OCT-GAN: Single Step Shadow and Noise Removal from Optical Coherence Tomography Images of the Human Optic Nerve Head

Haris Cheong[1, 2, 3], Sripad Krishna Devalla[1], Thanadet Chuangsuwanich[1, 2, 3], Tin A. Tun[3], Xiaofei Wang[4], Tin Aung[3, 5, 13], Leopold Schmetterer[3, 6, 7, 8, 9, 10, 13], Martin L. Buist[2], Craig Boote[1, 11], Alexandre H. Thiéry[12], and Michaël J. A. Girard[1, 3, 13] ⋆

[1] Ophthalmic Engineering & Innovation Laboratory (OEIL), Singapore Eye Research Institute, Singapore National Eye Centre, Singapore

[2] Department of Biomedical Engineering, Faculty of Engineering, National University of Singapore, Singapore

[3] Singapore Eye Research Institute, Singapore National Eye Centre, Singapore

[4] Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing, China

[5] Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[6] School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore, Singapore

[7] SERI-NTU Advanced Ocular Engineering (STANCE), Singapore, Singapore

[8] Centre for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria

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

[10] Institute of Clinical and Experimental Ophthalmology, Basel, Switzerland

[11] Structural Biophysics Group, School of Optometry & Vision Sciences, Cardiff University, UK

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

[13] Duke-NUS Medical School, Singapore

⋆ mgirard@ophthalmic.engineering

February 9, 2021

Abstract

Abstract: Speckle noise and retinal shadows within OCT B-scans occlude important edges, fine textures and deep tissues, preventing accurate and robust diagnosis by algorithms and clinicians. We developed a single process that successfully removed both noise and retinal shadows from unseen single-frame B-scans within 10.4ms. Mean average gradient magnitude (AGM) for the proposed algorithm was 57.2% higher than current state-of-the-art, while mean peak signal to noise ratio (PSNR), contrast to noise ratio (CNR), and structural similarity index metric (SSIM) increased by 11.1%, 154% and 187% respectively compared to single-frame B-scans. Mean intralayer contrast (ILC) improvement for the retinal nerve fiber layer (RNFL), photoreceptor layer (PR) and retinal pigment epithelium (RPE) layers decreased from 0.362 ± 0.133 to 0.142 ± 0.102, 0.449 ± 0.116 to 0.0904 ± 0.0769, 0.381 ± 0.100 to 0.0590 ± 0.0451 respectively. The proposed algorithm reduces the necessity for long image acquisition times, minimizes

expensive hardware requirements and reduces motion artifacts in OCT images.

1   Introduction

Optical coherence tomography (OCT) is a well-established, noninvasive clinical imaging tool for in vivo viewing of cross-sectional images of optical nerve head (ONH) tissues with micrometer resolution [1]. Although there have been vast improvements in imaging resolution, speed, and depth of OCT imaging, some limitations exist. Since OCT uses coherent illumination, speckle noise is a major source of noise that degrades the image quality of OCT B-scans [2].

Speckle noise is a multiplicative noise inherent in coherence imaging and is caused by multiple forward and backward scattering of light waves. It frequently reduces contrast and the grainy speckle noise pattern has been found to limit both the axial and lateral effective image resolution [3]. Subtle but important morphological details, such as individual tissue layers [4, 5, 6] are prevented from being identified and observed [7], making speckle noise detrimental to clinical diagnosis [8].

The most common speckle removal approach adopted in commercial OCT machines is B-scan averaging [9]. Spectralis machines (Heidelberg Engineering, Heidelberg, Germany) use an algorithm called automatic real time (ART) to combine multiple B-scans which have been captured at the same location [10]. In the ART algorithm, the signal-to-noise-ratio of the image is continuously increasing with approximately the square root of the number of averaged single B-scans. ART is used with active eye tracking (TruTrack) which detects motion in the scanning laser ophthalmoscopy (SLO) image and repositions the OCT beam so that the OCT image is precisely aligned even in cases with some eye movement. We refer to all B-scans that have been processed using B-scan averaging as "ART" images.

Although high quality images can be produced using this technique, the longer scan durations (3.5 minutes for a standard OCT scan) can result in the presence of image artifacts such as registration errors [10] and motion artifacts [11] on the final image. This is mostly due to eye or head motions during scanning [12]. The inability of elderly or young patients to remain fixated for long periods of time further render this technique difficult to obtain 3D scans of the ONH [13] of relatively good quality.

Furthermore, ART does not prevent OCT signals obtained from locations beneath retinal blood vessels from being significantly diminished due to the scattering at the blood flowing through retinal blood vessels. This phenomenon produces artifacts in OCT images known as retinal shadows. These artifacts appear perpendicular to retinal layers, interrupting tissue layer continuity and causing errors in segmentation [14]. This in turn leads to inaccurate extraction of important structural metrics such as thickness of the retinal nerve fiber layer (RNFL), which is important in glaucoma monitoring [15]. Retinal shadows also reduce visibility of deep structures such as the anterior and posterior boundaries of the lamina cribrosa (LC), as weak, reflected signals from these structures are further attenuated by the lower incident light intensity within retinal shadows [16].

Recently, deep learning techniques have shown promise in reducing speckle noise. Mao et al. used a deep, fully convolutional encoding-decoding framework to suppress noise and perform super resolution analysis of input images [17]. Later in 2018, Ma et al. proposed an edge-sensitive generative adversarial network (GAN) to remove speckle noise from OCT images produced by commercial scanners [18]. Devalla et al. leveraged deep neural networks (DNNs), residual learning, and dilated convolutions to extract multi-scale features and contextual information to recover information lost due to speckle noise in OCT images of the ONH [19]. Many other works attempted to remove speckle noise with varying success, with a common recognition of the major quality degrading factor that speckle noise inflict on OCT images [20, 21, 22].

Some have attempted to remove retinal shadows as well. In 2011, Girard et al. developed two OCT modelling approaches to be used in conjunction, one to compensate for light attenuation and the other to enhance contrast in OCT images [16]. Later, in 2018, Vupparaboina et al. illustrated an improvement in choroid representation after shadow compensation [23]. Our more recent work [24] used a weighted custom loss function that removed shadows from ART images and illuminated faint features within retinal shadows. However, the above-mentioned algorithms require high quality images free from speckle noise and motion artifacts to function well, preventing users in possession of single-frame images and low-cost hardware from availing themselves to this technology.

The presence of speckle noise, motion artifacts, and retinal shadows often interact and overlap, complicating processes that attempt to alleviate and remove these quality degrading phenomena [4, 25]. Such attempts are often tedious and prone to errors, because multiple separate processes need to work together to remove each artifact individually, with the ordering of artifact removal causing issues for the other processes. In this study, we aimed to develop an algorithm to remove both speckle noise and retinal shadows within a single step. By doing so, we will be able to reduce the cost of OCT devices by using simpler OCT imaging hardware enhanced by software.

2   Methods

2.1   Patient Recruitment

24 healthy subjects (average age: 25.5 ± 2.5 years) were recruited at the Singapore National Eye Centre (SNEC). All subjects gave written informed consent. This study adhered to the tenets of the Declaration of Helsinki and was approved by the institutional review board of the hospital. The inclusion criteria for healthy subjects were an intraocular pressure (IOP) of less than 21mmHg and healthy optic nerves with a vertical cup-to-disc ratio of ≤ 0.5.

2.2   OCT Imaging

Recruited subjects were seated and imaged in dark room conditions by a single operator (TAT). A standard spectral domain OCT system (Spectralis; Heidelberg Engineering, Heidelberg, Germany) was used to image both eyes of each subject. Each volume contained 97 horizontal B-scans (32-µm distance between B-scans; 384 A-scans per B-scan) from a rectangular area 15° × 10° centered on the ONH. We obtained 2328 ART B-scans (clean, averaged over 75 frames). This constituted our training dataset. Another 300 ART B-scans and 300 single-frame B-scans(noisy, without signal averaging) were independently obtained and used as our testing dataset. Enhanced depth imaging [26] and eye tracking [27, 28] modalities were used during the acquisition.

2.3   Overall description

Our algorithm was a single step approach to removing both speckle noise and retinal blood vessel shadows simultaneously. It had two actively-trained networks competing with one another. The first network was referred to as the shadow detector network, and it predicted which pixels would be considered as shadowed pixels. The second network was referred to as the image processor, and it aimed to remove shadows and speckle noise simultaneously from single-frame OCT images such that the first network (shadow detection network) could no longer identify shadowed pixels. Briefly, we trained the shadow detection network once on ART images with added Gaussian noise with their corresponding manually segmented shadow mask as the ground truth. We added Gaussian noise instead of speckle noise as training deep learning models to denoise

B-scans with Gaussian noise provided empirically better results. It is unfortunately, difficult to ascertain the reason for this phenomenon.

First, binary segmentation masks (size 496 × 384) were manually created for all 2328 training ART images using ImageJ [29] by one observer (HC) where shadowed pixels were labelled as 1 and shadow-free pixels were labelled as 0. Next, we attempted to model single-frame images by creating "noisy" images. This was done by adding Gaussian noise to ART images. To extract comprehensive feature representations of each image, we required capable, pre-trained networks [30, 31, 32] for feature extraction. Seven feature representations were extracted from each noisy image and its ART counterpart using three pre-trained perceptual networks [30, 31, 32] in order to train the image processor network to output ART quality images from input noisy images. Finally, we trained the image processor network by passing the ART image (with artificial Gaussian noise) as input and using the predicted binary masks as part of the loss function. More details about the overall algorithm can be found below (Fig. 1).
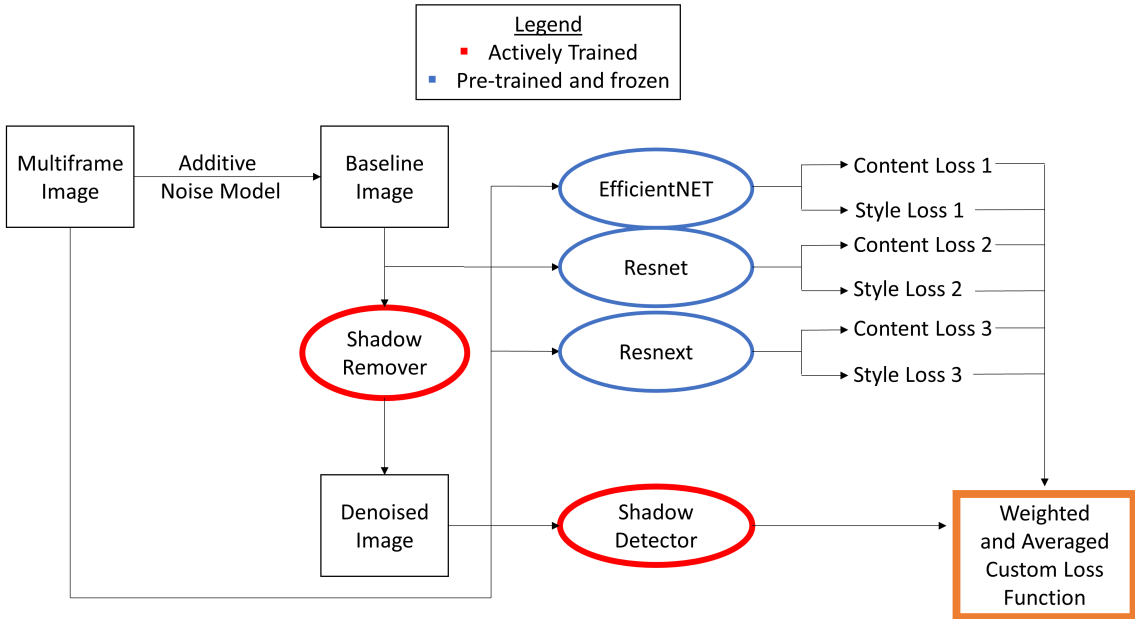
Figure 1: Overview of the proposed deep learning framework.

2.4   Shadow Detection Network and Image Processor Network Architecture

Both the shadow detection and image processor network architectures were created by modifying the standard UNet architecture [33] (Fig. 2). The shadow detection network was trained with a simple binary cross entropy loss [34], using the noisy images (ART image + Gaussian noise) as inputs and the manually segmented masks as ground truths. Each modified UNet had a sigmoid layer as its final activation, making it a per-pixel binary classifier. Each modified UNet first performed two convolutions with kernel size 3 and stride 1, followed by a ReLU activation [35] after each convolution. after each convolution. Then, images were downsampled with a 2×2 kernel, halving the height and width of the feature maps. This occurred four times, with the number of feature maps at each smaller size increasing from 1 to 64, 128, 256, and 512, respectively. The shadow detection network was comprised of two towers. A downsampling tower halved the dimensions of the input image (size 512 × 512) via maxpooling to capture contextual information such

as the spatial arrangement of tissues, and an upsampling tower sequentially restored it back to its original resolution to capture the local information such as tissue texture [19]. Output images were then linearly scaled to values between 0 and 1 by subtraction of its minimum value and division by its maximum value.
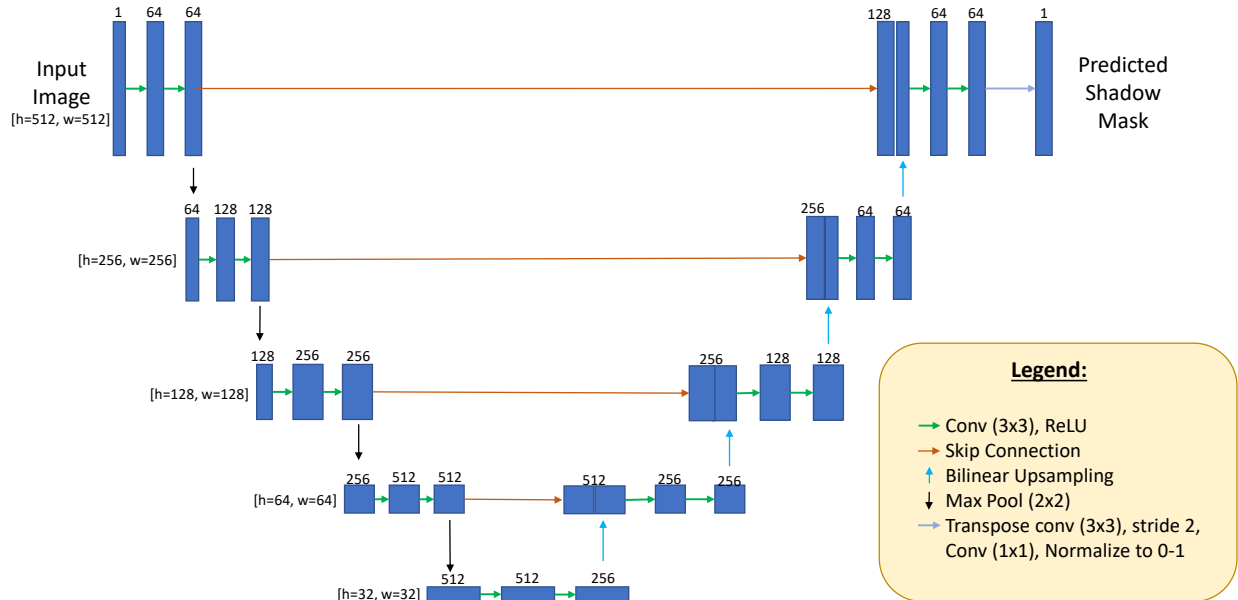

Figure 2: UNet Architecture used in the shadow detector and image processing network.

2.5 Image Augmentation

To ensure that our algorithm was robust and functioned on single-frame images with varying levels of noise and retinal shadows, we implemented online image rotation (-45° to 45°), XY translation (-50% to 50% of image size), image scaling (-50% to +50% of image size) and random horizontal flip during our training.

2.6 Speckle Noise Modelling

We needed to add noise to ART images to simulate the speckle noise found in single-frame images. The goal was to train the image processor to remove this artificial noise and in turn enable the image processor network to remove genuine speckle noise found in single-frame OCT images. We found through experiments that speckle noise was able to be modelled as Gaussian noise ($\mu = 0$, $\sigma = 1$) multiplied with a uniform distribution (range 0.02 to 0.5). In addition, including a large range for the Gaussian model helped the algorithm to perform robustly on single-frame images, which had varying levels of noise. These numbers were experimentally obtained by qualitative assessment of test images generated from single-frame images. A new noise sample was created for every ART image during training to encourage robust training of the image processor network.

2.7 Feature Extraction

As using mean squared error (MSE) directly on processed images as a loss function was found to produce blurring effects on processed images, we instead applied MSE onto extracted feature representations of noisy images and their corresponding ART B-scans. To extract comprehensive feature representations of each image, we required capable, pre-trained networks [30, 31, 32] for feature extraction. Our framework consisted of three pre-trained and frozen feature extraction networks [30, 31, 32] (Fig. 1). These frozen networks were used to extract features from input images and will henceforth be referred to as perceptual networks. We used the three classification networks trained on ImageNet as our perceptual networks, namely, EfficientNet-B4 [30], WideResnet101_2 [31], and Resnext101_32x8d [32]. We leveraged the "ensemble effect" whereby gradients were averaged from three different highly accurate perceptual networks to produce a more accurate backpropagation update [36] for the image processor network. High-level feature representations were extracted from the final convolutional layer of EfficientNet-B4, while both intermediate and high level feature representations were extracted from residual block 2, 4, 6, and 8 for WideResnet101_2 and Resnext101_32x8d for computation of content and style losses. Each feature representation of a processed image was compared to (using MSE) the feature representation of its corresponding ART image. These comparisons were then included in a custom loss function that we describe in the next section.

2.8 Loss Function for Training the Shadow Detector and Image Processor networks

We successfully trained the image processor network and simultaneously removed speckle noise and retinal shadows using a combination of different loss functions. These losses were:

2.8.1 Shadow Loss

The shadow loss was defined to ensure that all shadows were effectively removed so that they become indistinguishable from surrounding tissues. When a given image X had been processed, it was passed to the shadow detector network to produce a predicted shadow mask, M_{X} (with maximum pixel intensities equal to 1). All pixel intensities were then summed, and then normalized by dividing this sum with the sum of the pixels within the ground truth manually segmented mask. This normalized sum was defined as the shadow loss.

2.8.2 Content Loss

We used the content loss to ensure that critical information within all non-shadowed regions of a given image was retained after shadow correction. To compute content loss, we compared intermediate and high-level feature representations between a given processed image D and its corresponding ART image C. Note that the content loss had been used in Style Transfer [37] with great success at maintaining fine details and edges. We first applied the manually segmented shadow mask to the processed image, D_{masked} and its corresponding ART image, C_{masked}. This blocked out pixels in the retinal shadows so that the content loss would not be affected by any shadow removed. Next, we extracted feature representations from all perceptual networks for the processed image, and its corresponding ART image. The content loss was then defined as:

$$L_{\mathrm{content}}(D_{\mathrm{masked}}, C_{\mathrm{masked}}) = \sum_{i=2,4,6,8} \frac{1}{C_i H_i W_i} |P_i(D_{\mathrm{masked}}) - P_i(C_{\mathrm{masked}})|^2 \tag{1}$$

where P_i is a feature representation of the ith selected residual block of a perceptual network. Note that i=2,4,6,8 for the WideResnet101_2 and Resnext101_32x8d perceptual network and i refers to the last convolutional layer of the EfficientNet-B4 perceptual network.

2.8.3 Style Loss

To ensure that image textures remained the same in non-shadowed regions after shadow correction, we computed the style loss for masked processed image D_{masked} and its corresponding ART image, C_{masked}. To compute the style loss, we first calculated the Gram matrix of an image to find a representation of its style. Then, the style loss for each image pair $(D_{\text{masked}}, C_{\text{masked}})$ was defined to be the Euclidian norm between its Gram matrices:

$$L_{\text{style}}(D_{\text{masked}}, C_{\text{masked}}) = \sum_i |G_i(D_{\text{masked}}) - G_i(C_{\text{masked}})|^2 \tag{2}$$

where $G_i(x)$ is a $C_i \times C_i$ matrix defined as:

$$G_i(x) = P_i(x)_{C_i W_i H_i} \times P_i(x)_{H_i W_i C_i} \tag{3}$$

2.8.4 Total Loss

The total loss was computed as a weighted sum of the content, style, and shadow losses to ensure all losses were of the same order of magnitude. The shadow loss was set as the reference (as already being normalized) and no weight was assigned. The total loss was defined as:

$$L_{\text{total}} = \sum_j (w_j L_{\text{content,j}} + k_j L_{\text{style,j}}) + L_{\text{shadow}} \tag{4}$$

where w_j and k_j are the weights to be derived experimentally; j summed over the type of perceptual network, i.e. EfficientNet-B4, WideResnet101_2, and Resnext101_32x8d. To obtain the weight values, we first trained the image processor network without style loss (k=0) to determine all w. We then introduced all style losses and normalized them so that their magnitudes were on the same scale as the content losses. Through this process the weights $w_{\text{EfficientNet-B4}}$, $w_{\text{WideResnet101\_2}}$, $w_{\text{Resnext101\_32x8d}}$, $k_{\text{EfficientNet-B4}}$, $k_{\text{WideResnet101\_2}}$, $k_{\text{Resnext101\_32x8d}}$, were given the following values: $2.86, 4, 6.67, 6.67 \times 10^{-5}, 1.8 \times 10^{-5}, 2.1 \times 10^{-5}$ respectively.

2.9 Training Parameters

We used 2328 ART B-scans during training and 300 single-frame B-scans with its corresponding ART B-scan during testing. These ART images were used as the ground truth images for the content and style losses, but not for the shadow loss such images still contained shadows. Each B-scan was added with a randomly generated Gaussian noise model (created according to section F). During training, the image processor network learnt how to remove the randomly generated Gaussian noise using content and style losses, and it simultaneously learnt to remove retinal blood vessel shadows through the use of the shadow loss.

All training and testing were performed on five Nvidia GTX 1080 Ti cards with CUDA V10.1.105, paired with Nvidia driver V436.48 and cuDNN v7.6.5. Using these hardware specifications, each image took an average of 10.3 ms to be processed. The total training time was 4 days using the Adam optimizer at a learning rate of 1×10^{-5} and a batch size of 6. A learning rate decay was implemented to halve learning rates every 10 epochs. We stopped the training when no improvements in output images could be observed.

2.10 Noise and Retinal Shadow Removal Metrics

We used average gradient magnitudes (AGM), the peak-signal-to-noise-ratio (PSNR), the contrast-to-noise-ratio (CNR) and the mean-structural-similarity-index (SSIM) to quantify the noise removal capabilities of our proposed algorithm. All noise removal metrics were normalized with respect to their corresponding ART image for easy comparison. All noise removal metrics were extracted from regions of interest (ROIs) that

did not contain retinal shadows to prevent shadow removal from affecting noise removal metrics. We also used the intra-layer contrast (ILC) and the layer-wise pixel intensity (LPI) profiles to assess the proposed algorithm's effectiveness in removing shadows. During testing, we obtained all metrics on noisy, non-averaged single-frame B-scans. All ART B-scans were then aligned to their corresponding single-frame B-scan using rigid translation/rotation transformations using 3D software (Amira, version 5.6; FEI) before noise and shadow removal metrics were extracted.

2.10.1 Noise Removal Quantitative Assessment

The AGM was used to quantify the sharpness of output images. We used the AGM implementation found in the Python package Numpy [38], defined as:

$$AGM = \frac{1}{H \times W} \sum_x \sum_y \frac{G(x, y)}{\sqrt{2}} \tag{5}$$

where $G(x, y)$, H and W were the gradient vector, height and width of the B-scan respectively.

The PSNR (expressed in dB) was used to quantify the noise levels in an image relative to its true signal strength. We used the scikit-image [39] implementation of PSNR defined as:

$$PSNR = -10 \times \log_{10} \frac{|f_0 - \tilde{f}|^2}{|f_0|^2} \tag{6}$$

where f_0 was the pixel-intensity values of the registered ART B-scan, and $\tilde{f}$ was the pixel-intensity of the processed B-scan. A higher PSNR suggested that the processed images contained less noise and were of higher quality than images with lower PSNR.

The CNR provided an indication of how visible a retinal tissue layer is. It was defined as:

$$CNR = \frac{|\mu_r - \mu_b|}{\sqrt{0.5 \times (\sigma_r^2 + \sigma_b^2)}} \tag{7}$$

where μ_r, μ_b, σ_r^2 and σ_b^2 represented the means and variances of pixel intensities for a selected ROI within the tissue 'i' and a randomly chosen ROI from the background, respectively. Each ROI was chosen as a 20×384 pixels region at the top of the selected B-scan. A higher CNR suggested superior visibility of tissue 'i' within a given B-scan. We computed the CNR for the RNFL and compared them between single-frame, processed, and ART B-scans. We computed the CNR as a mean of 25 randomly selected ROIs per tissue for each given B-scan, each of size 8×8 pixels. All ROIs were manually chosen in each tissue by an expert observer (HC) using a custom Python script using the OpenCV [40] package.

The SSIM was computed to quantify changes in tissue structures (i.e., edges) between a given single-frame/processed image with its corresponding ART image as a reference. The SSIM was based on the computation of three terms: luminance, contrast, and structure, respectively. We used the implementation of the SSIM in the scikit-image package in Python defined as:

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

where $\sigma_x, \mu_y, \sigma_x, \sigma_y,$ and σ_{xy} were the local means, standard deviations, and cross-covariance for images x and y respectively.

2.10.2 Shadow Removal Quantitative Assessment

We computed the ILC to assess the performance of the proposed algorithm in removing shadows. The ILC was defined as:

$$ILC = |\frac{I_1 - I_2}{I_1 + I_2}|$$
(9)

where I_1 was the mean pixel intensity from five manually selected ROIs (size 5 ×5 pixels) that are shadow free in a given retinal layer, and I_2 was the corresponding value from five neighboring shadowed regions of the same tissue layer. The ILC ranged between 0 and 1, where values close to 0 indicated the absence of retinal shadows and values close to 1 indicated strongly visible blood vessel shadows.

We computed the intralayer contrast for multiple tissue layers of the ONH region, namely the RNFL, the photoreceptor layer (PR) and the retinal pigment epithelium (RPE) – before and after application of the proposed algorithm. Results for all metrics were recorded in the form of mean ± standard deviation.

3   Results

When trained on 2328 ART B-scans with online data augmentation, our deep learning framework was able to successfully remove noise and retinal shadows from unseen single-frame B-scans (Fig. 3). An independent test set of 300 single-frame B-scans was used to evaluate the noise and retinal shadow removal performance of the proposed deep learning framework qualitatively and quantitatively. The mean PSNR, the CNR and the SSIM increased with respect the input single-frame B-scans from 18.5 ± 0.46 dB to 20.5 ± 0.38 dB, 3.66 ± 0.92 to 8.97 ± 2.60 and 0.177 ± 0.004 to 0.45 ± 0.09 respectively. The ILC for the RNFL, the PR, and the RPE decreased from 0.362 ± 0.133 to 0.142 ± 0.102, 0.449 ± 0.116 to 0.090 ± 0.077, 0.381 ± 0.100 to 0.059 ± 0.045 respectively.
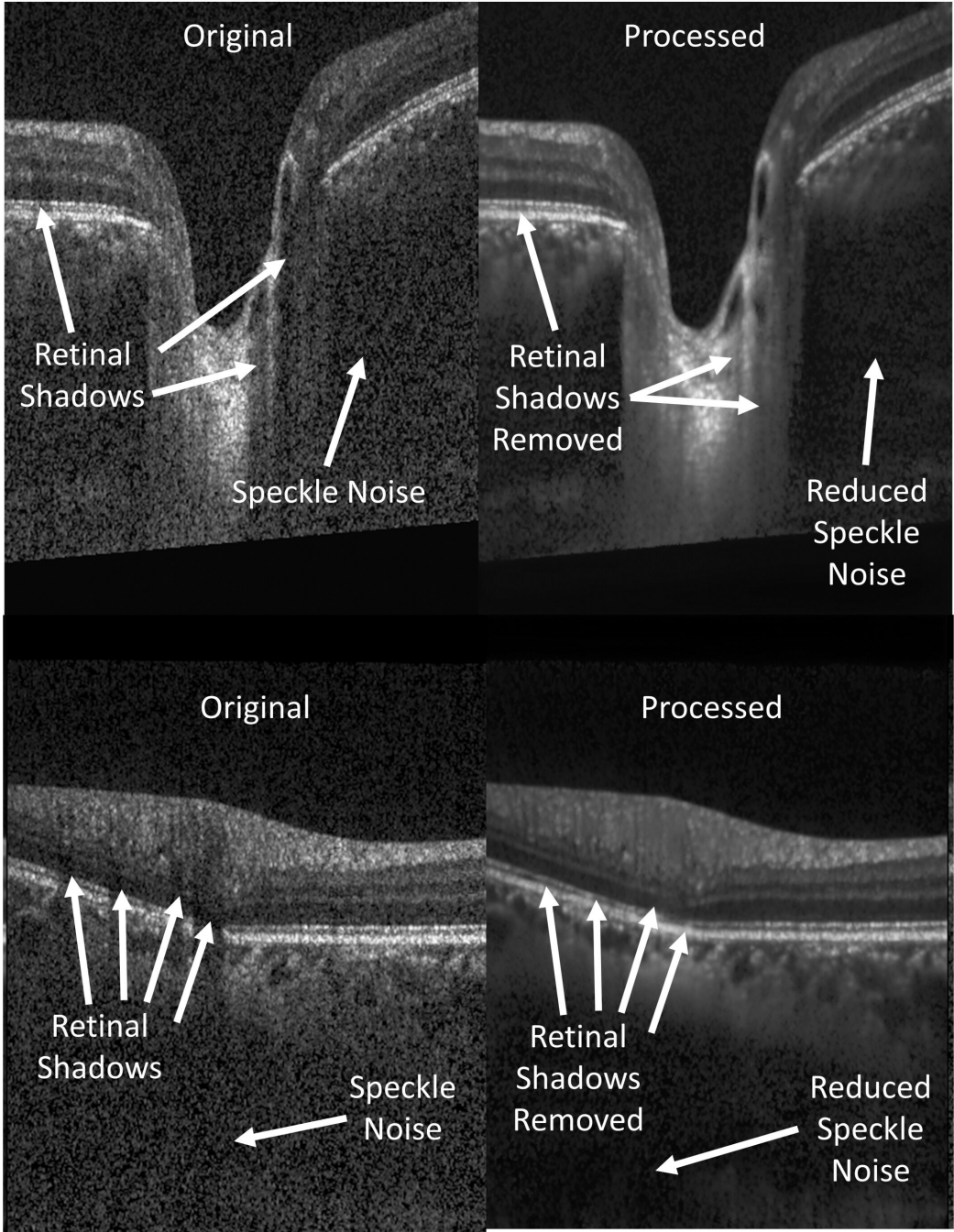
Figure 3: Samples of typical B-scans before (left) and after (right) being processed by our algorithm.

3.1   Denoising Performance – Qualitative Analysis

The proposed algorithm produced images without common artifacts of images processed by current deep learning frameworks, including blurring and checkerboard patterns. Single-frame B-scans processed by our algorithm looked qualitatively sharper (Fig. 4a, Fig. 4b) and visually closer to ART B-scans than images produced by the current state-of-the-art algorithm [19]. Qualitative analysis of preliminary results (Fig. 4b) suggested that our algorithm performed well on low ART mean (averaged over 20 frames) OCT images from an elderly patient with primary open angle glaucoma (POAG). Retinal shadows were effectively removed, improving visible information within retinal shadows. Overall sharpness was retained and visibility of all the ONH tissues was enhanced after being processed by the proposed algorithm.
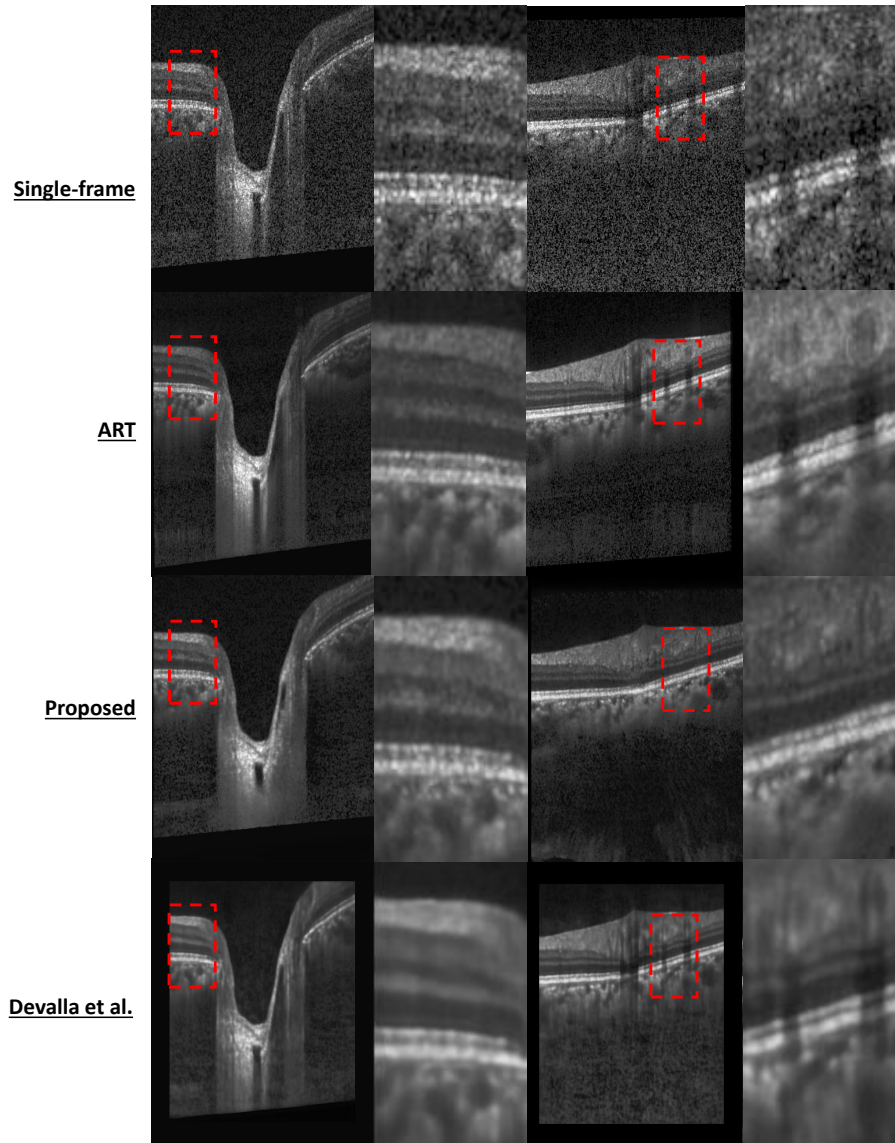
Figure 4a: Qualitative analysis of the proposed noise removal and shadow removal algorithm. Blurring can be seen in the current state-of-the-art (fourth row) [19].
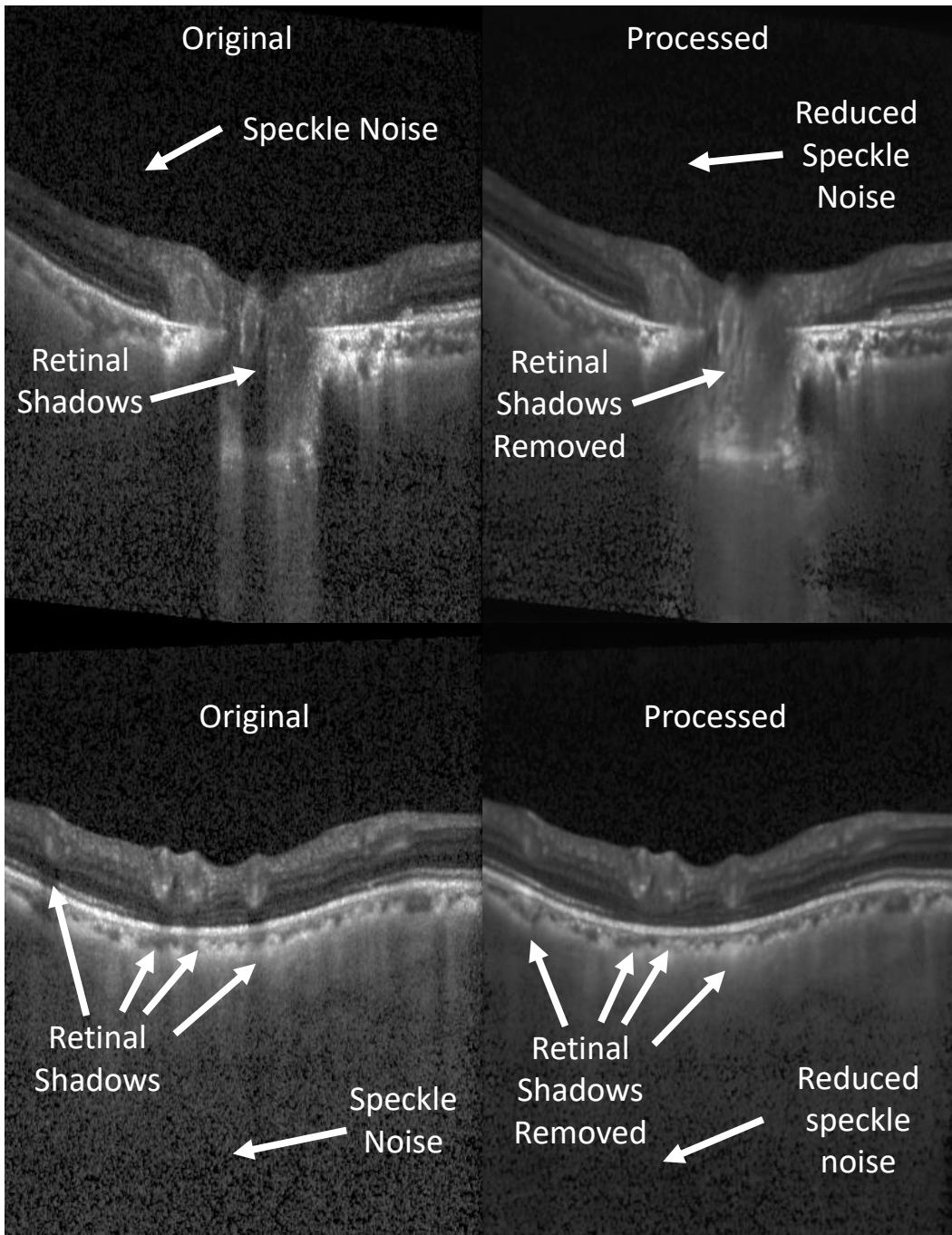
Figure 4b: Original (left) and processed (right) OCT B-scans of an elderly patient (77 years old) with primary open angle glaucoma (POAG).

3.2 Denoising Performance – Quantitative Analysis

In Fig. 5, we compared the AGM, the CNR, the PSNR, and the SSIM of images processed by the proposed algorithm with single-frame images. When evaluated on 300 single-frame B-scans, we observed that the proposed algorithm consistently produced images that are both qualitatively and quantitatively sharper than images produced by the current state-of-the-art (Fig. 4a). On average, each image produced by the proposed algorithm were 154%, 187%, 11.1% better than single-frame B-scans in terms of the CNR, the SSIM, the PSNR, respectively. The AGM was also 57.2% higher than images produced by the current state-of-the-art denoising algorithm [19].
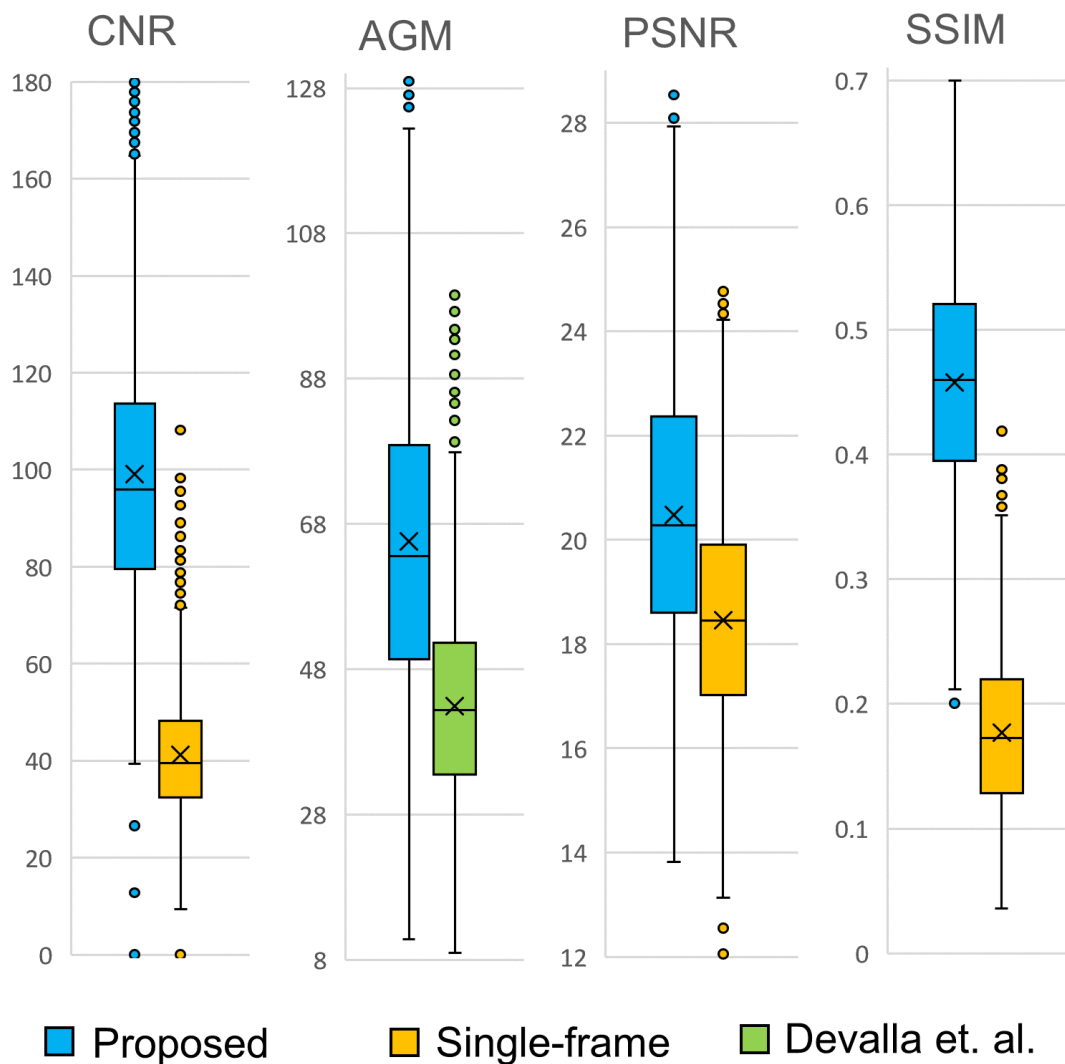

Figure 5: (from left) CNR, SSIM, and PSNR values improved relative to single-frame images. AGM of the proposed algorithm compared to that of the current state-of-the-art algorithm by Devalla et. al [19].

3.3 Deshadowing Performance – Quantitative Analysis

Our proposed algorithm produced images that had improved visibility within retinal shadows. The ILC for the RNFL, the PR, and the RPE improved by $60.0 \pm 29.3\%$, $79.0 \pm 19.4\%$ and $83.4 \pm 15.4\%$ respectively. On average, the ILC improved by $72.9 \pm 25.2\%$ (Fig. 6). The LPI profiles were also significantly flattened in the RNFL, the PR and the RPE layers (Fig. 7).
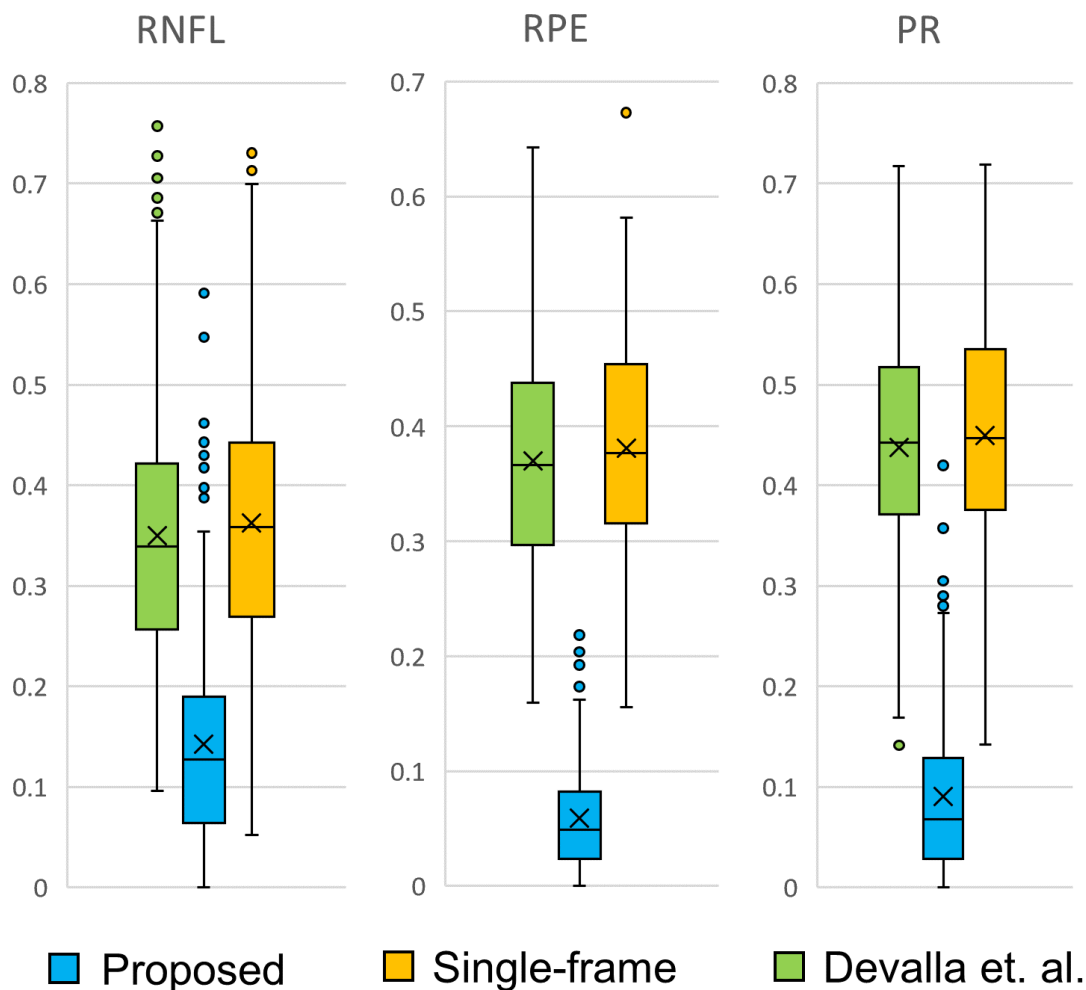

Figure 6: ILC comparison between images denoised by state-of-the-art [19] vs proposed algorithm.
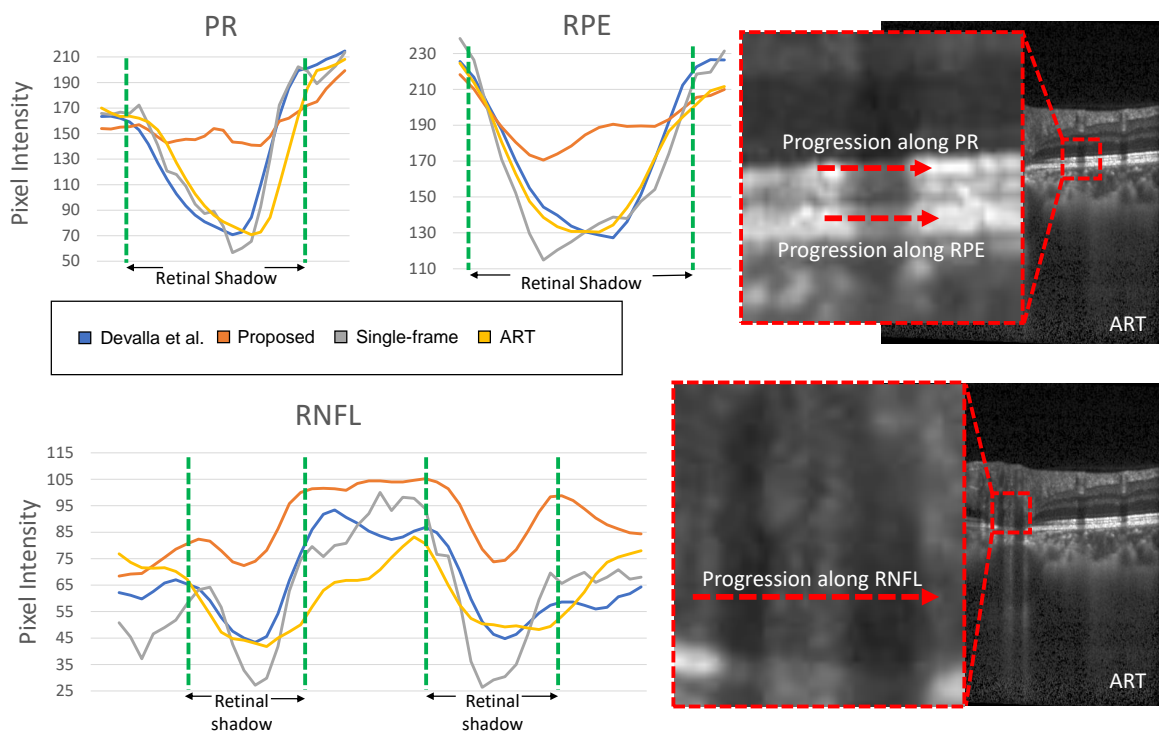
Figure 7: LPI profiles of output images along RNFL, PR and RPE layers were significantly flatter than ART B-scans, or B-scans denoised by the current state-of-the-art [19].

4 Discussion

In this study we present a custom deep learning approach that can remove noise and retinal shadows simultaneously from single-frame OCT B-scans of the ONH. All noise removal performance metrics such as the PSNR, the CNR, and the SSIM values consistently showed significant improvements compared to single-frame images. Thus, we may be able to offer a robust deep learning framework to obtain high quality OCT B-scans with reduced scanning duration and minimized patient discomfort.

B-scans processed by our algorithm were qualitatively similar to their corresponding ART B-scans, with the added benefit of improved visibility within retinal shadows (Fig. 3, Fig. 4a, Fig. 4b). Processed B-scans were also devoid of motion and registration artifacts that commonly plagued ART averaging. We accomplish this by learning from features extracted from ART averaged images. Our training technique allowed our algorithm to improve the PSNR while simultaneously avoiding the motion and registration artifacts as single-frame images are far less likely to have those due to shorter acquisition times. We postulate that the algorithm is unlikely to be able to replicate these artifacts even while learning from features extracted from ART images as these artifacts are random and inconsistent in nature, and thus is unlikely to be learnt by neural networks. Rather, it is likely that these artifacts are treated as "noise" during training and improve robustness instead of adversely affecting our results.

The SSIM and the CNR were significantly improved with respect to single-frame images by 154% and 187%, respectively. The mean AGM was also 57.2% higher than the current state-of-the-art [19], providing clinicians a markedly sharper image. Image sharpness is critical given that many pathologies require sharp layer boundaries for accurate retinal layer thickness measurements. One example would be quantifying macular edema, which require measurements of retinal thickness in response to therapy [41]. Given the significance of retinal layers and connective tissues in the prognosis and diagnosis of ocular pathologies such as glaucoma and age-related macular degeneration, enhanced visibility would improve automated algorithms downstream of the post processing pipeline, namely alignment, registration, segmentation, diagnosis and, ultimately, prognosis.

The proposed algorithm did not require any further segmentation, delineation, or identification of shadows by the user. Similar to our previous work [24], the ILC mean and standard deviation decreased with the depth of the retinal layer of interest (Fig. 6), suggesting that performance of the proposed algorithm was consistently better in deeper layers. The proposed algorithm substantially recovered the visibility of the anterior lamina cribrosa (LC) boundary and anterior LC insertion, which may result in a more confident prediction of early glaucoma [42]. Moreover, the main load bearing tissues of the eye in the ONH region, such as the LC and adjacent peripapillary sclera, could be monitored for pre-disease biomechanical and morphological changes. Changes in these tissues have been previously identified as risk factors for glaucoma [43]. Measurements of the anatomy of such tissues could be more robust and substantially improved after application of the proposed algorithm.

In this study, several limitations warrant further discussion. While we did not find any evidence of pathology being obscured or introduced into output images, it is extremely important to validate this in pathological cases. However, we would need to image the exact same tissue region with and without the presence of blood flow (to remove retinal blood vessel shadows). Such experiments would be extremely complex to carry out in vivo, especially in humans, even if blood vessels were to be flushed with saline during experiments. Such validations may be required for full clinical acceptance of this methodology. Furthermore, it would be critical to also confirm that the proposed algorithm would not interfere with another AI algorithm (especially those aimed at diagnosis and prognosis). Nevertheless, it is possible that the proposed algorithm might improve diagnosis and prognosis algorithms by improving the quality of the input data. We aim to test this hypothesis in the future.

Furthermore, although the proposed algorithm functioned well on single-frame images from healthy in-

dividuals, more work is required to ensure that it can reproduce similar performance on B-scans of eyes with pathophysiological conditions such as glaucoma. While our results (Fig. 4b)) indicated that our algorithm may not be sensitive to the age or POAG status of the patient, more work is required to show that there is no significant effect of age and other disease factors on the efficacy of our algorithm. This is especially critical for deep learning approaches, which respond unpredictably to input data that is different from images used during training. As this algorithm was trained on single-frame images from a Spectralis OCT device, it is unknown if it can maintain this performance on OCT images from other devices. Each scenario stated above may require a separate training set. Our future studies will therefore focus on validating the performance of the proposed algorithm across devices and between healthy and pathological eyes.

5 Conclusion

The proposed algorithm successfully removed both noise and retinal shadows from single-frame B-scans. The algorithm also drastically reduced the time needed (3.5 minutes to 10.6s) for medical professionals and patients to obtain ART quality B-scans (75 times signal averaged). This could have significant economic benefits for hospitals by allowing less money reducing expenditure on expensive, high quality OCT machines. Patients would also benefit by a reduction in the time needed to remain in a fixated position during OCT image acquisition. Automated segmentation and diagnosis algorithms could also benefit clinical diagnostics by providing increased structural clarity, improved layer continuity and enhanced visibility both within shadows and retinal layers. The combination of both noise removal and retinal shadow removal algorithms in a single step will improve latency and be a step toward the goal of real-time OCT image processing.

Disclosures

Haris Cheong: None, Sripad Krishna Devalla: None, Thanadet Chuangsuwanich: None, Tin A. Tun: None, Xiaofei Wang: None, Tin Aung: None, Leopold Schmetterer: None, Martin L. Buist: None, Craig Boote: None, Alexandre H. Thiéry: Abyss Processing (Co-Founder), Michaël J. A. Girard: Abyss Processing (Co-Founder)

References

[1] Carmen A Puliafito, Michael R Hee, Charles P Lin, Elias Reichel, Joel S Schuman, Jay S Duker, Joseph A Izatt, Eric A Swanson, and James G Fujimoto. Imaging of macular diseases with optical coherence tomography. *Ophthalmology*, 102(2):217–229, 1995.

[2] Alexander Wong, Akshaya Mishra, Kostadinka Bizheva, and David A Clausi. General bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery. *Optics express*, 18(8):8338–8352, 2010.

[3] Maciej Szkulmowski, Iwona Gorczynska, Daniel Szlag, Marcin Sylwestrzak, Andrzej Kowalczyk, and Maciej Wojtkowski. Efficient reduction of speckle noise in optical coherence tomography. *Optics express*, 20(2):1337–1359, 2012.

[4] Mitsuro Sugita, Stefan Zotter, Michael Pircher, Tomoyuki Makihira, Kenichi Saito, Nobuhiro Tomatsu, Makoto Sato, Philipp Roberts, Ursula Schmidt-Erfurth, and Christoph K Hitzenberger. Motion artifact and speckle noise reduction in polarization sensitive optical coherence tomography by retinal tracking. *Biomedical optics express*, 5(1):106–122, 2014.

[5] Jing Wu, Bianca S Gerendas, Sebastian M Waldstein, Georg Langs, Christian Simader, and Ursula Schmidt-Erfurth. Stable registration of pathological 3d-oct scans using retinal vessels. In *Proceedings of the Ophthalmic Medical Image Analysis First International Workshop*, 2014.

[6] Iwona Gorczynska, Justin V Migacz, Robert J Zawadzki, Arlie G Capps, and John S Werner. Comparison of amplitude-decorrelation, speckle-variance and phase-variance oct angiography methods for imaging the human retina and choroid. *Biomedical optics express*, 7(3):911–942, 2016.

[7] Fei Shi, Ning Cai, Yunbo Gu, Dianlin Hu, Yuhui Ma, Yang Chen, and Xinjian Chen. Despecnet: a cnn-based method for speckle reduction in retinal optical coherence tomography images. *Physics in Medicine & Biology*, 64(17):175010, 2019.

[8] Vivek J Srinivasan, Maciej Wojtkowski, Andre J Witkin, Jay S Duker, Tony H Ko, Mariana Carvalho, Joel S Schuman, Andrzej Kowalczyk, and James G Fujimoto. High-definition and 3-dimensional imaging of macular pathologies with high-speed ultrahigh-resolution optical coherence tomography. *Ophthalmology*, 113(11):2054–2065. e3, 2006.

[9] Atsushi Sakamoto, Masanori Hangai, and Nagahisa Yoshimura. Spectral-domain optical coherence tomography with multiple b-scan averaging for enhanced imaging of retinal diseases. *Ophthalmology*, 115(6):1071–1078. e7, 2008.

[10] Giovanni J Ughi, Matilda Larsson, Christophe Dubois, Peter R Sinnaeve, Walter Desmet, Jan D'Hooge, Tom Adriaenssens, and Mark Coosemans. Automatic three-dimensional registration of intravascular optical coherence tomography images. *Journal of biomedical optics*, 17(2):026005, 2012.

[11] Shaozhen Song, Zhihong Huang, and Ruikang K Wang. Tracking mechanical wave propagation within tissue using phase-sensitive optical coherence tomography: motion artifact and its compensation. *Journal of biomedical optics*, 18(12):121505, 2013.

[12] SH Yun, GJ Tearney, JF De Boer, and BE Bouma. Motion artifacts in optical coherence tomography with frequency-domain ranging. *Optics Express*, 12(13):2977–2998, 2004.

[13] Yali Jia, Steven T Bailey, David J Wilson, Ou Tan, Michael L Klein, Christina J Flaxel, Benjamin Potsaid, Jonathan J Liu, Chen D Lu, and Martin F Kraus. Quantitative optical coherence tomography angiography of choroidal neovascularization in age-related macular degeneration. *Ophthalmology*, 121(7):1435–1444, 2014.

[14] Cong Ye, Marco Yu, and Christopher Kai-shun Leung. Impact of segmentation errors and retinal blood vessels on retinal nerve fibre layer measurements using spectral-domain optical coherence tomography. *Acta Ophthalmologica*, 94(3):e211–e219, 2016.

[15] Jehn-Yu Huang, Melike Pekmezci, Nisreen Mesiwala, Andrew Kao, and Shan Lin. Diagnostic power of optic disc morphology, peripapillary retinal nerve fiber layer thickness, and macular inner retinal layer thickness in glaucoma diagnosis with fourier-domain optical coherence tomography. *Journal of glaucoma*, 20(2):87–94, 2011.

[16] Michaël JA Girard, Nicholas G Strouthidis, C Ross Ethier, and Jean Martial Mari. Shadow removal and contrast enhancement in optical coherence tomography images of the human optic nerve head. *Investigative ophthalmology & visual science*, 52(10):7738–7748, 2011.

[17] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in neural information processing systems*, pages 2802–2810, 2016.

[18] Yuhui Ma, Xinjian Chen, Weifang Zhu, Xuena Cheng, Dehui Xiang, and Fei Shi. Speckle noise reduction in optical coherence tomography images based on edge-sensitive cgan. *Biomedical optics express*, 9(11):5129–5146, 2018.

[19] Sripad Krishna Devalla, Giridhar Subramanian, Tan Hung Pham, Xiaofei Wang, Shamira Perera, Tin A. Tun, Tin Aung, Leopold Schmetterer, Alexandre H. Thiéry, and Michaël J. A. Girard. A deep learning approach to denoise optical coherence tomography images of the optic nerve head. *Scientific Reports*, 9(1):14454, 2019.

[20] Neha Gour and Pritee Khanna. Speckle denoising in optical coherence tomography images using residual deep convolutional neural network. *Multimedia Tools and Applications*, pages 1–17, 2019.

[21] Min Xu, Chen Tang, Fugui Hao, Mingming Chen, and Zhenkun Lei. Texture preservation and speckle reduction in poor optical coherence tomography using the convolutional neural network. *Medical Image Analysis*, page 101727, 2020.

[22] Zailiang Chen, Ziyang Zeng, Hailan Shen, Xianxian Zheng, Peishan Dai, and Pingbo Ouyang. Dn-gan: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images. *Biomedical Signal Processing and Control*, 55:101632, 2020.

[23] Kiran Kumar Vupparaboina, Kunal K Dansingani, Abhilash Goud, Mohammed Abdul Rasheed, Fayez Jawed, Soumya Jana, Ashutosh Richhariya, K Bailey Freund, and Jay Chhablani. Quantitative shadow compensated optical coherence tomography of choroidal vasculature. *Scientific reports*, 8(1):1–9, 2018.

[24] Haris Cheong, Sripad Krishna Devalla, Tan Hung Pham, Liang Zhang, Tin Aung Tun, Xiaofei Wang, Shamira Perera, Leopold Schmetterer, Tin Aung, and Craig Boote. Deshadowgan: a deep learning approach to remove shadows from optical coherence tomography images. *Translational Vision Science & Technology*, 9(2):23–23, 2020.

[25] Qinqin Zhang, Fang Zheng, Elie H Motulsky, Giovanni Gregori, Zhongdi Chu, Chieh-Li Chen, Chunxia Li, Luis De Sisternes, Mary Durbin, and Philip J Rosenfeld. A novel strategy for quantifying choriocapillaris flow voids using swept-source oct angiography. *Investigative ophthalmology & visual science*, 59(1):203–211, 2018.

[26] Ian Wong, Hideki Koizumi, and Wico Lai. Enhanced depth imaging optical coherence tomography. *Ophthalmic surgery, lasers & imaging : the official journal of the International Society for Imaging in the Eye*, 42 Suppl:S75–84, 2011.

[27] R Daniel Ferguson, Daniel X Hammer, Lelia Adelina Paunescu, Siobahn Beaton, and Joel S Schuman. Tracking optical coherence tomography. *Optics letters*, 29(18):2139–2141, 2004.

[28] Daniel X Hammer, R Daniel Ferguson, Nicusor V Iftimia, Teoman Ustun, Gadi Wollstein, Hiroshi Ishikawa, Michelle L Gabriele, William D Dilworth, Larry Kagemann, and Joel S Schuman. Advanced scanning methods with tracking optical coherence tomography. *Optics express*, 13(20):7937–7947, 2005.

[29] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682, 2012.

[30] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

[31] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

[32] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2016.

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[34] Shie Mannor, Dori Peleg, and Reuven Rubinstein. The cross entropy method for classification. In *Proceedings of the 22nd international conference on Machine learning*, pages 561–568, 2005.

[35] Matthew D Zeiler, M Ranzato, Rajat Monga, Min Mao, Kun Yang, Quoc Viet Le, Patrick Nguyen, Alan Senior, Vincent Vanhoucke, and Jeffrey Dean. On rectified linear units for speech processing. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3517–3521. IEEE, 2013.

[36] Sean Tao. Deep neural network ensembles. In *International Conference on Machine Learning, Optimization, and Data Science*, pages 1–12. Springer, 2019.

[37] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

[38] Travis E Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

[39] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.

[40] Gary Bradski and Adrian Kaehler. Opencv. *Dr. Dobb's journal of software tools*, 3, 2000.

[41] Glenn J Jaffe and Joseph Caprioli. Optical coherence tomography to detect and manage retinal disease and glaucoma. *American journal of ophthalmology*, 137(1):156–169, 2004.

[42] Kyoung Min Lee, Tae-Woo Kim, Robert N Weinreb, Eun Ji Lee, Michaël JA Girard, and Jean Martial Mari. Anterior lamina cribrosa insertion in primary open-angle glaucoma patients and healthy subjects. *PLoS One*, 9(12):e114935, 2014.

[43] Hongli Yang, J Crawford Downs, Christopher Girkin, Lisandro Sakata, Anthony Bellezza, Hilary Thompson, and Claude F Burgoyne. 3-d histomorphometry of the normal and early glaucomatous monkey optic nerve head: lamina cribrosa and peripapillary scleral position and thickness. *Investigative ophthalmology & visual science*, 48(10):4597–4607, 2007.