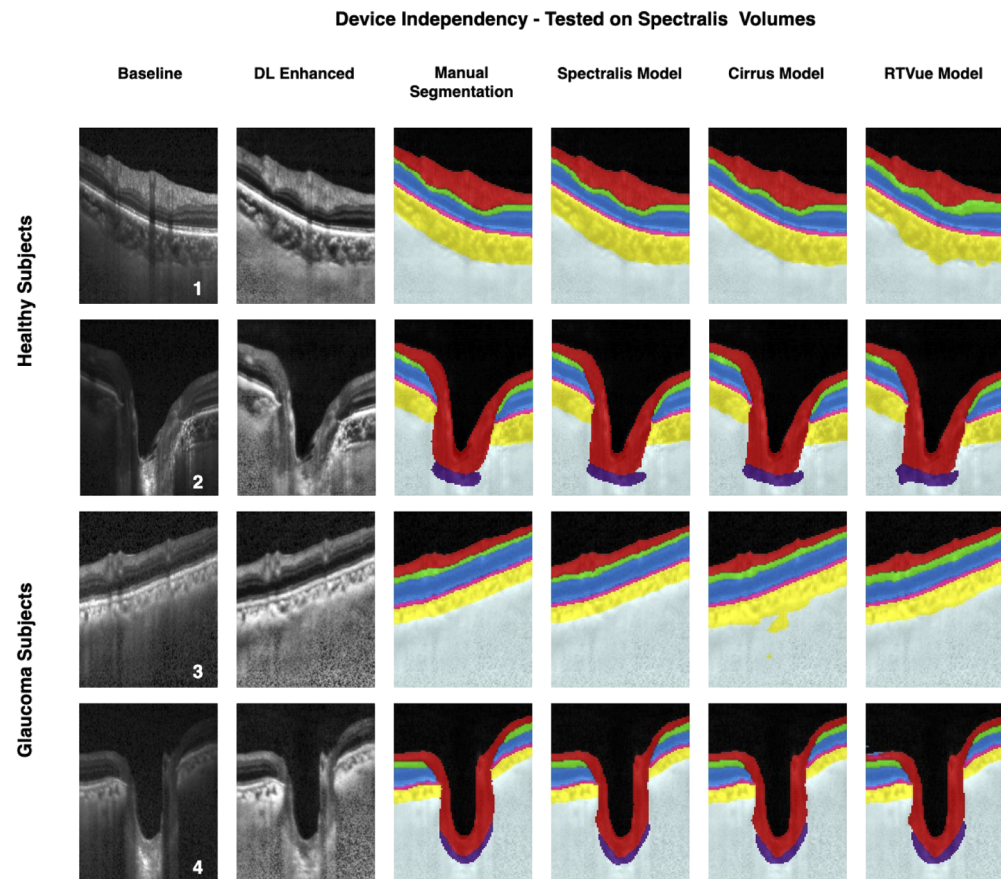

Fig. 7. The device independent segmentation performance of the proposed ONH-Net is shown. The segmentation performance on four randomly chosen (1-2 healthy; 3-4 glaucoma) Spectralis volumes from the test set are shown (one B-scan per volume). The 1st, 2nd, and 3rd columns represent the baseline, DL enhanced and the corresponding manual segmentation for the chosen B-scan. The 4th, 5th, and 6th columns represent the segmentations obtained when tested using the Spectralis, Cirrus, and RTVue trained segmentation model.

However, with image enhancement (DL-enhanced dataset), ONH-Net trained with one device was able to accurately segment all tissue layers on the other two devices with mean Dice coefficients and sensitivities > 0.92 (**Fig. 4–6, 5th column**). In addition, when trained and tested on the same device, it performed better for several ONH layers (p < 0.05). The tissue wise quantitative metrics for the aforementioned cases can be found in Supplement 1 (**Tables S1-S6**).

Further, when trained and tested with the DL-enhanced OCT volumes, irrespective of the device used for training, there were no significant differences (p<0.05) in the segmentation performance for all tissues (**Fig. 7, 8, 9**), except for the LC. The tissue wise quantitative metrics for the individual cases can be found in **Supplement 1 (Tables S7-S12)**. Finally, we observed so significant differences in the segmentation performance between glaucoma and healthy subjects.
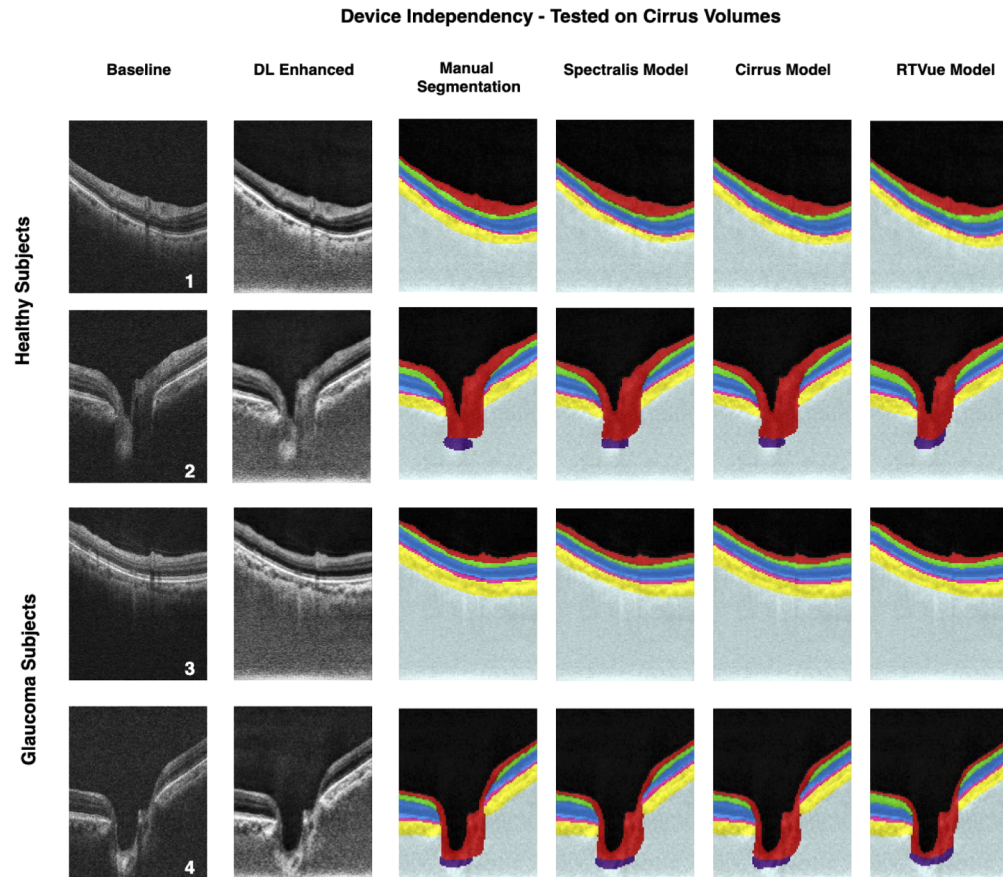

Fig. 8. The device independent segmentation performance of the proposed ONH-Net is shown. The segmentation performance on four randomly chosen (1-2 healthy; 3-4 glaucoma) Cirrus volumes from the test set are shown (one B-scan per volume). The 1st, 2nd, and 3rd columns represent the baseline, DL enhanced and the corresponding manual segmentation for the chosen B-scan. The 4th, 5th, and 6th columns represent the segmentations obtained when tested using the Spectralis, Cirrus, and RTVue trained segmentation model.
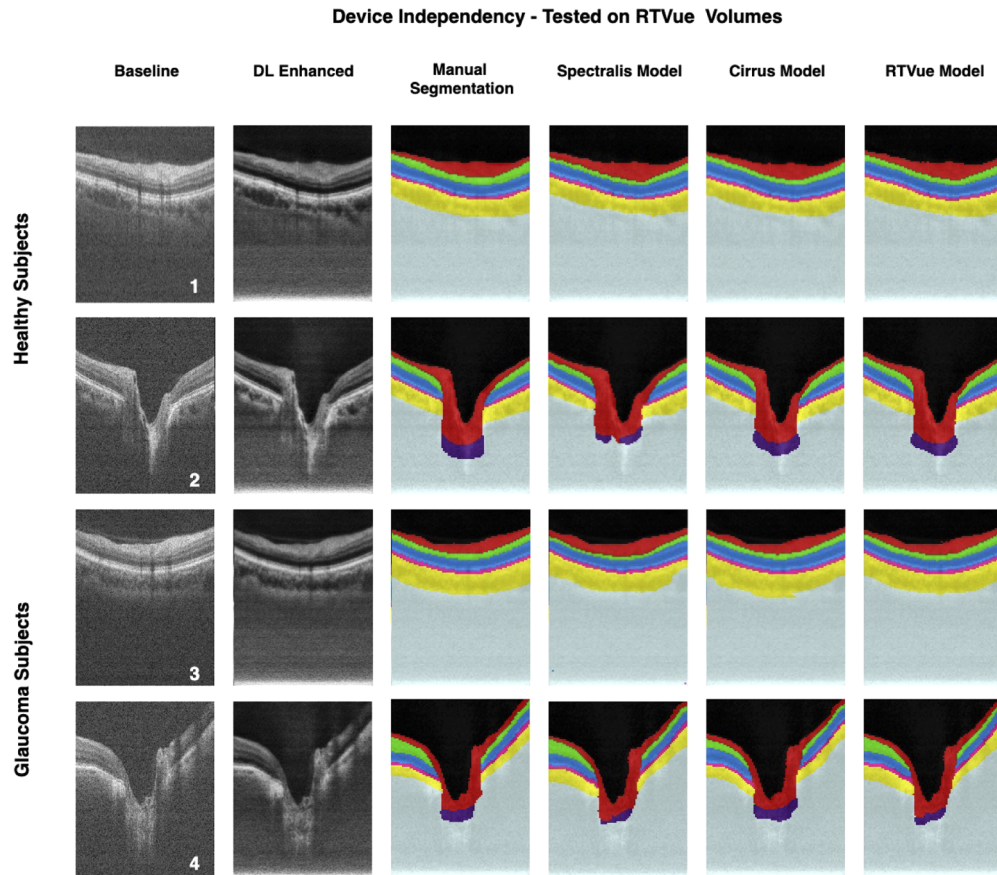
Fig. 9. The device independent segmentation performance of the proposed ONH-Net is shown. The segmentation performance on four randomly chosen (1-2 healthy; 3-4 glaucoma) RTVue volumes from the test set are shown (one B-scan per volume). The 1st, 2nd, and 3rd columns represent the baseline, DL enhanced and the corresponding manual segmentation for the chosen B-scan. The 4th, 5th, and 6th columns represent the segmentations obtained when tested using a Spectralis, Cirrus, and RTVue trained segmentation model.

4. Discussion

In this study, we proposed a 3D segmentation framework (ONH-Net) that is easily translatable across OCT devices in a label-free manner (i.e. without the need to manually re-segment data for each device). Specifically, we developed 2 sets of DL networks. The first (referred to as the 'enhancer') was able to enhance OCT image quality from 3 OCT devices, and harmonized image-characteristics across these devices. The second performed 3D segmentation of 6 important ONH tissue layers. We found that the use of the 'enhancer' was critical for our segmentation network to achieve device independency. In other words, our 3D segmentation network trained on any of 3 devices successfully segmented ONH tissue layers from the other two devices with high performance.

Our work suggests that it is possible to automatically segment OCT volumes from a new OCT device without having to re-train ONH-Net with manual segmentations from that device. Besides existing commercial SD-OCT manufacturers, the democratization and emergence of OCT as the clinical gold-standard for *in vivo* ophthalmic examinations [81] has encouraged the

entry of several new manufacturers to the market as well. Further, owing to advancements in imaging technology, there has been a rise of the next generation devices: swept-source [82], polarization sensitive [83], and adaptive optics [84] based OCTs. Given that preparing reliable manual segmentations (training data) for OCT-based DL algorithms requires months of training for a skilled technician, and that it would take more than 8 hours of manual work to accurately segment just a single 3D volume for just a limited number of tissue layers (here 6), it will soon become practically infeasible to perform manual segmentations for all OCT brands, device models, generations, and applications. Furthermore, only a few research groups have successfully managed to exploit DL to fully-isolate ocular structures from 3D OCT images [25–29,32], and only for a very limited number of devices. There is therefore a strong need for a single DL segmentation framework that can easily be translated across all existing and future OCT devices, thus eliminating the excruciating task of preparing training datasets manually. Our approach provides a high-performing solution to that problem. Eventually, we believe, this could open doors for multi-device glaucoma management.

While classical image processing frameworks can indeed be used to improve the quality of OCT images, the resulting enhanced images would still retain the device-specific image characteristics (i.e., intensity and contrast profile). In this study, we hypothesized that by reducing the device-specific characteristics of the enhanced images, it might be possible to 'deceive' DL networks that subsequently use them into perceiving images from multiple devices in a similar manner. Given that this might not be possible to achieve using simple mathematical operations, we proposed the idea of a DL approach that didn't require explicit or hardcoded functions, but rather learnt to do the same organically. During the training, when repeatedly exposed to images from multiple devices, the enhancer network constantly refined its weights to best suit all of them. As a result, we visually observed that the DL enhanced images had characteristics (i.e., intensity and contrast profiles) that were less-specific to any one device. However, we were unable to quantify this observation yet, and further research is required to understand the same.

In this study, we found that the use of enhancer was crucial for ONH-Net to achieve device independency, in other words, the ability to segment OCT volumes from devices it had not been trained with earlier. This can be attributed to the design of the proposed DL networks that allowed a perception of visual information through a host of low-level (e.g. tissue boundaries) and high-level abstract features (e.g. speckle pattern, intensity, and contrast profile). When image enhancement was used as a pre-processing step, the enhancer not only improved the quality of low-level features, but also reduced differences in high-level abstract features across OCT devices, thus 'deceiving' ONH-Net into perceiving volumes from all three devices similarly. This enabled ONH-Net trained on the DL-enhanced OCT volumes from one device to successfully isolate the ONH tissues from the other two devices with very high performance (mean Dice coefficients > 0.92). Note that such a performance is superior to that of our previous 2D segmentation framework that also had the additional caveat that it only worked on a single device [26]. In addition, irrespective of the device used for training, there were no significant differences (p>0.05) in segmentation performance. In all cases, our DL segmentations were deemed clinically reliable (refer to **Supplement 1**).

To confirm the hypothesis on the need for the enhancer network, we also trained and tested the ONH-Net with only the digitally enhanced images. Although the quality of the digitally enhanced images was comparable to that of the DL enhanced images, the segmentation performance when tested on unseen devices was still poor (refer **Table S13** in **Supplement 1**). This can be attributed to the fact that the digitally enhanced OCT images still retained their device specific image characteristics, thus, the re-iterating the necessity to obtain harmonized images as a precursor to achieve device independency.

In a recent landmark study, De Fauw et al. [71] proposed the idea of using device-independent representations (segmentation maps) for the diagnosis of retinal pathologies from OCT images.

However, the study was not truly device-independent, as, even though the diagnosis network was device-independent, the segmentation network was still trained with multiple devices. Similarly, our approach may not truly be considered as device-independent. While ONH-Net is device-independent, the enhancer (on which ONH-Net relies on) needs to be trained with data for all considered devices. But this is a still a very acceptable option, because the enhancer only requires un-labeled images (i.e. non-segmented; ~100 OCT volumes) for any new device that is being considered. After which, automated segmentation can still be performed without ever needing manual segmentation for that new device. Such a task would require a few minutes rather than several weeks/months needed for manual segmentations.

Finally, the proposed approach should not be confused with 'transfer learning' [85], a DL technique gaining momentum in medical imaging [74,86–89]. In this technique, a DL network is first pre-trained on large-size datasets (e.g. ImageNet [51]), and when subsequently fine-tuned on a smaller dataset for the task of interest (e.g. segmentation), it re-uses the pre-trained knowledge (high-level representations [e.g. edges, shapes]) to generalize better. In our approach, the generalization of ONH-Net was achieved using the enhanced images, and not the actual knowledge of the enhancer network, thus keeping the learning of both the networks mutually exclusive, yet necessary.

There are several limitations to this study that warrant further discussion. First, we used only 20 volumes in total to test the segmentation performance for each device. Second, the study was performed only using spectral-domain OCT devices, but not swept-source. Third, although the enhancer simultaneously addressed multiple issues affecting image quality, we were unable to quantify the effect of each. Also, we were unable to quantify the extent to which the 'DL-enhanced' B-scans were harmonized. Fourth, we observed slight differences in LC curvature and LC thickness when the LC was segmented using ONH-Net trained on different devices (**Fig. 7, Fig. 8, Fig. 9; 2nd and 4th rows**). Given the significance of LC morphology in glaucoma [90], this subjectivity could affect glaucoma diagnosis. This is yet to be tested. Further, in a few B-scans (**Fig. 7, Fig. 8, Fig. 9; 6th column**), we observed that the GCC segmentations were thicker when the ONH-Net was trained on volumes from RTVue device. These variabilities might limit a truly multi-device glaucoma management. We are currently exploring the use of advanced DL concepts such as semi-supervised learning [91] to address these issues that may have occurred as a result of limited training data.

Finally, although ONH-Net was invariant to volumes with glaucoma, it is unclear if the same will be true in the presence of other conditions such as cataract [92], peripapillary atrophy [93], and high-myopia [94] that commonly co-exist with glaucoma.

To summarize, we demonstrate as a proof of concept that it is possible to develop DL segmentation tools that are easily translatable across OCT devices without ever needing additional manual segmentation data. The core contributions of this study to achieve the same included: (1) the development of ONH-Net – a highly modular DL approach for the segmentation of 3D OCT volumes of the ONH; and (2) the development of the enhancer – a DL approach to enhance the OCT image quality from multiple devices and simultaneously reduce the differences in the device-specific image characteristics. Through our core contributions, we were able to address (as a proof of concept) the device-specific nature of DL algorithms, an important factor that limits the translations and wide-spread adoption of DL algorithms in clinics. Finally, we hope the proposed framework can potentially help patients for the longitudinal follow-up on multiple devices, and encourage multi-center glaucoma studies also.

Funding

Disclosures

Dr. Michaël J. A. Girard and Dr. Alexandre H. Thiéry are co-founders of Abyss Processing.

See Supplement 1 for supporting content.

References

1. J. S. Schuman, "Spectral domain optical coherence tomography for glaucoma (an AOS thesis)," Trans Am Ophthalmol Soc **106**, 426–458 (2008).
2. C. Bowd, R. N. Weinreb, J. M. Williams, and L. M. Zangwill, "The retinal nerve fiber layer thickness in ocular hypertensive, normal, and glaucomatous eyes with optical coherence tomography," Arch. Ophthalmol. **118**(1), 22–26 (2000).
3. A. Miki, F. A. Medeiros, R. N. Weinreb, S. Jain, F. He, L. Sharpsten, N. Khachatryan, N. Hammel, J. M. Liebmann, C. A. Girkin, P. A. Sample, and L. M. Zangwill, "Rates of retinal nerve fiber layer thinning in glaucoma suspect eyes," Ophthalmology **121**(7), 1350–1358 (2014).
4. Z. Lin, S. Huang, B. Xie, and Y. Zhong, "Peripapillary Choroidal Thickness and Open-Angle Glaucoma: A Meta-Analysis," J Ophthalmol **2016**, 5484568 (2016).
5. 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–OCT584 (2016).
6. K. J. Halupka, B. J. Antony, M. H. Lee, K. A. Lucy, R. S. Rai, H. Ishikawa, G. Wollstein, J. S. Schuman, and R. Garnavi, "Retinal optical coherence tomography image enhancement via deep learning," Biomed. Opt. Express **9**(12), 6205–6221 (2018).
7. S. C. Park, J. Brumm, R. L. Furlanetto, C. Netto, Y. Liu, C. Tello, J. M. Liebmann, and R. Ritch, "Lamina cribrosa depth in different stages of glaucoma," Invest. Ophthalmol. Visual Sci. **56**(3), 2059–2064 (2015).
8. F. A. Almobarak, N. O'Leary, A. S. C. Reis, G. P. Sharpe, D. M. Hutchison, M. T. Nicolela, and B. C. Chauhan, "Automated Segmentation of Optic Nerve Head Structures With Optical Coherence Tomography," Invest. Ophthalmol. Visual Sci. **55**(2), 1161–1168 (2014).
9. K. X. Cheong, L. W. Lim, K. Z. Li, and C. S. Tan, "A novel and faster method of manual grading to measure choroidal thickness using optical coherence tomography," Eye **32**(2), 433–438 (2018).
10. S. L. Mansberger, S. A. Menda, B. A. Fortune, S. K. Gardiner, and S. Demirel, "Automated Segmentation Errors When Using Optical Coherence Tomography to Measure Retinal Nerve Fiber Layer Thickness in Glaucoma," Am. J. Ophthalmol. **174**, 1–8 (2017).
11. B. Al-Diri, A. Hunter, and D. Steel, "An Active Contour Model for Segmenting and Measuring Retinal Vessels," IEEE Trans. Med. Imaging **28**(9), 1488–1497 (2009).
12. M. A. Mayer, J. Hornegger, C. Y. Mardin, and R. P. Tornow, "Retinal Nerve Fiber Layer Segmentation on FD-OCT Scans of Normal Subjects and Glaucoma Patients," Biomed. Opt. Express **1**(5), 1358–1383 (2010).
13. S. Niu, Q. Chen, L. de Sisternes, D. L. Rubin, W. Zhang, and Q. Liu, "Automated retinal layers segmentation in SD-OCT images using dual-gradient and spatial correlation smoothness constraint," Comput. Biol. Med. **54**, 116–128 (2014).
14. J. Tian, P. Marziliano, M. Baskaran, T. A. Tun, and T. Aung, "Automatic segmentation of the choroid in enhanced depth imaging optical coherence tomography images," Biomed. Opt. Express **4**(3), 397–411 (2013).
15. L. Zhang, K. Lee, M. Niemeijer, R. F. Mullins, M. Sonka, and M. D. Abramoff, "Automated segmentation of the choroid from clinical SD-OCT," Invest. Ophthalmol. Visual Sci. **53**(12), 7510–7519 (2012).
16. Z. Hu, M. D. Abràmoff, Y. H. Kwon, K. Lee, and M. K. Garvin, "Automated Segmentation of Neural Canal Opening and Optic Cup in 3D Spectral Optical Coherence Tomography Volumes of the Optic Nerve Head," Invest. Ophthalmol. Visual Sci. **51**(11), 5708–5717 (2010).
17. H. Ishikawa, J. Kim, T. R. Friberg, G. Wollstein, L. Kagemann, M. L. Gabriele, K. A. Townsend, K. R. Sung, J. S. Duker, J. G. Fujimoto, and J. S. Schuman, "Three-Dimensional Optical Coherence Tomography (3D-OCT) Image Enhancement with Segmentation-Free Contour Modeling C-Mode," Invest. Ophthalmol. Visual Sci. **50**(3), 1344–1349 (2009).
18. R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, "Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map," Med. Image Anal. **17**(8), 907–928 (2013).
19. K. Lee, H. Zhang, A. Wahle, M. D. Abràmoff, and M. Sonka, "Multi-layer 3D Simultaneous Retinal OCT Layer Segmentation: Just-Enough Interaction for Routine Clinical Use," in *VipIMAGE 2017*, (Springer International Publishing, 2018), 862–871.
20. Y. Sun, T. Zhang, Y. Zhao, and Y. He, "3D Automatic Segmentation Method for Retinal Optical Coherence Tomography Volume Data Using Boundary Surface Enhancement," arXiv:1508.00966 [cs.CV] (2015).
21. C. Wang, Y. Wang, D. Kaba, H. Zhu, Y. Lv, Z. Wang, X. Liu, and Y. Li, "Segmentation of Intra-retinal Layers in 3D Optic Nerve Head Images," in *Image and Graphics*, (Springer International Publishing, 2015), 321–332.

22. D. Alonso-Caneiro, S. A. Read, and M. J. Collins, "Automatic segmentation of choroidal thickness in optical coherence tomography," Biomed. Opt. Express **4**(12), 2795–2812 (2013).
23. R. A. Alshareef, S. Dumpala, S. Rapole, M. Januwada, A. Goud, H. K. Peguda, and J. Chhablani, "Prevalence and Distribution of Segmentation Errors in Macular Ganglion Cell Analysis of Healthy Eyes Using Cirrus HD-OCT," PLoS One **11**(5), e0155319 (2016).
24. J. Chhablani, T. Krishnan, V. Sethi, and I. Kozak, "Artifacts in optical coherence tomography," Saudi. J. Ophthalmol. **28**(2), 81–87 (2014).
25. S. K. Devalla, K. S. Chin, J.-M. Mari, T. A. Tun, N. G. 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**(1), 63–74 (2018).
26. 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).
27. 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), 2732–2744 (2017).
28. D. Lu, M. Heisler, S. Lee, G. W. Ding, E. Navajas, M. V. Sarunic, and M. F. Beg, "Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network," Med. Image Anal. **54**, 100–110 (2019).
29. 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).
30. X. Sui, Y. Zheng, B. Wei, H. Bi, J. Wu, X. Pan, Y. Yin, and S. Zhang, "Choroid segmentation from Optical Coherence Tomography with graph-edge weights learned from deep convolutional neural networks," Neurocomputing **237**, 332–341 (2017).
31. T. H. Pham, S. K. Devalla, A. Ang, S. Zhi Da, A. H. Thiery, C. Boote, C.-Y. Cheng, V. Koh, and M. J. A. Girard, "Deep Learning Algorithms to Isolate and Quantify the Structures of the Anterior Segment in Optical Coherence Tomography Images," arXiv:1909.00331 [eess.IV] (2019).
32. F. G. Venhuizen, B. van Ginneken, B. Liefers, M. J. J. P. van Grinsven, S. Fauser, C. Hoyng, T. Theelen, and C. I. Sánchez, "Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks," Biomed. Opt. Express **8**(7), 3292–3316 (2017).
33. T. C. Chen, A. Hoguet, A. K. Junk, K. Nouri-Mahdavi, S. Radhakrishnan, H. L. Takusagawa, and P. P. Chen, "Spectral-Domain OCT: Helping the Clinician Diagnose Glaucoma: A Report by the American Academy of Ophthalmology," Ophthalmology **125**(11), 1817–1827 (2018).
34. D. Romo-Bucheli, P. Seeböck, J. I. Orlando, B. S. Gerendas, S. M. Waldstein, U. Schmidt-Erfurth, and H. Bogunović, "Reducing image variability across OCT devices with unsupervised unpaired learning for improved segmentation of retina," Biomed. Opt. Express **11**(1), 346–363 (2020).
35. M. J. Girard, N. G. Strouthidis, C. R. Ethier, and J. M. Mari, "Shadow removal and contrast enhancement in optical coherence tomography images of the human optic nerve head," Invest. Ophthalmol. Visual Sci. **52**(10), 7738–7748 (2011).
36. W. Wu, O. Tan, R. R. Pappuru, H. Duan, and D. Huang, "Assessment of frame-averaging algorithms in OCT image analysis," Ophthalmic Surg. Lasers Imaging **44**(2), 168–175 (2013).
37. S. M. P. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," Comput. Gr. Image Process **39**(3), 355–368 (1987).
38. B. S. Min, D. K. Lim, S. J. Kim, and J. H. Lee, "A Novel Method of Determining Parameters of CLAHE Based on Image Entropy," IJSEIA **7**(5), 113–120 (2013).
39. S. K. Devalla, G. Subramanian, T. H. Pham, X. Wang, S. Perera, T. A. Tun, T. Aung, L. Schmetterer, A. H. Thiéry, and M. J. A. Girard, "A Deep Learning Approach to Denoise Optical Coherence Tomography Images of the Optic Nerve Head," Sci. Rep. **9**(1), 14454 (2019).
40. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, (Springer International Publishing, 2015), 234–241.
41. E. V. Michal Drozdzal, Gabriel Chartrand, Samuel Kadoury, and Chris Pal, "The Importance of Skip Connections in Biomedical Image Segmentation," arXiv:1608.04117 [cs.CV] (2016).
42. X. Z. Kaiming He, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," arXiv:1512.03385 [cs.CV] (2015).
43. F. Yu and V. Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," arXiv:1511.07122 [cs.CV] (2015).
44. Y. Liu, M. M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer Convolutional Features for Edge Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, 5872–5881.
45. C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," arXiv:1811.03378 [cs.LG] (2018).

46. Justin Johnson, Alexandre Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," arXiv:1603.08155 [cs.CV] (2016).

47. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *3rd International Conference on Learning Representations (ICLR 2015)*, (San Diego, CA, USA, 2015).

48. H. Cheong, S. K. Devalla, T. H. Pham, Z. Liang, T. A. Tun, X. Wang, S. Perera, L. SchmeŠerer, A. Tin, C. Boote, A. H. Thiery, and M. J. A. Girard, "DeshadowGAN: A Deep Learning Approach to Remove Shadows from Optical Coherence Tomography Images," arXiv:1910.02844v1 [eess.IV] (2019).

49. Karim Armanious, Chenming Jiang, Marc Fischer, Thomas Küstner, Konstantin Nikolaou, Sergios Gatidis, and B. Yang, "MedGAN: Medical Image Translation using GANs," arXiv:1806.06397 [cs.CV] (2018).

50. Haofu Liao, Wei-An Lin, Jianbo Yuan, S. Kevin Zhou, and J. Luo, "Artifact Disentanglement Network for Unsupervised Metal Artifact Reduction," arXiv:1906.01806 [eess.IV] (2019).

51. J. Deng, W. Dong, R. Socher, L. Li, L. Kai, and F.-F. Li, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009), 248–255.

52. J. B. Diederik and P. Kingma, "Adam: A Method for Stochastic Optimization," arXiv:1412.6980 [cs.LG] (2014).

53. W. Zhou and A. C. Bovik, "A universal image quality index," IEEE Signal Process. Lett. **9**(3), 81–84 (2002).

54. W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Process. **13**(4), 600–612 (2004).

55. J. M. Mari, N. G. Strouthidis, S. C. Park, and M. J. A. Girard, "Enhancement of Lamina Cribrosa Visibility in Optical Coherence Tomography Images Using Adaptive Compensation," Invest. Ophthalmol. Visual Sci. **54**(3), 2238–2247 (2013).

56. F. Milletari, N. Navab, and S.-A. Ahmadi, *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation* (2016), pp. 565–571.

57. C. M. Deniz, S. Xiang, R. S. Hallyburton, A. Welbeck, J. S. Babb, S. Honig, K. Cho, and G. Chang, "Segmentation of the Proximal Femur from MR Images using Deep Convolutional Neural Networks," Sci. Rep. **8**(1), 16485 (2018).

58. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, (Springer International Publishing, 2016), 424–432.

59. H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori, "An application of cascaded 3D fully convolutional networks for medical image segmentation," Comput. Med. Imag. Grap. **66**, 90–99 (2018).

60. Y. Huang, Q. Dou, Z.-X. Wang, L.-Z. Liu, Y. Jin, L. Chaofeng, L. Wang, H. Chen, and R.-H. Xu, *3D RoI-aware U-Net for Accurate and Efficient Colorectal Tumor Segmentation* (2019).

61. D. Müller and F. Kramer, "MIScnn: A Framework for Medical Image Segmentation with Convolutional Neural Networks and Deep Learning," arXiv:1910.09308 [eess.IV] (2019).

62. H. Roth, L. Lu, A. Farag, A. Sohn, and R. Summers, Spatial Aggregation of Holistically-Nested Networks for Automated Pancreas Segmentation (2016).

63. Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P. A. Heng, "3D deeply supervised network for automated segmentation of volumetric medical images," Med. Image Anal. **41**, 40–54 (2017).

64. A. Abbasi, A. Monadjemi, L. Fang, H. Rabbani, and Y. Zhang, "Three-dimensional optical coherence tomography image denoising through multi-input fully-convolutional networks," Comput. Biol. Med. **108**, 1–8 (2019).

65. S. Feng, W. Zhu, H. Zhao, F. Shi, D. Xiang, and X. Chen, VinceptionC3D: a 3D convolutional neural network for retinal OCT image classification, SPIE Medical Imaging (SPIE, 2019), Vol. 10949.

66. N. Eladawi, M. Elmogy, M. Ghazal, L. Fraiwan, A. Aboelfetouh, A. Riad, H. Sandhu, R. Keynton, and A. El-Baz, "Early Signs Detection of Diabetic Retinopathy Using Optical Coherence Tomography Angiography Scans Based on 3D Multi-Path Convolutional Neural Network," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019), 1390–1394.

67. M.-X. Li, S.-Q. Yu, W. Zhang, H. Zhou, X. Xu, T.-W. Qian, and Y.-J. Wan, "Segmentation of retinal fluid based on deep learning: application of three-dimensional fully convolutional neural networks in optical coherence tomography images," Int. J. Ophthalmol **12**, 1012–1020 (2019).

68. E. Noury, S. Sudhakaran, R. Chang, A. Ran, C. Cheung, S. Thapa, H. Rao, S. Dasari, M. Riyazuddin, S. Nagaraj, and R. Zadeh, *Detecting Glaucoma Using 3D Convolutional Neural Network of Raw SD-OCT Optic Nerve Scans* (2019).

69. S. Maetschke, B. Antony, H. Ishikawa, G. Wollstein, J. Schuman, and R. Garnavi, "A feature agnostic approach for glaucoma detection in OCT volumes," PLoS One **14**(7), e0219126 (2019).

70. A. Benou, R. Veksler, A. Friedman, and T. Riklin Raviv, "Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences," Med. Image Anal. **42**, 145–159 (2017).

71. J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, G. van den Driessche, B. Lakshminarayanan, C. Meyer, F. Mackinder, S. Bouton, K. Ayoub, R. Chopra, D. King, A. Karthikesalingam, C. O. Hughes, R. Raine, J. Hughes, D. A. Sim, C. Egan, A. Tufail, H. Montgomery, D. Hassabis, G. Rees, T. Back, P. T. Khaw, M. Suleyman, J. Cornebise, P. A. Keane, and O. Ronneberger, "Clinically applicable deep learning for diagnosis and referral in retinal disease," Nat. Med. **24**(9), 1342–1350 (2018).

72. N. Georgiev and A. Asenov, "Automatic Segmentation of Lumbar Spine MRI Using Ensemble of 2D Algorithms," in *Computational Methods and Clinical Applications for Spine Imaging*, (Springer International Publishing, 2019), 154–162.

73. K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, and B. Glocker, "Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, (Springer International Publishing, 2018), 450–462.

74. X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, J. R. Ledsam, M. K. Schmid, K. Balaskas, E. J. Topol, L. M. Bachmann, P. A. Keane, and A. K. Denniston, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," Lancet Glob Health **1**(6), e271–e297 (2019).

75. Q. Lyu, H. Shan, and G. Wang, *MRI Super-Resolution with Ensemble Learning and Complementary Priors* (2019).

76. L. Rokach, "Ensemble-based classifiers," Artif. Intell. Rev. **33**(1-2), 1–39 (2010).

77. T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," Array **3-4**, 100004 (2019).

78. F. Li, H. Chen, Z. Liu, X.-d. Zhang, M.-s. Jiang, Z.-z. Wu, and K.-q. Zhou, "Deep learning-based automated detection of retinal diseases using optical coherence tomography images," Biomed. Opt. Express **10**(12), 6204–6226 (2019).

79. N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," J. Mach. Learn. Res. **15**, 1929–1958 (2014).

80. S. Vaswani, F. Bach, and M. Schmidt, "Fast and Faster Convergence of SGD for Over-Parameterized Models and an Accelerated Perceptron," arXiv:1810.07288 [cs.LG] (2018).

81. J. Fujimoto and E. Swanson, "The Development, Commercialization, and Impact of Optical Coherence Tomography," Invest. Ophthalmol. Visual Sci. **57**(9), OCT1–OCT13 (2016).

82. A. Yasin Alibhai, C. Or, and A. J. Witkin, "Swept Source Optical Coherence Tomography: a Review," Curr. Ophthalmol. Rep. **6**(1), 7–16 (2018).

83. J. F. de Boer, C. K. Hitzenberger, and Y. Yasuno, "Polarization sensitive optical coherence tomography - a review [Invited]," Biomed. Opt. Express **8**(3), 1838–1873 (2017).

84. M. Pircher and R. J. Zawadzki, "Review of adaptive optics OCT (AO-OCT): principles and applications for retinal imaging [Invited]," Biomed. Opt. Express **8**(5), 2536–2562 (2017).

85. K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," J. Big. Data **3**(1), 9 (2016).

86. J. Chang, J. Yu, T. Han, H. Chang, and E. Park, "A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer," in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2017), 1–4.

87. M. Maqsood, F. Nazir, U. Khan, F. Aadil, H. Jamal, I. Mehmood, and O.-Y. Song, "Transfer Learning Assisted Classification and Detection of Alzheimer's Disease Stages Using 3D MRI Scans," Sensors **19**(11), 2645 (2019).

88. A. Hosny, C. Parmar, T. P. Coroller, P. Grossmann, R. Zeleznik, A. Kumar, J. Bussink, R. J. Gillies, R. H. Mak, and H. J. W. L. Aerts, "Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study," PLoS Med. **15**(11), e1002711 (2018).

89. M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging," arXiv:1902.07208 [cs.CV] (2019).

90. S. H. Lee, T. W. Kim, E. J. Lee, M. J. Girard, and J. M. Mari, "Diagnostic Power of Lamina Cribrosa Depth and Curvature in Glaucoma," Invest. Ophthalmol. Visual Sci. **58**(2), 755–762 (2017).

91. G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, and M. D. Bruijne, "Semi-Supervised Medical Image Segmentation via Learning Consistency under Transformations," arXiv:1911.01218 [cs.CV] (2019).

92. J. M. Heltzer, "Coexisting glaucoma and cataract," Ophthalmology **111**(2), 408–409 (2004).

93. J. B. Jonas, "Clinical implications of peripapillary atrophy in glaucoma," Curr. Opin. Ophthalmol. **16**(2), 84–88 (2005).

94. L. Xu, Y. Wang, S. Wang, Y. Wang, and J. B. Jonas, "High Myopia and Glaucoma Susceptibility: The Beijing Eye Study," Ophthalmology **114**(2), 216–220 (2007).