Accurate tissue interface segmentation via adversarial pre-segmentation of anterior segment OCT images

JIAHONG OUYANG,[1,4,5] TEJAS SUDHARSHAN MATHAI,[1,4,6] KIRA LATHROP,[2,3,7] AND JOHN GALEOTTI[1,2,8]

[1] *The Robotics Institute, Carnegie Mellon University, PA 15213, USA*
[2] *Department of Bioengineering, University of Pittsburgh, PA 15213, USA*
[3] *Department of Ophthalmology, University of Pittsburgh, PA 15213, USA*
[4] *Equal contribution*
[5] *jiahongo@andrew.cmu.edu*
[6] *tmathai@andrew.cmu.edu*
[7] *lathropkl@upmc.edu*
[8] *jgaleotti@cmu.edu*

Abstract: Optical Coherence Tomography (OCT) is an imaging modality that has been widely adopted for visualizing corneal, retinal and limbal tissue structure with micron resolution. It can be used to diagnose pathological conditions of the eye, and for developing pre-operative surgical plans. In contrast to the posterior retina, imaging the anterior tissue structures, such as the limbus and cornea, results in B-scans that exhibit increased speckle noise patterns and imaging artifacts. These artifacts, such as shadowing and specularity, pose a challenge during the analysis of the acquired volumes as they substantially obfuscate the location of tissue interfaces. To deal with the artifacts and speckle noise patterns and accurately segment the shallowest tissue interface, we propose a cascaded neural network framework, which comprises of a conditional Generative Adversarial Network (cGAN) and a Tissue Interface Segmentation Network (TISN). The cGAN pre-segments OCT B-scans by removing undesired specular artifacts and speckle noise patterns just above the shallowest tissue interface, and the TISN combines the original OCT image with the pre-segmentation to segment the shallowest interface. We show the applicability of the cascaded framework to corneal datasets, demonstrate that it precisely segments the shallowest corneal interface, and also show its generalization capacity to limbal datasets. We also propose a hybrid framework, wherein the cGAN pre-segmentation is passed to a traditional image analysis-based segmentation algorithm, and describe the improved segmentation performance. To the best of our knowledge, this is the first approach to remove severe specular artifacts and speckle noise patterns (prior to the shallowest interface) that affects the interpretation of anterior segment OCT datasets, thereby resulting in the accurate segmentation of the shallowest tissue interface. To the best of our knowledge, this is the first work to show the potential of incorporating a cGAN into larger deep learning frameworks for improved corneal and limbal OCT image segmentation. Our cGAN design directly improves the visualization of corneal and limbal OCT images from OCT scanners, and improves the performance of current OCT segmentation algorithms.

1. Introduction

Optical coherence tomography (OCT) is a non-invasive and non-contact imaging technique that has been widely adopted for imaging sub-surface tissue structures with micrometer depth resolution in clinical ophthalmology [1,2]. OCT is a popular method to visualize structures in the eye, especially those in the retina [1], cornea [3], and limbus [4]. Specific to the anterior segment of the eye, OCT has been clinically used to characterize the changes that occur during the progression

of Keratoconus [5,6], diagnose benign and malignant conjunctival and corneal pathologies, such as Ocular Surface Squamous Neoplasia [4,6], and monitor potential complications for many anterior segment surgical procedures, such as Deep Anterior Lamellar Keratoplasty (DALK) [7] and Descemet Membrane Endothelial Keratoplasty (DMEK) [6]. Furthermore, OCT has been used to image the limbus [4,8,9], and enabled the analysis of the Palisades of Vogt (POV) [10].

In all these applications, accurate estimation of the corneal or limbal tissue boundaries is required to determine a quantitative parameter for diagnosis or treatment. For example, in [5], the corneal tissue interfaces were identified to estimate corneal biometric parameters. In [10], the shallowest limbal interface was first identified, and then the tissue structure visualized in the image was "flattened" [10–12] to enable the measurement of the palisade density. However, precise estimation of the corneal and limbal tissue interface location is challenging in anterior segment OCT imaging. The low signal-to-noise ratio (SNR), increased speckle noise patterns, and predominant specular artifacts pose barriers towards automatic delineation of the tissue interfaces (see Fig. 1). Furthermore, datasets are typically acquired in a clinical setting using different OCT scanners (including custom-built OCT scanners for clinical research) from different vendors as shown in Fig. 1. The scan settings of these OCT machines are usually different, thereby resulting in datasets with different image dimensions, SNR, speckle noise patterns, and specular artifacts.
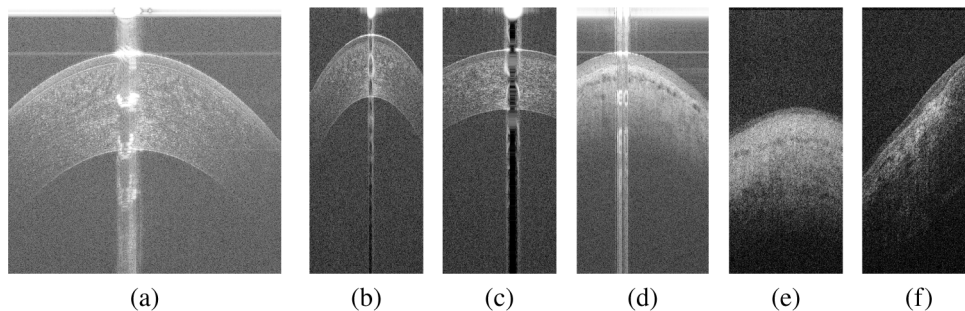

Fig. 1. Original B-scans from (a) a 6×6mm corneal volume acquired by a custom SD-OCT scanner, (b) a 6×6mm corneal volume and (c) a 3×3mm corneal volume acquired by a UHR-OCT scanner, (d) a 4×4mm limbal volume acquired by a hand-held Leica SD-OCT scanner, and (e)-(f) 4×4mm limbal volumes acquired by a UHR-OCT scanner. Specular artifacts in (a)-(d) and poor visibility in (e)-(f) affect the precise delineation of the tissue interfaces.

Speckle noise patterns and specular artifacts are major factors that influence the correct interpretation of anterior segment OCT images. To mitigate these degradations, there are many hardware- and software-based approaches that process each B-scan before they are analyzed in a segmentation pipeline. Hardware-based speckle noise reduction techniques [13–15] rely on the acquisition of multiple tomograms with decorrelated speckle patterns, such that they can be averaged to obtain images with lower speckle contrast. These techniques usually require modification of the OCT system's optical configuration and/or its scanning protocols. Software-based methods include wavelet transformations [16–21], local averaging and median filtering [22,23], percentile and bilateral filtering [12], regularization [24], local Bayesian estimation [25], and diffusion filtering [26]. Efforts were also made to remove artifacts by using the reference spectrum [11,27], and piezoelectric fiber stretchers [28] in the Fourier domain. However, these methods only work when a fixed type of artifact is encountered, such as the horizontal artifacts in [11,27], and they do not generalize to datasets where the assumption of the artifact presence is violated [29] as seen in Fig. 2. Furthermore, all the prior work are not robust when the SNR dropoff is substantial, which is typically the case while imaging the limbus; the anatomic

curvature (and thus orientation toward the OCT scanner) changes when moving away from the cornea and towards the limbus, thereby causing a significant decrease in visibility of tissue boundaries as seen in Figs. 1(d) – 1(f). Particularly in our case, datasets were acquired by OCT scanners that imaged the limbal junction; the OCT scanner commenced scanning at the limbus and crossed over to the cornea, thereby incorporating the limbal junction during image acquisition. At the limbus, often only the shallowest interface is visible, and as the scanner crosses the limbal junction to image the cornea, different interfaces are gradually seen, such as the Bowman's Layer etc. In this work, we focus on delineating the shallowest tissue interface in all corneal and limbal datasets.
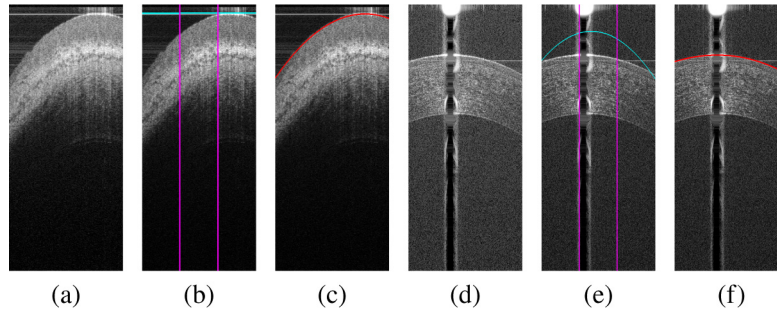

(a)	(b)	(c)	(d)	(e)	(f)

Fig. 2. (a),(d) Original B-scans from a 4×4mm limbal dataset acquired using a hand-held SD-OCT scanner and from a 3×3mm corneal dataset acquired using a UHR-OCT scanner respectively. As proposed in previous algorithms [5,7,11,64–71], vertical lines (magenta) denote the division of the image into three regions in order to deal with specular artifacts. (b),(e) Segmentation of the shallowest interface (cyan contour) by these algorithms failed due to presence of specular artifacts in different regions in the image. (c),(f) Segmentation result (red curve) from the proposed cascaded framework that accurately determined the location of shallowest tissue interface.

Towards the goal of mitigating these image degrading factors, a recent learning-based method featuring a conditional Generative Adversarial Network (GAN) [30,31] was proposed to remove the speckle noise patterns in retinal OCT images [32,33]. It also generalized to datasets acquired from multiple OCT scanners. Although qualitatively good results were obtained, the central premise in their approach was based on limited (little to none) eye motion between frames during imaging. The ground truth data was generated using a compounding technique; the same tissue area was imaged multiple times, and individual volumes were registered yielding averaged B-scans for training, which corresponded to the gold standard despeckled images. However, in our case, this methodology to generate ground truth data for training is not feasible as corneal datasets exhibit large motion when acquired in-vivo, which makes registration and compounding challenging. In addition, existing research databases, from which corneal datasets can be extracted for use in algorithmic development, rarely contain multiple scans of the same tissue area for compounding. Moreover, the authors in [33] opined that it was difficult to judge the efficacy of a despeckling algorithm using existing metrics, such as SNR or Contrast-to-Noise Ratio (CNR), as any one metric is not a good determining factor of the quality of the denoised image. They suggested that an alternate way to analyze the utility of a despeckling method was to estimate the improvement in segmentation accuracy following denoising.

To deal with these challenging scenarios, it is desirable for a tissue-interface segmentation algorithm to possess the following characteristics: 1) Robustness in the presence of speckle noise and artifacts, 2) Generalization capacity across datasets acquired from multiple OCT scanners with different scan settings, and 3) Applicability to different (anterior segment) anatomical regions. Currently, there are a myriad of prior approaches that directly segment corneal and

retinal tissue interfaces. They can be broadly grouped into four categories: 1) Traditional image analysis-based segmentation algorithms, 2) Graph-based segmentation methods, 3) Contour modeling-based segmentation methods, and 4) Machine learning-based (including deep learning-based) segmentation algorithms. Traditional image analysis-based approaches filter the individual B-scans to enhance the contrast of tissue interfaces, and then threshold the image to segment the corneal [12,34,35] and retinal [36–38] interface boundaries. These filters are typically hand-tuned and chosen for the explicit purpose of reducing speckle noise patterns and enhancing edges in the image for easier segmentation. Graph-based methods [39–47] pose the segmentation of the interfaces as an optimization problem, wherein tissue interfaces are detected subject to surface smoothness priors and the distance constraint between interfaces. Other graph-based methods [48,49] involve posing the boundary segmentation problem as a shortest-path finding approach, wherein the shortest path between a source node and sink node is deduced, given costs assigned to the nodes between them. Contour modeling approaches utilize active contours that dynamically change their shape based on shape metrics, such as deviation from a second order polynomial [50,51], edge gradients [52] underlying the contour etc.

Machine learning techniques express the segmentation problem as a classification task; features related to the tissue interfaces to be segmented are extracted, and then classified as belonging to the tissue boundary or background [53–55]. In other cases, learning-based methods are an element of a hybrid system [56,57], wherein the generated output, or the intermediate learned features, improve/assist the performance of traditional/graph-based/contour modeling approaches. Currently, deep neural networks are the state-of-the-art algorithms [57–63] of choice for the segmentation task as they can learn highly discriminative multi-scale features from training data, thereby outperforming all other segmentation approaches. These neural network models are alluring because key algorithm parameters are learned from the training data, which are often manually tuned in other approaches - for example, hand-crafted parameters in traditional image analysis-based [12,34–38] and active contour-based approaches [50–52]. They can also be applied to pathological patients if appropriate datasets were introduced during the training procedure.

However, among all the aforementioned methods, the majority of traditional methods [40,41,43–49] and learning-based methods [57–63,72] are focused on retinal interface segmentation. Corneal interface segmentation algorithms are predominately based on traditional approaches [5,7,11,64–71], with limited learning-based approaches [29,73] being proposed. Similarly, prior work on limbal interface segmentation is limited to a traditional image analysis-based approach [12]. Moreover, most of the prior work is suited towards the task of segmenting tissue interfaces of only one particular type of anatomy, such as retina or cornea, and these prior approaches are not easily generalizable across different types of anatomy. As shown in Fig. 2, most of the traditional approaches were not resilient when the methodology was transferred to our datasets obtained from different OCT scanners, which contained bulk tissue motion, severe specular artifacts and speckle noise patterns.

As seen in Figs. 2(b) and 2(e), previous segmentation approaches would divide (A-scan-wise) the OCT image into three sections, and assume that the location of the central specular artifact was limited to the center of the OCT image (region between the vertical magenta lines) [5,7,11,64–71]. But as seen in Fig. 2, this assumption can be violated when the central artifacts are located in different image regions [29]. In such cases, prior approaches failed to accurately segment the tissue interface as shown in Figs. 2(b) and 2(e). From our experiments, we postulated that most traditional algorithms are confounded by the presence of these strong specular artifacts and speckle noise patterns. Yet, once the shallowest interface is identified, these traditional approaches were able to delineate other interfaces, such as Bowman's Layer, Endothelium etc.

Furthermore, there were two independent and concurrently published deep learning-based corneal interface segmentation approaches [29,73]. One of these approaches [73] acquired data

from a single OCT scanner, and focused only on the region centered around the corneal apex in these OCT sequences as the drop in SNR was greater when moving away from this region. The other approach is our recent publication [29], where we utilized the entire OCT sequence from multiple scanners containing strong specular artifacts and low SNR regions, and successfully segmented three tissue interfaces. Yet, our previously proposed approach did not readily provide intermediate outputs, wherein the specular artifacts and speckle noise patterns were ameliorated, which could be used as input to the traditional approaches [5,7,11,12,64–71] for segmentation.

While we had hoped that the current state-of-the-art algorithms (traditional or deep learning) could directly denoise/segment these challenging corneal and limbal datasets with satisfactory results, it was not the case, nor has anyone else published anything on this task to the best of our knowledge. In this paper, we show the poor performance of these previous state-of-the-art approaches, which were originally intended for retinal tissue interface segmentation, on challenging corneal and limbal datasets.

Contributions: Towards the goal of accurate interface segmentation in OCT images acquired from different OCT scanners, in this paper, we propose the first approach (to the best of our knowledge) to accurately identify the shallowest tissue interface in OCT images by mitigating speckle noise patterns and severe specular artifacts. We propose the creation of an intermediate OCT image representation that can influence the performance of a segmentation approach. Our major contributions in this paper are:

1. A novel application of a conditional Generative Adversarial Network (cGAN) to remove speckle noise patterns and specular artifacts from the air gap of Anterior Segment OCT (AS-OCT) datasets, thereby making AS-OCT images potentially easier to interpret for ophthalmologists, trainees, non-experts, and segmentation algorithms.

2. **Cascaded Framework**: We present a cascaded neural network framework, which comprises of a cGAN and a Tissue Interface Segmentation Network (TISN). The cGAN pre-segments OCT images by removing undesired specular artifacts and speckle noise patterns just prior to the shallowest tissue interface. The pre-segmentation output of the cGAN is an intermediate output. Following pre-segmentation, the TISN predicts the final segmentation using both the original and pre-segmented images, and the shallowest interface is extracted and fitted with a curve.

3. **Hybrid Framework**: The intermediate pre-segmentation output yielded by the cGAN is used as the image input to another tissue-interface segmentation algorithm, e.g. [12]. In general, the pre-segmentation can be used by any segmentation algorithm, but in the Hybrid Framework the second-stage segmentation algorithm does not have access to the original OCT image.

4. **cGAN Weighted Loss**: We propose a task-specific weighted loss for the cGAN, which enforces the preservation of details related to the tissue structure, while removing specular artifacts and speckle noise patterns just prior to the shallowest interface in a context-aware manner.

5. This is the first application of training and testing our previously published CorNet [29] neural network architecture on challenging limbal datasets.

Our cascaded framework was first applied to corneal datasets, which were acquired using two different OCT systems and different scan protocols. Encouraged by our cascaded framework's performance on corneas, we diversified our training to also include limbal datasets (also acquired with different OCT systems). It seemed reasonable to seek generalized learning since the characteristics of limbal datasets are similar to corneal datasets in terms of low SNR, speckle

noise patterns, and specular artifacts. In all these datasets, we segmented the shallowest interface that could be extracted in each B-scan.

A key motivation for the proposed hybrid framework was to directly integrate the output of the cGAN into the image acquisition pipeline of custom-built OCT scanners. As we postulated earlier, the varying degrees of specular artifacts and speckle noise patterns confound traditional segmentation algorithms. If the cGAN were integrated into the imaging pipeline and OCT B-scans were pre-segmented after acquisition, then we hypothesized that previously proposed segmentation algorithms should benefit from the removal of specular artifacts and speckle noise patterns just above the shallowest interface. Thus, our goal with the development of the hybrid framework was to show that the pre-segmented OCT image enabled one of these segmentation algorithms [12] to generate lower segmentation errors.

To quantify the performance of our proposed frameworks, we compared the results of the following baselines: 1) A traditional image analysis-based algorithm [12] that directly segmented the tissue interface, 2) The hybrid framework, 3) A deep learning-based approach [29] that directly segmented the tissue interface, and 4) The cascaded framework. We provide a summary of the major results below:

1. We show that our approach is generalizable to datasets acquired from multiple scanners displaying varying degrees of specular noise, artifacts, and bulk tissue motion.

2. Our proposed frameworks segment the shallowest interface in datasets where the scanner starts by imaging the limbus, crosses over the limbal junction, and images the cornea.

3. By executing a traditional image analysis-based algorithm on the pre-segmentation, the segmentation error was reduced.

4. We always accurately segmented the shallowest interface in corneal datasets using our proposed frameworks.

5. In a majority of limbal datasets (15/18), we were able to precisely delineate the shallowest interface with our proposed frameworks.

2. Methods

2.1. Problem statement

Given an OCT image $\mathcal{I}$, the task of a conditional Generative Adversarial Network (cGAN) is to find a function $\mathcal{F}_G : \{\mathcal{I}, z\} \rightarrow \mathcal{P}$ that maps a pixel in $\mathcal{I}$ using a random noise vector z to a pre-segmented output image $\mathcal{P}$. The pixels in $\mathcal{P}$ just prior to the tissue interface are mapped to 0 (black), while those at and below the interface are retained.

Next, the task of the Tissue Interface Segmentation Network (TISN) is to determine a mapping $\mathcal{F}_O : \{\mathcal{I}, \mathcal{P}\} \rightarrow \mathcal{S}$, wherein every corresponding pixel in $\mathcal{I}$ and $\mathcal{P}$ is assigned a label $\mathcal{L} \in \{0, 1\}$ in the final segmentation $\mathcal{S}$. In this paper, we only segment the shallowest tissue interface in the image, and thus assign pixels in $\mathcal{S}$ as: (0) pixels just above the tissue interface, (1) pixels at and below the tissue interface. $\mathcal{P}$ can then be used in a hybrid framework by any other segmentation algorithm. Our frameworks are pictorially shown in Fig. 3.

2.2. Architecture

We first describe the neural network architecture that was used as the base for both the cGAN (generator), and the TISN. As mentioned in Sec. 1, images of the anterior segment of the eye acquired using OCT contain low SNR, strong specular artifacts, and faintly discernable interfaces that are corrupted by speckle noise patterns. In our previous work [29], we have shown that the CorNet architecture captures faintly visible features across multiple scales. The
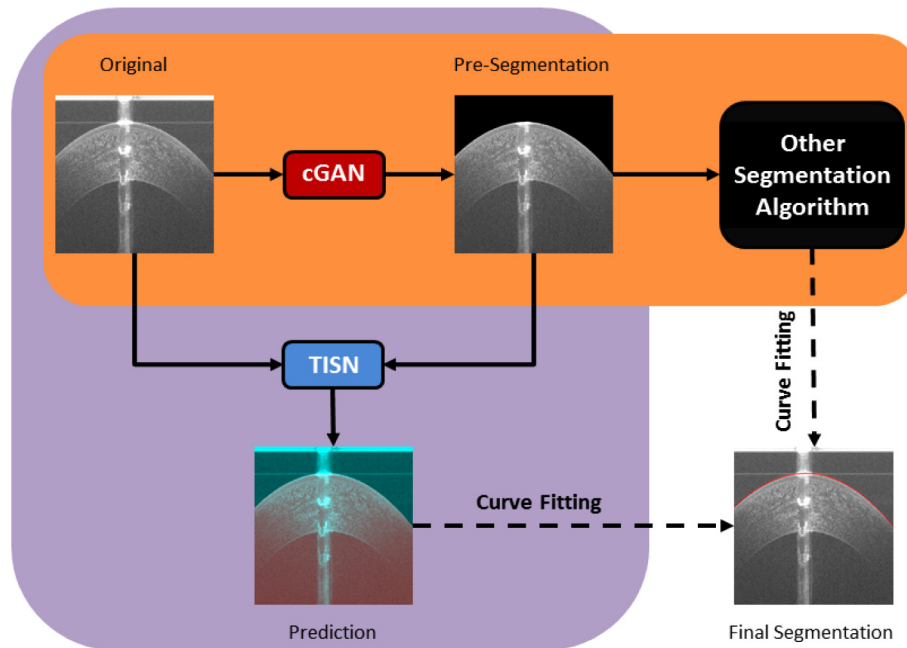
Fig. 3. Our proposed approach contains two frameworks: a cascaded framework (purple) and a hybrid framework (orange). First, a conditional Generative Adversarial Network (cGAN) takes an input OCT image, and produces an intermediate pre-segmentation image. In the pre-segmentation, pixels just prior to the shallowest tissue interface are set to 0 (black), while others are retained. In the cascaded framework, the pre-segmentation, along with the input image, are passed to a Tissue Interface Segmentation Network (TISN). The TISN predicts the location of shallowest interface by generating a binary segmentation mask (overlaid on the original image with a false color overlay; red - foreground, turquoise - background). In the hybrid framework, the pre-segmentation can be utilized by other segmentation algorithms. Ultimately, both frameworks fit a curve to the interface to produce the final segmentation.

network architecture is shown in Fig. 4, and it produced state-of-the-art results on corneal datasets acquired using different OCT systems and using different scan protocols. The errors were 2× lower than non-proprietary state-of-the-art segmentation algorithms, including traditional image analysis-based [11,71] and deep learning-based approaches [61,72,74].

The CorNet architecture was built upon the UNET [74] and BRUNET [72] architectures. The CorNet architecture enhanced the reuse of features generated in the neural network through residual connections [75], dense connections [76], and dilated convolutions [77–79]. The hypercolumn activation maps [80] of the CorNet architecture are contrasted against the UNET and BRUNET architectures for the task of segmenting corneal tissue interfaces in Fig. 5. As opposed to the UNET and BRUNET architectures, the CorNet design alleviated the vanishing gradient problem, and prevented the holes in the segmentation generated by current deep learning-based approaches [61,72,74]. By contrasting the activation maps of the CorNet in Figs. 5(m) – 5(o) with those of the UNET in 5(c) – 5(e) and the BRUNET in 5(h) – 5(j), it is clear that tissue interfaces are more clearly defined in the CorNet architecture indicating the advantages of using residual and dense connections with dilated convolutions. Holes at the corneal apex due to the strong central artifact are now filled in by the CorNet architecture. Similarly, the CorNet design could also accurately extract poorly defined corneal interfaces, such as the Endothelium (bottom-most layer), which is very common in anterior segment OCT imaging [29].
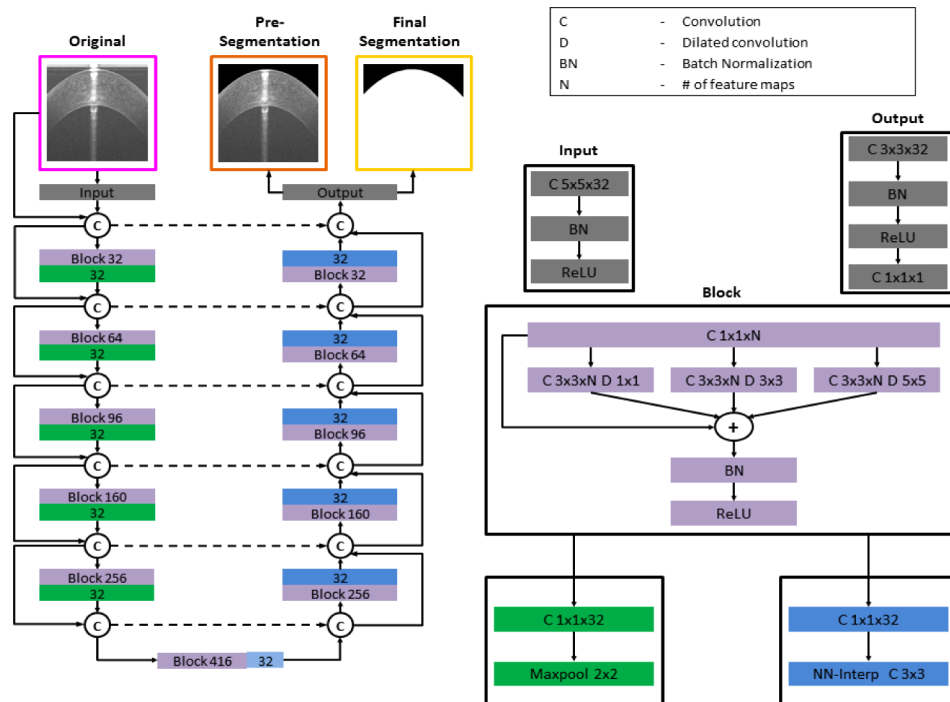
Fig. 4. The CorNet architecture is the base used for training both the cGAN and TISN. The input to the cGAN is a two-channel image: the input OCT image and binary mask *w* (see Sec. 3.1.2), and the output is a pre-segmented OCT image (orange box). The TISN gets a two-channel input (magenta and orange boxes), and the output is a binary mask (yellow box). The dark green blocks in the contracting path represent downsampling operations, while the blue blocks constitute upsampling computations. This model uses residual and dense connections to efficiently pre-segment the OCT image, and predict the location of the shallowest interface in the final output. The light blue module at the bottom of the model did not upsample feature maps, instead it functioned as a bottleneck to create outputs with the same size as those from the last layer.

As shown in Fig. 4, the CorNet architecture comprised of contracting and expanding branches; each branch consisted of a building block, which was inspired by the Inception block [79], followed by a bottleneck block. The building block extracted features related to edges and boundaries at different resolutions. The bottleneck block compactly represented the salient attributes, and these properties (even from earlier layers) were encouraged to be reused throughout the network. Thereby, faint tissue boundaries essential to our segmentation task were distinguished from speckle noise patterns, and pixels corresponding to the tissue interface and those below it were correctly predicted. In addition, extensive experiments were conducted in [29] to determine the right feature selection mechanisms [61,81–84] for segmentation, such as max-pooling [84] for downsampling and nearest neighbor interpolation + 3×3 convolution [83] for upsampling.

2.3. Conditional generative adversarial network (cGAN)

2.3.1. Original cGAN

Conditional Generative Adversarial Networks [31] are currently popular choices for image-to-image translation tasks, such as image super-resolution and painting style transfer. In these tasks, the cGAN learns to generate an output by being introduced to (conditioned on) an input image.
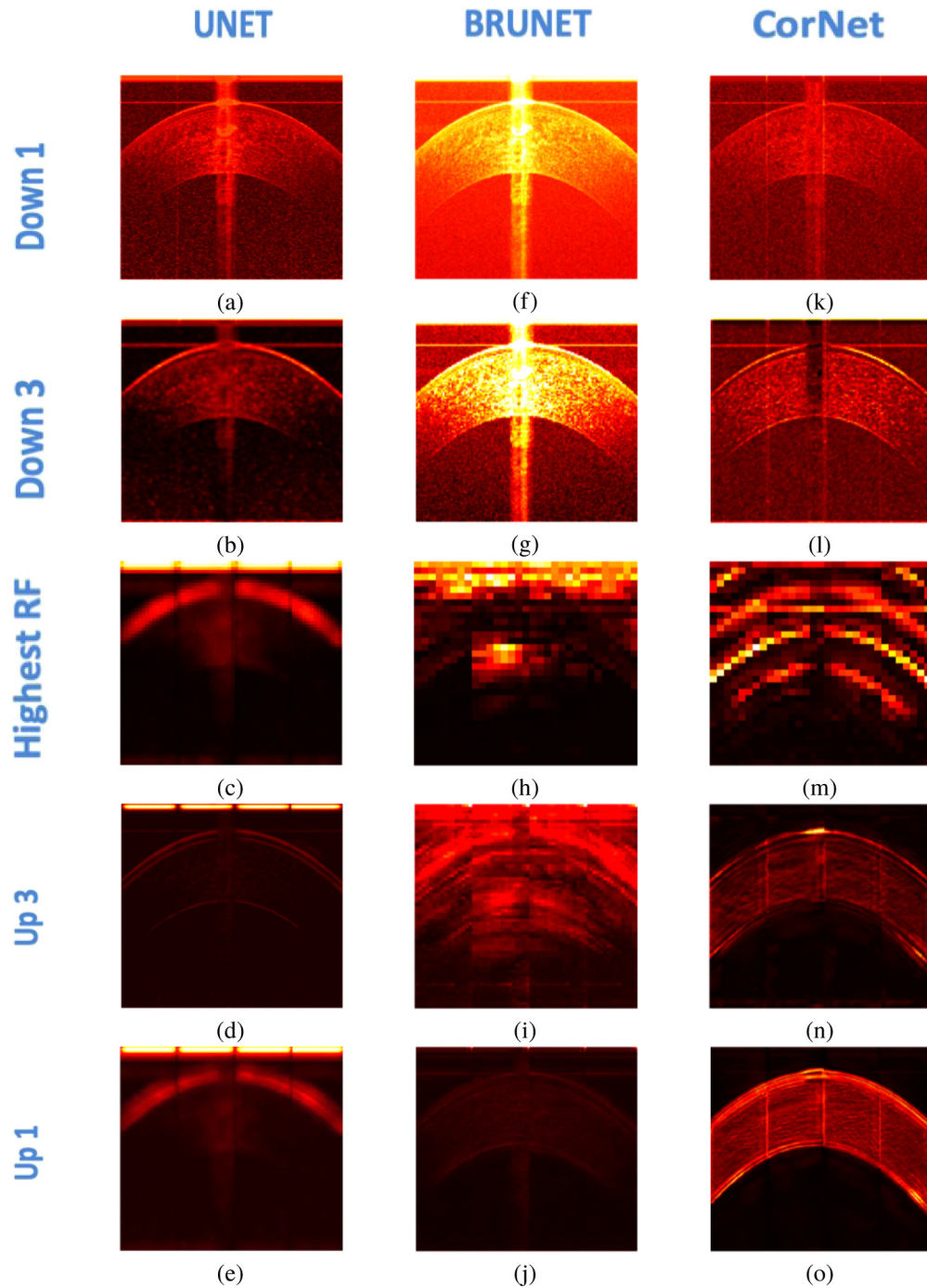
Fig. 5. Each column shows the hypercolumn activation maps [80] of the UNET, BRUNET and CorNet architectures respectively for a corneal OCT B-scan. The activations were extracted from the downsampling layers 1 and 3, upsampling layers 1 and 3, and the layer with the highest receptive field (RF). Notice the improved structural detail in the CorNet activations as opposed to UNET and BRUNET.

The cGAN framework consists of two entities: a Generator (G) and a Discriminator (D). The generator G takes an input image x and a random noise vector z, and generates a prediction y_f that is similar to the desired gold standard output y_t. Next, the input x is paired with y_t and y_f, thereby creating two pairs of images respectively; the true gold standard pair (x, y_t) and the predicted pair (x, y_f). Then, the discriminator D recognizes the pair that most accurately represents the gold standard output desired. These two entities are trained in conjunction, such that they compete with each other; G tries to fool D by producing an output that closely resembles the gold standard, while D tries improve its ability to distinguish the two pairs of images.

Initially, G generates a prediction y_f that poorly resembles y_t. It learns to produce more realistic predictions by minimizing an objective function shown in Eq. (1). On the other hand, D tries to maximize this objective by accurately distinguishing the generated prediction y_f from the true gold standard y_t. The objective function comprises of two losses: L_{cGAN} in Eq. (2), and L_1 in Eq. (3), with λ being a hyper-parameter. The L_1 loss is a "structured" loss [31] forcing the output of the generator to be close to the ground truth in the L_1 sense. This loss encouraged lesser blurring [31] in generated pre-segmentations as opposed to the original GAN formulation [30], which utilized an L_2 loss. The PatchGAN [31] discriminator was employed to output the probability of a pair of images being real or fake.

$$G^* = arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_1(G) \qquad (1)$$

$$L_{cGAN}(G, D) = E_{x,y_t} [log\ D(x, y_t)] + E_{x,z} [log(1 - D(x, G(x, z)))] \qquad (2)$$

$$L_1 = E_{x,y_t,z} [\|y_t - G(x, z)\|_1] \qquad (3)$$

Directly transferring the full cGAN implementation with the cGAN loss in Eq. (1) to our OCT datasets resulted in checkerboard artifacts [83] in the generated predictions. Moreover, as shown in Fig. 6, parts of the tissue boundary that needed to be preserved were removed instead. From our experiments, we made two empirical observations: 1) The UNET generator architecture [74] that was utilized in the cGAN paper [31] created checkerboard artifacts in the generated pre-segmentation and did not preserve tissue boundaries correctly; it has been shown in prior work [29,72,83] and in Fig. 5 that the original UNET implementation is not the optimal choice; 2) The L_1 loss in Eq. (3) penalizes all pixels in the image equally.

2.3.2. Modified cGAN with weighted loss

The required output of the cGAN is a pre-segmented OCT image, wherein the background pixels just prior to the shallowest tissue interface are to be eliminated, and the region at and below the interface is to be preserved. As mentioned before, the L1 loss in Eq. (3) equally penalized all pixels in the image without imparting a higher penalty to the background pixels, which contains specular artifacts and speckle noise patterns hindering segmentation, above the shallowest tissue interface. To mitigate this problem, a novel task-specific weighted L_1 loss, defined in Eq. (4), is proposed in this paper. In Eq. (4), $\circ$ denotes the pixel-wise product, and α is the hyper-parameter that imparts higher weight to the background pixels over the foreground pixels.

$$L_{w1} = E_{x,y,z} [\alpha w \circ \|y_t - G(x, z)\|_1 + (1 - w) \circ \|y_t - G(x, z)\|_1] \qquad (4)$$

As the preservation of pixels at and below the interface is paramount, our loss function incorporated a binary mask w, which imparted different weights to the foreground and background pixels. This mask was generated from the gold standard annotation of an expert grader for each image in the training dataset, and its design is further described in Sec. 3.1.2. We replaced the L_1 loss in Eq. (1) with our weighted L_1 loss in Eq. (4). Furthermore, we performed an experiment where we incorporated the weight mask in the original UNET-based cGAN implementation, and yet the tissue interfaces in the generated pre-segmentations were not preserved as shown in Fig. 6. This shows that the combination of the CorNet architecture and the modified cGAN loss function helped eliminate the speckle patterns and specular artifacts just prior to the shallowest interface.
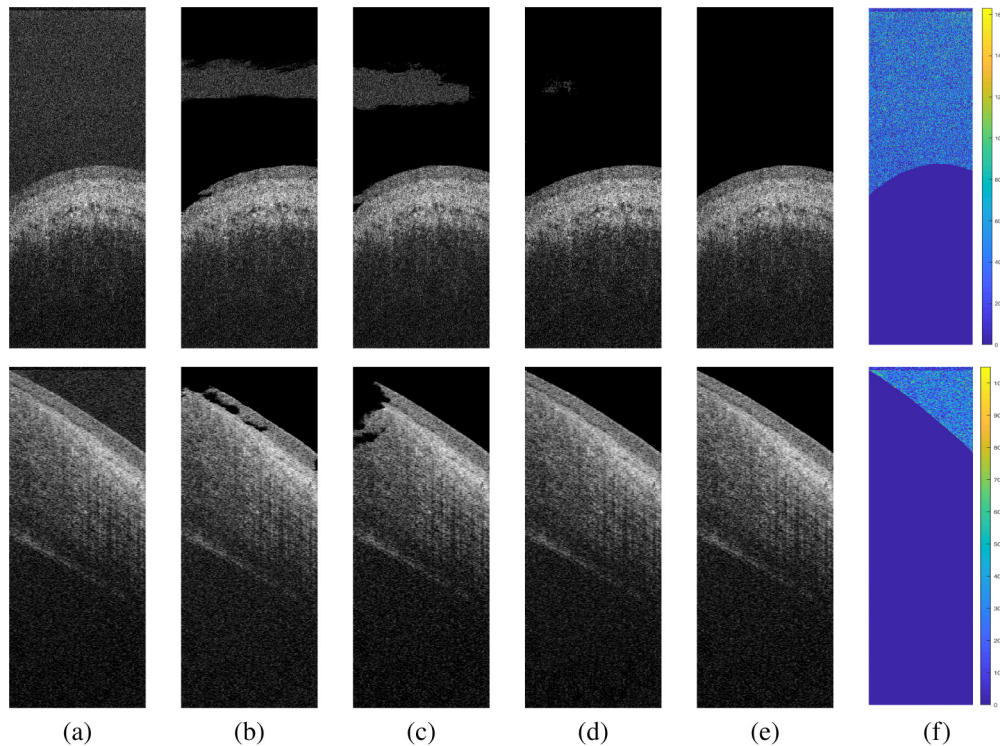
Fig. 6. Comparing generated pre-segmentations between the UNET architecture used in the original cGAN implementation [31] against those generated by the CorNet architecture [29]. Original B-scans for two different limbal datasets are shown in (a). The corresponding generated pre-segmentations from the original UNET-based cGAN implementation (without weighted mask) is shown in (b), while those pre-segmentations for the modified UNET-based cGAN (with the weighted mask) is shown in (c). Similarly, the generated pre-segmentations from the CorNet-based cGAN (without weighted mask) is shown in (d), while those pre-segmentations for the modified CorNet-based cGAN (with the weighted mask) is shown in (e). Heat maps showing the difference between the original B-scan and the corresponding pre-segmentation by CorNet (with the weighted mask) is shown in (f). Note that the UNET did not remove the speckle patterns above the shallow tissue interface, while also encroaching upon the tissue boundaries without preserving them accurately.

2.4. Tissue interface segmentation network (TISN)

As mentioned in Sec. 2.2, the CorNet architecture was used as the base model in order to segment the shallowest tissue interface. The intermediate pre-segmented OCT image from the cGAN, along with the original OCT image, is passed to the TISN to delineate the shallowest tissue interface. The output of the TISN is a binary mask, wherein pixels corresponding to the tissue interface and those below it were labeled as the foreground (1) and those above the interface were labeled as the background (0). As shown in Figs. 3 and 4, the shallowest interface was extracted from this binary mask [85] and fitted with a curve [86].

3.   Experiments and results

3.1.   Data

3.1.1.   Acquisition

25 corneal datasets and 25 limbal datasets, totaling 50 datasets, were randomly selected from an existing research database [29]. These datasets were acquired using different scan protocols from three different OCT scanners: a custom Bioptigen Spectral Domain OCT (SD-OCT) scanner (Device 1) that has been described before [87], a high-speed ultra-high resolution OCT (hsUHR-OCT) scanner (Device 2) [88], and a Leica (formerly Bioptigen) Envisu C2300 SD-OCT system (Device 3) [89]. Device 1 had a 3.4μm axial and 6μm lateral spacing, and it was used to scan an area of size 6 × 6mm on the cornea. Device 2 was used to scan two areas of sizes 6×6mm and 3×3mm respectively. This system had a 1.3μm axial and a 15μm lateral spacing while interrogating the 6×6mm tissue area. It had the same axial spacing, but a different lateral spacing of 7.5μm while imaging the 3×3mm area. Device 3 had a ~2.44μm axial and 12μm lateral spacing when fitted with the 18mm anterior imaging lens. Devices 1 and 2 were solely used to scan the cornea, with the former producing datasets of dimensions 1024×1000×50 pixels, and the latter generating datasets of dimensions 400×1024×50 pixels. Devices 2 and 3 were used to scan the limbus, resulting in volumes that had varying dimensions; the number of A-scans across all limbal datasets varied between 256 and 1024, with a constant 1024 pixels axial resolution, and the number of B-scans across all datasets varied between 25 and 375.

3.1.2.   Data preparation

From the 50 datasets, we had a total of 1250 corneal images and 4437 limbal images respectively. Of the 50 corneal and limbal datasets, 14 datasets were randomly chosen for training the cGAN, and the remaining were used for testing. From the total set, we chose the training set to comprise of a balanced number of limbal and corneal datasets (7 each) that exhibited different magnitudes of specular artifacts, shadowing, and speckle. The training set contained 350 corneal and 1382 limbal images respectively, and the remaining were set aside in the testing set. Among the 7 corneal training datasets, 3 corneal datasets from Device 1, 2 corneal datasets from Device 2 (size 6×6 mm) and 2 corneal datasets from Device 2 (size 3×3 mm) were used. Among the 7 limbal training datasets, 4 datasets from Device 2 and 3 datasets from Device 3 were used. Similarly, for the testing datasets, 18 datasets each from the cornea and limbus were used. Among the 18 corneal testing datasets, 6 datasets from Device 1, 6 datasets from Device 2 (size 6×6 mm) and 6 corneal datasets from Device 2 (size 3×3 mm) were used. Among the 18 limbal testing datasets, 12 datasets from Device 2, and 6 datasets from Device 3 were used.

These datasets were chosen such that they came from both eyes, and each dataset was acquired from a different patient. Considering the varying dimensions of the OCT images acquired from three OCT systems that were used in this work, along with the limited GPU RAM available for training, it was challenging to train a framework using full-width images while preserving the pixel resolution. Similar to previous approaches [29,61], we sliced the input images width-wise (c.f Fig. 7) to produce a set of images of dimensions 256×1024 pixels, and in this way, we preserved the OCT image resolution. We did not resize the image or crop the image in any other way except for slicing the image width-wise to preserve the image resolution. We used the same datasets that were selected in the training set for training both the cGAN and the TISN.

An example annotation by an expert is shown in Fig. 8(a). To generate the gold standard pre-segmentation images for training, we eliminated the speckle noise and specular artifacts by setting the region above the annotated surface to 0 (black), and kept the same pixel intensities corresponding to the tissue structure at the annotation contour and for all pixels below it - see Fig. 8(b). The binary mask *w* that was used in the Eq. (4) is shown in Fig. 8(c). Using the image in Fig. 8(d) as reference, we detail the process of obtaining *w*. In Fig. 8(d), the expert annotation
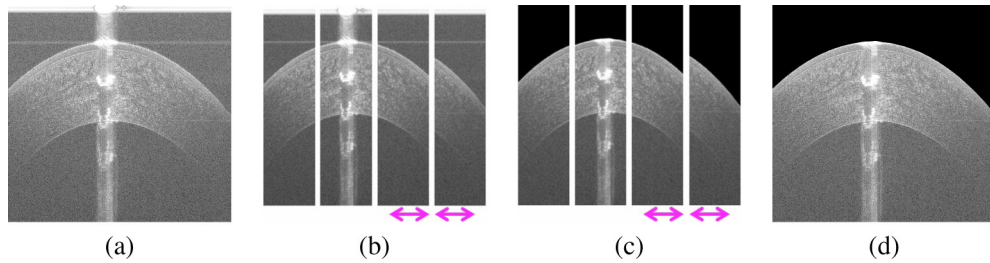
Fig. 7. (a) Original B-scan of dimensions 1000×1024 pixels from a 6×6mm corneal volume acquired by a custom SD-OCT scanner; (b) Four slices of dimensions 256×1024 pixels obtained by dividing the image in (a) width-wise; (c) Four corresponding predictions generated by the cGAN with the same dimensions as each input slice in (b); (d) The final pre-segmentation image recreated by putting together the slices in (c), with the final dimensions matching the 1000×1024 pixels of the original B-scan. In the case of non-square OCT images, the magenta arrows show the extent of the overlapping regions in slices 3 and 4.

of the tissue interface boundary is shown in red, and this red annotated contour was shifted down by 50 pixels to the position of the magenta contour. The magenta contour, along with the blue region below the contour, was considered the foreground, while all pixels above the magenta contour belong to the background. The background in the binary mask was set to 1 and the foreground was set to 0, with the background being weighted α times higher than the foreground.
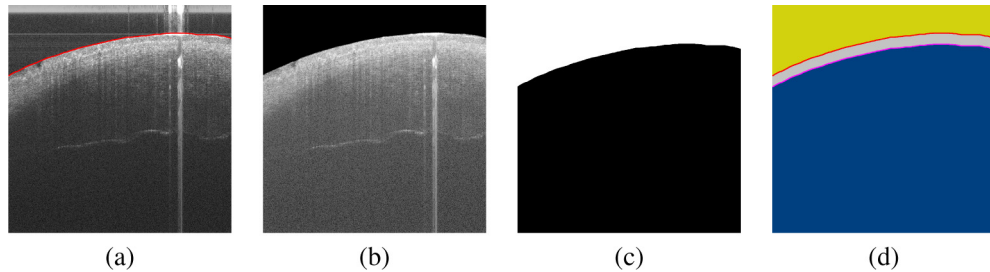

Fig. 8. (a) Expert annotation of an original B-scan in a 6×6mm limbal volume acquired by Device 3, (b) Gold standard pre-segmentation image for training, (c) Binary mask w used in Eq. (4) for training the cGAN, (d) Label map detailing the process of generating w (see Sec. 3.1.2).

In order to understand the effect of the proposed mask design, let us consider an alternate binary mask design w^*. Let w^* represent the mask of the expert annotation in Fig. 8(a), wherein the pixels above the annotation (without shifting it down/up) are the background and those at and below the annotation are the foreground, with the background weighted α times higher than the foreground. When the cGAN used this mask w^*, it mistakenly eroded the tissue interface and regions below it similar to the image in Figs. 6(b) and 6(c). In such a scenario, there is no large penalty applied to the erosion of pixels as detailed in Eq. (4). In order to correct this mistake, it would be necessary to impart a higher penalty to the region that was eroded. To do so, we measured the maximum extent of structural erosion (at the tissue interface and/or pixels below it) from the shallowest interface in the UNET pre-segmentation outputs. Using this value (rounded up to a nearest multiple of 10), we shifted expert annotation down (by 50 pixels) in our binary mask w, and conferred the same weight α to the regions (yellow + red + gray) to avoid the erosion of the tissue interface. The structural extent of erosion was measured from the

cGAN generator (UNET) outputs using the training datasets as input. The 50 pixel downward annotation shift was based off of the training datasets, and it satisfactorily prevented the erosion of interfaces in our experiments. However, as newer and different datasets might contain varying degrees of pathology, this shift value may vary while fine tuning network weights.

3.1.3. Data augmentation

As our training datasets were smaller in number in contrast to larger computer vision datasets [90], we augmented our datasets to increase the variety of the images that were seen during the training. These augmentations [91] included horizontal flips, gamma adjustment, elastic deformations, Gaussian blurring, median blurring, bilateral blurring, Gaussian noise addition, cropping, and affine transformations. The full set of augmented images was used to train the TISN as it required substantially larger amounts of data to generalize to new test inputs. On the other hand, the cGAN can be trained with smaller quantities of input training data as it has been shown to perform well on small training datasets [31]. For the cGAN, augmentation was done by simply flipping each input slice horizontally along the X-axis.

3.2. Experimental setup

3.2.1. cGAN training

Training of the cGAN commenced from scratch using the architecture shown in Fig. 4. The input to the generator was a two-channel image; the first channel corresponds to the input OCT image, and the second channel corresponds to the binary mask w. We used $\lambda = 100$, and $\alpha = 10$ in final objective function, and optimized the network parameters using the ADAM optimizer [92]. We used 90% of the input data for training, and the remaining 10% for validation. We trained the network for 100 epochs with the learning rate set to 2×10^{-3}. In order to prevent the network from over-fitting to the training data, early stopping was applied when the validation loss did not decrease for 10 epochs. At the last layer of the generator, a convolution operation, followed by a TanH activation, was used to convert the final feature maps into the desired output pre-segmentation with pixel values mapped to the range of $[-1, 1]$. A NVIDIA Tesla V100 16GB GPU was used for training the cGAN with a batch size of 4. During test time, the input OCT image is replicated to produce a two-channel input to the cGAN.

3.2.2. TISN training

The same datasets from cGAN training were used for training the TISN from scratch. The input to the TISN is a two-channel image; the first channel corresponds to the original input image, and the second channel corresponds to the predicted pre-segmentation obtained from the cGAN. The two-channel input allowed the TISN to focus on the high frequency regions, corresponding to the interface, in the image. The Mean Squared Error (MSE) loss, along with the ADAM optimizer [92], was used for training. In this work, we used MSE loss to be consistent with the original CorNet implementation [29], but the MSE loss can easily be substituted for the cross entropy loss [74] or the dice loss [93]. The batch size used for training was set to 2 slices as we fully wanted to utilize memory on a NVIDIA Titan Xp GPU. Validation data comprised of 10% of the training data. We trained the network for a total of 150 epochs with the learning rate set to 10^{-3}. When the validation loss did not improve for 5 epochs, the learning rate was decreased by a factor of 2. Finally, in order to prevent over-fitting, the training of the TISN was halted through early stopping when the validation loss did not improve for 10 consecutive epochs. The feature maps in the final layer of the network are activated using the softmax function to produce a two-channel output. Once the network was trained, it was used to segment the shallowest interface in our testing datasets. At test time, the TISN yielded a two-channel output; the first channel corresponded to the foreground tissue segmentation, and the second channel corresponded to the background pixel segmentation (above the tissue interface). The foreground pixels corresponded to the

boundary of the interface and those pixels below it, while the pixels above the tissue boundary denoted the background. Finally, the predicted segmentation was fitted with a curve [86] after the tissue interface was identified using a fast GPU-based method [85]. We show our final results in Figs. 10, 11 and 20 along with Visualization 1, Visualization 2, and Visualization 3. The predicted segmentation results obtained from the traditional and deep learning algorithms with and without the pre-segmentation are shown in Fig. 9.
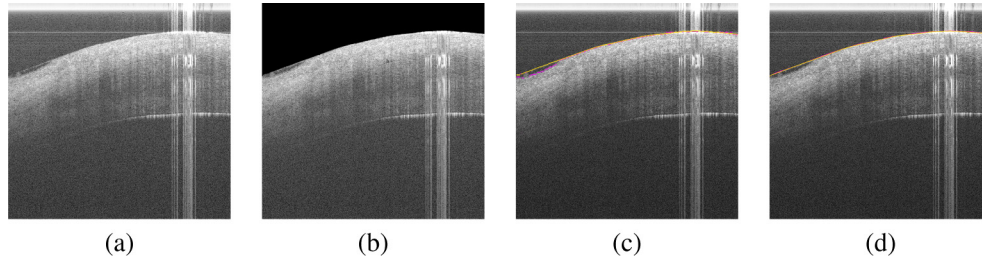

(a)   (b)   (c)   (d)

Fig. 9. (a) Original B-scan of dimensions 1000×1024 pixels from a 4×4mm limbal volume acquired by Device 3; (b) Corresponding pre-segmentation obtained from our cGAN that is fed to the traditional and deep learning algorithms for final segmentation; (c) Comparing the segmentation result of a traditional algorithm executed without the pre-segmentation (magenta contour) against the result of the algorithm executed with the pre-segmentation (gold contour); (d) Comparing the segmentation result of a deep learning algorithm executed without the pre-segmentation (magenta contour) against the result of the algorithm executed with the pre-segmentation (gold contour). Note that the segmentation result closely follows the tissue interface boundary when the pre-segmentation is used.

3.3. Baseline comparisons

Extensive evaluation of the performance of our approach was conducted across all the testing datasets. First, we wanted to investigate the accuracy of a traditional image analysis-based algorithm [12] that directly segmented the interface in our test datasets. Briefly, this algorithm filtered the OCT image to reduce speckle noise and artifacts, extracted the monogenic signal [94], and segmented the tissue interface. We denote this baseline in the rest of the paper by the acronym: Traditional WithOut Pre-Segmentation (TWOPS).

Second, we designed a hybrid framework, where the pre-segmented OCT image from the cGAN is used by the traditional image analysis-based algorithm [12] to segment the shallowest interface. We wanted to determine the improvement in segmentation accuracy when the traditional algorithm used the pre-segmentation instead of the original OCT image. Going forward, we denote this baseline by the acronym: Traditional With Pre-Segmentation (TWPS).

Third, we trained a CorNet architecture [29] to directly segment the foreground in the input OCT image, *without* including the cGAN pre-segmentation as an additional input channel. We compared the direct segmentation result against our cascaded framework. Henceforth, in the remainder of the paper, we refer to the direct deep learning-based segmentation approach by the acronym: Deep Learning WithOut Pre-Segmentation (DLWOPS). Finally, we call our cascaded framework as: Deep Learning With Pre-Segmentation (DLWPS).

To summarize, the following baseline methods were considered for performance evaluation:

1. TWOPS - A traditional image analysis-based algorithm [12] that directly segmented the tissue interface.

2. TWPS - The hybrid framework.

3. DLWOPS - A deep learning-based approach [29] that directly segmented the interface.
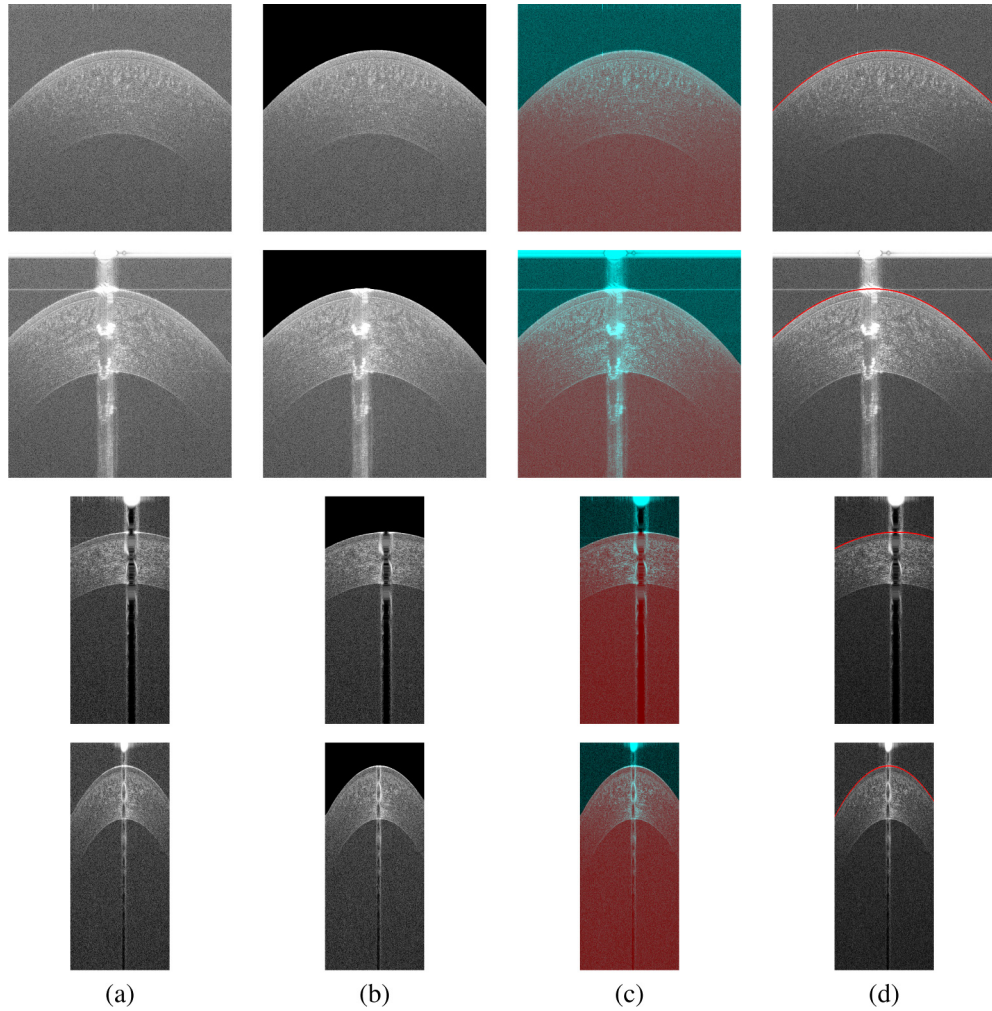
(a)　　　　　　　(b)　　　　　　　(c)　　　　　　　(d)

Fig. 10. Corneal interface segmentation results for datasets acquired using Devices 1 and 2. Columns from left to right: (a) Original B-scans in corneal OCT datasets, (b) Pre-segmented OCT images from the cGAN with the specular artifact and speckle noise patterns removed just prior to the shallowest tissue interface, (c) Binary segmentation from the TISN overlaid in false color (red - foreground, turquoise - background) on the original B-scan, (d) Curve fit to the shallowest interface (red contour).
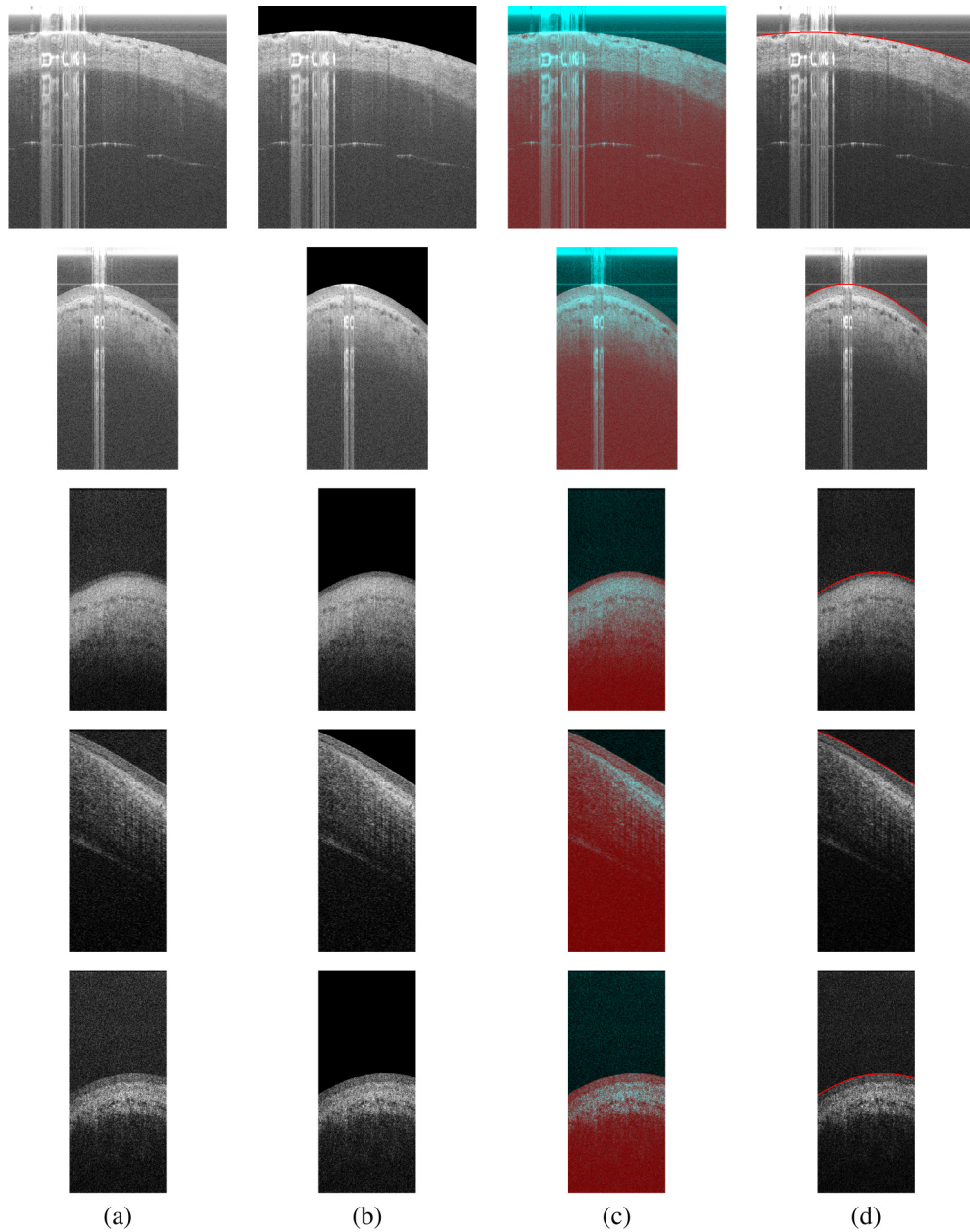
(a) (b) (c) (d)

Fig. 11. Limbal interface segmentation results for datasets acquired using Devices 2 and 3. Columns from left to right: (a) Original B-scans in the limbal OCT datasets, (b) Pre-segmented OCT images from the cGAN with the specular artifact and speckle noise patterns removed above the shallowest tissue interface, (c) Binary segmentation from the TISN overlaid in false color (red - foreground, turquoise - background) on the original B-scan, (d) Curve fit to the shallowest interface (red contour).

4. DLWPS - The cascaded framework.

3.4. Evaluation

3.4.1. Annotation

Each corneal dataset was annotated by an expert grader (G1; Grader 1) and a trained grader (G2; Grader 2). However, only expert annotations were available for the limbal datasets in the research database. The graders were asked to annotate the shallowest interface in all test datasets. For each dataset, the graders annotated the interface using a 5-pixel width band with an admissible annotation error of 3 pixels. All the annotations were fitted with a curve for comparison with the different baselines. We also estimated the inter-grader annotation variability for the corneal datasets, and refer to it in the rest of the paper by the acronym: IG.

3.4.2. Metrics

In order to compare the segmentation accuracy across the different baselines, we calculated the following metrics: 1) Mean Absolute Difference in Layer Boundary Position (MADLBP) and 2) Hausdorff Distance (HD) between the fitted curves. These metric values were determined over all testing datasets, and only for the shallowest interface. In Eqs. (5) and (7), the sets of points that represent the gold standard annotation and the segmentation to which it is compared (each fitted with curves) are denoted by G and S respectively. We denote by $y_G(x)$ the Y-coordinate (rounded down after curve fitting) of the point in G whose X-coordinate is x, and $y_S(x)$ is the Y-coordinate (rounded down) of the point in S. $d_S(p)$ is the distance of a point p in G to the closest point in S, and similarly for $d_G(p)$.

　We chose MADLBP in Eq. (5) as one of our error metrics since it was used in [12] to compare the segmentation accuracy between the automatic segmentations and grader annotations. Although MADLBP quantifies error in pixels, it did not measure the Euclidean distance error; instead, it simply measured the positional distance between the detected boundary location and the annotation along the same A-scan. On the other hand, the Hausdorff distance in Eq. (7) captured the greatest of all distances between the points in the segmentation and annotation. Therefore, it quantitatively describes the worst segmentation error in microns as it is more clinically relevant (e.g. to detect structural changes over time). In this work, we did not compute Dice similarity as it did not provide segmentation error in microns.

$$\text{MADLBP} = \frac{1}{X} \sum_{x=0}^{X-1} |y_G(x) - y_S(x)| \tag{5}$$

$$\text{HD} = \max \left(\max_{p \in G} d_S(p), \ \max_{p \in S} d_G(p) \right) \tag{6}$$

In Fig. 12, the HD error and the MADLBP error across all baselines for the corneal datasets acquired from devices 1 and 2 were compared. In Fig. 13, the benefit of pre-segmenting the OCT image was verified by first grouping the baselines into two categories - Traditional Comparison (TC; TWOPS vs TWPS) and Deep Learning Comparison (DLC; DLWOPS vs DLWPS) - and then contrasting the maximum HD error per dataset for each category and for each grader. We also provide the mean and standard deviation errors for all images in each corneal dataset in Fig. 14; we display these segmentation errors compared against the expert grader for the TC and DLC categories. We also determined the HD and MADLBP error across the limbal datasets in Figs. 15 and 19. Again in Fig. 16, we estimated the benefit of pre-segmenting limbal datasets by grouping baselines into two categories, TC and DLC, and comparing maximum HD error per dataset for each category. Moreover, we found a few instances where our cascaded framework failed to correctly segment the tissue interface as seen in Fig. 16 (results after the red vertical

line). Similarly, we provide the mean and standard deviation errors for all images in each limbal dataset in Fig. 17; we display these segmentation errors compared against the expert grader for the TC and DLC categories.
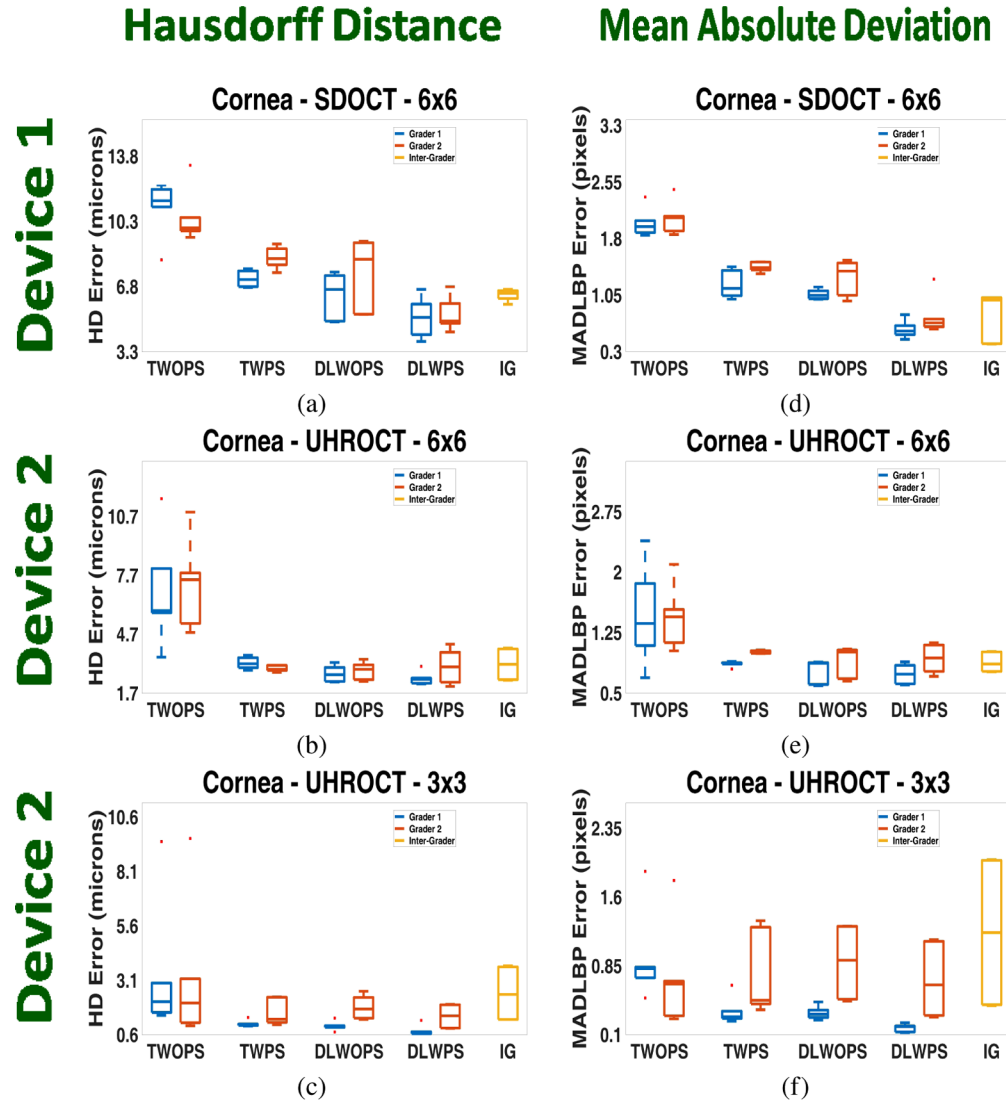

Fig. 12. (a)-(c) HD error and (d)-(f) MADLBP error comparison for the corneal datasets acquired with Devices 1 and 2 respectively. In the boxplots, the segmentation results obtained for each baseline method are contrasted against expert grader (blue) and trained grader (red) annotations, while the Inter-Grader (IG) variability is shown in yellow.
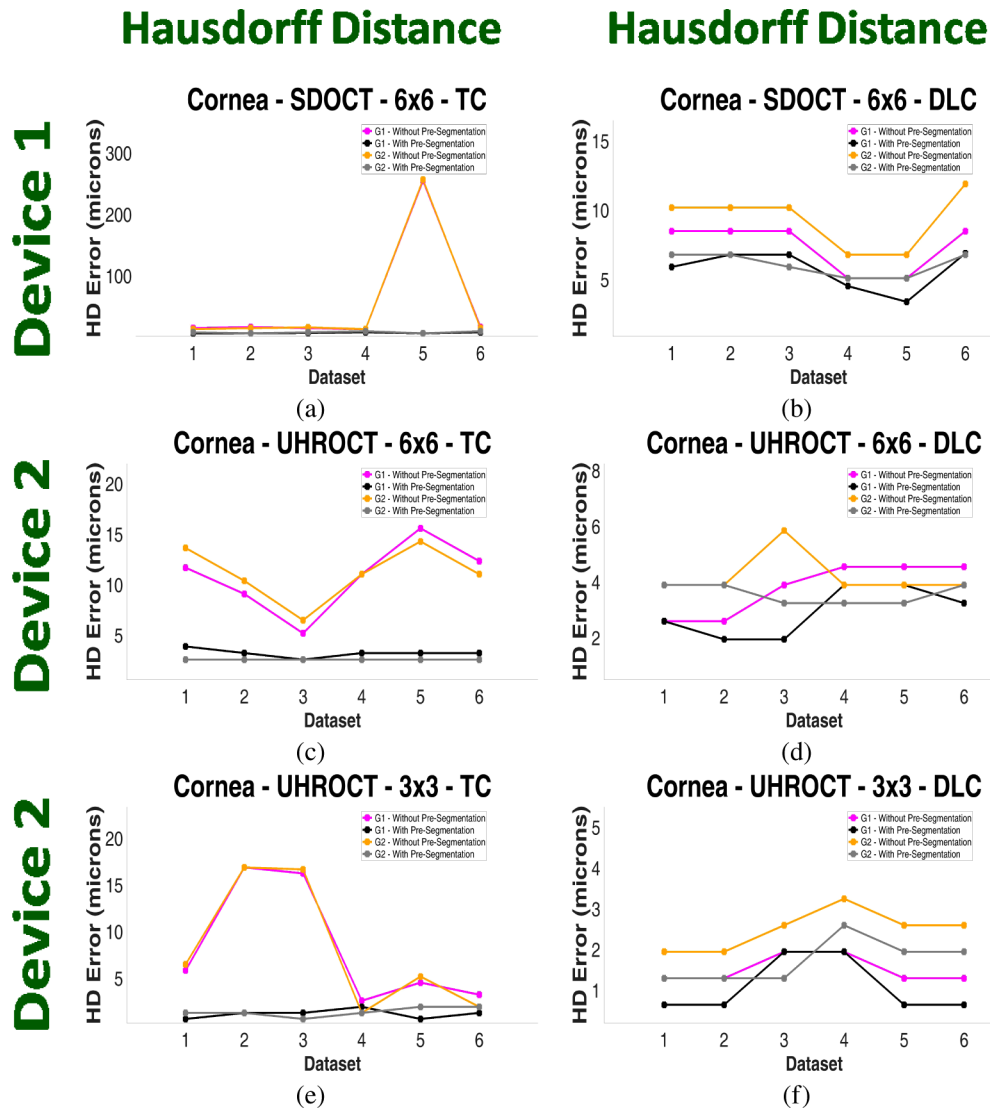
Hausdorff Distance

Hausdorff Distance

Device 1

Device 2

Device 2

Fig. 13. Quantitative estimation of the benefit of pre-segmenting the corneal OCT image. All the baselines were grouped into two categories: Traditional Comparison (TC; TWOPS vs TWPS), and Deep Learning Comparison (DLC; DLWOPS vs DLWPS). The first column corresponds to the former, and the second column corresponds to the latter. For each corneal test dataset, the image with the maximum HD error was found over all images in the sequence, and the image location in the sequence was stored. This was done only for the TWOPS and DLWOPS baselines respectively. The stored location indicies were then used to retrieve the corresponding HD errors from the TWPS and DLWPS baselines respectively. This procedure was repeated for each grader and plotted. G1 : without pre-segmentation (purple curve), with pre-segmentation (black curve). G2 : without pre-segmentation (yellow curve), with pre-segmentation (gray curve).
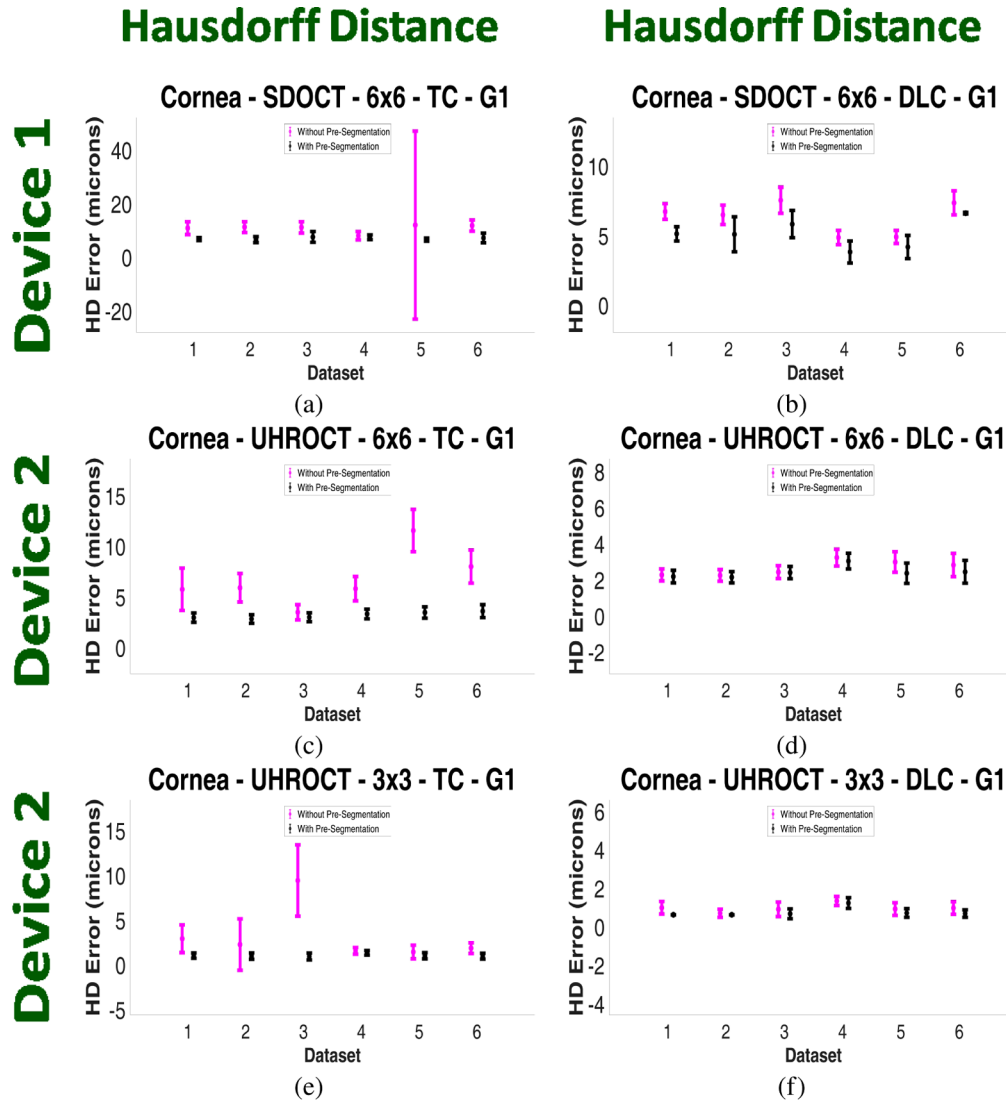
Fig. 14. Mean (circle) and Standard Deviation (error bars) of the Hausdorff Distance error for corneal datasets acquired with Devices 1 and 2 respectively. All the baselines were grouped into two categories: Traditional Comparison (TC; TWOPS vs TWPS), and Deep Learning Comparison (DLC; DLWOPS vs DLWPS). The mean and standard deviation error bars are shown only for the expert grader (G1) annotation. The With Out Pre-Segmentation (WOPS) results are shown by magenta error bars, while the With Pre-Segmentation (WPS) results are shown in black.
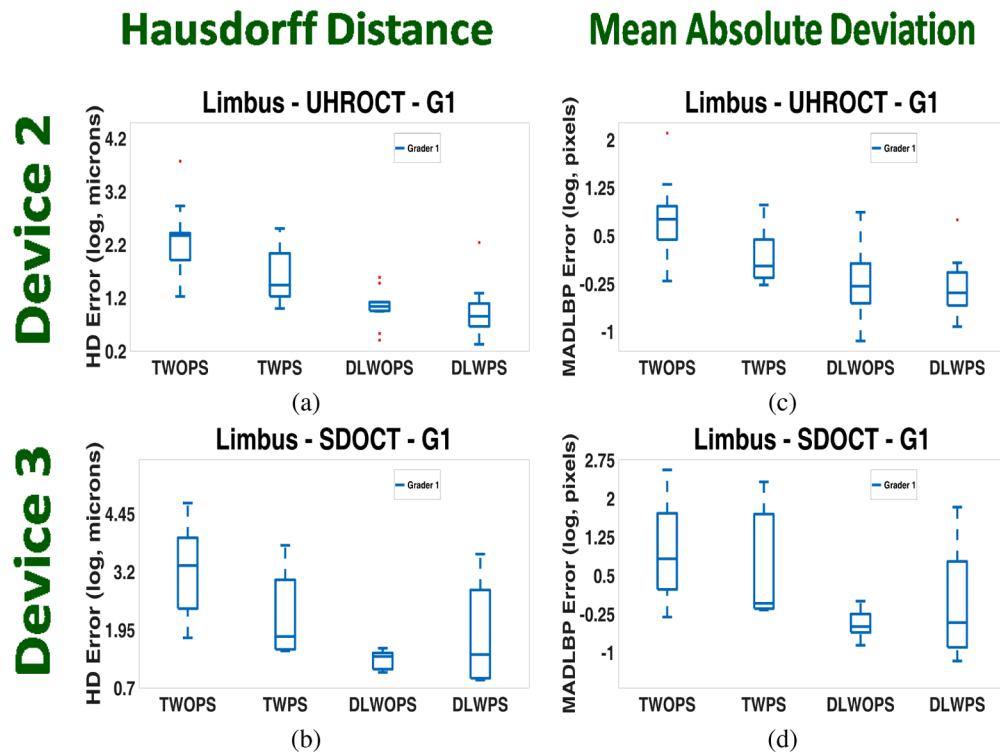
Fig. 15. (a)-(b) HD error and (c)-(d) MADLBP error comparison for the limbal datasets acquired with Devices 2 and 3 respectively. For the limbal datasets, the segmentation results obtained for each baseline method were contrasted exclusively against the expert annotations (G1). This graph plots the errors across all limbal datasets, including the failure cases. In contrast to Fig. 19, note the increased segmentation error in the DLWPS baseline due to imprecise pre-segmentations.
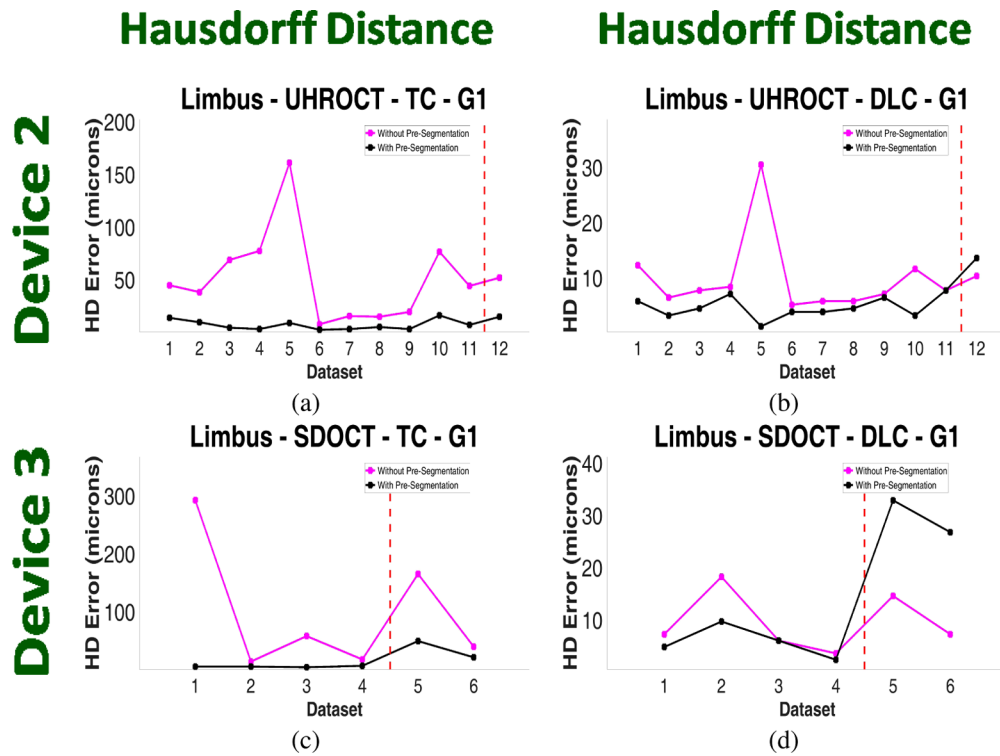
Fig. 16. Quantitative estimation of the benefit of pre-segmenting the limbal OCT image. All the baselines were grouped into two categories: TC (TWOPS vs TWPS), and DLC (DLWOPS vs DLWPS). The first column corresponds to the former, and the second column corresponds to the latter. For each test dataset, the image with the maximum HD error was found over all images in the sequence, and the image location in the sequence was stored. This was done only for the TWOPS and DLWOPS baselines respectively. The stored location indicies were then used to retrieve the corresponding HD errors from the TWPS and DLWPS baselines respectively. This procedure was done for only the expert grader and plotted. G1 : without pre-segmentation (purple curve), with pre-segmentation (black curve). Errors shown after red vertical line correspond to the failure cases of our approach.
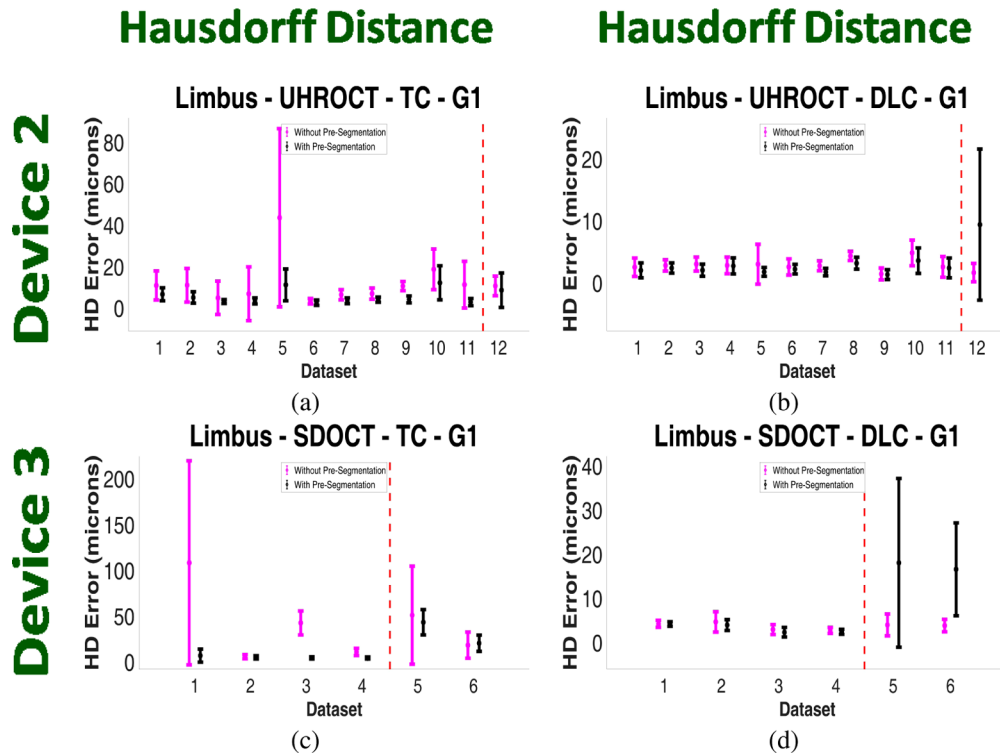
Fig. 17. Mean (circle) and Standard Deviation (error bars) of the Hausdorff Distance error for limbal datasets acquired with Devices 2 and 3 respectively. All the baselines were grouped into two categories: Traditional Comparison (TC; TWOPS vs TWPS), and Deep Learning Comparison (DLC; DLWOPS vs DLWPS). The mean and standard deviation error bars are shown only for the expert grader (G1) annotation. The With Out Pre-Segmentation (WOPS) results are shown by magenta error bars, while the With Pre-Segmentation (WPS) results are shown in black. Errors shown after red vertical line correspond to the failure cases of our approach.

4. Discussion

4.1. Segmentation accuracy of corneal interface

From the HD and MADLBP errors in Figs. 12, the error is worse for the TWOPS baseline method, where the traditional algorithm [12] used the original OCT image (without the pre-segmentation) to directly segment the interface. The hand-crafted features in this baseline algorithm failed to handle severe specular artifacts and noise patterns as seen in Fig. 2. In contrast, the TWPS baseline (hybrid framework), which uses the pre-segmented image instead of the original OCT image, produced a lower segmentation error. This observation is reflected in Figs. 14(a), 14(c) and 14(e). To quantify these observations, a paired t-test between the TWOPS and TWPS baselines was computed for each error metric, and we estimated that the results were statistically significant (p_{HD} = 4.2747e-05, p_{MADLBP} = 1.2859e-05). From these results, we concluded that the traditional algorithm fared better in the hybrid framework when the pre-segmented OCT image was used to segment the corneal tissue interface.

The DLWOPS baseline in Fig. 12 had lower HD and MADLBP errors than the TWPS baseline for the expert grader annotations. However, the errors were higher for the trained grader especially on the 3×3mm datasets from Device 2, as seen in Figs. 12(c) and 12(f), due to the large inter-grader variability. On the other hand, our DLWPS baseline approach, which used the pre-segmented image, fared better in contrast to the other three baselines. This observation is reinforced in Figs. 14(b), 14(d) and 14(f). Again, we computed paired t-tests between the DLWPS approach and all other baselines to determine the improvement in segmentation accuracy for each error metric. From the p-values in Table. 1 and Fig. 12, the cascaded framework generated results that were an improvement upon the other baselines, and indicated statistically significant results across all corneal datasets (p < 0.05).

Table 1. Statistical significance between our cascaded framework (DLWPS) against each baseline method for all the corneal datasets from Devices 1 and 2.

	TWOPS	TWPS	DLWOPS
p_{HD}	5.1929e-06	2.2079e-04	5.1454e-04
p_{MADLBP}	2.6848e-06	1.9264e-04	2.0734e-04

We also wanted to determine the improvement in segmentation accuracy on an per-image basis in each of the corneal test datasets. To do so, we first grouped the baselines into two categories: only traditional image analysis-based approaches (TWOPS vs. TWPS), and only deep learning-based approaches (DLWOPS vs. DLWPS). Next, we searched for the image in each corneal dataset that had the maximum HD error over all images in that dataset, and noted its index in the sequence. This was done only for the TWOPS and DLWOPS baselines respectively, and we plotted these maximum HD errors for each grader in Fig. 13 (purple and yellow colored curves). Then, we queried the errors for the same images (using the image indicies) in the TWPS and DLWPS baseline approaches respectively, and plotted the corresponding HD errors for each grader in Fig. 13 (black and gray curves). From Fig. 13, we noted that the baselines incorporating the pre-segemented OCT image performed better than one that did not include the pre-segmentation - the error bars in Fig. 14 support our observations. The pre-segmentation always improved the segmentation performance of the traditional image-analysis based approach when incorporated into a hybrid framework, and also improved the accuracy of a deep learning-based approach in a majority of corneal datasets when used in the cascaded framework. This statement is supported by the mean and standard deviation errors in This observation is reflected in Figs. 14(a), 14(c) and 14(e). This quantitatively attests to the benefit of utilizing the pre-segmented OCT image as part of a segmentation framework.

4.2. Segmentation accuracy of limbal interface

We plotted the segmentation error for the baseline methods executed on limbal datasets in Figs. 15, 16 and 19. In Fig. 15, we plotted the errors across all limbal test datasets, including the instances when the cascaded and hybrid frameworks failed to accurately segment the shallowest interface. In Fig. 19, we plot the errors only for the successful instances of interface segmentation. From Figs. 15 and 19, the error for the TWOPS baseline is the worst amongst all baselines as it failed to handle strong specular artifacts and severe speckle noise. This can also be seen from the mean and standard deviation errors across all images in each limbal dataset in Figs. 17(a) and 17(c). On the other hand, the TWPS baseline fared better with lower errors than the TWOPS baseline. Similar to Sec. 4.1, we also assessed the improvement in segmentation accuracy on a per-image basis for each of the 18 limbal datasets. We plotted these errors in Fig. 16. From the errors (after the red vertical dashed line) in Figs. 16(a), 16(c), 17(a), and 17(c), the hybrid framework (TWPS baseline) was able to reduce the segmentation error even with an incorrect OCT image pre-segmentation in a majority of datasets. Therefore, the incorporation of the pre-segmented OCT image in the hybrid framework lead to lower errors for the traditional image analysis-based approach in a majority of limbal datasets.

The DLWOPS baseline had lower errors as shown in Figs. 15 and 19 as compared to the TWOPS and TWPS baselines. This can also be seen from the mean and standard deviation segmentation errors across all images in each dataset in Fig. 17. But, at an image level, it sometimes yielded higher segmentation errors as seen in Figs. 16(b) and 16(d). On the other hand, the DLWPS baseline (cascaded framework) improved the segmentation error in a majority of the datasets, with the exception of three datasets, which are our failure cases. As shown in Fig. 18, two datasets presented with saturated tissue regions, which were washed out by specular artifacts. Another dataset contained regions where the interface was barely visible due to being obfuscated by speckle noise of the same amplitude. Due to these reasons, the incorrect pre-segmented OCT image degraded the segmentation performance of the TISN. Consequently, the segmentation error of the TWPS (hybrid framework) and DLWPS (cascaded framework) baselines was increased. As seen in Fig. 16 and Fig. 17 (after the red vertical dashed line), the DLWOPS baseline performed the best among all other baselines for these datasets.

We expound on the aforementioned reasons for segmentation failure. First, the contextual information available to the cGAN to remove the speckle noise patterns and specular artifacts is hindered when the pixel intensities on the tissue interface are either washed out due to saturation of the line scan camera [11,12,29] as shown in Fig. 18(a) (top two rows), or blend in with the background and specular artifacts of the same amplitude [11] as seen in Fig. 18(a) (bottom). In such outlier cases, the boundary becomes difficult to delineate across multiple scales through downsampling and upsampling operations in the encoder and decoder blocks, such that even the dilated convolutions and dense connections employed in the network are insufficient to recover context from surrounding boundary regions when localizing the interface.

Second, the TISN over-relied on the pre-segmentation in order to generate the final segmenation. During training of the TISN, the original image was coupled with the gold standard pre-segmentation output (see Fig. 8) into a two-channel input. The TISN learned that the tissue boundary in the gold standard pre-segmentation was the location of the start of the true boundary. However, the TISN was not trained with gold standard pre-segmented images that were artificially induced to be corrupted and noisy, such as the images shown in Fig. 18(b). Hence, the performance of the TISN on such incorrectly pre-segmented OCT images is poor.

One way to address this issue is to re-train the framework with gold standard pre-segmentations that have corrupted boundaries. In this pilot work, we did not introduce any corruption to the gold standard pre-segmentation used during training as we wanted to directly measure the performance of the TISN when provided with a pre-segmentation from the cGAN (without regard to any imprecise pre-segmentation). Another option is to exploit the temporal correlation between
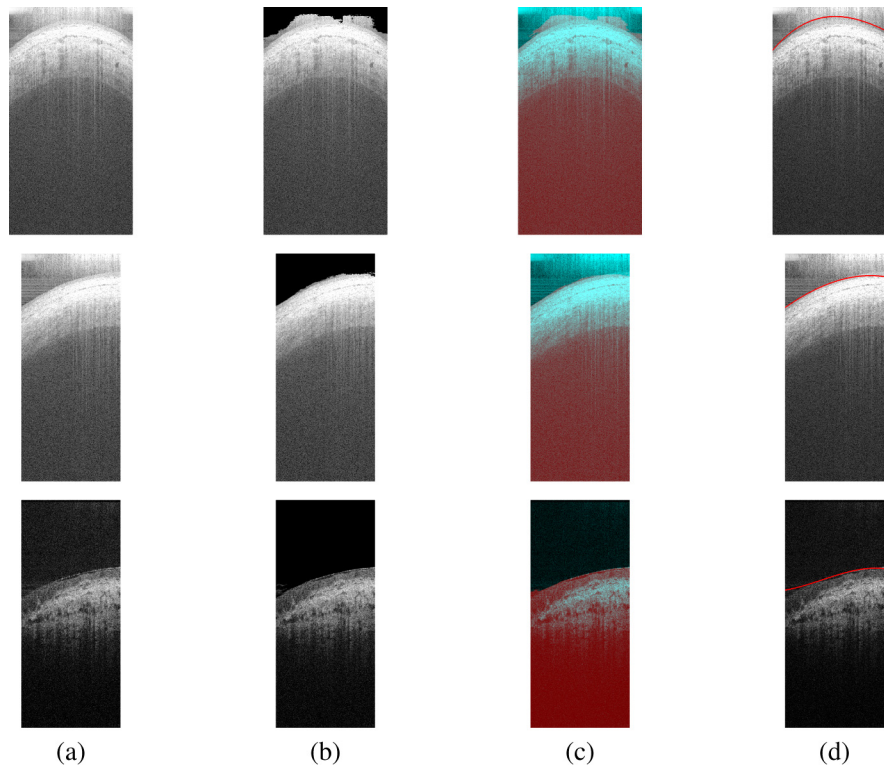
(a)   (b)   (c)   (d)

Fig. 18. Failure cases of our cascaded framework on three challenging limbal OCT datasets. Columns from left to right: (a) Original B-scans in the limbal OCT volumes, (b) cGAN pre-segmentation results that imprecisely removed speckle noise patterns and specular artifacts above the shallowest tissue interface, (c) The binary segmentation masks from the TISN overlaid in false color (red - foreground, turquoise - background) on the original B-scans, (d) Curve fit to the shallowest interface (red contour).

B-scans in the dataset through recurrent neural networks. We intend to pursue these ideas in our future work.

In this work, we set aside these three challenging failure cases, and estimated the improvement in segmentation accuracy across the remaining 15 limbus datasets. We conducted a paired t-test between the TWOPS and TWPS baselines for each error metric, and determined that our errors were statistically significant ($p_{HD} = 0.0471$, $p_{MADLBP} = 0.0313$). We also calculated paired t-tests between the DLWPS baseline and all other baselines to determine the statistical significance of our results for each error metric. As seen in Table. 2, our DLWPS cascaded framework generated statistically significant results ($p < 0.05$).

Table 2. Statistical significance between our cascaded framework (DLWPS) against each baseline method for 15 (out of 18) limbal datasets acquired from Devices 2 and 3.

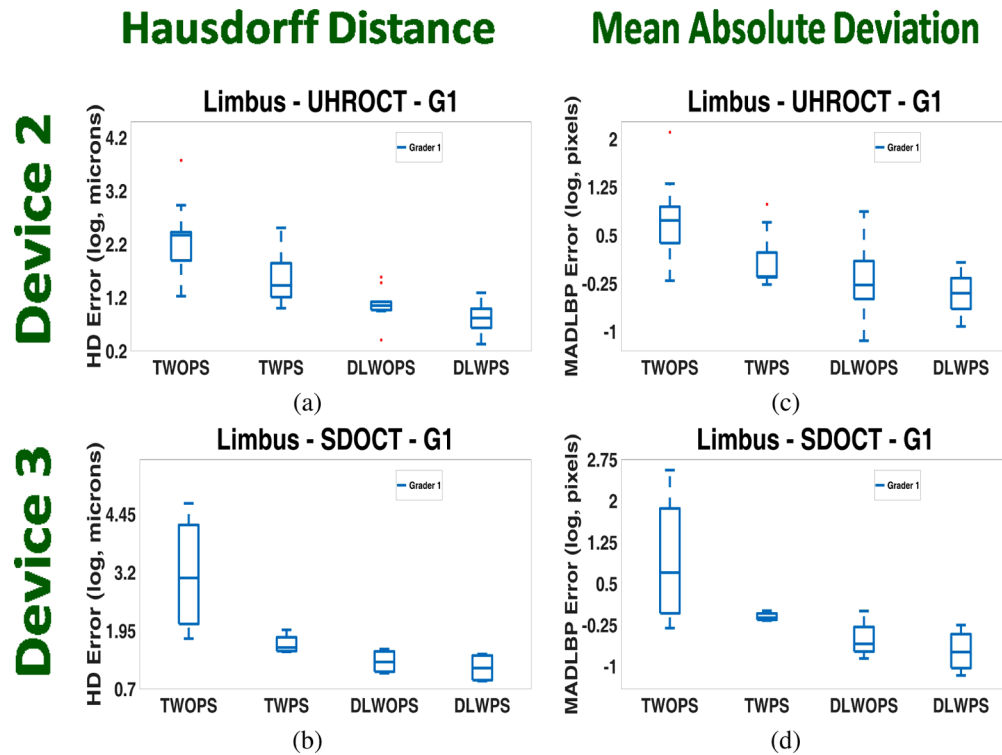|	TWOPS	TWPS	DLWOPS
p_{HD}	0.0240	0.0014	1.0335e-04
p_{MADLBP}	0.0126	0.0012	0.0344

Fig. 19. (a)-(b) HD error and (c)-(d) MADLBP error comparison for the limbal datasets acquired with Devices 2 and 3 respectively. For the limbal datasets, the segmentation results obtained for each baseline method were contrasted exclusively against the expert annotations (G1). These graphs plot errors for the successful segmentation results on 15 limbal test datasets.

4.3. Interface segmentation at limbal junction

During imaging of the limbal region, it is very common to acquire B-scans of the cornea and the limbus in the same dataset. This is because the scan pattern of the OCT scanner that is used to acquire the dataset will sometimes encompass sections of the limbus and the cornea. Bulk tissue motion between B-scans in a dataset is also customary during image acquisition. Therefore, it is crucial to capture the shallowest tissue interface of the limbus and the cornea as it enables distinguishing between these two distinct regions in the future. By correctly locating these interfaces, a registration algorithm can be used to potentially align regions at and below these interfaces, while compensating for bulk tissue motion. To the best of our knowledge, we believe our approach is the first to accurately detect the shallowest corneal and limbal interface in OCT images acquired at the limbal junction even in the presence of severe speckle noise patterns and specular artifacts. Results of our approach are shown in Fig. 20, wherein the shallowest interface is identified in B-scans that partially overlap both the cornea and the limbus.

4.4. Necessity of the cGAN

Most of the prior applications of a cGAN are for denoising/segmenting interfaces in retinal OCT images [33]. Moreover, a cGAN has never been applied to denoise/segment interfaces in corneal and limbal datasets. While we hoped that the cGAN could directly denoise with its original loss function, it was not the case and there is currently no published literature to this regard. As
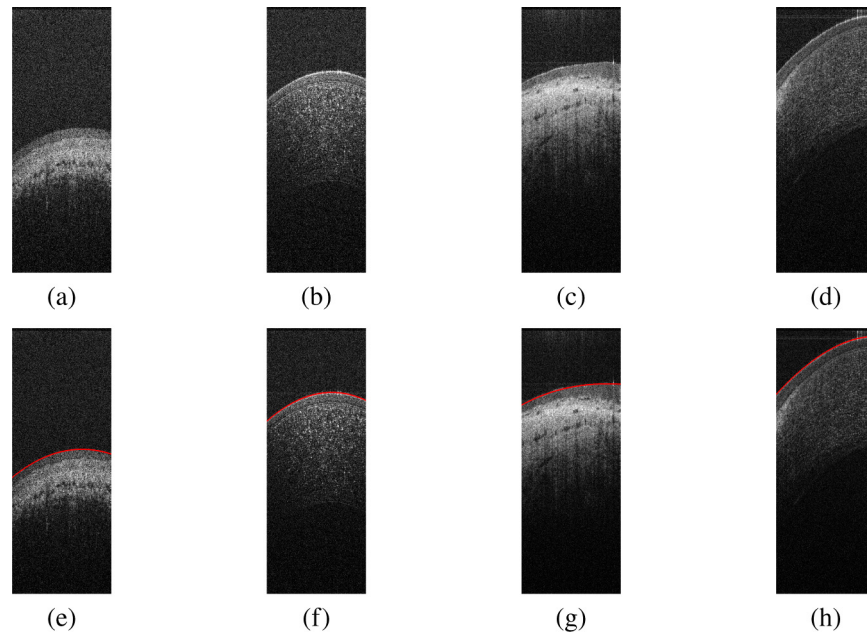
Fig. 20. Segmenting the shallowest tissue interface in OCT datasets, wherein the OCT scanner commenced imaging from the limbus and crossed over into the cornea, thereby encompassing the limbal junction. (a),(b) B-scans #1 and #300 in an OCT dataset corresponding to the limbus and the cornea respectively. (c),(d) B-scans #1 and #220 in a different OCT dataset corresponding to the limbus and the cornea respectively. (e),(f),(g),(h) Segmentation (red curve) of the shallowest tissue interface in images shown in (a),(b),(c) and (d) respectively. Note the partial overlap of the limbal (left) and corneal (right) region in the B-scan in (d), and the correct identification of the shallowest interface in (h).

mentioned in our paper, denoising would require alignment of the B-scans in order to generate gold standard denoised OCT images for training.

In this paper, we are the first to attempt segmentation of tissue structures in corneal and limbal OCT images using a cGAN by putting forward the novel notion of a (incomplete) pre-segmentation, which is further processed by a final segmentation algorithm. This final segmentation algorithm can either be a traditional segmentation algorithm (ex. Graph-based) or it could be another deep neural network, and this algorithm in turn benefits from the pre-segmentation.

The cGAN provides a pre-segmentation with the speckle noise and artifacts above the shallowest interface eliminated. It may be easy to use a simple thresholding algorithm to segment the corneal/limbal tissue interface at that point. But, corneal and limbal interfaces can be modeled with second/third/fourth order polynomials [1-5]. This is in contrast to retinal tissue interfaces as they cannot be easily modeled with curves since their boundary can be convex in certain regions and concave in other regions. It is for this reason that curve fitting is necessary to model corneal and limbal tissue interfaces.

Furthermore, after generating the pre-segmentation, we could have measured the similarity of the pre-segmentation to the original image. However, this similarity measure is subjective [33]. Instead, we wanted to quantitatively measure the performance of various segmentation algorithms (traditional and deep learning). This was also suggested in [33] as a better validation method for a "denoising" algorithm, and we have chosen this validation scheme.

Moreover, we wanted to measure the improvement in segmentation accuracy when a traditional or deep learning algorithm used the pre-segmentation instead of the original OCT B-scan.

Furthermore, it is easy to incorporate the pre-segmentation as an input to a traditional algorithm, but it is challenging to incorporate the pre-segmentation into a deep learning framework. This is the reason for the need of the Tissue Interface Segmentation Network (TISN). We detail this necessity in the next section.

4.5. *Choice of framework design*

In this work, we proposed to generate an intermediate representation of the OCT image, i.e. the cGAN pre-segmentation, that can influence the performance of a segmentation algorithm. To this end, we proposed a cascaded and hybrid segmentation framework. However, we theorized that there are other framework designs that can be implemented instead of the proposed approaches. Amongst other approaches, for example, we could have utilized a GAN directly for segmenting the tissue interface from the OCT image, or trained a multi-task neural network framework (CNN, GAN, etc.) to provide both the pre-segmentation and the final interface segmentation. We reiterate our motivations next that laid the groundwork for the proposed frameworks over the aforementioned design choices. As mentioned before, our motivations were: 1) To generate a pre-segmentation that could be utilized in a hybrid framework, 2) Integrate the pre-segmentation into the image acquisition pipeline of custom-built OCT scanners, and 3) Incorporate the pre-segmentation in a cascaded framework and compare its segmentation performance against that of a state-of-the-art CNN-based segmentation method [29].

Utilizing the GAN to directly yield the final interface segmentation does not provide an intermediate output, which can be integrated in a hybrid framework. Similarly, the multi-task framework would provide both the pre-segmented OCT image and the final interface segmentation. The pre-segmentation can be directly used in the hybrid framework and imaging pipeline respectively. However, the final segmentation may only be influenced by the shared weights of the multi-task network, and not by the pre-segmentation, which will be different from the final segmentation. Thus, if the pre-segmentation must influence the final interface segmentation (as it should), it may be necessary to train a new framework again (in a cascaded fashion with the multi-task framework) that would include the pre-segmentation. Therefore, we believe that our choice of framework design was warranted.

5. Conclusion

In this paper, we generated an intermediate OCT image representation which can influence the performance of a segmentation algorithm. The intermediate representation is a pre-segmentation, generated by a cGAN, wherein speckle noise patterns and specular artifacts are eliminated just prior to the shallowest tissue interface in the OCT image. We proposed two frameworks that incorporate the intermediate representation: a cascaded framework and a hybrid framework. The cascaded framework comprised of a cGAN and a TISN; the cGAN pre-segmented the OCT image by removing the undesired specular artifacts and speckle noise patterns that confounded boundary segmentation, while the TISN segmented the final tissue interface by combining the original image and the pre-segmentation as inputs. The hybrid framework contained an image analysis-based segmentation method, among other state-of-the-art methods, that exploited the cGAN pre-segmentation to generate the final tissue interface segmentation. The frameworks were trained on corneal and limbal datasets acquired from three different OCT scanners with different scan protocols. They were able to handle varying degrees of specular artifacts, speckle noise patterns, and bulk tissue motion, and deliver consistent segmentation results. We compared the results of our frameworks against those from the state-of-the-art image analysis-based and deep learning-based algorithms. To the best of our knowledge, this is the first approach for OCT-based tissue interface segmentation that integrated the cGAN component of our framework in a hybrid fashion. We have shown the benefit of pre-segmenting the OCT image through the lower segmentation errors that were yielded. Finally, we have shown the utility of our

algorithm in being able to segment the tissue interface at the limbal junction. We believe that the cGAN pre-segmentation output can be easily integrated into the image acquisition pipelines of custom-built OCT scanners.

Funding

Acknowledgements

Disclosures

International patent pending, US 62/860,392 & US 62/860,415. The authors declare that there are no other conflicts of interest related to this article.

References

1. D. Huang, E. Swanson, C. Lin, J. Schuman, W. Stinson, W. Chang, M. Hee, T. Flotte, K. Gregory, and C. Puliafito, *et al.*, "Optical coherence tomography," Science **254**(5035), 1178–1181 (1991).
2. J. G. Fujimoto, W. Drexler, J. S. Schuman, and C. K. Hitzenberger, "Isp focus issue: Optical coherence tomography (oct) in ophthalmology," Opt. Express **17**(5), 3978–3979 (2009).
3. J. A. Izatt, M. R. Hee, E. A. Swanson, C. P. Lin, D. Huang, J. S. Schuman, C. A. Puliafito, and J. G. Fujimoto, "Micrometer-scale resolution imaging of the anterior eye in vivo with optical coherence tomography," JAMA Ophthalmol. **112**(12), 1584–1589 (1994).
4. K. L. Lathrop, D. Gupta, L. Kagemann, J. S. Schuman, and N. SundarRaj, "Optical Coherence Tomography as a rapid, accurate, noncontact method of visualizing the palisades of vogt," Invest. Ophthalmol. Visual Sci. **53**(3), 1381–1387 (2012).
5. A. N. Kuo, R. P. McNabb, M. Zhao, F. LaRocca, S. S. Stinnett, S. Farsiu, and J. A. Izatt, "Corneal biometry from volumetric sdoct and comparison with existing clinical modalities," Biomed. Opt. Express **3**(6), 1279–1290 (2012).
6. N. Venkateswaran, A. Galor, J. Wang, and C. L. Karp, "Optical coherence tomography for ocular surface and corneal diseases: a review," Eye Vis. **5**(1), 13 (2018).
7. B. Keller, M. Draelos, G. Tang, S. Farsiu, A. N. Kuo, K. Hauser, and J. A. Izatt, "Real-time corneal segmentation and 3d needle tracking in intrasurgical oct," Biomed. Opt. Express **9**(6), 2716–2732 (2018).
8. K. Bizheva, N. Hutchings, L. Sorbara, A. A. Moayed, and T. Simpson, "In vivo volumetric imaging of the human corneo-scleral limbus with spectral domain oct," Biomed. Opt. Express **2**(7), 1794–1802 (2011).
9. K. Bizheva, B. Tan, B. MacLellan, Z. Hosseinaee, E. Mason, D. Hileeto, and L. Sorbara, "In-vivo imaging of the palisades of vogt and the limbal crypts with sub-micrometer axial resolution optical coherence tomography," Biomed. Opt. Express **8**(9), 4141–4151 (2017).
10. M. Haagdorens, J. Behaegel, J. Rozema, V. Van Gerwen, S. Michiels, S. Ní Dhubhghaill, M.-J. Tassignon, and N. Zakaria, "A method for quantifying limbal stem cell niches using oct imaging," Br. J. Ophthalmol. **101**(9), 1250–1255 (2017).
11. F. LaRocca, S. J. Chiu, R. P. McNabb, A. N. Kuo, J. A. Izatt, and S. Farsiu, "Robust automatic segmentation of corneal layer boundaries in sdoct images using graph theory and dynamic programming," Biomed. Opt. Express **2**(6), 1524–1538 (2011).
12. T. S. Mathai, J. Galeotti, and K. L. Lathrop, "Visualizing the palisades of vogt: limbal registration by surface segmentation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, (2018), pp. 1327–1331.
13. M. Szkulmowski, I. Gorczynska, D. Szlag, M. Sylwestrzak, A. Kowalczyk, and M. Wojtkowski, "Efficient reduction of speckle noise in optical coherence tomography," Opt. Express **20**(2), 1337–1359 (2012).
14. A. E. Desjardins, B. J. Vakoc, W. Y. Oh, S. M. R. Motaghiannezam, G. J. Tearney, and B. E. Bouma, "Angle-resolved optical coherence tomography with sequential angular selectivity for speckle reduction," Opt. Express **15**(10), 6200–6209 (2007).
15. M. Hughes, M. Spring, and A. Podoleanu, "Speckle noise reduction in optical coherence tomography of paint layers," Appl. Opt. **49**(1), 99–107 (2010).

16. D. C. Adler, T. H. Ko, and J. G. Fujimoto, "Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter," Opt. Lett. **29**(24), 2878–2880 (2004).

17. A. Ozcan, A. Bilenca, A. E. Desjardins, B. E. Bouma, and G. J. Tearney, "Speckle reduction in optical coherence tomography images using digital filtering," J. Opt. Soc. Am. A **24**(7), 1901–1910 (2007).

18. P. Puvanathasan and K. Bizheva, "Speckle noise reduction algorithm for optical coherence tomography based on interval type ii fuzzy set," Opt. Express **15**(24), 15747–15758 (2007).

19. M. Gargesha, M. W. Jenkins, A. M. Rollins, and D. L. Wilson, "Denoising and 4d visualization of oct images," Opt. Express **16**(16), 12313–12333 (2008).

20. S. Chitchian, M. A. Fiddy, and N. M. Fried, "Denoising during optical coherence tomography of the prostate nerves via wavelet shrinkage using dual-tree complex wavelet transform," J. Biomed. Opt. **14**(1), 014031 (2009).

21. Z. Hongwei, L. Baowang, and F. Juan, "Adaptive wavelet transformation for speckle reduction in optical coherence tomography images," in *2011 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, (2011), pp. 1–5.

22. M. Pircher, E. Götzinger, R. Leitgeb, A. F. Fercher, and C. K. Hitzenberger, "Measurement and imaging of water concentration in human cornea with differential absorption optical coherence tomography," Opt. Express **11**(18), 2190–2197 (2003).

23. J. Rogowska and M. E. Brezinski, "Image processing techniques for noise removal, enhancement and segmentation of cartilage OCT images," Phys. Med. Biol. **47**(4), 641–655 (2002).

24. D. L. Marks, T. S. Ralston, and S. A. Boppart, "Speckle reduction by i-divergence regularization in optical coherence tomography," J. Opt. Soc. Am. A **22**(11), 2366–2371 (2005).

25. A. Wong, A. Mishra, K. Bizheva, and D. A. Clausi, "General bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery," Opt. Express **18**(8), 8338–8352 (2010).

26. R. Bernardes, C. Maduro, P. Serranho, A. Araújo, S. Barbeiro, and J. Cunha-Vaz, "Improved adaptive complex diffusion despeckling filter," Opt. Express **18**(23), 24048–24059 (2010).

27. S. Moon, S. W. Lee, and Z. Chen, "Reference spectrum extraction and fixed-pattern noise removal in optical coherence tomography," Opt. Express **18**(24), 24395–24404 (2010).

28. S. Vergnole, G. Lamouche, and M. L. Dufour, "Artifact removal in fourier-domain optical coherence tomography with a piezoelectric fiber stretcher," Opt. Lett. **33**(7), 732–734 (2008).

29. T. S. Mathai, K. L. Lathrop, and J. Galeotti, "Learning to segment corneal tissue interfaces in oct images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, (2019), pp. 1432–1436.

30. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds. (Curran Associates, Inc., 2014), pp. 2672–2680.

31. P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), pp. 5967–5976.

32. A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," CoRR (2016).

33. Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, "Speckle noise reduction in optical coherence tomography images based on edge-sensitive cgan," Biomed. Opt. Express **9**(11), 5129–5146 (2018).

34. B. R. Davidson and J. K. Barton, "Application of optical coherence tomography to automated contact lens metrology," J. Biomed. Opt. **15**(1), 016009 (2010).

35. M. Shen, L. Cui, M. Li, D. Zhu, M. Wang, and J. Wang, "Extended scan depth optical coherence tomography for evaluating ocular surface shape," J. Biomed. Opt. **16**(5), 056007 (2011).

36. D. C. Fernández, H. M. Salinas, and C. A. Puliafito, "Automated detection of retinal layer structures on optical coherence tomography images," Opt. Express **13**(25), 10200–10216 (2005).

37. H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, and J. S. Schuman, "Macular segmentation with optical coherence tomography," Invest. Ophthalmol. Visual Sci. **46**(6), 2012–2017 (2005).

38. T. Fabritius, S. Makita, M. Miura, R. Myllylä, and Y. Yasuno, "Automated segmentation of the macula by optical coherence tomography," Opt. Express **17**(18), 15659–15669 (2009).

39. K. Li, X. Wu, D. Chen, and M. Sonka, "Optimal surface segmentation in volumetric images-a graph-theoretic approach," IEEE Trans. Pattern Anal. Mach. Intell. **28**(1), 119–134 (2006).

40. P. Dufour, L. Ceklic, H. Abdillahi, S. Schroeder, S. De Zanet, U. Wolf-Schnurrbusch, and J. Kowal, "Graph-based multi-surface segmentation of oct data using trained hard and soft constraints," IEEE Trans. Med. Imaging **32**(3), 531–543 (2013).

41. A. Shah, J. Bai, Z. Hu, S. Sadda, and X. Wu, "Multiple surface segmentation using truncated convex priors," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, (Springer International Publishing, 2015), pp. 97–104.

42. Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-d image segmentation," Int. J. Comput. Vis. **70**(2), 109–131 (2006).

43. M. Garvin, M. Abramoff, X. Wu, S. Russell, T. L. Burns, and M. Sonka, "Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images," IEEE Trans. Med. Imaging **28**(9), 1436–1447 (2009).

44. F. Shi, X. Chen, H. Zhao, W. Zhu, D. Xiang, E. Gao, M. Sonka, and H. Chen, "Automated 3-d retinal layer segmentation of macular optical coherence tomography images with serous pigment epithelial detachments," IEEE Trans. Med. Imaging **34**(2), 441–452 (2015).

45. K. Lee, M. Niemeijer, M. K. Garvin, Y. H. Kwon, M. Sonka, and M. D. Abramoff, "Segmentation of the optic disc in 3-d oct scans of the optic nerve head," IEEE Trans. Med. Imaging **29**(1), 159–168 (2010).

46. Q. Song, J. Bai, M. K. Garvin, M. Sonka, J. M. Buatti, and X. Wu, "Optimal multiple surface segmentation with shape and context priors," IEEE Trans. Med. Imaging **32**(2), 376–386 (2013).

47. A. Shah, J. Wang, M. K. Garvin, M. Sonka, and X. Wu, "Automated surface segmentation of internal limiting membrane in spectral-domain optical coherence tomography volumes with a deep cup using a 3-d range expansion approach,", in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, (2014), pp. 1405–1408.

48. J. Tian, B. Varga, G. M. Somfai, W.-H. Lee, W. E. Smiddy, and D. C. DeBuc, "Real-time automatic segmentation of optical coherence tomography volume data of the macular region," PLoS One **10**(8), e0133908–20 (2015).

49. S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation," Opt. Express **18**(18), 19413–19428 (2010).

50. A. Yazdanpanah, G. Hamarneh, B. Smith, and M. Sarunic, "Intra-retinal layer segmentation in optical coherence tomography using an active contour approach," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, (Springer International Publishing, 2009), pp. 649–656.

51. S. Niu, L. de Sisternes, Q. Chen, T. Leng, and D. L. Rubin, "Automated geographic atrophy segmentation for sd-oct images using region-based c-v model via local similarity factor," Biomed. Opt. Express **7**(2), 581–600 (2016).

52. L. Sisternes, G. Jonna, J. Moss, M. F. Marmor, T. Leng, and D. Rubin, "Automated intraretinal segmentation of sd-oct images in normal and age-related macular degeneration eyes," Biomed. Opt. Express **8**(3), 1926 (2017).

53. A. Lang, A. Carass, M. Hauser, E. S. Sotirchos, P. A. Calabresi, H. S. Ying, and J. L. Prince, "Retinal layer segmentation of macular oct images using boundary classification," Biomed. Opt. Express **4**(7), 1133–1152 (2013).

54. Z. Ma, J. M. R. S. Tavares, and R. M. N. Jorge, "A review on the current segmentation algorithms for medical images," in *IMAGAPP*, (2009).

55. R. Kafieh, H. Rabbani, and S. Kermani, "A review of algorithms for segmentation of optical coherence tomography from retina," J. medical signals sensors **3**, 45–60 (2013).

56. B. J. Antony, M. D. Abràmoff, M. M. Harper, W. Jeong, E. H. Sohn, Y. H. Kwon, R. Kardon, and M. K. Garvin, "A combined machine-learning and graph-based framework for the segmentation of retinal surfaces in sd-oct volumes," Biomed. Opt. Express **4**(12), 2712–2728 (2013).

57. 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).

58. M. Chen, J. Wang, I. Oguz, B. L. VanderBeek, and J. C. Gee, "Automated segmentation of the choroid in edi-oct images with retinal pathology using convolution neural networks," in *Fetal, Infant and Ophthalmic Medical Image Analysis*, (Springer International Publishing, 2017), pp. 177–184.

59. 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).

60. F. G. Venhuizen, B. van Ginneken, B. Liefers, M. J. 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).

61. 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).

62. A. Shah, M. D. Abramoff, and X. Wu, "Simultaneous multiple surface segmentation using deep learning," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (Springer International Publishing, 2017), pp. 3–11.

63. C. S. Lee, A. J. Tyring, N. P. Deruyter, Y. Wu, A. Rokem, and A. Y. Lee, "Deep-learning based, automated segmentation of macular edema in optical coherence tomography," Biomed. Opt. Express **8**(7), 3440–3448 (2017).

64. L. Ge, M. Shen, A. Tao, J. Wang, G. Dou, and F. P. S. Lu, "Automatic segmentation of the central epithelium imaged with three optical coherence tomography devices," Eye & contact lens **38**(3), 150–157 (2012).

65. D. Williams, Y. Zheng, F. Bao, and A. Elsheikh, "Automatic segmentation of anterior segment optical coherence tomography images," J. Biomed. Opt. **18**(5), 056003 (2013).

66. Y. Li, R. Shekhar, and D. Huang, "Corneal pachymetry mapping with high-speed optical coherence tomography," Ophthalmology **113**(5), 792–799.e2 (2006).

67. D. Williams, Y. Zheng, P. G. Davey, F. Bao, M. Shen, and A. Elsheikh, "Reconstruction of 3d surface maps from anterior segment optical coherence tomography images using graph theory and genetic algorithms," Biomed. Signal Process. Control. **25**, 91–98 (2016).

68. H. Rabbani, R. Kafieh, M. K. Jahromi, S. Jorjandi, A. M. Dehnavi, F. Hajizadeh, and A.-R. Peyman, "Obtaining thickness maps of corneal layers using the optimal algorithm for intracorneal layer segmentation," Int. J. Biomed. Imaging **2016**, 1–11 (2016).

69. M. K. Jahromi, R. Kafieh, H. Rabbani, A. M. Dehnavi, A. Peyman, F. Hajizadeh, and M. Ommani, "An automatic algorithm for segmentation of the boundaries of corneal layers in optical coherence tomography images using gaussian mixture model," J. medical signals sensors **4**, 171–180 (2014).

70. T. Schmoll, A. Unterhuber, C. Kolbitsch, T. D. Le, A. D. Stingl, and R. A. Leitgeb, "Precise thickness measurements of bowman's layer, epithelium, and tear film," Optom. Vision Science **89**(5), E795–E802 (2012).

71. T. Zhang, A. Elazab, X. Wang, F. Jia, J. Wu, G. Li, and Q. Hu, "A novel technique for robust and fast segmentation of corneal layer interfaces based on spectral-domain optical coherence tomography imaging," IEEE Access **5**, 10352–10363 (2017).

72. S. Apostolopoulos, S. De Zanet, C. Ciller, S. Wolf, and R. Sznitman, "Pathological oct retinal layer segmentation using branch residual u-shape networks," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, eds. (Springer International Publishing, 2017), pp. 294–301.

73. V. A. dos Santos, L. Schmetterer, H. Stegmann, M. Pfister, A. Messner, G. Schmidinger, G. Garhofer, and R. M. Werkmeister, "Corneanet: fast segmentation of cornea oct scans of healthy and keratoconic eyes using deep learning," Biomed. Opt. Express **10**(2), 622–641 (2019).

74. 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), pp. 234–241.

75. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), pp. 770–778.

76. G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), pp. 2261–2269.

77. F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *ICLR*, (2016).

78. 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).

79. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015), pp. 1–9.

80. B. Hariharan, P. A. Arbeláez, R. B. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015), pp. 447–456.

81. H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, (2015), pp. 1520–1528.

82. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015), pp. 3431–3440.

83. A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," Distill **1**(10), 00003 (2016).

84. N. Khosravan and U. Bagci, "S4ND: single-shot single-scale lung nodule detection," CoRR **abs/1805.02279** (2018).

85. T. S. Mathai, J. Galeotti, S. Horvath, and G. Stetten, "Graphics processor unit (gpu) accelerated shallow transparent layer detection in optical coherence tomographic (oct) images for real-time corneal surgical guidance", in *Augmented Environments for Computer-Assisted Interventions*, (Springer International Publishing, 2014), pp. 1–13.

86. W. Cleveland, "Lowess: A program for smoothing scatterplots by robust locally weighted regression," Am. Stat. **35**(1), 54 (1981).

87. B. Wang, L. Kagemann, J. S. Schuman, H. Ishikawa, R. A. Bilonick, Y. Ling, I. A. Sigal, Z. Nadler, A. Francis, M. G. Sandrian, and G. Wollstein, "Gold nanorods as a contrast agent for doppler optical coherence tomography," PLoS One **9**(3), e90690 (2014).

88. V. J. Srinivasan, M. Wojtkowski, A. J. Witkin, J. S. Duker, T. H. Ko, M. Carvalho, J. S. Schuman, A. Kowalczyk, and J. 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).

89. Leica, "Leica envisu C2300 system specifications," https://www.leica-microsystems.com/fileadmin/downloads/Envisu%20C2300/Brochures/Envisu_C2300_EBrochure_2017_en.pdf.

90. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," Int. J. Comput. Vis. **115**(3), 211–252 (2015).

91. P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, (2003), pp. 958–963.

92. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," CoRR (2015).

93. F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, (2016), pp. 565–571.

94. M. Felsberg and G. Sommer, "The monogenic signal," IEEE Trans. Acoust., Speech, Signal Process. **49**(12), 3136–3144 (2001).