Analysis of Spatio-Temporally Varying Features in Optical Coherence Tomographic (OCT) and Ultrasound (US) Image Sequences

Tejas Sudharshan Mathai

May 2019

CMU-RI-TR-19-39

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Thesis Committee:
John Galeotti, Chair
George Stetten,
Ruslan Salakhutdinov
Ajay Gopinath, Abbott

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Robotics

Abstract

Optical Coherence Tomography (OCT) and Ultrasound (US) are non-ionizing and non-invasive imaging modalities that are clinically used to visualize anatomical structures in the body. OCT has been widely adopted in clinical ophthalmology due to its micron-scale resolution to visualize in-vivo structures of the eye. Ultra-High Frequency Ultrasound (UHFUS) captures images of tissue at a depth of $\sim$1cm with 30 micron resolution.

One key application area for OCT is assessing the stem cell distribution residing inside the Palisades of Vogt (POV) in the limbus. The limbus is located at the intersection of the clear cornea and the white sclera in the eye. Another application area is identifying corneal tissue interfaces for surgical procedures, such as Laser In-Situ Keratomileusis (LASIK). As for UHFUS, the key application in this thesis is vascular measurements, including monitoring for possible chronic rejection of hand transplants.

Achieving higher resolutions with OCT and UHFUS increases the speckle noise during imaging due to smaller resolution cells. In addition to the speckle noise present at every pixel, other localized imaging artifacts, such as shadowing and specular saturation artifacts, substantially diminish the visibility of tissue interface boundaries. These boundaries and edges are crucial for diagnosing particular pathological conditions or diseases, and for developing a pre-operative surgical plan. Moreover, abrupt tissue motion obfuscates the analysis of image sequences, including volumetric sweeps. An ideal approach to circumventing these issues would be validated across multiple devices and acquisition parameters, produce measurements that are clinically/surgically relevant, and be amenable to real-time implementation.

The main contribution of this thesis is in developing classical computer vision-based and deep learning-based approaches to address the problems faced in these clinical application areas. We have also developed a hybrid framework that removes OCT speckle noise from the air gap using a learning-based method, thereby dramatically improving the performance of classical approaches in situations where they previously failed. The presented classical and learned approaches all utilize consumer Graphics Processor Units (GPUs) for high-speed parallel computation to advance the clinical capabilities of OCT and UHFUS.

In this thesis, we have enabled visualizing changes in tissue structures over time and have begun quantifying some of these changes as well. Our improved OCT volumetric re-

constructions allow us to extract cross-sectional images of the Palisades of Vogt (POV) from OCT sequences acquired by different OCT scanners with different scan settings to quantify the POV structural changes over time. We have designed state-of-the-art deep learning-based methods for scanner-agnostic segmentation of multiple corneal tissue interfaces. We have developed the first artificial intelligence system for measuring the intimal vessel wall thickness in human subjects imaged using UHFUS. Our methods have identified salient regions of image sequences for additional analysis of application-specific features and quantitative metrics.

Acknowledgements

My advisor, John Galeotti; My mentor, George Stetten; Other knowledgeable people: Ruslan Salakhutdinov, Ajay Gopinath, Kira Lathrop, Vijay Gorantla, My parents, Kusuma Mathai and Mathai Thomas; My sister, Sheetal Mathai; Haewon Jeong; Divya Kothandapani and family; My lab mates: Samantha, Vikas, Jihang, Randy, Bruce, Mary, Jiahong; My friends from school and undergrad: Brijesh, JC, Rathi, Rawoof, Ram, Raunak, Siddharth, Varun, Arnold; My friends at Carnegie Mellon: Srinivas, Sibi, Ming, Zhi, Chris, Shaurya, Arun, Elena, Lucas, Lair, Nirajan, Dragan, Ale, Sara, Drew, Sylwia, Jay, Laura, Arvind, Emily, Vince, Swati and family, Trevor. Finally, The Robotics Institute and Carnegie Mellon University for providing me with the best experiences in my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Various non-invasive imaging modalities are used by clinicians to determine the existence of pathological conditions and diseases. Aside from cameras (which image surfaces) and X-rays/fluoroscopes (which integrate projections through tissue), most medical imaging modalities are able to image at multiple distinct depths within tissue. These tomographic imaging modalities include Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Optical Coherence Tomography (OCT), and Ultrasound (US). However, only MRI, OCT and US are considered safe and non-ionizing. Compared to MRI, OCT and US imaging systems are substantially smaller, cheaper, and capable of more rapid image acquisition, including video-rate imaging. Thus, OCT and US are uniquely suited for widespread tomographic imaging for both diagnosis and for real-time feedback in surgical guidance.

The key motivation of this thesis work is to advance the clinical capabilities of OCT and

ultrasound. We have identified key application areas for each modality, which are amenable to the approaches that have been developed in this thesis. We provide a brief description of these applications below:

1. **Application 1:** Identifying corneal interfaces to enhance corneal surgical procedures using OCT

    - Corneal interfaces need to be identified for the measurement of clinical parameters, such as corneal refractive power/index.
    - These measurements are needed for corneal surgical procedures, such as Laser In-Situ Keratomileusis (LASIK) and Deep Anterior Lamellar Keratoplasty (DALK).

2. **Application 2:** Assessing stem cell deficiency using OCT

    - Stem cells reside in a region of the limbus, which is the intersection of the clear cornea and white sclera of the eye, called the Palisades of Vogt (POV).
    - Surgical procedures and/or certain types of diseases cause the stem cell population in these POV regions to change over time.
    - Quantifying the morphological changes in the POV over time aids in the assessment of limbal stem cell deficiency.

3. **Application 3:** Monitoring and tracking vessels using ultrasound

    - Measurement of vessel-based clinical parameters is an important indicator of risk in different applications, such as hand-transplant chronic rejection and cardiovascular risk. Segmentation and 3D visualization of the vessel contours enables the quantifiable and repeatable measurement of vascular parameters.

- Vessel-based measures, such as lumen diameter, intima-media thickness and intimal wall thickness, are crucial to determination of vessel wall hardening (chronic rejection), vascular blockage, and potential for future aneurysms.
- Automatic identification and tracking of vessels also enables future automated and semi-automated robotic approaches to inserting catheters for vascular access.

1.2 Contributions

The key contributions in this thesis are as follows:

1. *Segmentation and 3D visualization of multiple corneal tissue interfaces from OCT video sequences.*

    - **Classical approach:** A real-time image analysis algorithm has been developed to segment four corneal tissue interfaces in OCT sequences, using a classical (not automatically learned) approach.
    - **Deep learning approach:** A Convolutional Neural Network (CNN) titled *CorNet* has been designed for the segmentation of three corneal tissue interfaces in OCT sequences obtained from different OCT machines with different scan settings.
    - **Hybrid approach:** The deep learning framework provides an intermediate output, wherein the air-gap's speckle noise and specular artifacts are removed prior to the shallowest corneal interface. The intermediate output is utilized by the classical approach for segmentation of the shallowest corneal interface.

2. *Segmentation of the shallowest limbal interface and 3D visualization of limbal structures from OCT video sequences.*

    - **Classical approach:** A real-time algorithm (incorporating local phase analysis) segments the shallowest limbal interface in each image of an OCT sequence. Subsequently, the segmented interfaces are aligned to reconstruct a 3D view of the limbal region.
    - **Deep learning approach:** The *CorNet* CNN architecture was extended to segment the shallowest limbal interface in OCT sequences, and 3D reconstructed

views of the limbus were generated.

- **Hybrid approach:** The deep learning framework generates an intermediate output with the air-gap's speckle noise and specular artifacts removed just above to the shallowest limbal interface. This intermediate output is used by the classical algorithm in order to segment the shallowest limbal interface.

3. *Segmentation of vessel contours from ultrasound video sequences.*

   - **Classical approach:** A real-time algorithm (incorporating local phase analysis and a level-set) delineates vessel contours, and tracks them using an Extended Kalman Filter (EKF).
   - **Deep learning approach:** A Convolutional Long Short Term Memory (ConvLSTM) neural network robustly identifies the vessel contours in US video sequences acquired from different US machines with different scan settings.

1.3 Publications and Patents

We list the publications that have resulted as part of this thesis, some of which have led to invention disclosures and the possibility of one or more patents being filed by CMU.

1. MICCAI AECAI 2014 [1] - Corneal surface segmentation (classical method) in OCT images.

2. Methods 2017 [2] - Review of registration algorithms commonly employed in ultrasound.

3. ARVO 2017 (poster) - Limbal surface segmentation (classical method) in OCT images for POV cross-section visualization.

4. ISBI 2018 [3] - Limbal interface segmentation (classical method) in OCT images for POV cross-section visualization.

5. MICCAI 2018 [4] - Vessel contour segmentation and tracking (classical method) in ultrasound sequences.

6. ISBI 2019 [5] - Segmentation of corneal interfaces (deep learning method) in OCT sequences.

7. Biomedical Optics Express 2019 (submitted) [6] - Speckle noise pattern and specular artifact removal (deep learning method) in OCT sequences.

8. MICCAI 2019 (submitted) - Segmentation of vessel contours (deep learning method) in ultrasound sequences.

- Invention disclosures have been filed for all 2019 work as well as MICCAI 2018. CMU is seeking to file patent(s).

1.4 Organization

In this section, the organization of the thesis is described as follows:

1. In Section 2.1, we identify relevant clinical applications with open research problems that require new solutions.

2. In Section 2.2, we provide brief introductions to OCT and US, which are the two modalities utilized for imaging in these application areas.

3. In Section 2.3, we describe the unique problems encountered specifically within each application that require innovative solutions to solve them.

4. Following the introduction, in Chapter 3, prior work that has been undertaken to solve the challenging problems in the clinical application areas is described.

5. Finally, in Chapters 4, 5 and 6, we describe the classical and learning-based approaches that have been developed in this thesis to address the open research problems in each application area.

6. In Chapter 7, we provide a discussion of the future work that can be undertaken to further the ideas proposed in this thesis.

Chapter 2

Clinical Background

2.1 Clinical Background and Applications

2.1.1 Identification of Corneal Tissue Interfaces

The anterior segment of the eye is comprised of different structures, such as the cornea, sclera (white region of the eye), lens, iris, cilliary body, etc. There are many micro-structures that are inherently present inside these regions, which regulate and maintain the function of the eye. One of the major components of the eye that contributes to and sustains human vision is the cornea. As shown in Figs. 2.1 and 2.2, the cornea is the part of the eye that is transparent, and covers the lens, iris, pupil, and other anterior structures. It works in conjunction with the lens in the eye to refract light and focus it onto the retina. The cornea accounts for 65-75 % of the eye's total optical power [7].

The configuration of the cornea includes five different entities: the corneal epithelium, Bowman's layer, Stromal interface, Descemet's Membrane, and the corneal endothelium. These five layers are shown in Fig. 2.2. The refractive power of the cornea is facilitated by

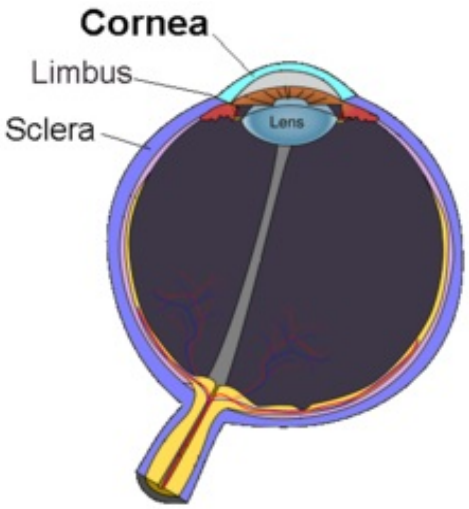


Figure 2.1: Cross-section of the eye. Figure by Mikael Haggstrom, used with permission.

two internal structures in the cornea: the epithelium and the endothelium. The shape of the corneal epithelium is the most significant contributing element to the optical power of the eye; it refracts the light entering the eye. The endothelium's shape plays a supporting role in focusing the refracted light through the lens onto the retina, where it is converted into electrical impulses that are sent to the brain to be processed [7].

However, the function of the cornea can deteriorate over time due to age, congenital conditions, infection, or due to injury etc. Changes in the curvature of the cornea can affect the refractive power of the cornea [8], and can lead to loss of visual acuity. The refractive power of the cornea can be measured using a technique called Keratometry. To accurately calculate the refractive power, the radii of curvature of the anterior (epithelium) and posterior (endothelium) surfaces of the cornea need to be estimated. The refractive power can then be quantified in Diopters, which is the inverse of the focal length of the cornea when the focal length is measured in meters.

The only way to measure the radii of curvature and corneal refractive power, before

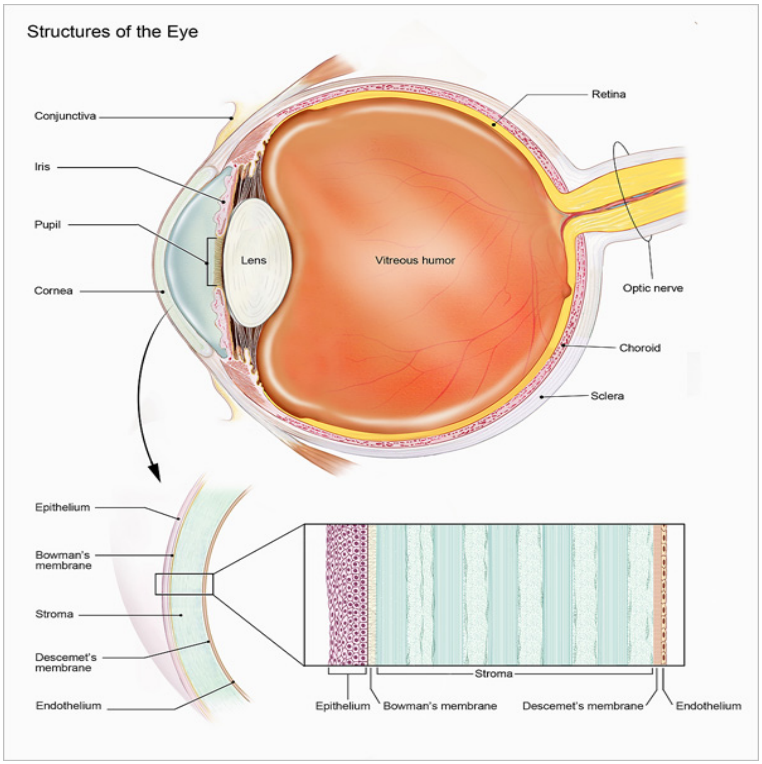


Figure 2.2: Corneal cross-section. Used with permission from the National Eye Institute (NEI), USA.

non-invasive imaging modalities were invented, was through corneal topography. Corneal topographers are now currently sold by many vendors (for example: Atlas 9000, Carl Zeiss), but they can only measure the radius of curvature of the anterior surface (epithelium) of the eye. To characterize the optical properties of the entire cornea, assumptions were made regarding the radius of curvature of the posterior surface (endothelium), and its contribution to corneal refractive power [8–10]. Utilizing the radius of curvature of the epithelium was satisfactory for the clinical outcomes of patients with normal corneas, who were undergoing surgery to restructure the cornea.

A particular method to compute the corneal refractive index is to use the thick lens equation to describe the refractive power of both the epithelium and endothelium:

$$\Phi_{EFL} = \frac{\eta_c - \eta_{air}}{r_{ant}} + \frac{\eta_{aq} - \eta_c}{r_{post}} - CCT\frac{(\eta_c - \eta_{air})(\eta_{aq} - \eta_c)}{\eta_c r_{ant} r_{post}} \tag{2.1}$$

$\eta_c = 1.376$, $\eta_{air} = 1.000$, and $\eta_{aq} = 1.336$ are the refractive indicies for the cornea, air, the aqueous humor in the eye respectively. CCT is the central corneal thickness, and r_{ant} and r_{post} are the radii of curvature of the anterior (epithelium) and posterior (endothelium) surfaces of the cornea respectively. The refractive index values are the Gullstrand values found in literature [11]. With the placido-disc corneal topography systems, only the anterior corneal surface is measured, with an assumption on a constant ratio that exists between the anterior and posterior corneal surfaces [8]:

$$r_{post} = k_c r_{ant} \tag{2.2}$$

Utilizing the Gullstrand model of the eye [8,11], the radius of curvature of the epithelium r_{ant} is 7.7mm, while the radius of curvature of the endothelium r_{post} is 6.8mm [8]. Thus, the constant k_c is 0.883. Gullstrand also provides a value of 0.5mm for the central corneal thickness (CCT) [8]. Substituting Eq. (2.2) into Eq. (2.1), we obtain:

$$\Phi_{EFL} = \frac{1}{r_{ant}}\left(0.376 - 0.0453 + \frac{0.00000618}{r_{ant}}\right) \tag{2.3}$$

As the central corneal thickness (CCT) value is small in relation to the corneal refractive power, a final assumption of 7.7mm for the epithelial radius of curvature is applied [8]:

$$\Phi_{EFL} = \frac{1}{r_{ant}}(0.3307 + 0.0008) \tag{2.4}$$

$$\Phi_{EFL} = \frac{0.3315}{r_{ant}} \tag{2.5}$$

$$\Phi_{EFL} = \frac{1.3315 - 1.0000}{r_{ant}} \tag{2.6}$$

$$\implies \Phi_{EFL} = \frac{\eta_k - \eta_{air}}{r_{ant}} \tag{2.7}$$

The calculated values for the Keratometric refractive index $\eta_k = 1.3315$ (and $\eta_{air} = 1.0000$) are based on a specific set of assumptions, the Gullstrand model of the eye, and the thick lens equation [8]. However, there are many vendors in the market who utilize different values for the Keratometric refractive index η_k, and at present, the ANSI standard specifies an η_k = 1.3375 [12]. This value was obtained by using the formula to calculate the back focal length of a lens model [13]:

$$\Phi_{BFL} = \frac{\eta_c r_{ant} \Phi_{EFL}}{(\eta_c r_{ant}) - CCT(\eta_c - \eta_{air})} \tag{2.8}$$

After substituting the values for Φ_{EFL} in Eq. (2.5), the $r_{ant} = 7.7$mm from the Gullstrand eye model, and the CCT value of 0.5 into Eq. (2.8), we end up with:

$$\Phi_{BFL} = \frac{0.3375}{r_{ant}} \tag{2.9}$$

$$\Phi_{BFL} = \frac{1.3375 - 1.0000}{r_{ant}} \tag{2.10}$$

$$\implies \Phi_{BFL} = \Phi_K = \frac{\eta_k - \eta_{air}}{r_{ant}} \tag{2.11}$$

From Eq. (2.11), we obtain the ANSI specified Keratometric Refractive Index (KRI) of η_k = 1.3375, and the corneal refractive power Φ_K. Utilizing KRI and the radius of curvature of the anterior surface (epithelium), the corneal refractive power can be measured, and this is the current standard of care clinically. This measurement was satisfactory for the clinical outcomes of patients with normal corneas, who were undergoing surgery to restructure the cornea. But, this model did not hold for patients who have already had corneal restructuring surgical procedures, such as those who have undergone Laser Refractive Surgery (LRS). For these patients who underwent LRS first, then underwent additional surgery, such as cataract surgery, the outcome of the surgery was erroneous and undesired [14].

Desired Clinical Measurements

There is an urgent need to develop tomographic imaging modalities that can measure both the anterior and posterior surface curvatures without a relationship between the two.

LRS is a surgical procedure that is used to help patients reduce the need for corrective eyewear, and it is one of the most common forms of outpatient surgeries. 7 million people in the United States have already undergone some form of LRS, such as Laser In-Situ Keratomileusis (LASIK), Photo-Refractive Keratectomy (PRK) etc [15], and at least 700,000 patients will undergo LASIK to reduce their dependence on glasses. It is typically performed on adults between the ages of 20-40 years old, and they typically have 20/20 uncorrected vision on average after LRS [16–18].

Even after these patients have undergone LRS, deterioration of vision with age will result in the lens of the eye becoming cloudy, and cataract surgery will be necessitated. From a longitudinal study from 1995 to 2002 [19], it was estimated that the annual rate of cataract surgery for individuals older than 62 years of age was 5.3%. Thus, there will be nearly

370,000 patients every year who will have cataracts in atleast one eye.

Cataract surgery is the surgical procedure where the natural aging lens in the human eye is replaced by an artificial Intra-Ocular Lens (IOL). The IOL rectifies the refractive power which was lost from the removal of the natural aging lens (cataract). The accuracy in predicting the lost refractive power, and compensating for it using the IOL allows a patient to be less dependent on eye wear, such as corrective glasses. This accuracy of the predicted refractive power is dependent on accurate estimates of the corneal refractive power Φ_K. For patients who have already undergone LRS or other types of corneal restructuring surgeries, the previous method of estimating the Keratometric Refractive Index does not hold. This is because altering the corneal structure through surgery, such as LRS, violates Eq. (2.2), where there exists an assumption of a constant ratio between the radii of curvatures of the anterior (epithelium) and posterior (endothelium) surfaces. In such cases, measurement of both the corneal surfaces is needed.

To recapitulate, the corneal refractive power (paraxial corneal power) for normal corneas is computed using the thick lens equation as shown in Eq. (2.1). However, current corneal topography systems use the posterior vertex behind the cornea (inside the eye) as a reference for the rear focal point instead of the rear principal plane used in the computation of the paraxial corneal power [13]. This means that in order to convert corneal refractive power (paraxial) from a tomographic imaging modality, such as OCT, to a value that is in the range of corneal biometry measurement devices, such as corneal topographers, it is necessary to have a "Keratometric Equivalent Power" [20] formula that performs this conversion. We have defined this formula in Eq. (2.8) as:

$$\Phi_{KEP} = \Phi_{BFL} = \frac{\eta_c r_{ant} \Phi_{EFL}}{(\eta_c r_{ant}) - CCT(\eta_c - \eta_{air})} \tag{2.12}$$

Once the position of both the corneal surfaces have been determined in an image obtained by a tomographic imaging modality, such as OCT, the radii of curvature (r) and asphericity (k) of the anterior and posterior surfaces of the cornea can be measured through least squares fitting of optically-corrected [20–22] three-dimensional corneal surfaces to a formula for the conicoid of revolution of a sphere [20, 23–26]:

$$(z - z_o) = \frac{c\Big((x - x_o)^2 + (y - y_o)^2\Big)^2}{1 + \sqrt{1 - kc^2\Big((x - x_o)^2 + (y - y_o)^2\Big)^2}} \tag{2.13}$$

where (x, y, z) are the positions of the surface points, (x_o, y_o, z_o) is the posterior vertex behind the cornea as a reference for the focal point, and c is the reciprocal of the radius of curvature r.

Summary of Clinical Significance:

- There is a need to develop a method to measure the radii of curvatures of both the anterior (corneal epithelium) and posterior (corneal endothelium) surfaces without a relationship between the two.
- The measurement of the corneal refractive index and the corneal refractive power is contingent upon the identification of the epithelium, endothelium, and other tissue interfaces in the cornea of the eye.

2.1.2 Visualizing the Palisades of Vogt (POV)

In the cornea, epithelial stem cells reside in between ridge-like structures called the Palisades of Vogt (POV). As shown in Fig. 2.3, the POV are present 50-100 μm below the surface of the eye and are located at the intersection of the clear region (cornea) and white region (sclera) of the eye, with this intersection being called the limbus. The POV have unique finger-like configurations in each individual (akin to fingerprints), and they change over time and in response to congenital conditions, injury, surgery etc. The POV were first observed in 1866, and described in 1921 [27]. Recent attention to the POV has shown that it is the location of the stem cells that are needed for maintaining corneal homeostasis, clarity of vision, and regeneration of the corneal epithelium [28–34].

Clinically, there is a need for the quantification of progression of diseases affecting the cornea, such as corneal neoplasia (corneal cancer), lupus vulgaris, limbal stem cell deficiency (LSCD) [34,36], and injury. These conditions cause the depletion of stem cells residing in the POV [36]. The stem cells are crucial to the renewal of damaged tissue, and there is a compelling need to determine the total number and distribution of the cells residing within the POV [37,38], and assess the changes over time. Unfortunately, it is enormously challenging to reliably and consistently determine the stem cell distribution in the POV across images obtained from the same or different OCT imaging systems [34, 36–40]. By improving visualization of the POV and analysis of the depletion of the stem cells residing in the POV over time, stem cell deficiency and corneal disease progression can be better quantified [34,36–40]. Improving quantification beyond the current ad-hoc approaches will advance the development of new stem cell therapies to assist in the treatment of stem cell deficiency [34,36–40].

The POV are difficult to visualize due to their location and varying configurations in

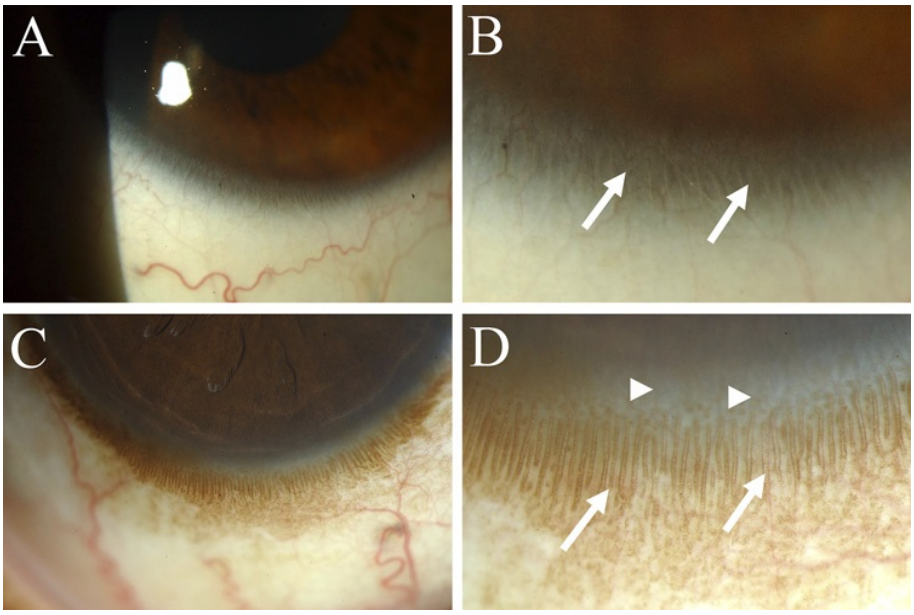


Figure 2.3: Slit lamp images of the cornea. A. Outside view of the corneo-scleral junction in a patient whose limbus is not pigmented. The inferior limbus is imaged, and the POV are not pigmented. B. Zoomed view of the same patient's eye at the inferior limbus, with the arrows showing the POV as linear structures in the limbus. C. In a different patient, the same region of the inferior limbus and POV are hyper-pigmented. D. A zoomed view of the POV in the same patient's eye; the arrows show the clearly visibile hyper-pigmented POV, with the arrow heads showing the projection of the POV to the corneal side. Figure used with permission, British Medical Journal, license 4324901418118 [35].

each individual [34, 36, 39]. Conventional imaging methods, such as slit lamps and in-vivo confocal microscopy, provide some details about the structures in some patients, but they are inconsistent and in some cases require direct contact with the eye [34, 36, 39]. The premise in previously published approaches has been to show inter-modality agreement in imaging and visualizing the same POV configuration, by comparing OCT cross-sections of the POV against histological slices [36, 37, 39], or confocal microscopy images [34, 38, 40].

Recently, Optical Coherence Tomography (OCT), a non-invasive imaging modality with a micrometer scale resolution, has been shown to be capable of visualizing the POV [3, 34, 36–40]. Using OCT, 3D volumes of the limbal region can be obtained, where a volume con-

sists of sequentially acquired images (B-scans). From the reconstructed 3D volumes of the limbal region, cross-sections (C-mode images) of the tissue can be extracted. These cross-sections will show the ridge-like structures of the POV, from which the Palisade Density (PD) [37] can be measured:

$$PD = \frac{1000}{\nabla_{PR} + \nabla_{ERP}} \tag{2.14}$$

where ∇_{PR} is the mean palisade ridge width (measured in μm) and ∇_{ERP} is the mean interpalisade epithelial rete peg width (measured in μm) [37]. The epithelial rete pegs are the cavities at the peripheral intersection of the cornea and sclera; the corneal epithelial cells at the periphery of the cornea extend into the stromal region, and form a cavity/basin like region. The numerator value of 1000 corresponds to the standardized 1000 A-scans (columns) in a single OCT B-scan corresponding to a 6mm scanning range over the cornea.

Summary of Clinical Significance:

- There is a need to visualize a 3D reconstruction of the imaged limbal area.
- Cross-sections of the limbal region showing the POV need to be extracted from volumes.
- The extracted cross-sections will need to be compared across volumes acquired by different imaging systems, specifically OCT imaging scanners.
- From the extracted cross-sections, the Palisade Density can be estimated using Eq. (2.14).

2.1.3 Monitoring Hand Transplant Rejection

Composite Tissue Allotransplantation (CTA), such as hand or face transplants, is the surgical procedure wherein multiple inter-connected tissues, such as skin, muscles, bone, nerves, vessels etc., are transplanted from a donor to a recipient. In contrast, traditional organ transplantation deals with a single (separate) organ as a whole. CTA improves the quality of life of a patient by restoring the anatomic, functional, and cosmetic integrity of the transplanted tissue. Hand transplantation is a form of CTA, wherein the hand from a cadaveric donor is transplanted to the forearm of a recipient, who is typically an amputee. There have been many successful hand transplant surgical procedures that have occurred all over the world, and the outcomes after surgery are typically encouraging. The extended focus of the clinical trials happening today are focused on improving the safety, reliability, and efficacy of these reconstructive surgeries. However, monitoring the transplant post surgery is one of the most challenging aspects of CTA that surgeons and clinicians face today.

Transplanting a hand from a donor to a recipient is difficult as the different tissues in the hand transplant all produce immune responses from the recipient's body to the newly transplanted graft [41]. The various tissues include skin, subcutaneous tissue, muscle, bone, nerve, and blood vessels [42]. It has been experimentally verified that skin produces an acute immune response once transplanted from a donor to a recipient.

As our bodies typically have immune systems that protect us from germs, they identify and remember the proteins that are present on the surface of germs, called antigens. When foreign germs (antigens) enter the human body, an immune response is triggered to attack these germs. As a transplant recipient receives a graft from a donor, it is almost impossible to have an exact match of the cells in the transplanted graft to the cells in the recipient. In response to the foreign antigens of the graft, the body produces an immediate response to

attack the graft, which can lead to transplant rejection.

To avoid such problematic scenarios, it is common for surgeons and clinicians to thoroughly match the tissue type of the donor and recipient as much as possible before transplantation has occurred. Even after transplanting the graft onto the recipient, the patient is provided with immuno-suppressant drugs to reduce or mitigate the response to the foreign graft. The goal in such scenarios is to minimize the immune response of the recipient to reject the graft, especially when the graft does not closely match. Without these immuno-suppressants, the recipient's immune system will almost certainly reject the transplant [41].

Even with an immuno-suppressive drug load, the recipient will face some rejection of the graft. Acute Rejection (AR) of the transplant can occur anywhere between a week after transplantation to three months after transplantation [41]. On the other hand, Chronic Rejection (CR) can occur over many years, and slowly damages the grafted tissue by launching an immune response against the new transplant [41]. A typical rash due to AR seen on the skin of the transplanted graft in the recipient is shown in Fig. 2.4.

Clinicians currently use a whole host of methods to monitor hand transplant rejection including clinical and functional examinations, skin biopsies, donor specific antibody measurments (DSA), and standard vascular imaging [41]. Vascular imaging is a way to monitor chronic hand transplant rejection by measuring the changes in the thickness of the arterial wall and the patency of the lumen (open cross-section) of the vessel. In normal subjects (no transplant), the arteries and veins retain normal function and patency. Usually, the physiologic function of an artery in a normal subject is to deliver oxygenated blood to tissues, and as shown in Fig. 2.5, its structure is made up of three major components: the inner vessel wall (intima), the outer vessel wall (adventitia), and the smooth muscle region between the two (media). The vessel lumen is the part of the vessel through which the blood flows.

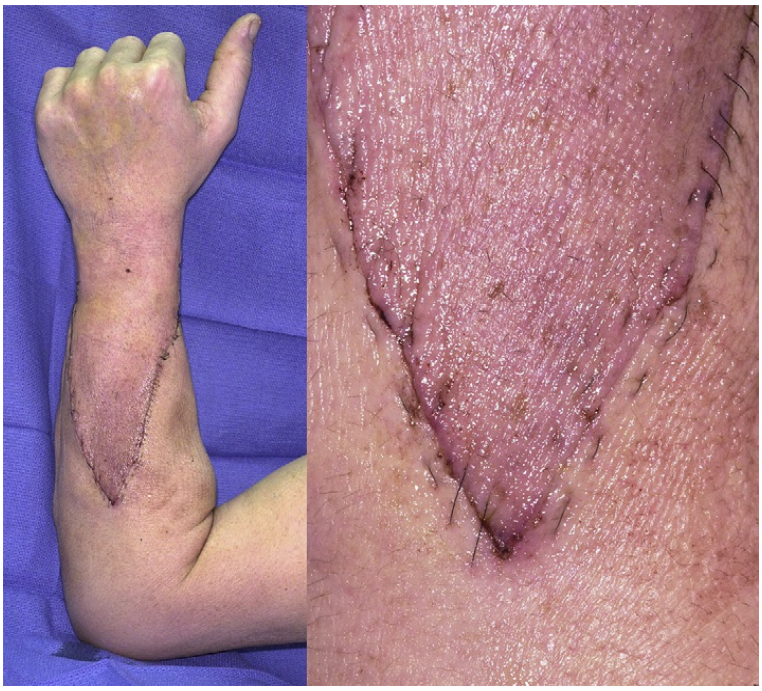


Figure 2.4: A rash seen on the skin of the recipient. It is limited to the line where sutures were made during transplantation. Figure used with permission - Hand Clinics, Elsevier, license 4325111357417 [41].

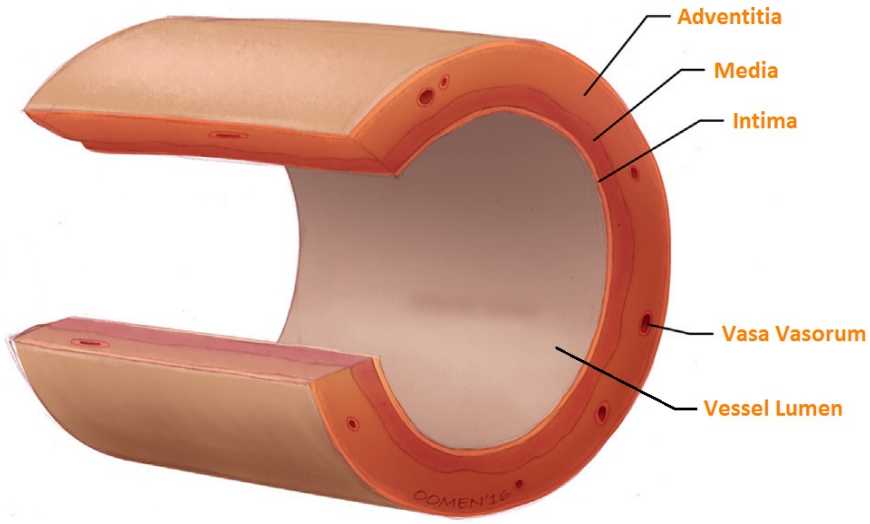


Figure 2.5: Artery in a normal human subject. Figure used with permission - Journal of the American College of Cardiology, Elsevier, license 4325400758996 [43].

However, with AR and CR, it is common to see changes in the micro-vasculature of the transplant recipient. Over time, the immune response of the recipient to the donor graft will cause the transplant to experience a reduction in the vessel wall compliance, changes in the intimal wall thickness, and in other cases variations in atherosclerotic plaque buildup. Presently, the most accurate modality to confirm AR and CR is through histopathology [41] (see Fig. 2.6). This means that a biopsy of the tissue sample is taken from deep inside the vascular structure of the recipient, and the micro-structures are analyzed by a pathologist for the markers and different manifestations of rejection of the hand transplant.

In patients who have had hand transplants, the changes in the vessel wall progress over time. When the hand transplant is completely rejected, it has been reported that it was due to a condition called Intimal Hyperplasia (IH) [41]. Hyperplasia is the condition where there is an excess in the reproduction of the cells in the tissue, and specifically in the case of hand transplants, the intimal cells are reproducing quickly. This abnormal reproduction of intimal cells leads to thickening of the intimal vessel wall, and reduction in the area of the vessel lumen. When histopathology was conducted, Intimal Hyperplasia was confirmed as the cause of the rejection of a hand transplant in at least one known case [41].

There was concern in 2011 that conventional methods were unable to monitor and diagnose Intimal Hyperplasia as the primary indicator of chronic rejection of the hand transplant [41]. But, it has been shown that CR could be successfully monitored using high-resolution ultrasound imaging [44]. High resolution ultrasound can be used to measure the thickness of the intimal wall, and the patency of the vessel lumen. It has been postulated that measuring the intimal wall thickness over time will provide a way to monitor the hand transplant, and allow clinicians to change immuno-suppressive drugs to prevent the hand transplant from being rejected eventually.

Figure 2.6: Histopathological staining of the arterial wall post biopsy showing intimal hyperplasia. There is a narrowing in the opening of the vessel (yellow stain) due to thick intimal vessel wall (light pink) encroachment into the vessel lumen. The media (light brown) and the adventitia (darker pink) show some signs of inflammation. Figure used with permission - Hand Clinics, Elsevier, license 4325111357417 [41].

Summary of Clinical Significance:

- Chronic rejection of the hand transplant is correlated with and may be caused by Intimal Hyperplasia, where the intimal wall thickens and encroaches into the vessel lumen, thereby causing it to become narrow and reduce blood flow.
- Clinicans presently need a way to measure the intimal vessel wall thickness, and monitor its changes over time.
- Segmenting the boundaries of the vessels in ultrasound images will help clinicians monitor a hand transplant.
- Measuring the intimal wall thickness and vessel lumen patency will enable clinicians to monitor hand transplant rejection, and change the course of immuno-suppressive drugs for a recipient to mitigate transplant rejection.

2.2 Real-Time Imaging Modalities

2.2.1 Optical Coherence Tomography (OCT)

Optical coherence tomography (OCT) was first introduced in the early 1990's as a non-ionizing and non-invasive imaging modality [45]. It was shown to be useful in imaging retinal tissues and vessels at the scale of microns [45]. Light waves are used in OCT to probe tissue structures, and a light source emits a beam of light to image the sample. Before the beam is detected at the receiver, the light waves are scattered, reflected, and refracted in multiple directions inside the tissue. OCT uses low-coherence interferometry to filter out most scattered light by selecting the optical path length or the time delay between the transmission and reception of the light beam, thereby allowing OCT to reconstruct a depth-profile of the tissue sample [45,46]. Two- or three-dimensional scans can be obtained by shifting the beam such that it scans laterally across the tissue sample.

Since its inception, three main categories of OCT imaging have been developed: Time Domain OCT (TD-OCT), Spectral Domain OCT (SD-OCT), and Swept Source OCT (SS-OCT). Basic introductions to these categories in OCT imaging are presented next, which will help the reader have an understanding of the principles of OCT-based imaging and the challenges facing OCT image analysis.

Time Domain Optical Coherence Tomography (TD-OCT)

First-generation OCT systems were based on time-domain OCT with different configurations for imaging. Research grade systems were the first to be built with TDOCT as the basic optical configuration, which was shown to be useful for bio-medical applications, such as imaging the anterior and posterior segment of the eye. TDOCT was also commercialized

by Carl Zeiss Meditec Inc. The working mechanism of TDOCT is illustrated in Fig. 2.7.

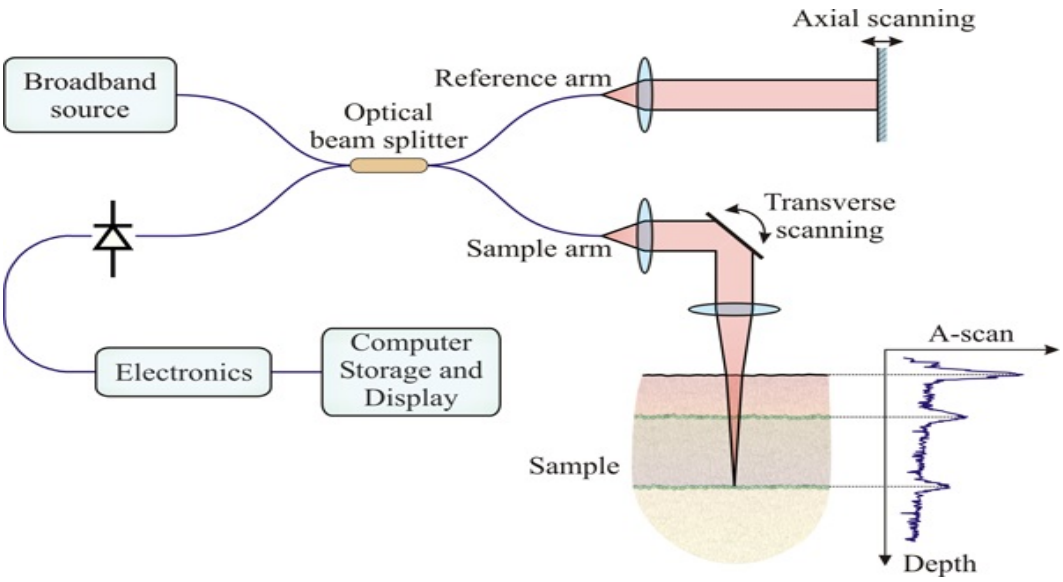


Figure 2.7: Optical configuration in TDOCT (reproduced from Dr. David Sampson's illustration - OBEL laboratory).

The broadband light source produces a beam of light, spanning a large range of optical wavelengths, which is then split by a Michelson interferometer (beam splitter) into two paths: a reference arm and a sample arm. The light following the reference arm path is reflected by a mirror at the end of the arm, which is can be translated in the axial direction. The light following the sample arm interrogates the tissue sample at a specific point, undergoing backscattering, before being focused back into the interferometer. The two light paths then undergo superposition in the interferometer and are recorded; the light in the sample arm (after backscattering) and reference arm (after reflection) destructively interfere, through superposition, producing a pattern called an interferogram. The interference signal is only strongly detected if the light in both the arms have traveled the same distance in each arm (optical path length). This means that significant measurable constructive interference is only observed when the difference in the path lengths between the two arms

is less than the coherence length of the broadband light source.

By axially translating the mirror in the reference arm, the desired optical path length of the backscattered light can be controlled, thus providing information on the optical properties of the tissue sample at different depths. Moving the mirror in the reference arm has the effect of slicing (or sectioning) through the tissue structure, layer by layer, thereby providing a measurement of the optical reflectivity of the tissue sample at each layer at that depth. The optical reflectivity versus depth profile is acquired through a specific transverse point on the tissue sample, and this profile is referred to as an axial scan (A-scan). As a single A-scan does not provide much information, a series of such A-scans are acquired by transversely scanning a tissue sample using the mirror(s) attached to galvanometer(s) in the sample arm, thereby providing 2D and 3D information of the tissue structure.

Spectral Domain Optical Coherence Tomography (SD-OCT)

Spectral Domain OCT systems are identical to TDOCT systems in many ways. However, the key exception in SDOCT systems is that the mirror in the reference arm remains fixed during scanning as shown in Fig. 2.8. The depth information in SDOCT is not obtained by axially translating the mirror in the reference arm. Rather, the output light signal from the interferometer is passed to a spectrometer for analysis of the interferogram. The spectrometer uses a diffraction grating to split the received light into its spectral components. A lens focuses the light onto a linear detector array (a line scan camera), which measures the spectrum, and a Fourier Transform applied to the measured spectrum yields the optical reflectance properties of the tissue sample against the depth. The optical reflectivity versus depth information has been shown to be the same as the information obtained with TDOCT, except that with TDOCT, the reference arm mirror is translated axially, while a

spectrometer analyzes the interferometer output in SDOCT.

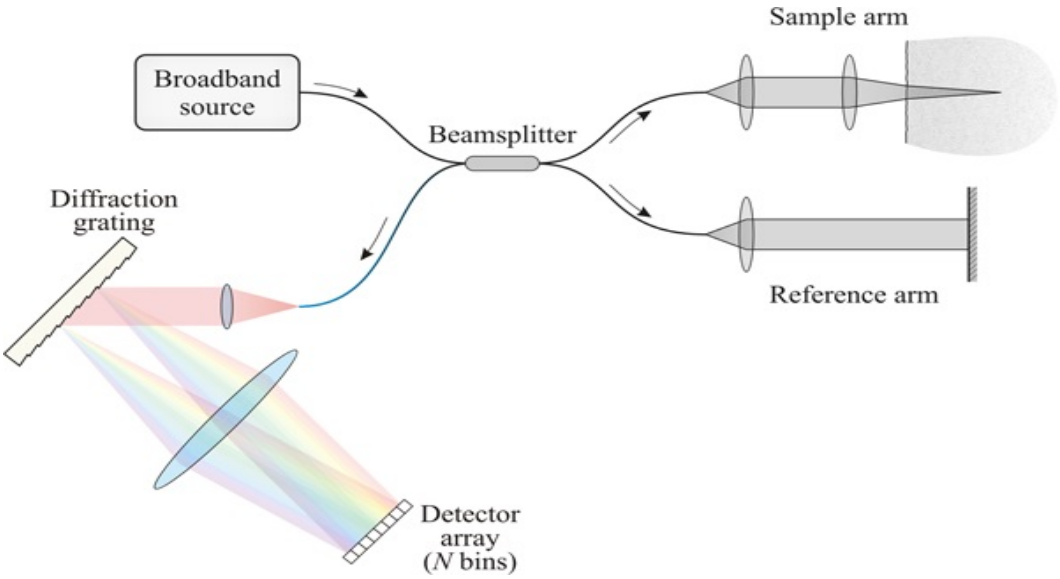


Figure 2.8: Optical configuration in SDOCT (reproduced from Dr. David Sampson's illustration - OBEL laboratory).

SDOCT captures and processes all the pixels in an A-scan in parallel, with a dramatic improvement in the signal-to-noise ratio (SNR) of up to 15-20 dB [47–49]. The SNR improvement allows deeper interrogation of tissue samples, acquisition of higher resolution images, and faster imaging [47]. However, the spectrometer has finite spectral resolution, and the fringe interference pattern decreases exponentially with depth [50]. Specifically, the depth at which there is a 6 dB decrease in the SNR can be defined in terms of the wavelength λ or the wavenumber k [50]:

$$z_{6dB} = \frac{2ln(2)}{\delta_r k} = \frac{ln(2)\lambda_o^2}{\pi \delta_r \lambda} \tag{2.15}$$

where λ_o is the center frequency, and $\delta_r \lambda$ is the Full Width at Half Maximum (FWHM) spectral resoltuion [49–51]. As such, there is a limit to the depth of imaging for a selected configuration of the source and detector. The maximum depth is limited by the Nyquist

criterion [50]. At the detector, the interferometer output is sampled with finite spectral resolution due to the pixel binning of the line scan camera. This sampling pattern is linear in the wavenumber k, which separates the interferogram into M channels [50]. This translates to a sampling interval of $\delta_s k$ for each of the M channels. Therefore, the total wavenumber range is $\Delta k = M\delta_s k$. This, in turn, sets the sampling interval in depth as $\delta_s z = 2\pi/(2\Delta k)$ [50]. Using the above values, the Nyquist criterion provides us with the maximum imaging depth in wavelengths and wavenumbers as:

$$z_{max} = \frac{\pi}{2\delta_s k} = \frac{\lambda_o^2}{4\delta_s \lambda} \tag{2.16}$$

In this thesis, we mainly concentrate on the 2D B-scans and 3D volumes that have been obtained from SD-OCT scanning systems.

Swept Source Optical Coherence Tomography (SS-OCT)

In SS-OCT, a laser that is narrow band and wavelength-tunable is used to sweep across a broad range of wavelengths. Instead of a diffraction grating, a photodiode is used as the detector, and the reference arm is fixed in SS-OCT as shown in Fig. 2.9. The spectrum of the interferometer output is recorded for each of these wavelengths sequentially in time using a single detector, thus allowing the spectrum to be recorded as a function of time. The spectral pattern encodes in its frequency content the depth-resolved structure of the imaged tissue sample at the waist of the light beam.

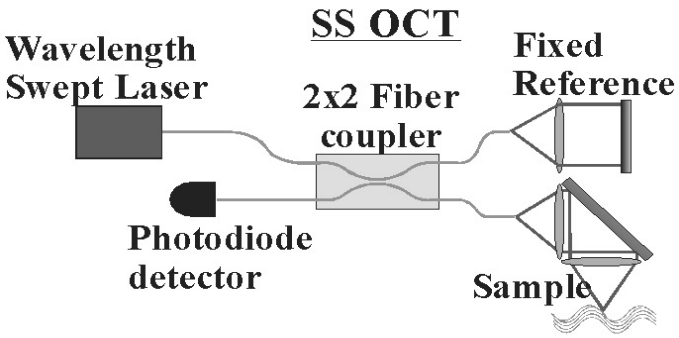


Figure 2.9: Optical configuration in SS-OCT (reproduced from Dr. Marinko Sarunic's illustration - BORG laboratory).

2.2.2 Ultrasound (US)

Ultrasound is a non-invasive and non-ionizing imaging modality that uses ultrasonic acoustic signals to acquire cross-sectional images of internal anatomy, such as the liver, kidney, vessels etc. The mechanism for imaging involves piezo-electric transducers housed in a plastic casing, which are excited by electric signals to produce acoustic waves at particular ultrasonic frequencies. These acoustic signals are focused, either using a lens arrangement or a phased array setup (beamforming technique), on the sample. Similar to OCT, these acoustic signals travel through the anatomic structure, wherein they undergo back-scattering and reflection by the different tissue layers.

Some part of the back-scattered acoustic signal is detected by the same transducers that produced the acoustic signal. The received signal, called the echo, is converted by the transducers into electric signals. These electric signals are then amplified, processed, and analyzed to produce cross-sectional ultrasound images of the internal anatomy. Ultrasound image formation depends on two factors: the intensity of the received echo signal (pixel intensity), and the time delay between the propagation of the acoustic waves and reception after back-scattering (pixel location).

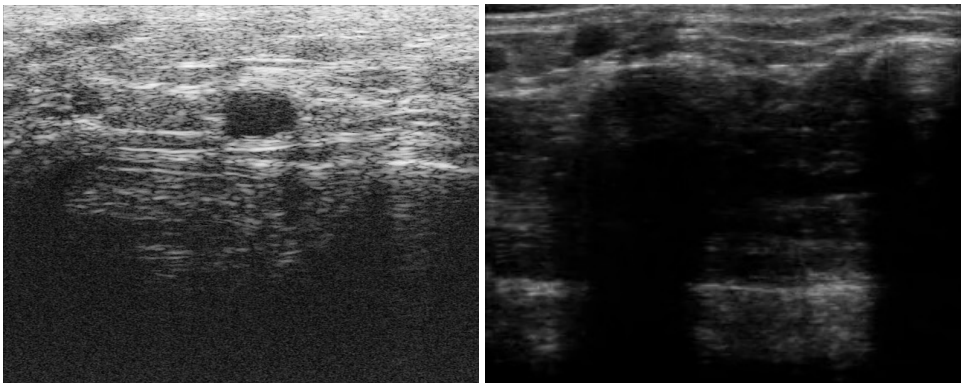
The intensity of the pixel at a location in the ultrasound image depends on the amplitude of the echo signal, which in turn depends on the change in the acoustical impedance of the material through which the acoustic waves are propagating. For example, if air is encountered in the path of the focused acoustic signal, then a large difference in the acoustic impedance is witnessed. This large impedance mis-match causes all the acoustic waves to be immediately reflected back to the detector. Accordingly, the ultrasound transducer must always be in contact with tissue sample being imaged. Furthermore, as ultrasound signals travel relatively easily in liquids, a viscous gel-based medium is typically introduced dur-

ing conventional imaging to ensure that the acoustic impedance is minimized between the transducer and outer tissue surface.

Traditionally, ultrasound transducers have operating frequencies in the range of 1 - 20 MHz. This frequency is inversely proportional to the spatial resolution, which is the distance between two scatterers at which the scatterers can be resolved and visually differentiated. This means that as the frequency is increased, the axial resolution of the ultrasound transducer improves, with the resolvable distance between to scatterers becoming smaller. Furthermore, as the ultrasound beam travels through the imaged tissue sample, attenuation will be encountered. Attenuation is strongly dependent on the frequency, and it has been shown that the relationship between the attenuation coefficient and frequency is approximately linear [52].

For soft tissue, the attenuation coefficient is 1 dB $cm^{-1}MHz^{-1}$. Thus, a trade-off between the depth of penetration and frequency exists; lower penetration is seen with higher frequency ultrasound transducers. Therefore, low frequency ultrasound transducers are used to image organs and structures that lie deep inside the body, such as liver and kidney, whereas higher frequency ultrasound is used to image fine details in superficial structures, such as muscles and breasts. Ultrasound machines that operate at frequencies higher than 20 MHz are categorized into two classes: High Frequency Ultrasound (HFUS), and Ultra-High Frequency Ultrasound (UHFUS).

As mentioned before, HFUS is used to image to image superficial structures. They have a frequency range of 15-30 MHz, and a penetration depth of 5-10 cm inside the tissue. The UHFUS designation becomes appropriate at frequencies higher than 30 MHz, and a recently FDA-approved ultrasound scanner (VisualSonics Vevo MD, Fujifilm Inc.) can operate at frequencies of 50-70 MHz. The Vevo MD ultrasound machine can acquire ultrasound



(a) (b)

Figure 2.10: (a) Example UHFUS image of the proper palmar digital artery in the hand. Image obtained with a 50 MHz ultrasound probe (b) Example HFUS image of the same vessel obtained with a 22 MHz ultrasound probe. In contrast to the HFUS image, note the higher resolution and increased speckle noise in the UHFUS image.

images at 30 microns resolution, with an imaging depth of $\sim$1 cm, making it useful for imaging nerves, vessels, and skin. For example, individual vessels can be imaged with greater resolution in UHFUS as compared to HFUS as seen in Fig. 2.10. UHFUS can also be used to measure several vascular substructures including Intimal Thickness (IT), Intima-Media Thickness (IMT), the Lumen Diameter (LD), and vessel lumen patency etc. Vessel based measurements was extensively reported in [53] as shown in Fig. 2.11, and a clinician (Dr. Vijay Gorantla, Wake Forest Health/UPMC) marked the measurements in an UHFUS image as shown in Fig. 2.12. These measurements have been shown to be possible using the Vevo 2100 ultrasound machine [44], the experimental ultrasound machine that was the precursor to the Vevo MD ultrasound machine. Evidence has suggested that both arteries and veins are primary targets of chronic hand transplant rejection [41], making these vessel measures potentially valuable for detecting Chronic Rejection (CR) of a hand transplant.

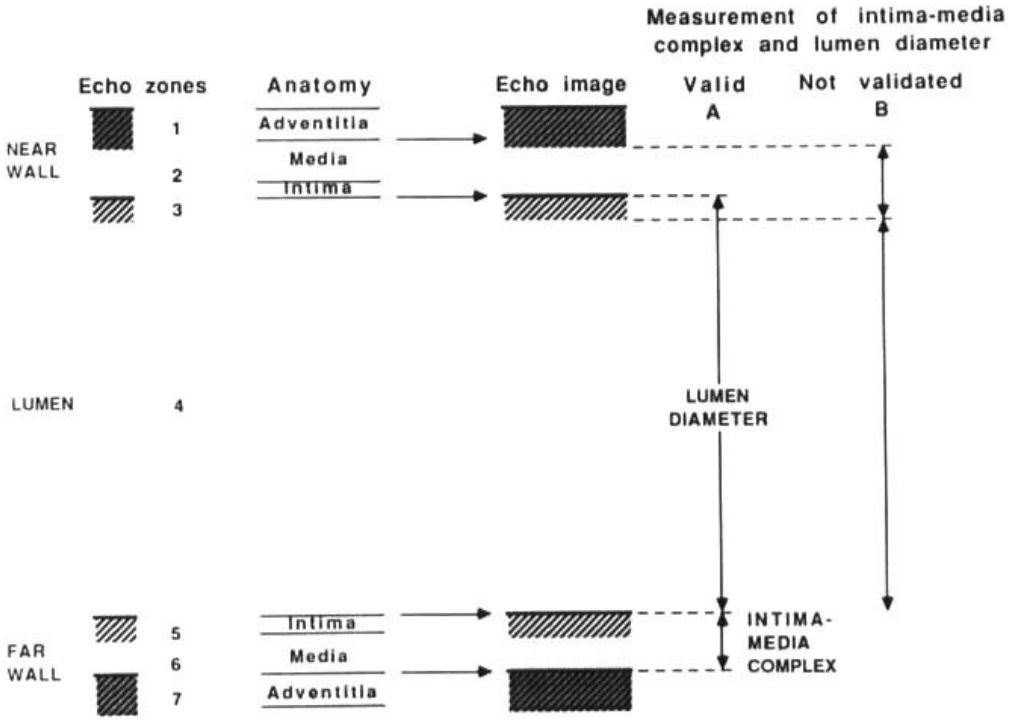


Figure 2.11: Vessel measurements first reported by [53]. Figure used with permission - Clinical Physiology and Functional Imaging, John Wiley and Sons, license 4336130506730 [53].

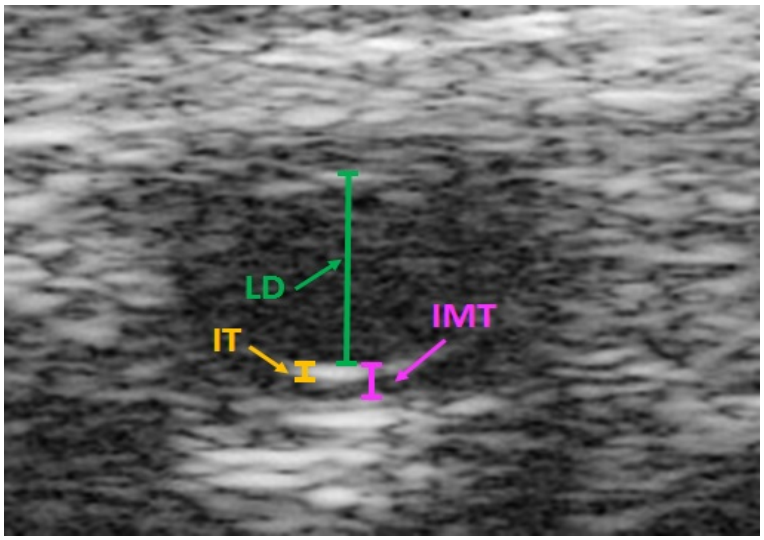


Figure 2.12: UHFUS image of an artery acquired by the Vevo 2100 ultrasound scanner. Measurements shown are the Intimal Thickness (IT), Intima-Media Thickness (IMT), and the Lumen Diameter (LD)

2.2.3 Modes of Imaging in OCT and Ultrasound

There are many different types of imaging modes in ultrasound and OCT. However, the following basic imaging modes are common among both modalities:

- A-Mode: The Amplitude (A) mode, also commonly referred as A-scan, is a one-dimensional scan, where the amplitude of the backscattered wave (acoustic or light) is displayed against the tissue depth.

- B-mode: The Brightness (B) mode, also commonly referred to as a B-scan, generates a two-dimensional image. Each column in this 2D image is an A-scan with the amplitude of the backscattered acoustic or light wave plotted as the brightness for each pixel in the image.

- C-mode: The cross-sectional (C) mode scans the tissue sample in a plane that is perpendicular to the B-mode image. To create a C-mode image, a gate is used to select a particular depth at which to receive the echoes from. Next, these echoes in the plane at that particular depth, which is at a constant distance from the transducer, are only received and recorded. This recorded signal is converted into an image, which represents a cross-section of the tissue sample imaged at a selected depth.

- M-mode: The Motion (M) mode allows for a video sequence of B-mode images to be acquired from imaging the tissue sample. This involves the acquisition of a series of B-mode images, where each column in the B-mode image represents a specific A-scan displayed vertically. For example, in ultrasound, the probe can either held stationary at a fixed position, or it can be swept horizontally. When the probe is stationary, the underlying tissue structure undergoes motion and ultrasound captures this motion. When the probe is swept horizontally, B-scans are sequentially acquired, resulting in

a video sequence. The M-mode is used to analyze the motion of internal organs; the time-varying displacement of the internal organs allows the estimation of the velocity of the moving organs.

- Doppler mode: In the Doppler mode, moving objects allow the Doppler effect to measure the velocity of these objects. The most popular use of the Doppler mode is to measure the velocity of blood flow. For example, when the blood flow through a vessel is imaged, the Doppler effect color codes the velocity of the blood flow, and overlays it on the B-mode image.

3D ultrasound is also an important diagnostic tool as it provides additional information over conventional 2D ultrasound. 3D anatomical views in ultrasound of the tissue sample provide detailed structure of the tissue, from which 2D images can be generated by arbitrarily slicing through the 3D volume. It is also easy to estimate quantitative values relating to the imaged sample using the 3D dataset. Presently, there are two different types of 3D probes used to acquire 3D ultrasound volumes:

- Probes containing a 2D matrix array to capture the 3D structure of the tissue sample. These probes have lower SNR.

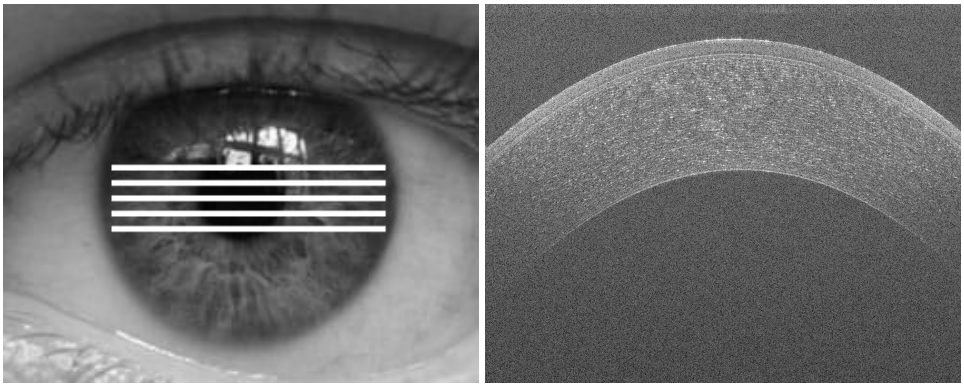
- Probes containing a smaller 2D probe that is mechanically swept along the third-dimension. However, these 3D probes can only scan a small region at any given time.

The following imaging modes are reserved specifically for OCT:

- Raster Scanning Mode: In the OCT raster scanning mode, a galvanometer in the sample arm controls the pattern of scanning, which is roughly a saw-tooth pattern. The light beam is first swept horizontally, from left to right, before blanking, and rapidly

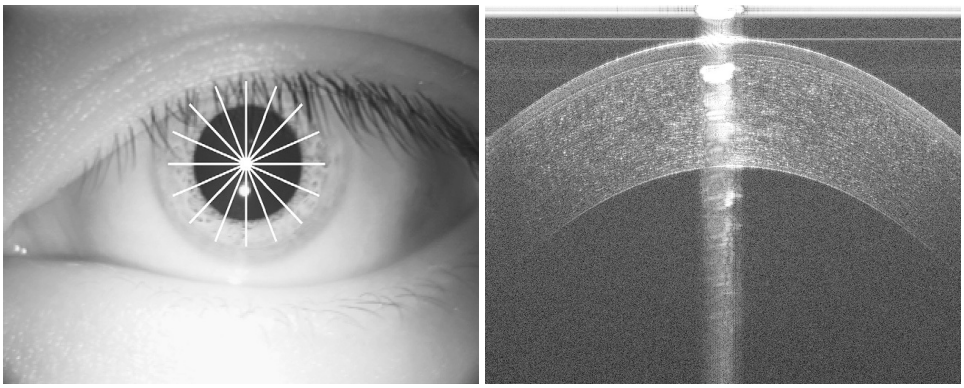
moving back to the left (see Fig. 2.13). This sweep along the horizontal direction is very fast, and this axis is dubbed the "fast scanning axis". A series of A-scans are obtained by sweeping along the fast axis, thus constituting a B-scan image. During this time, the vertical position of the scan is also changing (increasing or decreasing), depending on the desired direction along the tissue sample. The vertical scanning position increment is slower, and this axis is called the "slow scanning axis". As the tissue sample along the slow axis is scanned, a sequence of B-scans are obtained.

- Radial Scanning Mode: A radial scan mode is typically used to to generate a pachymetry map for measuring the thickness of the cornea. It usually consists of a scan pattern of radial lines, centered on the apex of the cornea, as shown in Fig. 2.14. These radial lines are equally spaced, and they provide a meridional cross-sectional B-scan of the cornea [54].



(a) (b)

Figure 2.13: (a) White lines represent the raster scan pattern. (b) The corresponding cross-sectional B-scan of the cornea along one horizontal line. At each white line in (a), a B-scan image (b) is obtained.



(a) (b)

Figure 2.14: (a) Radial scan pattern. Figure used with permission - Journal of Cataract and Refractive Surgery, Elsevier, license 4325451407563 [54]. (b) The corresponding cross-sectional B-scan of the cornea along one radial line. At each white radial line in (a), a B-scan image (b) is obtained.

2.3 Challenges in Imaging and Image Analysis

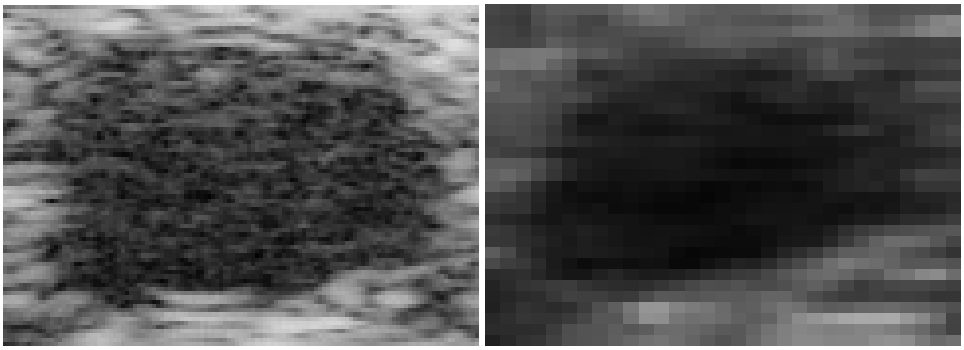
Non-ionizing imaging modalities such as OCT and Ultrasound lend themselves to safe real-time imaging and subsequent image analysis. When used at (or below) the established clinically safe power levels, they have no known long-term side effects, making them the ideal modalities for repeated longitudinal imaging a particular section of anatomy. These modalities are relatively inexpensive compared to other modalities, such as MRI and CT, making them more readily accessible for many different clinical applications.

While OCT and Ultrasound are similar in many ways, each modality has particular difficulties associated with imaging tissue samples. Furthermore, there are additional challenges in the analysis of the resulting images. Although, ultrasound uses sound waves for imaging and OCT uses light waves, they are both subject to similar imaging artifacts that affect the quality of the reconstructed images. For example, images acquired by both modalities exhibit shadows, attenuation, and specular artifacts.

Ultrasound

There are three sources of image corruption in ultrasound imaging [52]. The first source of noise is from the electronics of the ultrasound system. The second source is termed ”speckle” noise, which arises from the interference pattern originating from within a single resolution cell when different back-scattered signals from sub-resolution structures within that cell arrive at the detector with random phases and amplitudes. These back-scattered signals undergo superposition, and produce an interferogram, where the range of the speckle noise produced extends between a minimum of zero to a maximum value that is dependent on whether the interference is constructive or destructive. As seen in Fig. 2.15, speckle noise causes the acquired image to appear granular and noisy, even though the underlying tissue

is mostly homogeneous. The third type of image corruption that arises in imaging with ultrasound is artifacts, such as shadowing, mirroring, posterior enhancement, refraction, side lobes, grating lobes, and reverberation. For a detailed study of the different artifacts affecting ultrasound, the reader is referred to [55].



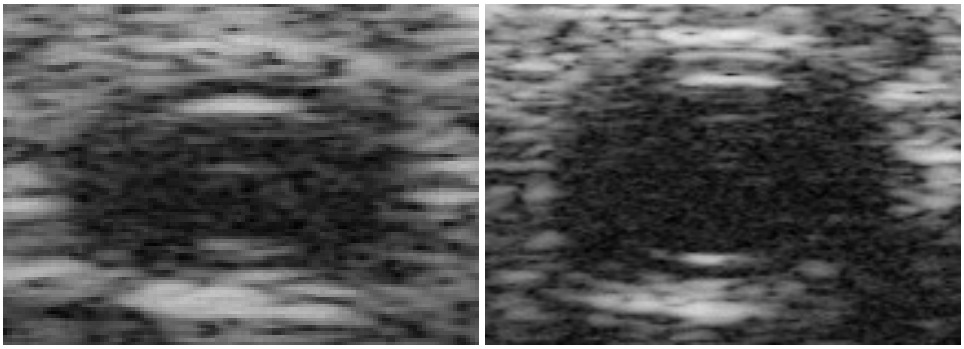
(a) (b)

Figure 2.15: (a) Vessel imaged by a UHFUS machine. (b) Vessel imaged by a HFUS machine. Note the increased granularity of the speckle noise inside the vessel lumen in (a), as opposed to the relatively homogeneous lumen in (b).

In the case of UHFUS images, tissue structures are axially well resolved, as seen in Fig. 2.16. But, the lateral resolution of tissue is poor. While imaging vessels, the intimal and the adventitial wall are sometimes resolved axially, however both the vessel walls are mostly invisible in the lateral direction (see Fig. 2.16). Due to this inherent nature of UHFUS imaging, it is enormously challenging to estimate the position of the vessel wall boundaries laterally. Since the vessel boundaries in the lateral direction cannot be reliably estimated, it is difficult to determine clinical measurements such as Intimal Thickness (IT) (see Fig. 2.12) accurately.

In addition to the above mentioned challenges, skilled operators, who have had training with imaging using ultrasound and who are also equipped with the knowledge of the human anatomy, are typically needed in order to acquire ultrasound images of a reasonable quality that are useful for analysis. It is also very difficult to maintain contact with

the patient when imaging with a linear probe in anatomical regions of high curvature. As ultrasound does not pass through bones, the location and orientation of the probe during imaging behind tissues, such as bones, is critical. Crucially however, once a sequence of ultrasound images of the hand for monitoring hand transplant rejection have been obtained, the exact location that they were acquired from needs to be logged, or they are lost and thus generating the same image during a follow-up imaging session will be difficult. The location of the scan needs to be registered to a 3D coordinate system, otherwise the vascular data cannot provide information on the progression of Intimal Hyperplasia over time.



(a) (b)

Figure 2.16: (a) and (b) show vessels imaged by a UHFUS machine.

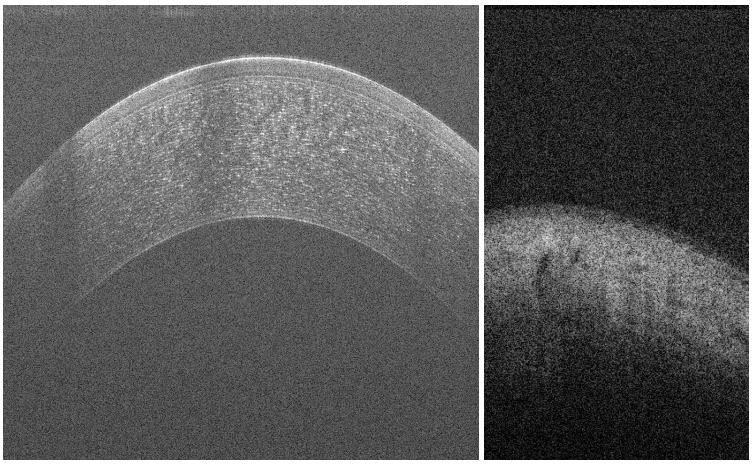
Thus, it is necessary to segment the vessel wall boundaries to calculate the vessel-based measurements, such as lumen diameter and intimal thickness. Furthermore, in order to allow for comparisons of these measurements across time, the vessel boundary needs to be segmented and tracked over time in sequential images acquired by the ultrasound machine.

OCT

With anterior-segment SD-OCT, there are many different artifacts that corrupt the reconstructed SD-OCT image. The most common artifacts that occur are missing sections in the

B-scan and the edges of the B-scan being cut off [56]. As seen in Fig. 2.17, this results in a part of the B-scan appearing black with missing tissue structure definition, or the signal-to-noise ratio at these areas slowly fading away. These are usually caused by changes in curvature of the tissue, opaque intervening tissue structure etc.

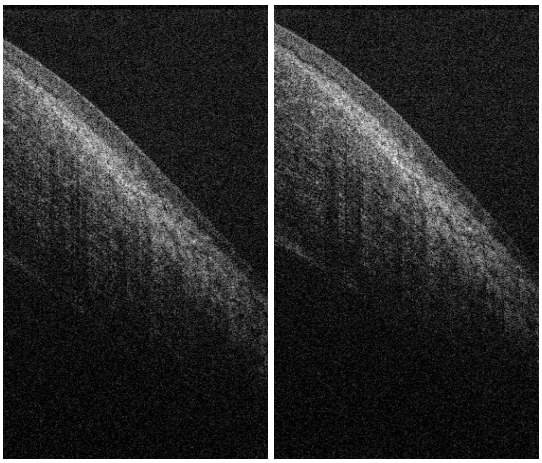
Other types of artifacts include patient motion [56] during the scan. It is typically very hard for a patient to be still during imaging. As SD-OCT scans a small region (usually 3×3, 4×4, or 6×6), micro-movements, such as movement of the patient's eye (saccades), heartbeat, respiration etc., can result in jittery B-scan images as seen in Fig. 2.18.



(a) (b)

Figure 2.17: (a) Eyelashes of a subject cast a shadow on the tissue structure during imaging, thereby rendering the left side of the image without structure. (b) Low intensity of the tissue structure to the right of the image makes it difficult to delineate the tissue boundaries correctly.

The most confounding artifacts that affect the analysis of SD-OCT images are the horizontal and central artifacts [57]. The central artifact occurs when the line scan camera (detector) is saturated by the back-scattered SD-OCT light beam, which undergoes scattering and reflections at the corneal apex, and the low line-scan camera exposure time (sensitivity) [57]. Since the Fourier transform is used to reconstruct the image from the SD-OCT



(a) (b)

Figure 2.18: Compared to frame 129 of a volume (a), the tissue structure in frame 130 (b) in the same volume undergoes significant motion in both the axial and lateral directions.

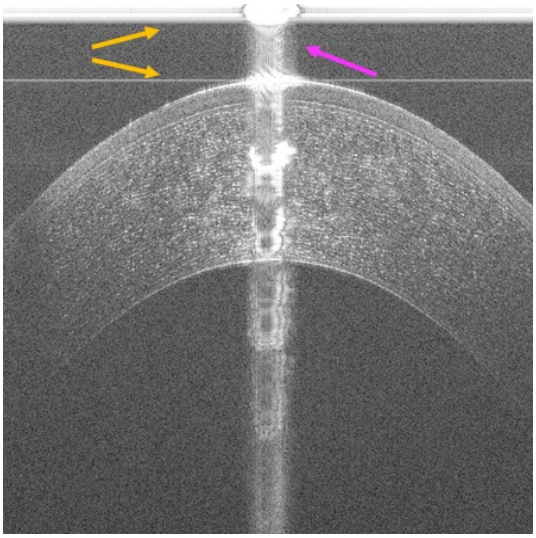


Figure 2.19: The gold arrows represent the horizontal artifacts, while the magenta arrow represents the central artifact.

signal, these saturated regions appear as periodic sinc functions in the B-scans. The horizontal artifacts are an indirect result of the DC subtraction algorithm applied by different vendors (Bioptigen, Carl Zeiss, Heidelberg) in their software to reduce the strong central artifact [57]. As shown in Fig. 2.19, both the aforementioned artifacts affect the delineation

of the corneal tissue interfaces in the image, and hence affect the segmentation of corneal interfaces. Any abnormality in the curvature of the tissue structure, which is usually seen in pathological eyes, also causes the light to be reflected straight back to the detector, saturating it, and creating these horizontal and central artifacts.

Acquiring pathological data is quite challenging, and there are a variety of scenarios that need to be considered. Usually, pathological datasets obtained with SD-OCT are very sparse datasets with poor image quality. Tissue structures in these patients are quite subtle, and can be very hard to visualize. They may only appear in a few images and disappear, or not be present at all in other images of a dataset. Furthermore, if the images where pathology is present are of low resolution, then they will be insufficient to model that particular condition. In such cases, finer detail of the underlying condition is needed, which is typically hard to acquire. For example, the position of the Palisades is quite ambiguous in many images of a dataset; they will only be present in 50-100 images in a 400 image dataset. Moreover, it is also very difficult to establish general a-priori models for pathological cases, as there are numerous known clinical cases that would need to be modeled. It is also hard to model the optical physics that would result from each specific type of disease and pathological symptom.

Chapter 3

Technical Background

3.1 Segmentation of Corneal Tissue Interfaces

Initial approaches towards segmenting the corneal tissue interfaces largely followed the work of [58], where the Epithelium, Bowman's layer, and Endothelium were segmented in 830nm and 1310nm SD-OCT images. Other approaches for retinal layer segmentation [59–61] were also used as the basis for segmentation of the corneal tissue interfaces. However, to date, graph-based segmentation approaches [57,62,63] have been the most popular methods to segment the corneal tissue interfaces. Other types of segmentation algorithms, such as those that use active contours [58,64,65], Hough Transforms [66], or Gaussian Mixture Models (GMM) [67,68], have also been proposed. Most of the approaches [57,58,63–67] are also restricted to only 2D image based segmentation of the corneal tissue interfaces. Some of these approaches also segment only a few layers, and/or segment each layer individually. The approach in [62], although only implemented on corneal OCT images of mice, simultaneously segments the corneal tissue interfaces in 3D OCT volumes. However,

all of these approaches are implemented to only segment corneal interfaces in OCT images acquired from a single OCT scanner.

The approaches in [58, 64, 65] used an active contour as the basis for segmenting the various layers in the OCT image. However, active contours are very sensitive to the initial location from which the level set is allowed to evolve, and incorrectly choosing this initialization position can lead to incorrect results. The method proposed in [66] utilized a training set of manually labeled images to generate an initial quadratic curve shape model of the cornea. This curve model was then optimized based on filtering the input OCT image with a Prewitt edge detector, and the Hough Transform was used to find the different corneal layers. However, the optimization of the curve shape model would immediately fail when there are large specular artifacts (central and horizontal) present in the OCT images.

The approach presented in [57] made ad-hoc assumptions on the position of the central artifact, and relied on pre-processing of the input OCT images before being fed to a Djikstra's algorithm-based segmentation approach to estimate the location of the corneal tissue interfaces. It is not guaranteed however that the central artifact would always be present in the central section of the OCT scan, and an estimation of the total thickness of the cornea does not necessarily extend to pathological cases. Furthermore, this approach was validated on only small datasets, and would not reliably perform well on datasets acquired by different OCT systems.

The approach in [69] was the first to transfer a consistent methodology to volumes and images acquired from three different OCT systems. However, the pre-processing steps that were followed were the same as those steps in [57,67,68]. These pre-processing steps are ad-hoc, for example, estimating the position of the central artifact by computing the difference in average intensities of pairs of columns in the image, and locating the central artifact if the

difference was larger than 6 [68]. These approaches [57, 67–69] do not transfer to different datasets where the characteristics of the artifacts are different. For example, we applied the first pre-processing step in [57, 69] to remove the horizontal artifact in images acquired by a Ultra High Resolution OCT (UHR-OCT) system [70] and a Bioptigen SD-OCT system [71]. This involved computing the mean intensity of each row in the image, and subtracting this mean value from each pixel in the corresponding row. As seen in the UHR-OCT images in Figs. 3.1(c) and 3.1(d), the amplitude of corneal tissue interfaces was drastically reduced in the left and right sides of the image. This SNR drop was significant when compared against the preprocessed result of the Bioptigen system acquired SD-OCT image in Figs. 3.1(a) and 3.1(b), where the tissue interfaces were still visible. Any previous algorithm would fail to segment the boundary in these images due to the lack of structural information.

In [67, 68], a Gaussian Mixture Model (GMM) was used to localize the Epithelial and Endothelial boundaries. Next, the OCT image was contrast enhanced, and the gradient of the contrast enhanced image was computed in order to estimate the location of the Bowman's layer. The methodology involved in this work, which is quite complicated, incorporated a GMM, a normalized cut graph segmentation algorithm, and a level set to refine the results. Furthermore, it was only tested on datasets acquired from a single SD-OCT system.

The algorithm published in [63, 65] used graph cuts, which incorporated regional and shape terms, to segment the corneal tissue interfaces. They pre-processed the OCT images to remove the central artifact and the Iris of the eye, followed by generating a plausible ellipsoidal region for segmentation using a graph cut through the use of an Entropy filter. As mentioned before, this approach was only proposed for datasets that were acquired by a single OCT system. The approach in [62] proposed a graph based segmentation approach to segment corneal tissue interfaces in OCT images of mice. Although this approach is the

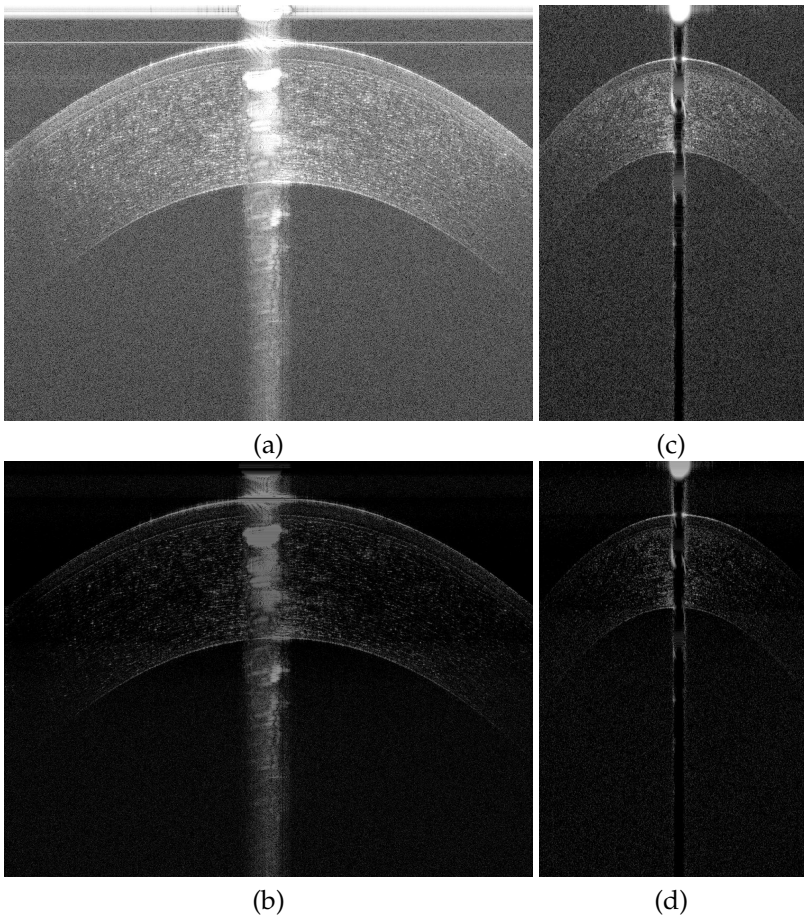


Figure 3.1: (a) and (c) show the original OCT images acquired by a Bioptigen SDOCT system [71] and a hsUHR-OCT system [70] respectively. (b) and (d) show the result of the pre-processing step in [57, 69] after subtracting the mean value of each row in the image from the pixel values in those rows. Notice the drastic drop in intensity of the tissue structure in (d) on the left and right sides of the image, with no boundaries clearly visible.

first to directly segment the Epithelium, Bowman's layer, and Endothelium in the cornea in 3D, it only focuses on segmenting the corneal interfaces in the middle of the radially acquired OCT images. Furthermore, it does not deal with datasets with large amounts of noise and visible artifacts. Moreover, prior algorithms do not segment the interfaces in images acquired by different OCT scanners with different settings.

While it would be useful to incorporate 3D information into an algorithmic approach

to segmentation, often times, patient motion can prohibit the utilization of such spatio-temporal cues. For example, if a Kalman filter is initialized with its state vector consisting of a parameterized curve representing a corneal tissue interface for an image at time t, and if large motion occurs at time step t+1, then the Kalman filter will incorrectly predict, track and segment the same corneal tissue at time step t+1. In such scenarios, it will be useful to know when large motion has occurred, ignore temporal cues, and focus on the spatially available information for segmenting the corneal tissue interfaces. Ultimately, it is vital that reliable and consistent segmentations of the corneal tissue interfaces be obtained for images across different OCT machines in order to measure the corneal thickness, corneal power, and other biometry measurements.

To date, there have been limited deep learning based approaches to segment the corneal tissue interfaces in OCT images of the eye. There have been two independent and concurrently published deep learning-based corneal interface segmentation approaches [5, 72]. One of these approaches [72] 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 [5], where we utilized the entire OCT sequence from multiple scanners containing strong specular artifacts and low SNR regions, and successfully segmented three tissue interfaces.

3.2 Visualizing the Palisades of Vogt

The premise in previously published approaches has been to show inter-modality agreement in imaging and visualizing the same POV configuration, meaning that OCT cross-sections of the POV were compared against histological slices [36, 37, 39] or confocal microscopy images [34, 38, 40, 73]. The main purposes of these previous approaches was to establish a correlation between the features seen by In-Vivo Confocal Microscopy (IVCM) or from histological slices, and the features visualized using SD-OCT. These approaches typically extracted C-mode cross-sections in the coronal plane (after the reconstruction of a volume of OCT images, which were acquired in the transverse plane), and compared them against images obtained using IVCM or with histological slices. The C-mode slices were typically automatically extracted by the SD-OCT system that was used for scanning the tissue in the first place. For example, in [73], the Avanti RTVue-XR SD-OCT system (Optovue Inc.) was used to image the limbal area, and extract C-mode cross-sectional images.

Typically, C-mode images were automatically pulled out of the volume by two boundaries, which were set to be parallel, in the coronal plane, to the ocular surface. The distance between these two boundaries corresponded to the number of slices that would be averaged to generate a C-mode image [73]. Depending on the SD-OCT system that was used, the default value of this thickness was different; for example, the Avanti RTVue-XR SD-OCT system had a default thickness value of $31\,\mu\mathrm{m}$. These two boundaries were manually adjusted to acquire C-mode images at different depths. Images were acquired $\sim 70\,\mu\mathrm{m}$ below the corneal/conjunctival surface, as it could be compared against prior work [35, 39, 73].

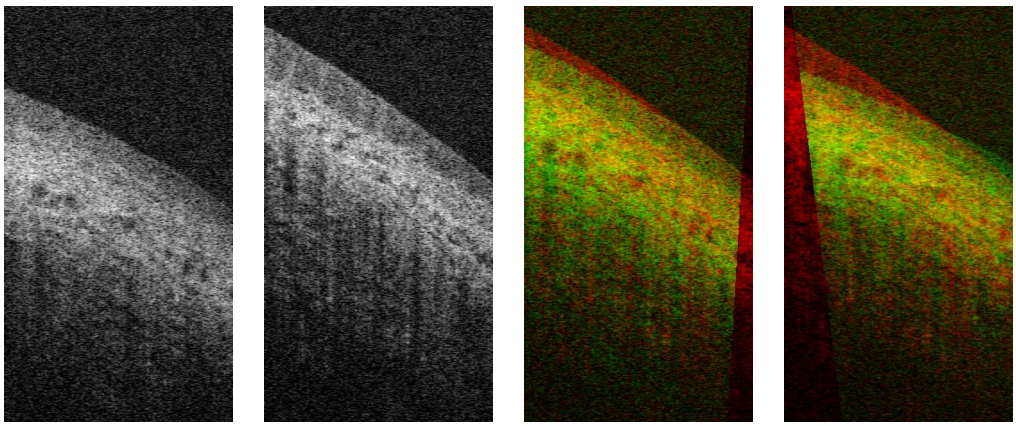
These prior approaches allowed easy visualization of the POV with SD-OCT, and cross-correlation with IVCM or histological slices. From these clinical approaches, observations were made such as: the POV were most predominantly visualized in the inferior and su-

perior quadrants of the eye [38, 73]; higher visibility of the POV was seen in subjects of younger ages, with hyper-pigmentation of the limbal area, over older patients; very low POV visibility in severe dry eye patients when compared against patients with moderate dry eye disease or normal subjects [73]; the anatomical status of the limbal region could be assessed to follow-up on the diseases/conditions that affect the limbal area, such as dry eye disease, contact lens wear, or ocular surface surgery [34, 36, 39, 73].

Currently, IVCM is the gold standard method to obtain information on vascular changes or cell size, inflammatory infiltrates, and morphological characteristics in the limbal region of the eye [34, 36, 73]. IVCM can be used to study the cellular integrity of the POV, the morphology of cells and smaller structures such as focal stromal projections. These cellular characteristics of the POV are presently unable to be studied by SD-OCT, as it can only provide an idea of its presence and density, but not the "quality" of the structures [73]. However, no work has been done before in comparing POV cross-sections extracted from volumes acquired by different OCT imaging systems. Doing so will reinforce the correlation between POV cross-sections extracted using the same imaging modality, namely OCT, and demonstrate the ability of OCT to reveal consistent structures in the limbus. Moreover, for an approach to be clinically useful, fast and consistent visualization of the POV imaged using SD-OCT is needed, which is currently unavailable.

Off-the-shelf built-in algorithms in SD-OCT scanners such as the Optovue Avanti scanner, to extract C-mode slices are not easily available for comparison due to patents on these algorithms. To date, there are no publicly available approaches to extract C-mode slices from SD-OCT volumes that image the ocular limbus and the POV. In general, visualization of the POV requires the extraction of cross-sections from registered volumes, which entails individual B-scan alignment [34]. While advances in POV visualization exist in lit-

erature [34, 36, 39], they employ highly laborious approaches. In [34], stacks of OCT volumes had to be aligned manually, and cross-sections were then extracted to visualize the Palisades. In [36], an automated algorithm *and manual alignment* was used to create 3D volumetric views of the POV in the limbal area; however, no details were provided on the registration algorithm adopted in their paper. To date, the only prior published alignment protocol [34] is not fully automatic, does not include a training protocol, and requires substantial subjective human judgement. Pure image-based registration approaches can also be used for the alignment of B-scans in a volume, such as the multi-resolution registration approach in [74] (available as the StackReg plugin in ImageJ), and the multi-resolution mutual information based registration approach in [75, 76]. However, both image-based approaches show some mis-registration in the B-scans (see Figs. 6.12(c) and 6.12(d)), and subsequent cross-sections extracted from volumes generated by these methods would be inaccurate. Through these approaches, neither the structures inside the limbus, nor the actual visible boundaries, were aligned consistently.



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

Figure 3.2: (a)-(b) The 1^{st} (target) and 150^{th} (reference) B-scans in a volume; (c) a false color overlay of the StackReg [74] registration result; (d) the mutual information-based [75] registration result. Note the lack of consistent alignment between (a) and (b) - neither the structures inside are aligned, nor the shallowest visible limbal tissue interface.

To date, there have been no deep learning based approaches to segment the outermost limbal tissue interface or the Palisades of Vogt in SD-OCT images of the eye. A lot of research has been devoted to segmenting multiple retinal layers and quantify the morphological changes the retinal layers undergo in patients with diseases such as diabetes [77], Age-related Macular Degeneration (AMD) [78, 79], Polypoidal Choroidal Vasculopathy (PCV) [80], and macular edema [81]. Most of these approaches [77,79,81] utilize an artifical neural network that is similar to [82], which uses a convolutional neural network (CNN) architecture called the U-Net. Other approaches [80] are based on the fully convolutional neural network (FCN) [83], while some are based on the LeNet [84]. In our experiments, the LeNet and U-Net do not have the discriminative power to segment corneal tissue interfaces.

To this end, we proposed the first rapid and fully automated method to visualize the POV imaged using SD-OCT [3]. The limbal interface was first segmented in each B-scan of a volume, and it was used to guide the registration of each B-scan to a reference B-scan. Doing so allowed a registered 3D volume to be obtained, from which POV cross-sections could be extracted. This preliminary work also incorporated the first fully validated protocol [3] to assess the similarity between POV cross-sections extracted automatically from OCT volumes, and POV cross-sections that were extracted after manual alignment of B-scans in the OCT volumes. This preliminary work was limited to volumes acquired from a single OCT imaging system; other OCT imaging systems exhibit different properties while imaging tissue, such as intensity changes, increased speckle noise, and specular reflection. This preliminary approach would not work well when the amplitude of the artifacts were high in the OCT images. Furthermore, a full en-face view of the POV around the entire limbus was not generated, nor has it been accomplished by anyone to the best of our knowledge.

3.3 Vessel Segmentation and Tracking

There have been several approaches to segmenting the boundaries of vessels in 2D ultrasound images [85–88]. However, most of these approaches are only suited to single 2D ultrasound images. For our purposes, it is necessary to segment and track the vessel boundary as it evolves in a video sequence that is usually acquired during an imaging session. To this end, traditional ultrasound based vessel segmentation and tracking has been researched before [89–93]. However, when tested on UHFUS images, these gradient-based edge detection approaches failed to detect and track the vessel boundaries in the presence of higher speckle noise. Furthermore, precise delineation of the deforming vessel is required for vessel-based measurements, whereas prior approaches [89–93] modeled the vessel as an ellipse without accounting for the deforming vessel contour.

A recent approach in [93] was designed for a specific imaging setting of 55% maximum gain, but when applied to UHFUS sequences, it failed to track vessels (wrong location with 0% overlap) regardless of gain settings as shown in Fig. 3.3. Furthermore, this approach was validated on limited data.

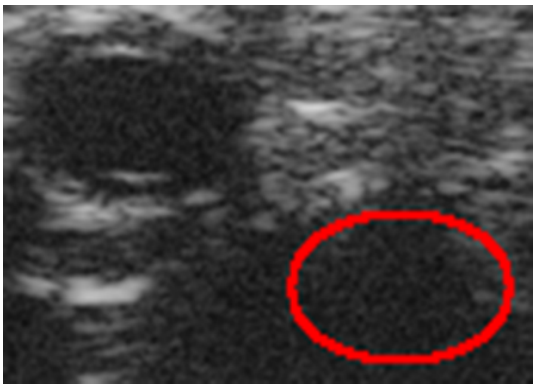


Figure 3.3: Failed vessel detection result (red ellipse) of algorithm in [93] on an UHFUS image. The actual vessel is present in the upper left corner.

Many approaches have been published in literature regarding the segmentation of ves-

sels from ultrasound sequences using level-set active contour frameworks [85–87, 94–97]. However, these frameworks are difficult to implement in real-time, and they are not robust to initialization problems. Initialization of a level set substantially constrains the range of possible solutions. The choice of a level set framework determines the complexity and associated runtime of the algorithm. For example, a recent image entropy based level-set based approach [88] designed for HFUS images ran slowly at 0.5 seconds per image on modern CPU hardware.

Learning-based vessel segmentation algorithms have also been proposed that take advantage of deep neural networks [98–101]. However, these methods make underlying assumptions about the methodology. For example, in [100], although the approach is sound, datasets were acquired from only one ultrasound machine, which had significantly lower speckle noise corrupting the vessel boundaries. Moreover, only longitudinal scans in the sagittal plane were considered in the estimation of the carotid intima-media thickness, and transverse cross-sections were not acquired and analyzed. In [98], candidate locations possibly containing vessels in ultrasound images, determined using their approach published in [93], were classified by a deep neural network as containing the vessel or not. However, as mentioned before, the approach in [93] only worked when the gain setting on the ultrasound machine was 55% of the maximum gain, and as it is the underlying basis approach for the method in [98], it would most likely fail for other settings or on datasets acquired from different machines. On the other hand, the methodology proposed in [101] centered the vessel candidate region before feeding the image into a neural network for detection of the vessel boundaries. It did not automatically take the entire image, and provide segmentation results for the vessel present in the image. Furthermore, both these approaches are significantly slower than the real-time vessel segmentation and tracking methods [89–93].

Chapter 4

Segmentation of Corneal Tissue Interfaces

4.1 Classical Approach

4.1.1 Problem Statement

For an anterior segment SD-OCT image I containing four corneal tissue interfaces in a volume of images, the problem can be considered as assigning every pixel in the image a specific label $L = 1 \dots K$. As shown in Fig. 4.1, we can define the segmentation task as assigning every pixel in the image I a label belonging to a particular tissue interface with $K = 5$. This includes the four corneal tissue interfaces (Epithelium, Bowman's layer, Bowman-Stroma interface, and Endothelium) and the background. Similar to Sec. 5.1.1 and 5.2.1, each B-scan is acquired independently in a SD-OCT volume, with no external tracking mechanism or a guaranteed global alignment available. Segmenting the corneal

tissue interfaces should allow corneal biometry measurements, such as corneal refractive power, to be estimated.

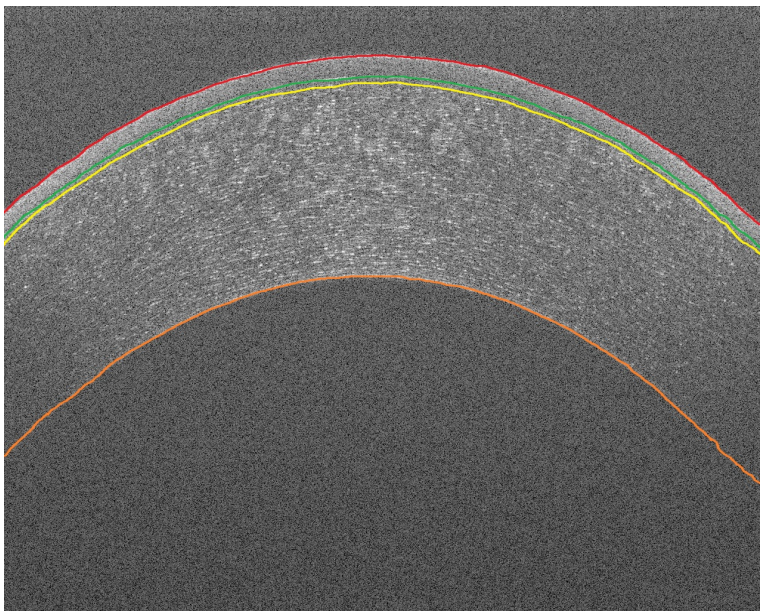


Figure 4.1: An expertly annotated anterior-segment SD-OCT B-scan showing the four different interfaces: Epithelium (red curve), Bowman's Layer (green curve), Bowman-Stroma Interface (yellow curve), and Endothelium (orange curve).

4.1.2 Data Acquisition

We obtained ten deidentified volumes consisting of 50 B-scans each from our clinical collaborators, and each B-scan's dimension was 1000×1024 pixels. The volumes contained B-scans from both eyes of human subjects. The volumes were acquired by a custom-built Bioptigen system [71] with a 100 nm bandwidth light source centered on 870 nm, yielding a theoretical axial resolution of 3.5μm, and scanning 6×6 mm areas of the cornea.

4.1.3 Noise Reduction

Gaussian Blurring: We applied Gaussian blurring to the original B-scan using a small kernel size of 3×3 pixels with σ_x=1.5 in the X-dimension and a small σ_y=0.1 in the Y-dimension. Small values were applied to only smooth out the noise along the X-dimension, while maintaining edge definition along the Y-dimension (see Fig. 4.2). A large value for σ_x removes a lot of structural detail in the X-dimension, and a large value for σ_y reduces the intensity of the corneal tissue interfaces, and thereby affects the detection of the interfaces after blurring. There is no noticeable difference in Gaussian blurring over median filtering for small kernel sizes as shown in Fig. 4.2. However, for large kernel sizes, median filtering significantly reduced the corneal tissue edge intensities. In our algorithm, we utilized Gaussian blurring as it is amenable to parallelization using GPU kernel convolutions.

Horizontal Artifact Reduction: We reduced the effect of the horizontal striations visualized in the B-scans, while trying to preserve the foreground tissue intensity by following the approach proposed in [57]. First, we computed the mean pixel intensity value of each row in the SDOCT image. Then, we subtracted the estimated mean of each row from all the pixel intensities of that row. Finally, after subtraction, we clamped all values below zero to the grayscale range of 0 to 255. All the pixels in the rows saturated by horizontal artifacts contained similar pixel intensities that were closer to the brighter end of the grayscale spectrum. In contrast, the rows not affected by the horizontal artifacts had lower mean intensity values as the pixel values in those rows would range the entire grayscale spectrum. This meant that the subtraction of the row mean from all pixels in each corresponding row significantly reduced the pixel intensities in rows affected by the horizontal artifact, while slightly lowering the pixel values of other rows, thereby preserving structural detail of the corneal tissue. Thus, the adoption of this approach resulted in efficient noise and horizontal

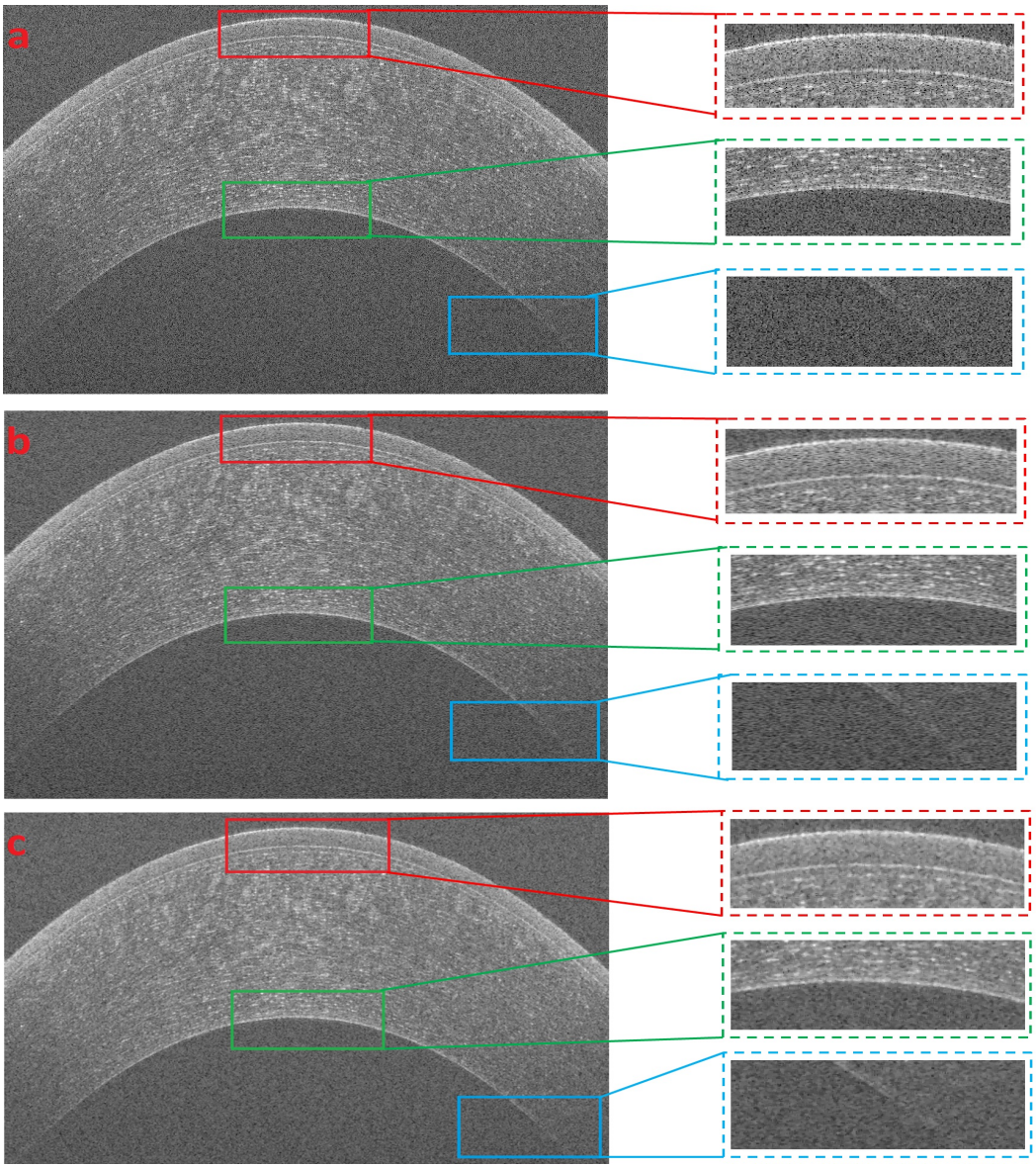


Figure 4.2: a) The original SDOCT B-scan. b) The Gaussian blurred image smoothed using a 3×3 pixel kernel with σ_x=1.5 and σ_y=0.1. c) The median filtered image smoothed with a 3×3 pixel kernel. The colored and dashed boxes show the zoomed regions in each image. There is no noticeable difference in the filtering done by a median filter over a Gaussian filter. However, the filtering effect is more profound in the median filtered image when the kernel size becomes larger.

artifact reduction as shown in Fig. 4.3(b) and 4.3(d).

Vertical Artifact Removal: The vertical artifact severely affects a few of the B-scans in a volume. This is due to the scanning pattern of the OCT light beam as it scans closer to the corneal apex, resulting in greater back-reflection from the structures beneath the cornea. We mitigated the effects of the vertical artifacts in the filtered image obtained from the previous step by adopting a variation of the approach put forth in [57]. First, we computed the mean pixel intensity of each A-scan (column) in the image. Next, we estimated the average of the mean values of all the A-scans in the image. Finally, we removed the vertical artifacts by setting to zero all the pixels in the columns whose mean values were greater than twice the average of means. The result of vertical artifact removal is shown in Fig. 4.4.

The A-scan means represented the columnar mean distribution across the image. Columns unaffected by the vertical artifact would have smaller mean values, in contrast to the columns affected by vertical artifacts, which would have mean values close to the higher intensity ranges of the grayscale spectrum. So, the value arising from the estimation of the average of the columnar means will be biased towards the columns that are not affected by the vertical artifact. By using this value, we can selectively detect columns in images affected by vertical artifacts whose means are higher than average of the means, and remove them as they do not contribute to the detection of the corneal tissue interfaces. Therefore, the vertical artifacts are robustly isolated and removed from the image, as seen in Fig. 4.4.

4.1.4 Gradient Computation

We computed the gradient of the filtered image after the horizontal artifacts are reduced and the vertical artifacts are removed. In particular, we use the method of central differences (see Eq. (4.1) below) to compute the gradient of the filtered image in the vertical direction

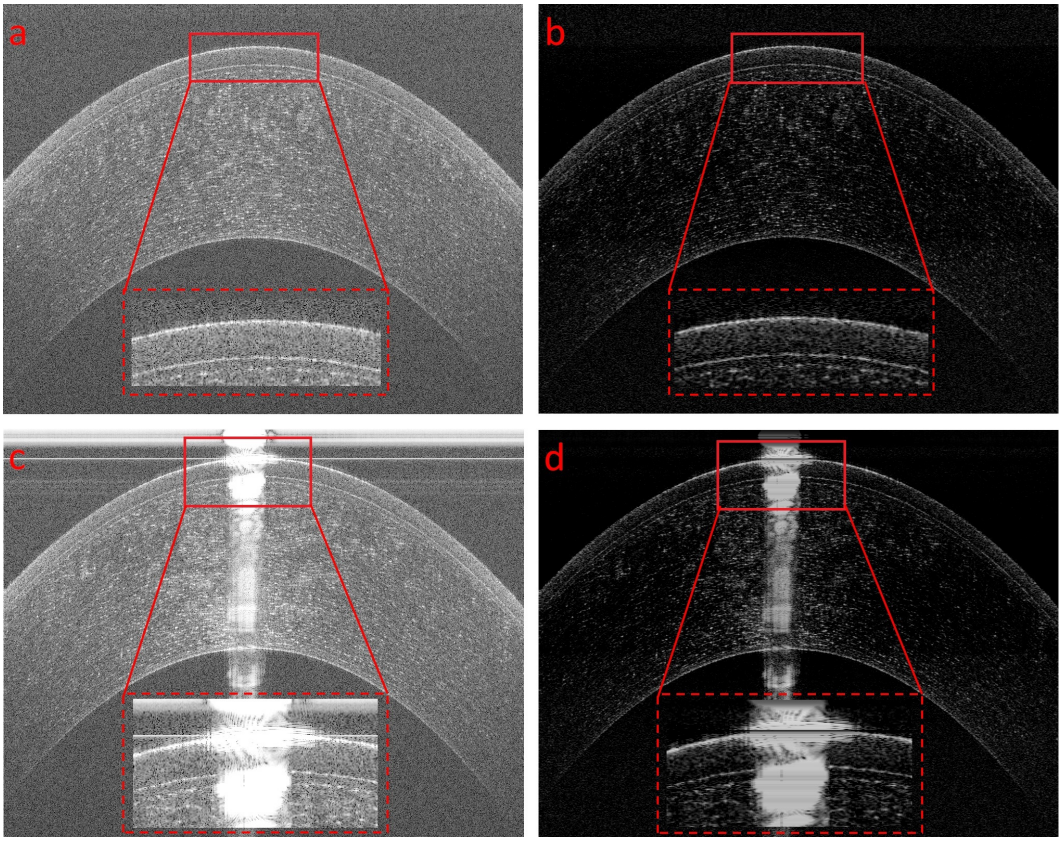


Figure 4.3: Reduction of the horizontal artifacts in the SDOCT image. The original SDOCT images are shown in a) and c), while the horizontal artifact reduced image is shown in b) and d). The zoomed regions inside the red boxes show the effect of the horizontal artifact reduction.

by convolving the image with a vertical kernel such as [-1; 0; 1]. Given the filtered image F of dimension W$\times$H, the value at a pixel [i,j] in the gradient image is estimated by:

$$G_{i,j} = \frac{F_{i,j+1} - F_{i,j-1}}{2}, i \in W and j \in H \tag{4.1}$$

We use the vertical gradient of the filtered B-scan (see Fig. 4.5) to detect two of the four corneal tissue interfaces. That is, of the four interfaces that we detected, the Epithelium and the Endothelium were detected using two segmentations of the filtered B-scan, specifi-

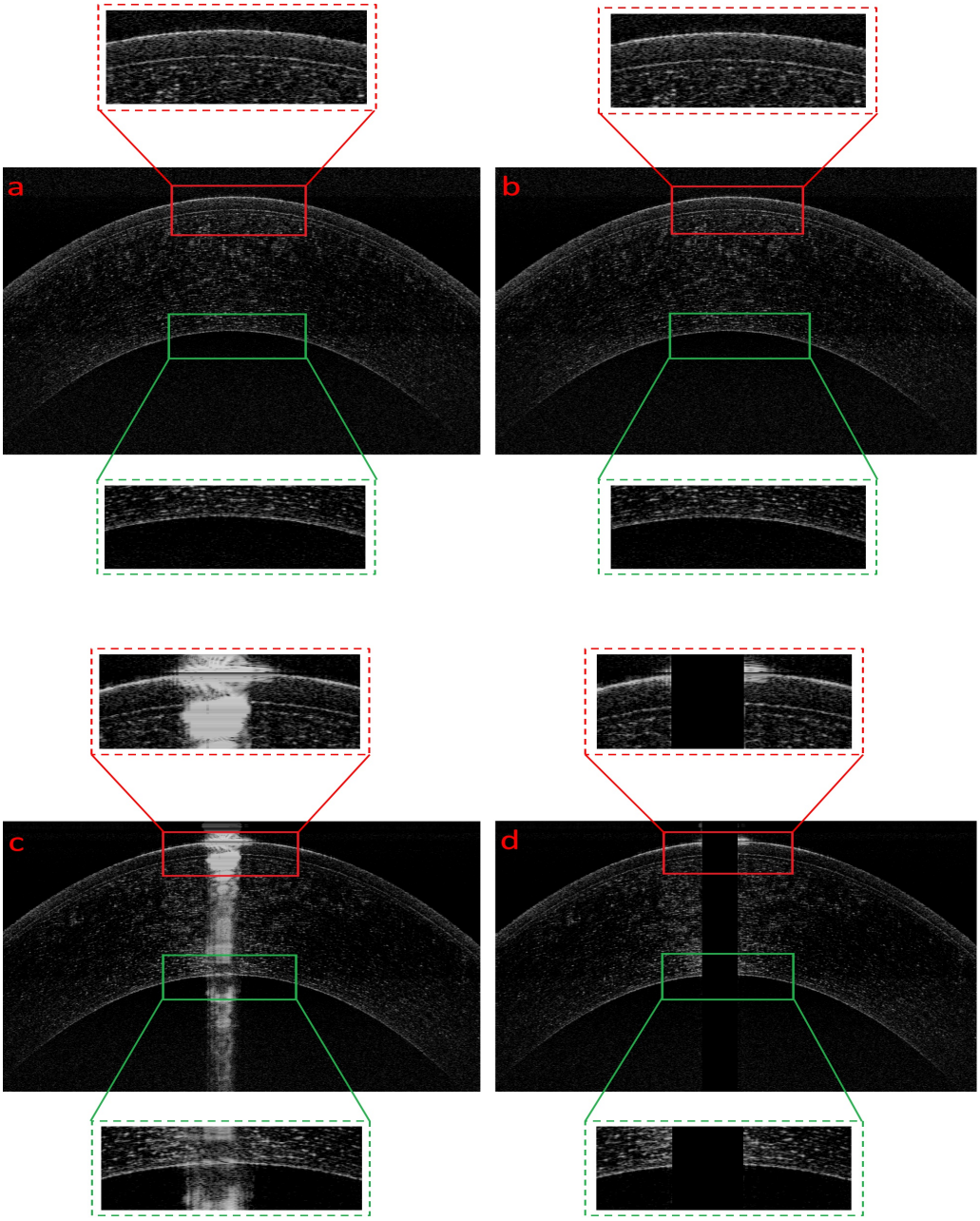


Figure 4.4: Removal of the vertical artifacts in the SDOCT image. The horizontal artifact reduced SDOCT images are shown in a) and c), while the vertical artifact removed images are shown in b) and d). Note that in a) and b), there are no vertical artifacts present in the horizontal artifact reduced image. Hence, the vertical artifact removed image is the same as the horizontal artifact reduced image.

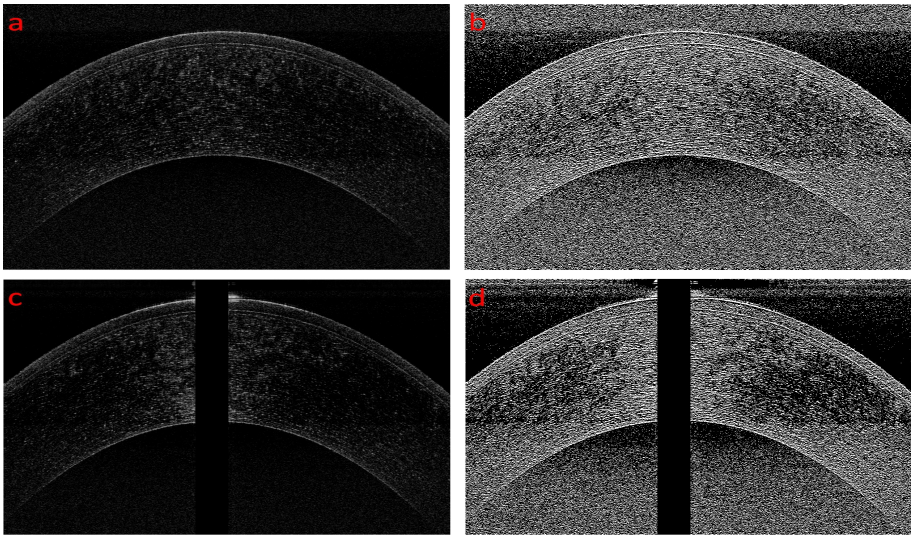


Figure 4.5: The filtered images are shown in (a) and (c), while the gradient of the filtered images are shown in (b) and (d).

cally an over-segmentation and an under-segmentation. However, Bowman's layer and the Bowman-Stroma interface were detected using the gradient of the filtered image. Henceforth, we represent a pixel at a location $[i, j]$ in an image with subscripts, as either $I_{i,j}$, $F_{i,j}$, or $G_{i,j}$.

4.1.5 Image Segmentation

We begin by computing two segmentations of the filtered image - an over-segmentation and an under-segmentation. To this end, we needed an estimate of the foreground's lower intensity range, and this was computed by using the vertical gradient of the filtered image. The vertical gradient of each A-scan in the filtered image will provide a distribution of intensity differences as we proceed axially along each A-scan. Specifically, as we proceed axially along an A-scan, many regions that may potentially represent a tissue interface boundary in the filtered image will transition from a lower pixel intensity to a higher pixel intensity. In the ideal case, at the exact location of an interface, there will be a corresponding high

gradient value. However, as tissue interfaces are often fuzzy in SDOCT images, a band of pixels in an A-scan (and indeed, across the image) in the filtered image will have higher pixel intensities and represent an interface. As these banded regions are encountered in each A-scan, they will have corresponding positive gradient values. So, when the maximum vertical gradient value for each A-scan in the image is extracted, the list of maximum values will represent the pixel intensity range of the foreground of the image. In this case, the foreground represents the tissue structure. Thereafter, we estimated the mean value of this range as it captures the mean of the distribution of the foreground pixel intensities. This mean value can be calculated as follows:

$$thresh_G = \frac{1}{W} \sum_{w=1}^{W} \max_{h \in H}(G_{w,h}) \tag{4.2}$$

where W and H represent the B-scan's width (number of A-scans) and height (pixels per A-scan). After we estimated $thresh_G$, the over-segmentation was generated by utilizing $thresh_G$ directly to segment the filtered B-scan. The under-segmentation was obtained by applying 50% of $thresh_G$ to the filtered image. The empirically determined under-segmentation percentile translated well across all volumes we were provided with.

$$thresh_{over} = thresh_G \tag{4.3}$$

$$thresh_{under} = 0.50 \times thresh_G \tag{4.4}$$

Outlier Removal: The two binary segmentations were then filtered to remove outliers as they affected the segmentation of the Epithelium and Endothelium. We removed outliers by considering the 3×3 neighborhood around each foreground pixel, and determining the number of pixels that were also part of the foreground in that neighborhood. We rejected

those detected pixel locations for which the potential member count was less than a specific number of segmented pixels. We applied this outlier removal method once for both segmented images, and the potential member count in a neighborhood was emperically set to 5, which worked robustly across all the volumes we used for validation.

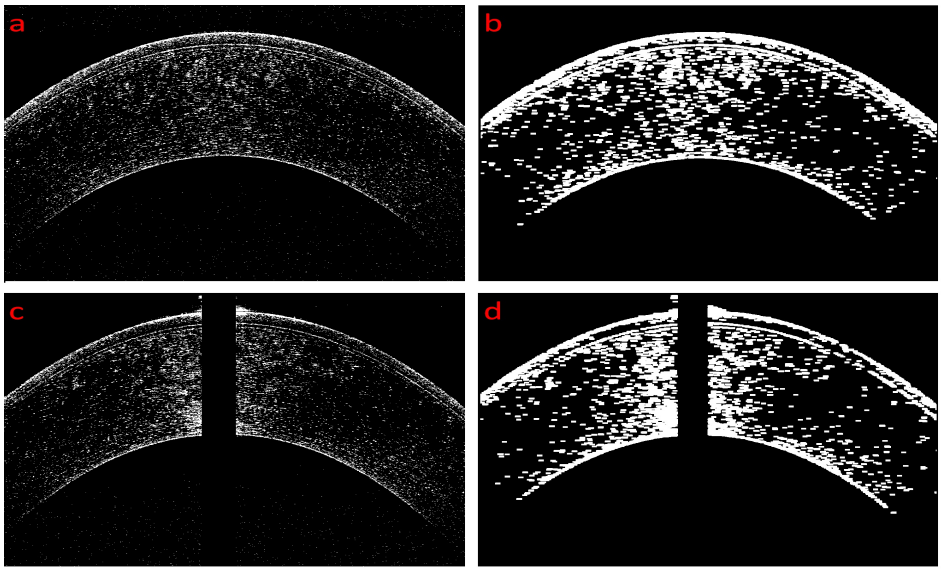


Figure 4.6: The over-segmentations of the filtered SDOCT image are shown in (a) and (c). After outlier removal, the binary image is dilated with a 5×11 kernel, and the dilated images are shown in (b) and (d).

Dilation: After outlier rejection, we dilated the filtered segmentations to fill in the disconnected regions. The segmentations were dilated in order to join regions that were originally detached, but belonged to the same region of tissue. Both binary segmentations were dilated with a 5×11 pixels kernel as we wanted to link regions that were disconnected by bridging across the columns of the image with a large kernel width. The results of segmentation, outlier removal, and dilation are shown in Figs. 4.6 and 4.7.

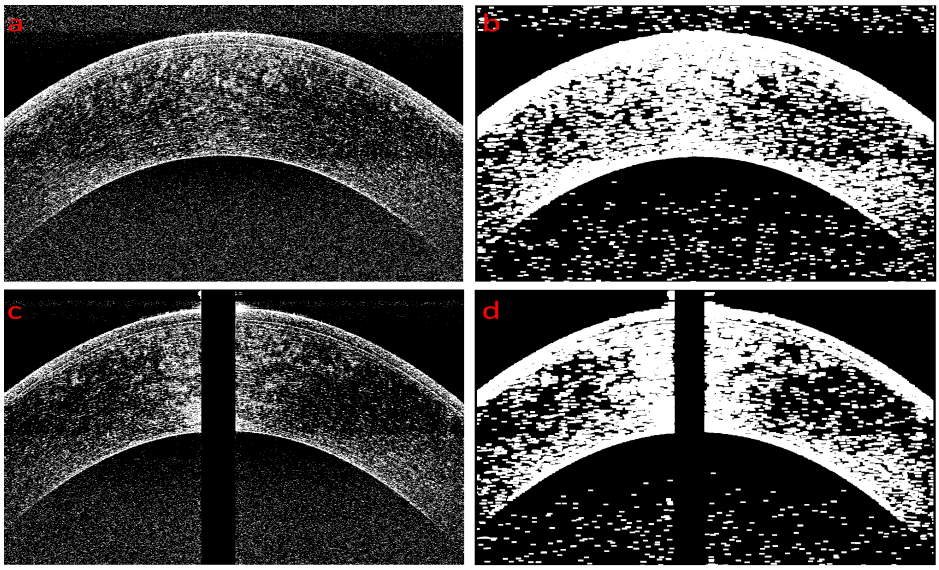


Figure 4.7: The under-segmentations of the filtered SDOCT image are shown in (a) and (c). After outlier removal, the binary image is dilated with a 5×11 kernel, and the dilated images are shown in (b) and (d).

4.1.6 Segmenting Corneal Tissue Interfaces

We use our fast A-scan based approach [1] to search for each tissue interface in the corresponding binary segmentation or gradient image. The Epithelium was detected using the dilation of the over-segmented image, while the Endothelium was detected using the dilation of the under-segmented image. Bowman's layer and the Bowman-Stroma interface were directly estimated from the vertical gradient of the filtered B-scan.

Epithelium

The Epithelium was detected by employing a top-to-bottom search in the dilated image of the over-segmentation. We employed a 7×7 kernel window, which was anchored at the center pixel in the first row of the kernel window. We iterated through all the pixels in each column in the dilated binary image, examined the 7×7 neighborhood around each pixel,

and found the location of the first bright pixel in each column that satisfied $F_{i,j} \geq thresh_{over}$ for all its neighbors. A 7×7 kernel is an adequate choice over a 3×3 or 5×5 kernel to detect this interface. This is because even after outlier removal, the over-segmented image is still affected by a small amount of noise. In addition, the 3×3 or 5×5 kernel sizes are smaller than the 5×11 kernel used to dilate the over-segmented image. So, the layer detection is disrupted when a 3×3 or 5×5 kernel is used. However, the 7×7 kernel is insensitive to this noise, and accurately detects the Epithelium.

Bowman's Layer

Bowman's layer lies underneath the Epithelium. We restricted our top-to-bottom search for this interface in the vertical gradient image to $40\,\mu\text{m}$ - $65\,\mu\text{m}$ below the Epithelium [102,103], and ignored all the pixels above the previously detected Epithelial interface. This heuristic was driven by the anatomical configuration, which was estimated by [102, 103]. As elaborated in the previous sections on image gradient and image segmentation, we searched for a gradual dark-to-bright transition in pixel intensity between the background region and the foreground across all A-scans in the filtered image. This transition represents the Bowman's layer, and it is demarcated in the gradient image by the locations of pixels where the sum of pixel intensities in a small 3×3 neighborhood around each pixel is the highest. The pixel intensities in the region exactly at the location of Bowman's layer in the filtered image will be large, and the gradient values in the neighborhood of the pixels that delineate this interface will consist of positive values. Thus, the sum of the gradients in the neighborhood of pixels that demarcate this interface will have a large positive values. At regions that are not close to the boundary, the sum of gradient values will be close to or lesser than zero as they are not indicative of a clear boundary presence. The 3×3 kernel is anchored at the cen-

ter pixel in the top row of the kernel, and we looked for the first pixel below the Epithelium that met the criterion of having the greatest sum of gradient values.

Bowman-Stroma Interface

This layer lies below the previously detected interfaces, and we again use the gradient image to detect it. The interface that separates Bowman's layer from the Stroma is very faintly visible in the filtered B-scan, and at times, it is very difficult to determine in the original B-scan even for a human observer. We continued with our top-to-bottom search for a gradual dark-to-bright transition in pixel intensity in the filtered B-scan. Again, we restricted our search region for this interface to the region $10\,\mu\text{m}$ - $20\,\mu\text{m}$ below Bowman's layer [102, 103]. As explained in the detection of Bowman's layer, this transition in the filtered image is represented in the gradient image by the locations of pixels where the sum of gradient pixel intensities in a small 3×3 neighborhood is the highest. The 3×3 kernel is anchored at the center pixel in the top row of the kernel, and we searched for the first pixel below the two previous interfaces that met the criterion of having the greatest sum of gradient values.

Endothelium

This tissue interface is the bottom-most interface in the filtered B-scan, and it is the most difficult tissue interface to detect as the SNR at the location of the Endothelium is very low. We used the dilated image of the under-segmentation as it provided the greatest structural detail of the Endothelium (see Fig. 4.7). We executed a top-to-bottom search of all A-scans in the dilated binary image that corresponded to a dark-to-bright transition in pixel intensity in the filtered image. We concentrated our search for this interface in the region $450\,\mu\text{m}$ - $750\,\mu\text{m}$ below the Epithelium [57, 102, 103]. We utilized a kernel of size 11×11, anchored

at the center pixel in the bottom row of the kernel. Through our empirical experiments, we were able to determine that the 11×11 kernel was an appropriate choice for detecting the Endothelium. This was because dilating the under-segmented binary image with a 5×11 kernel caused the region near the Endothelium to grow. As more pixel regions were connected together, we needed a larger kernel in order to identify this interface. Thus, we searched through each column to find the last pixel in that column whose neighbors all satisfied $F_{i,j} \geq thresh_{under}$.

Curve Fitting

Once the boundaries of the corneal tissue interfaces have been identified, they are fitted with a second or third order polynomial, as seen in Figs. 4.8 and 4.9, following the same curve fitting [104] approach in Sec. 5.1.4.

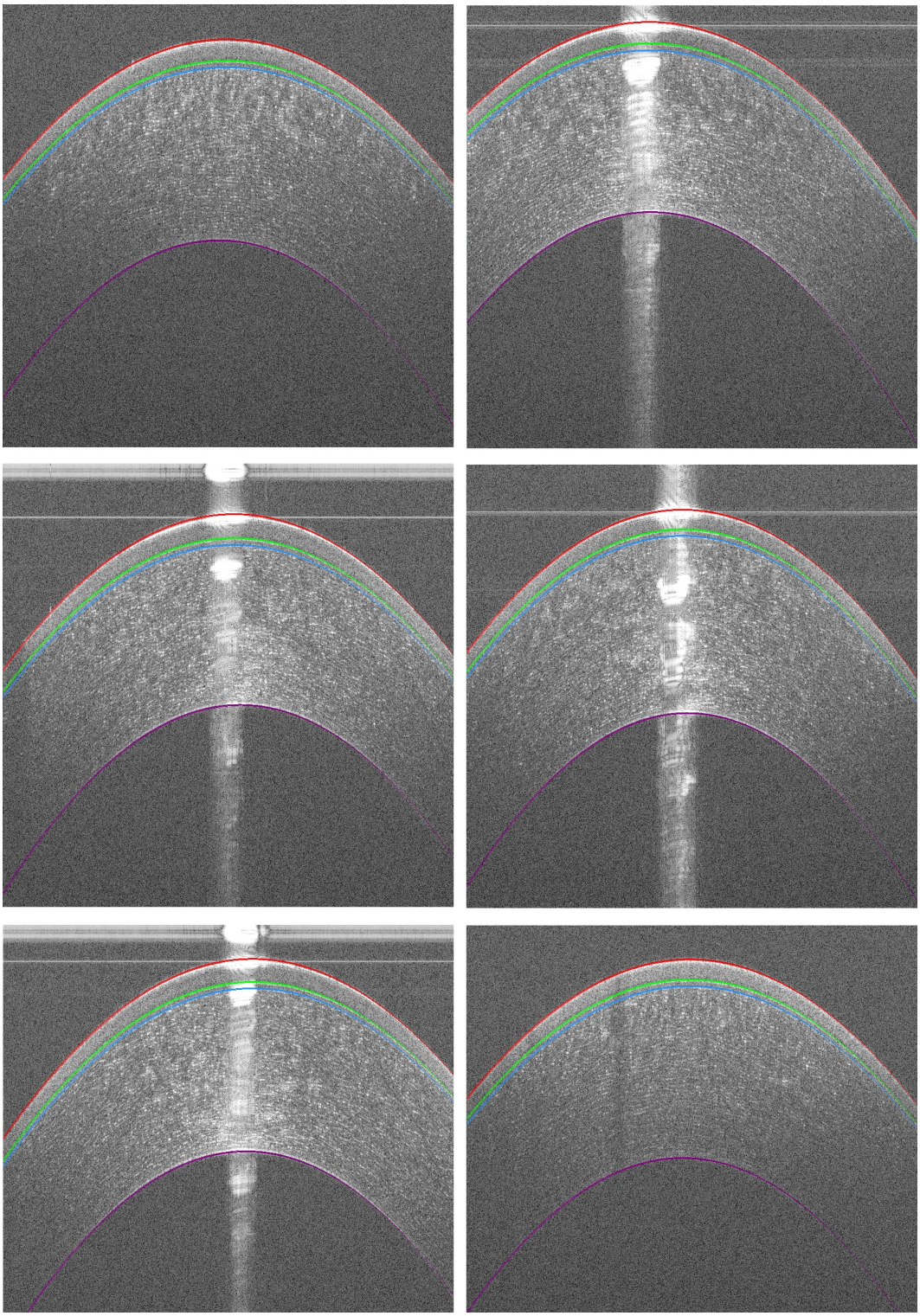


Figure 4.8: Detected corneal tissue interfaces overlaid on the corresponding original images.

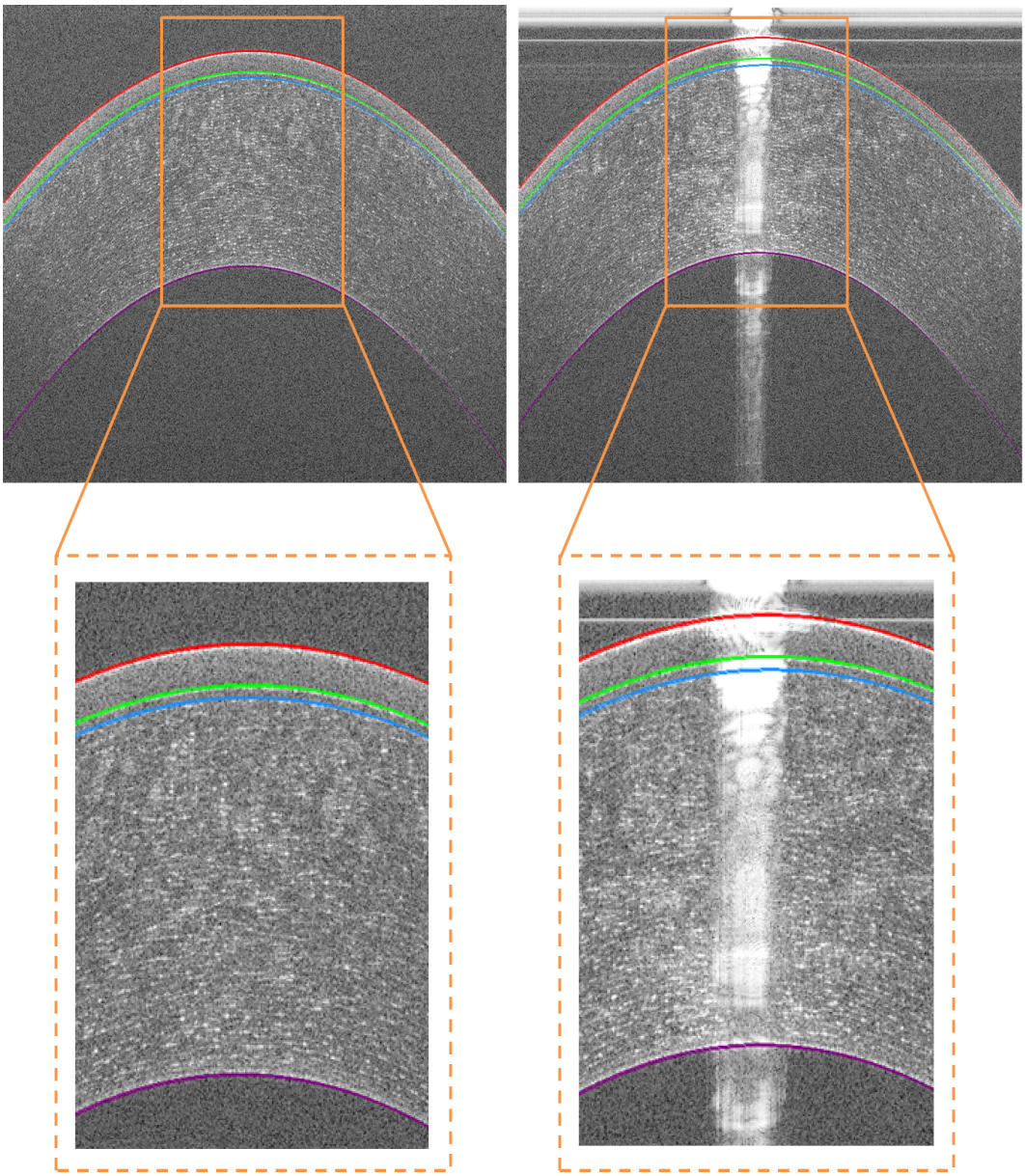


Figure 4.9: Detected corneal tissue interfaces overlaid on the images that have varying degrees of SNR and artifacts.

4.1.7 Results

Comparison of Automatic Versus Manual Segmentation

We validated our algorithm's performance by comparing the automated segmentations against manual annotations of two graders. Two experts selected a subset of 20 images from each of the four datasets, and manually annotated them. We also evaluated the inter-grader annotation variability, where we compared the annotations done by the two expert graders. The 20 B-scans were chosen from each dataset in such a way that they contained images that were severely affected by the horizontal and vertical artifacts. Similar to the human corneal SD-OCT images seen in [57], we believe these artifacts are detrimental to visualizing clear structures. Our quantitative metrics are shown in Tables 4.1, 4.2, and 4.3.

Table 4.1: Grader 1 - Mean and standard deviation differences in pixel locations of corneal interface segmentations between manual annotations done by the first grader against the automatic segmentation results (in pixels)

	Mean	Standard Deviation
Epithelium	1.14	0.35
Bowman's Layer	0.69	0.19
Bowman-Stroma Interface	1.07	0.31
Endothelium	1.36	0.33

Table 4.2: Grader 2 - Mean and standard deviation differences in pixel locations of corneal interface segmentations between manual annotations done by the second grader against the automatic segmentation results (in pixels)

	Mean	Standard Deviation
Epithelium	0.87	0.47
Bowman's Layer	0.61	0.31
Bowman-Stroma Interface	2.94	1.33
Endothelium	1.91	0.81

From Table 4.3, we can see that the two graders were very nearly in agreement with each other on the location of the Epithelium and Bowman's layer. However, the graders agreed

Table 4.3: Inter-Grader Variability - Mean and standard deviation difference (in pixels) between manual annotations done by two different graders

	Mean	Standard Deviation
Epithelium	0.89	0.34
Bowman's Layer	0.62	0.26
Bowman-Stroma Interface	1.95	1.04
Endothelium	1.44	0.59

less with each other on the location of the Bowman-Stroma Interface and Endothelium, as evidenced by the high mean and standard deviation differences for these two interfaces. This is can be attributed to the poor visibility of the boundary of these interfaces. As the annotations performed by the two human graders were reasonable representations of the true locations of each interface in the cornea, we expect that these annotations represent a reasonable target for any real-time or non real-time automated algorithm. Our real-time algorithm was able to achieve this goal.

Similar to [57,68], our real-time algorithm produced automated segmentations of corneal interfaces, which when compared against the manual annotations of each grader, provided mean and standard deviation differences that were close to the manual annotations as well as the inter-grader differences. The tissue interfaces of the cornea were detected correctly by our algorithm for all the validation images, and indeed for all images in all datasets we were provided with. Our algorithm did very well in estimating the corneal structure at regions in images where the tissue interfaces were simply not visible or were affected by artifacts as seen in Figs. 4.8 and 4.9.

Analysis, GPU Performance Metrics, and Run-time

We analyzed the performance of the algorithm on our NVIDIA Quadro K6000 GPU that was integrated in our desktop (Intel (R) Xeon (R) CPU, 3.10 GHz, 64 GB RAM, 64-bit Win-

dows 7 OS). We used NVIDIA's Visual Profiler v4.2 and NVIDIA's Nsight Visual Studio Edition v3.2 to measure timing statistics. Since a CUDA-enabled GPU is efficient for parallel programming and optimization, we allocated dedicated threads for each pixel in the SDOCT image, and each thread was allowed to process that pixel and its corresponding elements.

We define the real-time performance of an algorithm to be the capability of an algorithm to process $\geq$30 frames per second. We ran our algorithm 10 times to measure the runtime range of the program, and it was 3.801$\pm$0.5 seconds on a volume consisting of 50 B-scans. The timing range in which all four tissue interfaces were detected for a SDOCT image in the 50 B-scan volume was $\sim$24.97$\pm$4.32 milliseconds. Normally, our algorithm would allow real-time execution at $\geq$40 B-scans per second. At its best, our algorithm would allow real-time operation at $\geq$48 B-scans per second. If our algorithm takes longer to process each B-scan, then it would allow frame rates of $\geq$34 high-resolution (1000$\times$1024) B-scans per second. The runtime ranges incorporate the time taken to allocate memory on the CPU and the GPU, read images from the hard drive of our desktop, complete memory transfers to the GPU, process each B-scan, overlay the fit curves on the images, and release the previously allocated memory space on the CPU and GPU. We also implemented an optimized CPU-based version of the same algorithm, and the duration of execution of the CPU-based algorithm was 37 seconds. By utilizing the GPU, we achieve an $\geq$9.6$\times$ speedup in performance.

Discussion

We have presented an algorithm that can segment four corneal tissue interfaces on a per-image basis. We detected four tissue interfaces in an 1000$\times$1024 SD-OCT B-scan in $\sim$ 24.97

$\pm$ 4.32 milliseconds, and processed a volume consisting of 50 such B-scans in 3.801 $\pm$ 0.5 seconds. We achieved faster processing times with our algorithm in contrast to the approaches published in [57] and [68], wherein a single B-scan in a volume was processed in 1.13 seconds and 7.99 seconds respectively. In addition, it is not known if the timing measurements in [57] and [68] incorporate the memory allocation/deallocation times. Assuming that they do incorporate memory transfers, this means that the graph-based approach in [57] takes 56.5 seconds to segment three corneal layers in a dataset consisting of 50 B-scans with dimensions 1000$\times$1024, while the fastest method published in [68] takes 399.5 seconds to segment three corneal layers in a volume consisting of 50 B-scans. The size of the B-scans in a volume of 40 B-scans was not mentioned in [68]. This means that for a volume consisting of 50 B-scans, we segmented four corneal interfaces, and achieved a speedup of $\geq$ 14.6 $\times$ over the graph-based method in [57]. From the results in [68], our algorithm provides a speedup of $\geq$ 103.7 $\times$ over their fastest approach for a volume consisting of 50 B-scans. Moreover, we have achieved a $\geq$ 9.6 $\times$ speedup over an optimized CPU-based version of our algorithm. These statistics are shown in Fig. 4.10. Furthermore, the results of our validation are similar to the results obtained by the previously published methods. In our algorithm, we were loading each B-scan from the hard drive of the desktop to the GPU we were working with. In the case of off-the-shelf commercial SDOCT systems that employ GPU's for processing, the individual B-scans after Fourier reconstruction are already present on the GPU. Since the B-scans are already present in GPU memory, it will be straightforward to make efficient use of the GPU bandwidth in order to transfer data between different locations on the GPU, process each B-scan, and achieve greater throughput.

We have not utilized the information from adjacent B-scans in the volume in order to segment the corneal interfaces. This is because we were only provided with ten de-identified

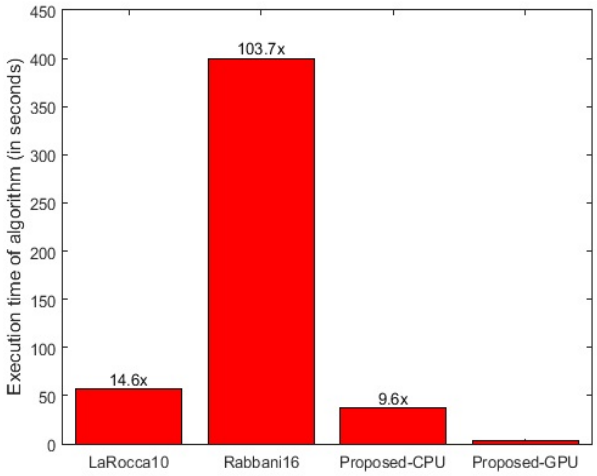


Figure 4.10: The bar graph shows the timing statistics of our algorithm compared against previously published methods.

SDOCT volumes of human subjects, and we did not have the opportunity to acquire multiple volumes at the same imaging location in the same imaging session. Typically, in clinical imaging, the subject's eye will move and cause jitter in the B-scans, thus requiring motion compensation in software. In Figs. 4.11, we show the results of our approach to mitigate the effects of motion in an example SDOCT volume that we were provided with. We registered the Epithelial tissue interface in each B-scan in the volume to the Epithelial tissue interface of the first B-scan in the volume (as seen in Fig. 4.11). Fig 4.12 shows the difference in the position of the corneal apex in a select number of B-scans before and after registration. By following this approach, we were able to align the B-scans in the volume better, and this is shown in Fig. 4.13. We also show the tissue interfaces, before and after alignment, rendered as volumes in Figs. 4.11 and 4.13. We implemented an Iterative Closest Point (ICP)-based [105] registration algorithm in MATLAB (Mathworks 2017a), and the Epithelial tissue interfaces detected from each B-scan in the volume were registered in $\sim$ 8.2 seconds.

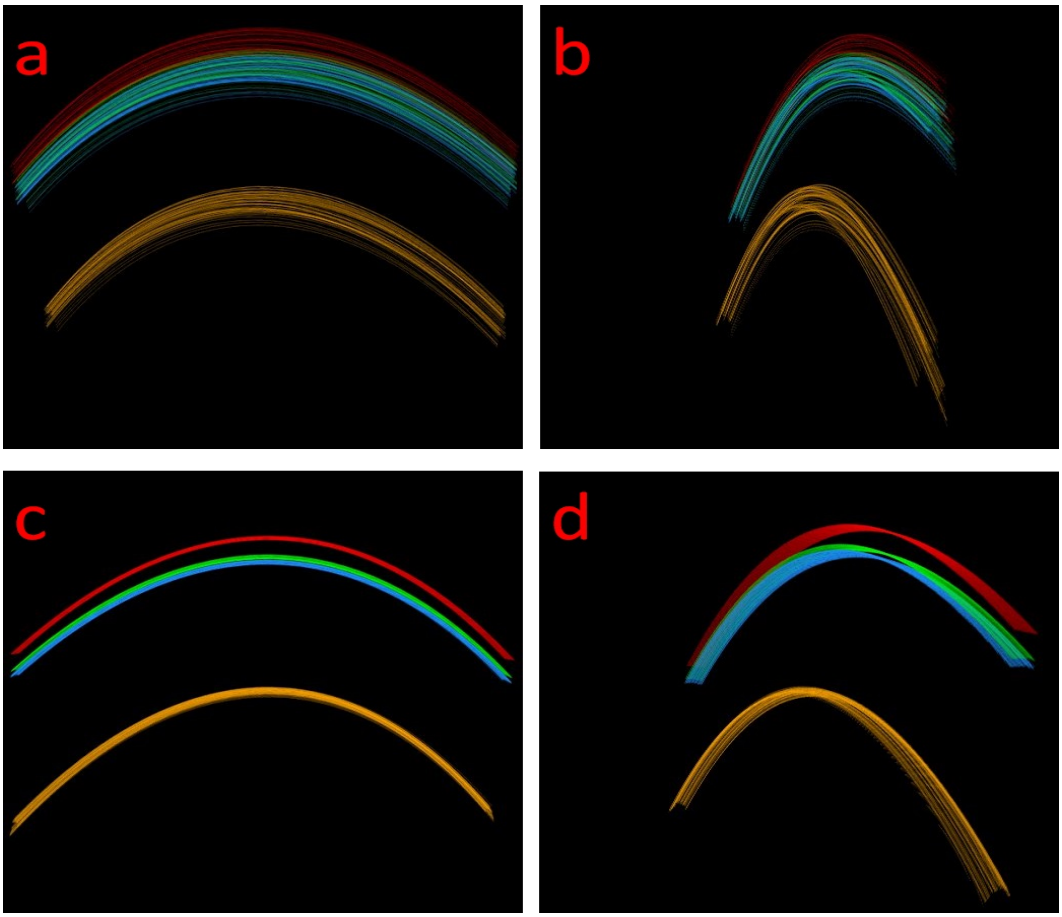


Figure 4.11: In each of the four images above, the detected tissue interfaces in an example volume containing 50 B-scans have been rendered. In (a) and (b), the detected corneal tissue interfaces from each B-scan are rendered together without registration. Note that in (a) and (b), due to the movement of the subject's eye, the tissue interfaces in the rendered volume are completely out of alignment. This is especially visible in (b), which shows that the B-scans in the volume are not registered with each other, with the motion of the subject's eye not being compensated for. In (c) and (d), we show example renderings of aligned tissue interfaces. The tissue interfaces were aligned by registering the Epithelial interface in all B-scans to the Epithelial interface in the first B-scan of the example volume. From (d), it is clearly seen that the tissue interfaces are better aligned than in (b).

Even a fast and simple ICP-based registration algorithm (which cannot preserve out-of-plane curvature (see Fig. 4.12), making the corneal surfaces appear cylindrical rather than spherical) still takes $\sim$ 8.2 seconds to register the 50 B-scans of a single volume. Although the results from our registration approach are promising, it is not amenable to our

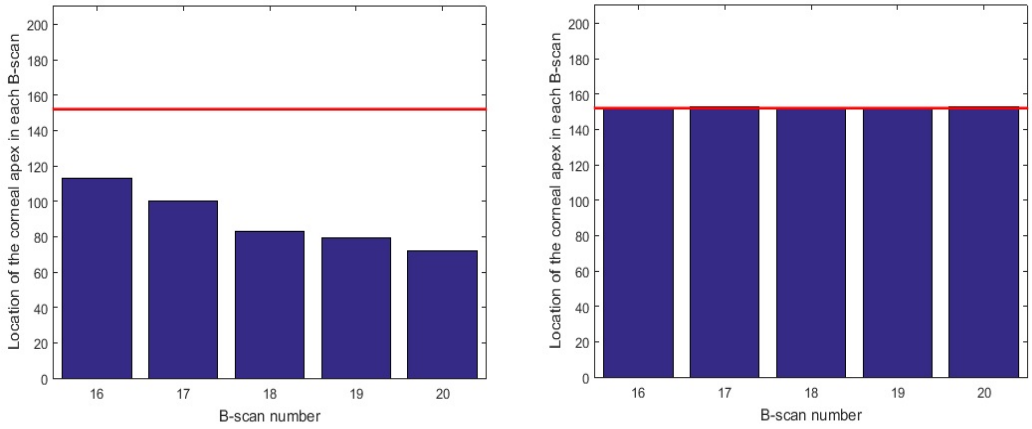


Figure 4.12: In each bar graph above, the location of the corneal apex in a set of five B-scans in the volume is shown; the five B-scans were from B-scans 16 through 20. The corneal apex was adjudged to be the shallowest pixel location in the image among all pixels representing the Epithelium. In the bar graph on the left, the location of the corneal apex in each B-scan was rising with the scanning pattern of the SDOCT system and the movement of the subject's eye. So, the corresponding row number in image that they were detected in is decreasing, as shown in the bar graph on the left. The red line represents the location of the corneal apex in the first B-scan in the volume. In order to compensate for the motion of the subject's eye, the Epithelial tissue interface detected in each B-scan of the volume was registered to the Epithelial tissue interface in the first B-scan of the volume. In the bar graph on the right, we see the effect of registration of Epithelial tissue interfaces. Here, the pixel location of the Epithelium in B-scans 16 through 20 is within one pixel of the location of the Epithelium in the first B-scan of the volume.

real-time approach currently. Accurate motion-compensation registration algorithms are not yet ready for real-time operation. Active research [106] [107] into motion compensation has put forward numerous registration approaches, with varying degrees of human interaction required for good results. An efficient method towards registration of B-scans in OCT volumes has been proposed in [106], but it requires multiple volumes to be acquired in the same imaging session. Other approaches, such as eye tracking [107] for motion compensation, require significant hardware and software modifications in terms of applying tracking signals for eye motion compensation along with signals that drive the OCT galvo mirrors. These approaches cannot be easily modified and integrated into our algorithm,

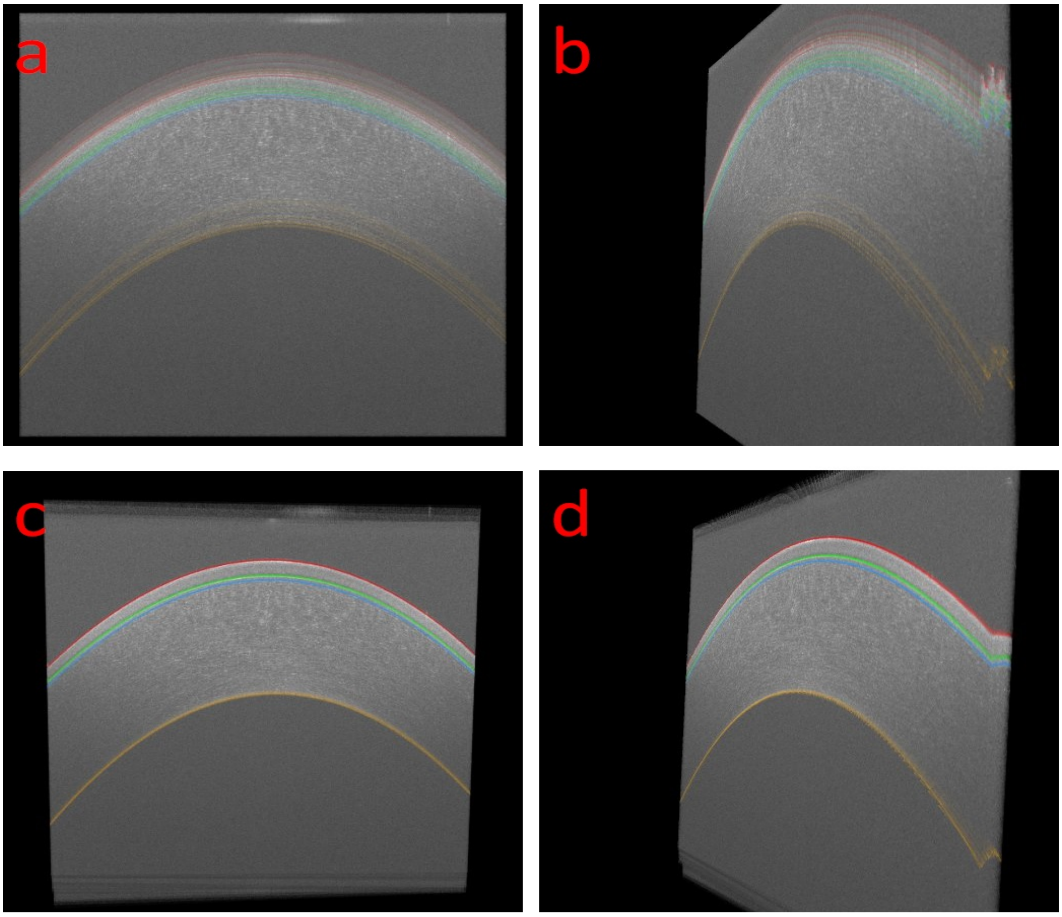


Figure 4.13: In each of the four images above, the detected tissue interfaces are overlaid onto the original B-scans in the volume, and then rendered as 3D volumes. In (a) and (b), we can see that the B-scans are not registered with each other, and the overlaid tissue interfaces are out of alignment. This is easily visible in (b). However, in (c) and (d), the B-scans are aligned by registering the Epithelial tissue interface of all B-scans to the Epithelial interface from the first B-scan in the volume. From (d), it is clearly seen that the tissue interfaces are aligned better than in (b).

and therefore, they are beyond the scope of this work.

The Endothelial curve in Fig. 4.13 is arched slightly, which is due to the registration process. Since we are only registering the Epithelial tissue interface across B-scans, the registration of other detected interfaces is unconstrained. Additionally, in the case of the Endothelium, the SNR decreases with depth, and thus makes registration a difficult task.

As achieving motion compensation through advanced registration techniques is not the core aim of this work, so we leave this for future experiments. In the future, we plan to exploit the positional information of the corneal interfaces in adjacent B-scans. By doing so, we can explore quick and efficient approaches towards registration of B-scans in SD-OCT volumes, such that the shape of the cornea after motion compensation is preserved. In addition, we also plan to build an optimal system that would stream all the B-scans to the GPU in such a way that while a B-scan is currently being processed, the previous and next B-scans in the volume would have been transferred through memory. This would allow our fast frame rates to be achieved.

4.2 Deep Learning Approach

Given a corneal OCT image $\mathcal{I}$, the task is to find a function $\mathcal{F} : \mathcal{I} \rightarrow \mathcal{L}$ that maps every pixel in $\mathcal{I}$ to a label $\mathcal{L} \in \{0, 1, 2, 3\}$. Similar to [57,108], the corneal interfaces to be segmented are: (1) Epithelium, (2) Bowman's Layer, and (3) Endothelium, with 0 being the background.

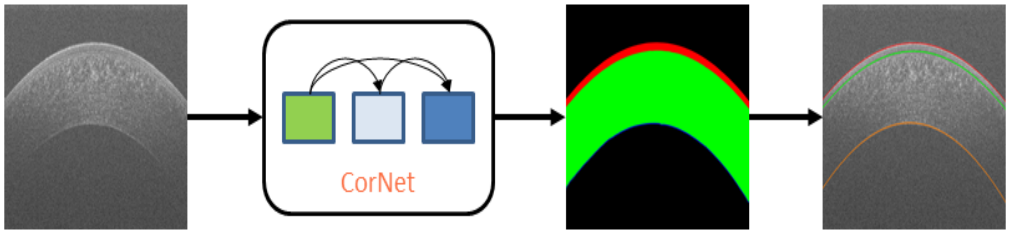


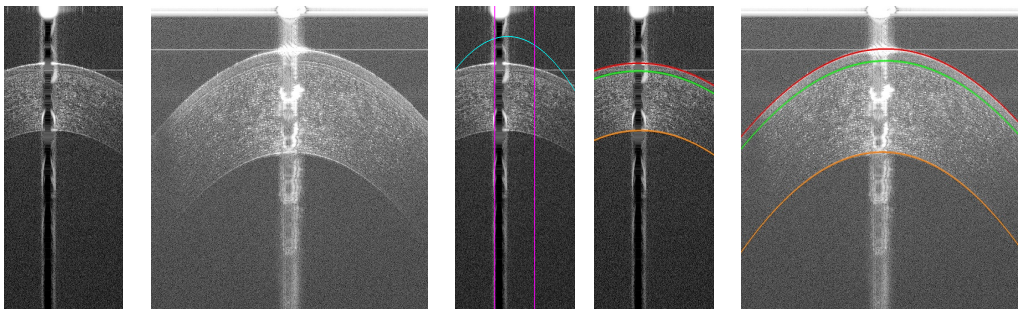
Figure 4.14: Our framework takes as input an OCT image, predicts the location of corneal interfaces using the CorNet architecture, and fits curves to the detected interfaces.

4.2.1 Data Acquisition

De-identified datasets that had been previously acquired for an existing research database were used [3]. 48 volumes from both eyes of 8 subjects were acquired with different scan sizes using two OCT scanners; a Bioptigen SD-OCT scanner (Device 1) [109], and a high-speed ultra-high resolution OCT (hsUHR-OCT) scanner (Device 2) [70]. Device 1 had a 3.4μm axial and 6μm lateral spacing when scanning a 6×6mm area, generating volumes of dimensions 1000×1024×50 (W×H×B-scans) pixels. Device 2 had a 1.3μm axial and a 15μm lateral spacing when scanning a 6×6mm area, and a 7.5μm lateral spacing when scanning a 3×3mm area respectively, yielding volumes of size 400×1024×50 pixels. Each dataset was annotated by an expert grader (Grader 1) and a trained grader (Grader 2).

4.2.2 Corneal Interface Segmentation Network (CorNet)

Fully convolutional networks, such as the UNET [77, 82] and BRUNET [79], are the state-of-the-art in retinal OCT segmentation. Such networks comprise of contracting and expanding branches, providing a dense output where each pixel is assigned the tissue class that it belongs to. The BRUNET architecture [79] overcame problems of the UNET, such as holes in the segmentation, by modifying the UNET architecture. First, dilated convolutions [110–112] were used in Inception-like blocks [112] to increase the receptive field of each layer. Next, batch normalization [113], residual [114] and bottleneck connections [112], and a feature map growth rate governed by a Fibonnaci sequence were incorporated. Finally, the input image was appropriately downsampled and connected to each layer. These changes greatly improved segmentation accuracy [79] over the UNET.



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

Figure 4.15: (a)-(b) Original B-scans from a 3×3mm UHR-OCT and 6×6mm SD-OCT volume; (c) Failed Epithelium segmentation result (cyan) from algorithms in [57,69,108]; (d)-(e) Our segmentation results for Epithelium (red), Bowman's layer (green), and Endothelium (orange) for images in (a) and (b).

However, when applied to corneal OCT images, the BRUNET under-segmented poorly defined corneal interfaces, which are very common in anterior segment OCT imaging. As seen in Figs. 4.15 and 5.21, these boundaries are corrupted by speckle noise, and have low signal-to-noise ratio (SNR). We empirically observed higher false positives in the final seg-

mentation; one explanation is that discriminative features related to these boundaries being learned in earlier layers are lost through the network, and residual connections are unable to recover this information.

One way to combine both coarse and fine image details is through the use of dense connections, which have been used to improve segmentation accuracy by encouraging heavy feature reuse through deep supervision [115–117]. With dense connections, each layer is connected to all its preceding layers by feature map concatenation, allowing discernible features of faint boundaries to be retrieved across multiple scales. But, this comes at a cost of increased computation [117, 118], and we empirically determined that a densely connected network at a depth of 6 levels provides a good balance between segmentation accuracy and computational efficiency [79, 118]. Additionally, max pooling was better at maintaining features of interest through the network over average pooling and convolutions of stride 2 [118]. Furthermore, nearest neighbor interpolation based upsampling followed by 3×3 convolution [119] performed better than bilinear interpolation based upsampling, bilinear interpolation + 3×3 convolution [119], unpooling [77, 120], and fractionally-strided convolutions [83].

In our experiments, we adopted the BRUNET architecture [79] as the base, and modified it based on our observations as shown in Fig. 6.13. Similar to [79], the number of output feature maps in each layer increased according to a capped Fibonacci sequence {32,64,96,160,256,416}, and limit the bottleneck feature map output to 32 to prevent feature map explosion.

Key modifications to the architecture, which we incorporated were: 1) Dense connections were used to improve gradient information flow and prevent over-fitting; 2) Max pooling was used to pick the most discriminative features at the end of each downsampling layer;

3) Nearest neighbor interpolation + 3×3 convolution was used to upsample feature maps in the expanding branch of the network. We name our corneal tissue interface segmentation architecture as *CorNet*.

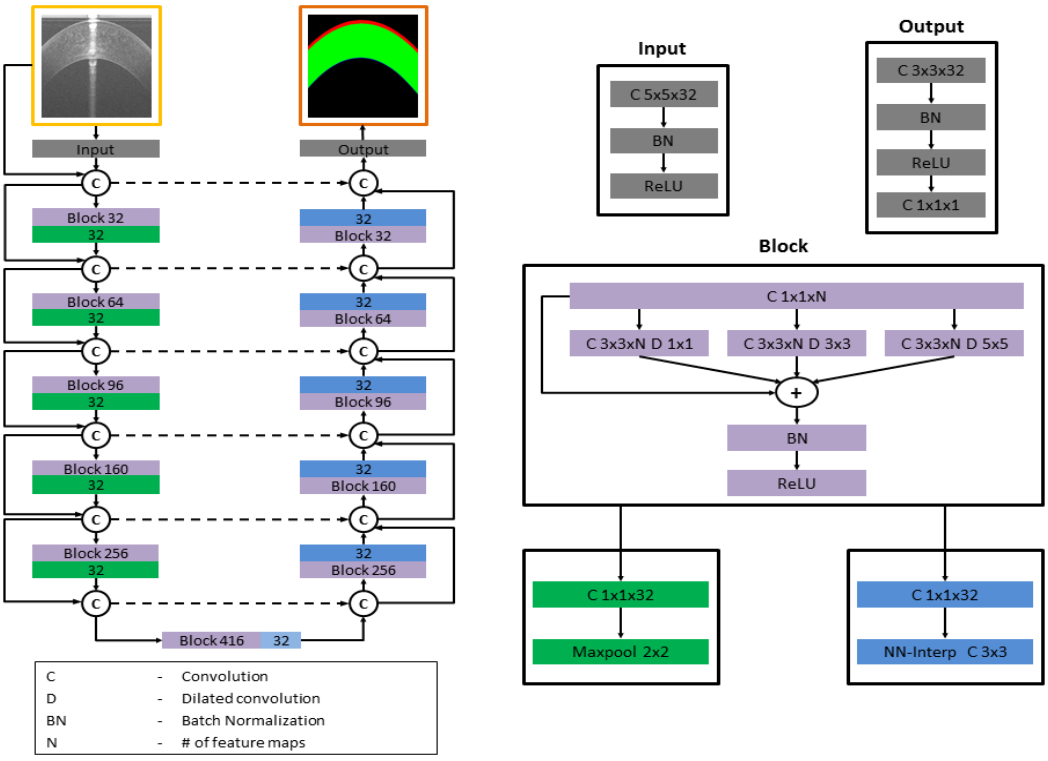


Figure 4.16: Our network architecture comprises of contracting and expanding branches. The dark green and blue blocks represent downsampling and upsampling computations respectively. Our network makes efficient use of residual and dense connections to generate the corneal interface segmentation in the final image, where each pixel is assigned the label of the tissue it belongs to. The input image is split width-wise into a set of slices of dimensions 256×1024 pixels, the network predicts an output for each slice, and the slices are aligned to recreate the original input dimension. Dense connections concatenate feature maps from previous layers. The light blue block at the bottom of the "U" (on the right end of the rectangle) does not perform upsampling, but as with the other blue blocks it functions as a bottleneck for the layer and generates feature maps of the same dimensions as the output feature maps from the previous layer.

4.2.3 Experiments

Setup

Of the 48 datasets, 18 datasets were chosen for training, such that it contained a balanced number of datasets from both devices, i.e., six 6×6mm datasets each from Device 1 and 2, and six 3×3mm datasets from Device 2. The testing dataset comprised of 30 datasets; ten 6×6mm datasets each from Device 1 and 2, and ten 3×3mm datasets from Device 2. 5-fold cross-validation was conducted, and the model from the fold with the lowest validation loss was chosen for testing.

Training

Training a CorNet model with full-width OCT images is limited by available RAM on the GPU and by the varying image sizes obtained from OCT scanners. To address these issues, the input images were sliced width-wise [77] into a set of images of dimensions 256×1024 pixels, thereby preserving the OCT image resolution. Data augmentation [121] is done through horizontal flips, gamma adjustment, Gaussian noise addition, Gaussian blurring, Median blurring, Bilateral blurring, cropping, affine transformations, and elastic deformations. Similar to [79], the loss function used was Mean Squared Error (MSE), and the network was trained using the ADAM optimizer [122]. The batch size was set to 2. The learning rate was set to 10^{-3}, and it was decreased by a factor of 2 if the loss did not improve for 5 epochs. Validation data comprised of 10% of the training data, and the network was trained until the loss did not improve for 10 epochs, at which point we executed early stopping. The network with the lowest validation loss among all the folds was chosen for evaluation on the testing set. The prediction for each interface was then fitted with a curve [3, 57, 108, 123] (see Fig. 4.17).

Baseline Comparisons

We extensively validated the performance of our CorNet architecture; first, we compared our results against those from the UNET [77, 82] and BRUNET [79] architectures as shown in Fig 4.18. Next, we compared our results against those obtained from [57, 108] in Table 4.4; only 6×6mm datasets from Device 1 were used as [57, 108] solely considered datasets of this dimension. Finally, in Tables 4.5 and 4.6, we compared our results against each grader, and also computed the inter-grader variability measures to quantify our deviation from the agreement in ground truth between graders.

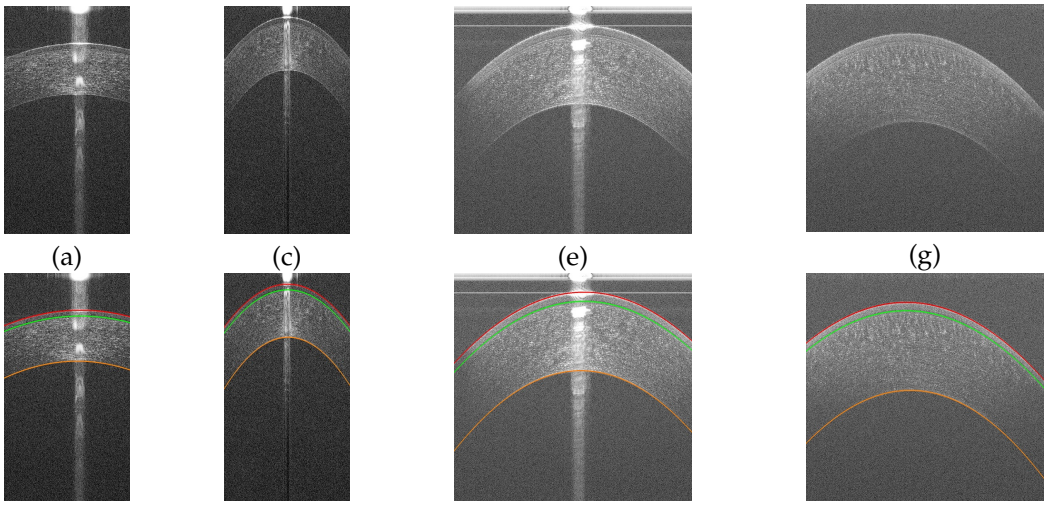


Figure 4.17: Original B-scans and segmented interfaces from different datasets: (a)-(b) 3×3mm UHR-OCT, (c)-(d) 6×6mm UHR-OCT, and (e)-(h) 6×6mm SD-OCT.

4.2.4 Results

Metrics

We computed the following metrics: 1) Mean Absolute Difference in Layer Boundary Position (MADLBP) and 2) Hausdorff Distance (HD) between the fitted curves. For consistency

in comparison, we computed MADLBP as it was the metric (in pixels) of choice in [57, 108]. However, MADLBP (Eq. 6.1) does not accurately quantify the distance error in microns between a particular pair of interfaces, which the Hausdorff distance (Eq. 4.6) captures instead. Dice similarity did not provide error in microns, and thus was not computed in this work. Metrics were computed for the Epithelium (EP), Bowman's Layer (BL), and Endothelium (EN). In Eqs. 6.1 and 4.6, G and S are the set of points in the ground truth annotation and segmentation (fitted with curves) respectively. $y_G(w)$ is the mean Y-coordinate (rounded down) of the points in G whose X-coordinate is w, and similarly for $y_S(w)$. $d_S(x)$ is the distance of a point x in G to the closest point in S, and similarly for $d_G(x)$.

$$\text{MADLBP} = \frac{1}{W}\sum_{w=0}^{W-1}|y_G(w) - y_S(w)| \tag{4.5}$$

$$\text{HD} = \max\left(\max_{x\in G} d_S(x),\ \max_{x\in S} d_G(x)\right) \tag{4.6}$$

From Fig. 4.18 and Table 4.4, our network outperformed the current deep learning [77, 79, 82] and traditional approaches [57, 108] respectively. Paired t-tests conducted between our approach and every baseline established that for each metric our results were statistically significant ($p < 0.05$).

The MADLBP error (in pixels) and mean Hausdorff distance (in microns) across 6×6mm datasets from Device 1 (Tables 4.5 and 4.6, top halves) for the expert grader is slightly lower when contrasted against the trained grader. We attribute this to the diffuse appearance of corneal interfaces [20,57,77] and lower axial resolution of Device 1 (3.4µm), thereby causing an expected deviation between the grader annotations, which is reflected in the inter-grader MADLBP error. Similar measures on the MADLBP error (in pixels) and mean Hausdorff

distance (in microns) across 3×3mm and 6×6mm datasets from Device 2 (Tables 4.5 and 4.6, bottom halves) were observed. Overall, we closely matched the inter-grader error across all datasets for the EP and BL interfaces, and in some cases, perform better than the agreement between graders.

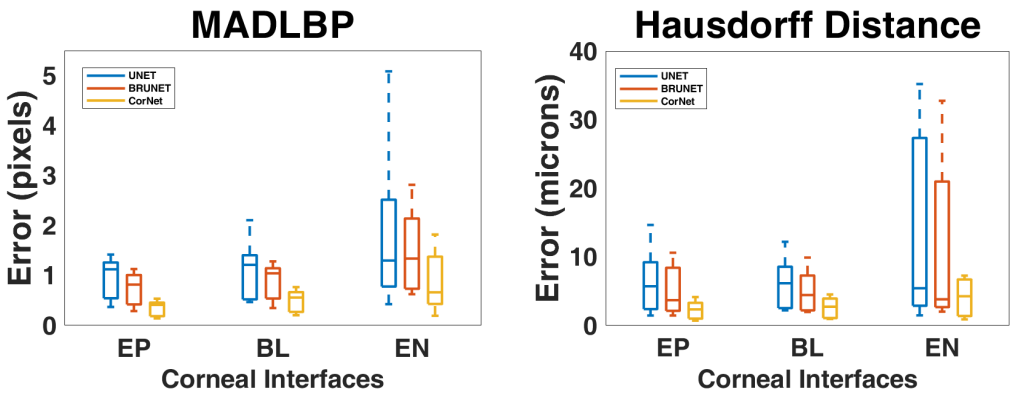


Figure 4.18: Error comparison between expert annotation and automated segmentation (fitted with curves) obtained from different deep learning based methods across all 30 testing datasets.

Table 4.4: Comparison of Mean Absolute Difference in Layer Boundary Position (MADLBP) error between traditional methods against the proposed deep learning based approach on ten 6×6mm volumes from Device 1. Only expert annotations were used for comparison. Errors are in pixels.

Approach	EP	BL	EN
LaRocca et al. [57]	0.84 ± 0.31	1.12 ± 0.4	1.97 ± 2.26
Zhang et al. [108]	0.69 ± 0.24	0.91 ± 0.35	1.73 ± 1.98
Proposed	$\mathbf{0.33 \pm 0.21}$	$\mathbf{0.42 \pm 0.13}$	$\mathbf{0.79 \pm 0.19}$

With respect to the EN, our errors were worse than the inter-grader agreement on the interface location. We attribute this to the low SNR in many corneal images, particularly at the left and right edges of the EN where the signal dropoff is substantial [57]. In these regions, the graders mentally extrapolated their annotations for this interface with poorly defined boundaries, which were usually obfuscated by speckle noise. When a curve is fitted

Table 4.5: Mean Absolute Difference in Layer Boundary Position (MADLBP) error across 6×6mm datasets from Device 1 (top half), and 3×3mm and 6×6mm datasets from Device 2 (bottom half). Errors are in pixels.

Layer	Grader 1	Grader 2	Inter-Grader
EP	0.33 ± 0.21	0.41 ± 0.14	0.49 ± 0.07
BL	0.42 ± 0.13	0.68 ± 0.17	0.51 ± 0.06
EN	0.79 ± 0.19	0.84 ± 0.34	0.56 ± 0.22
EP	0.32 ± 0.09	0.49 ± 0.13	0.49 ± 0.09
BL	0.41 ± 0.13	0.61 ± 0.15	0.5 ± 0.09
EN	0.93 ± 0.19	1.45 ± 0.39	0.61 ± 0.29

Table 4.6: Mean Hausdorff Distance (HD) error across 6×6mm datasets from Device 1 (top half), and 3×3mm and 6×6mm datasets from Device 2 (bottom half). Errors are in microns.

Layer	Grader 1	Grader 2	Inter-Grader
EP	3.17 ± 1.04	4.46 ± 1.23	3.21 ± 0.52
BL	3.52 ± 1.39	4.15 ± 1.05	3.22 ± 0.5
EN	5.55 ± 2.24	6.7 ± 3.78	4.05 ± 1.2
EP	1.52 ± 0.42	1.63 ± 0.42	1.21 ± 0.21
BL	1.89 ± 0.62	1.95 ± 0.68	1.23 ± 0.22
EN	3.05 ± 1.08	4.03 ± 1.34	1.76 ± 0.62

to both the annotation and prediction, there is a small degree of error during the comparison, which is unavoidable. This behavior has also been observed in [57, 108]. However, our EN errors were considerably better than the measured MADLBP and HD errors for the state-of-the-art image analysis-based and deep learning based approaches.

Major Observations

We made the following observations: 1) The proposed CorNet architecture consistently outperforms the state-of-the-art image analysis-based and deep learning-based approaches for the task of corneal tissue interface segmentation. 2) Maxpooling is optimal for feature selection across the common downsampling choices. 3) Nearest neighbor interpolation based feature map upsampling followed by 3×3 convolution improved segmentation over other upsampling operations. 4) Dense connections increased segmentation accuracy due

to greater gradient information flow through the network.

Performance

The CorNet took ∼15.1 s (Python) to segment an entire volume of 50 images of dimensions 1000×1024 pixels, at ∼302 ms per image. This is in contrast to 56.5 s for [57] (Matlab), ∼26.1 s for [108] (Matlab), ∼6.25 s for BRUNET (Python), and ∼10.75 s for UNET (Python); CorNet is slower than UNET or BRUNET due to dense connections. The results were calculated on a desktop using a 3.10 GHz Intel Xeon processor, 64 GB RAM, and a NVIDIA Titan Xp GPU. More advanced GPU architectures should be able to achieve real-time operation (a 10x speedup would be sufficient for 30 fps).

4.3 Hybrid and Cascaded Approaches

4.3.1 Problem Statement

Given a corneal 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.

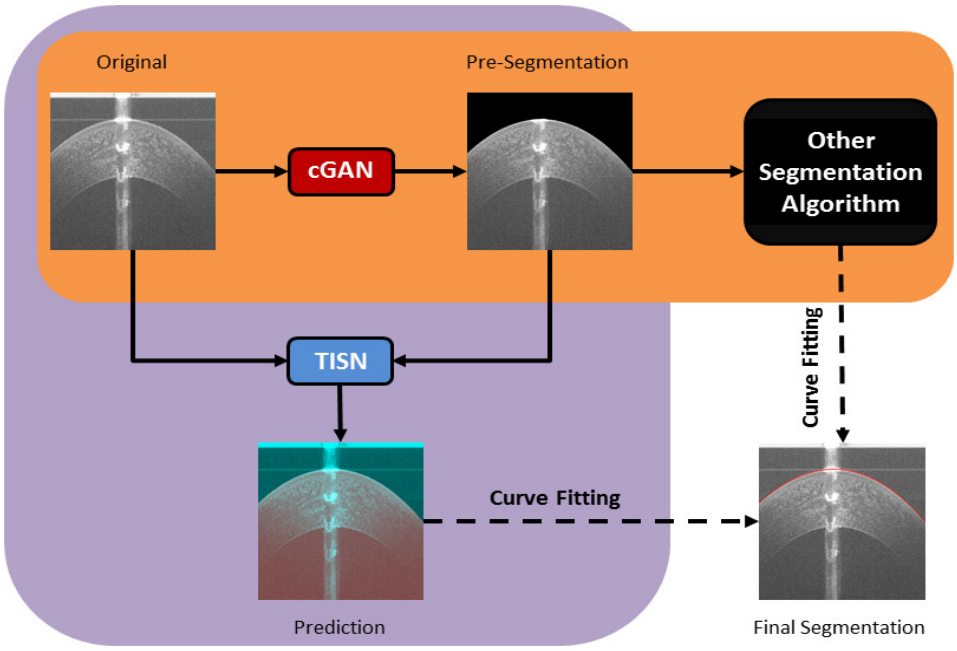


Figure 4.19: Our proposed approach contains two frameworks: a cascaded framework (purple) and a hybrid framework (orange). First, a conditional Generative Adversarial Network (cGAN) takes an input OCT image, and produces an intermediate pre-segmentation image. In the pre-segmentation, pixels just prior 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.

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 work, we only segment the shallowest tissue interface in the corneal OCT 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. 5.21.

4.3.2 Data Acquisition

25 corneal datasets were randomly selected from an existing research database [5]. These datasets were acquired using different scan protocols from two different OCT scanners: a custom Bioptigen Spectral Domain OCT (SD-OCT) scanner (Device 1) that has been described before [109], and a high-speed ultra-high resolution OCT (hsUHR-OCT) scanner (Device 2) [70]. Device 1 had a $3.4\mu m$ axial and $6\mu m$ 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. 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.

4.3.3 Adversarial Network Architecture

We first describe the neural network architecture that was used as the base for both the cGAN (generator), and the TISN. In this work, images of the anterior segment of the eye ac-

quired using OCT contain low SNR, strong specular artifacts, and faintly discernable interfaces that are corrupted by speckle noise patterns. In our previous work [5], we have shown that the CorNet architecture captures faintly visible features across multiple scales. It produced state-of-the-art results on corneal datasets acquired using different OCT systems and using different scan protocols. The errors were 2$\times$ lower than non-proprietary state-of-the-art segmentation algorithms, including traditional image analysis-based [57, 108] and deep learning-based approaches [77, 79, 82].

The CorNet architecture was built upon the BRUNET [79] model, and enhanced the reuse of features generated in the network through residual connections [114], dense connections [116], and dilated convolutions [110–112]. It alleviated the vanishing gradient problem, and prevented holes in the segmentation generated by current deep learning-based approaches [77, 79, 82]. It could accurately extract poorly defined corneal interfaces, such as the Endothelium, which is very common in anterior segment OCT imaging [5].

As shown in Fig. 5.22, the CorNet architecture comprised of contracting and expanding branches; each branch consisted of a building block, which was inspired by the Inception block [112], 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 [5] to determine the right feature selection mechanisms [77, 83, 118–120] for segmentation, such as max-pooling [118] for downsampling and nearest neighbor interpolation + 3$\times$3 convolution [119] for upsampling.

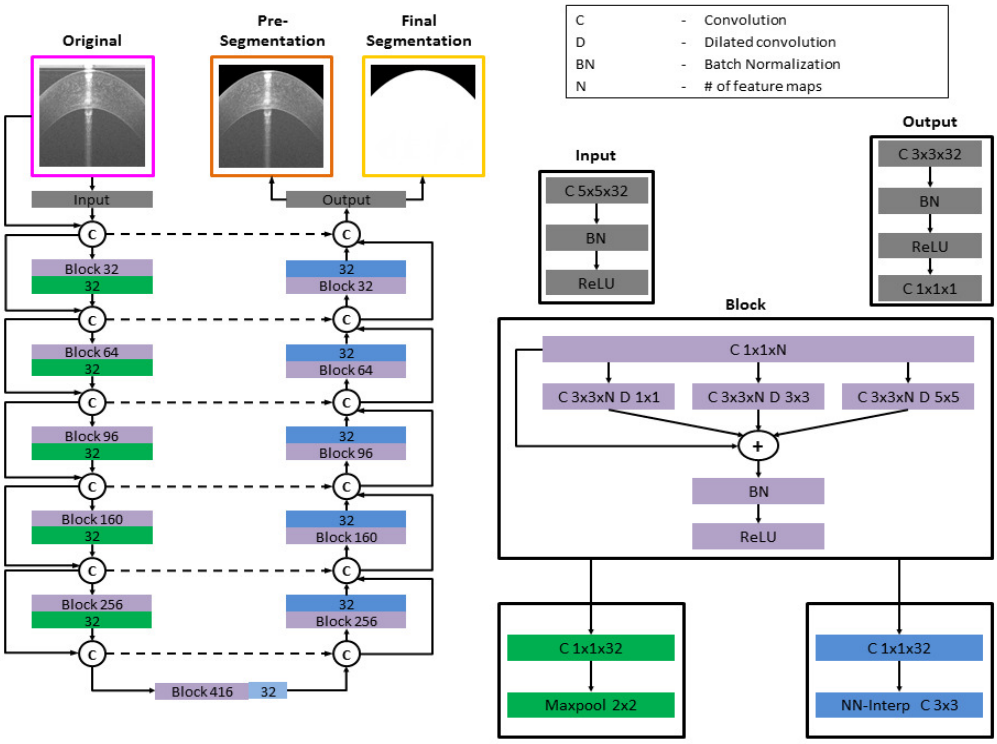


Figure 4.20: The CorNet model is the base architecture used for training both the cGAN and TISN. The input to the cGAN is a two-channel image, the input OCT image and binary mask w (see Sec. 4.3.4), and the output is a pre-segmented OCT image (orange box). The TISN gets a two-channel input (magenta and 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 block at the bottom of the "U" (on the right end of the rectangle) does not perform upsampling, but as with the other blue blocks it functions as a bottleneck for the layer and generates feature maps of the same dimensions as the output feature maps from the previous layer.

Conditional Generative Adversarial Network (cGAN)

Original cGAN

Conditional Generative Adversarial Networks [124] 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. 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 attempts to recognize which pair is the gold-standard pair. 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.

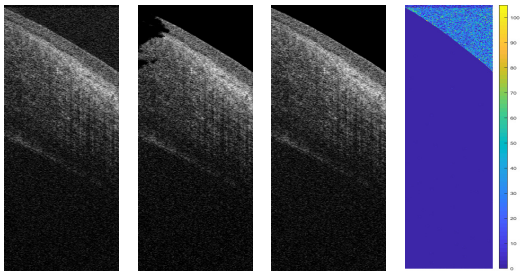
Initially, G generates a prediction y_f that poorly resembles y_t. It learns to produce more realistic predictions by minimizing an objective function shown in Eq. (4.7). On the other hand, D tries to maximize this objective by accurately distinguishing the generated prediction y_f from the true gold standard y_t. The objective function comprises of two losses: L_{cGAN} in Eq. (4.8), and L_1 in Eq. (4.9), with λ being a hyper-parameter. The L_1 loss penalizes regions in the generated output that differ from the ground truth image provided, thereby making the loss a "structured" loss [124]. It forces the output of the generator to be close to the ground truth in the L_1 sense. This loss resulted in less blurry outputs than the original GAN formulation [125], which utilized an L_2 loss. The PatchGAN [124] discriminator was employed to output the probability of a pair of images being real or fake.

$$G^* = arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_1(G) \tag{4.7}$$

$$L_{cGAN}(G, D) = E_{x,y_t}\Big[log\ D(x, y_t)\Big] + E_{x,z}\Big[log(1 - D(x, G(x, z))\Big] \tag{4.8}$$

$$L_1 = E_{x,y_t,z}\Big[\|y_t - G(x, z)\|_1\Big] \tag{4.9}$$

Directly transferring the full cGAN implementation with the cGAN loss in Eq. (4.7) to our OCT datasets resulted in checkerboard artifacts [119] in the generated predictions. Moreover, as shown in Fig. 4.21, parts of the tissue boundary that needed to be preserved were removed instead. From our experiments, we made two empirical observations: 1) The U-Net generator architecture [82] that was utilized in the cGAN paper [124] created checkerboard artifacts in the generated pre-segmentation and did not preserve tissue boundaries correctly; it has been shown in prior work [5,79,119] that the original U-Net implementation is not the optimal choice; 2) The L_1 loss in Eq. (4.9) penalizes all pixels in the image equally.



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

Figure 4.21: Comparing generated pre-segmentations between the U-Net architecture used in the original cGAN implementation [124] against those generated by the CorNet architecture [5]. (a) Original B-scan for a corneal dataset. (b) Generated pre-segmentation for the cGAN U-Net. (c) Generated pre-segmentation for the CorNet. Note that in (b), the U-Net did not remove the speckle patterns above the shallow tissue interface, while also encroaching upon the tissue boundaries without preserving them accurately. (d) Heat map of the difference between the original and pre-segmented OCT B-scans by CorNet.

Modified cGAN with Weighted Loss

The required output of the cGAN is a pre-segmented OCT image, wherein the air-gap 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. (4.9) 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. (4.10), is proposed in this paper. In Eq. (4.10), $\circ$ denotes the pixel-wise product, and α is the hyper-parameter that imparts higher weight to the background pixels over the foreground pixels.

$$L_{w1} = E_{x,y,z}\Big[\alpha w \circ \|y_t - G(x,z)\|_1 + (1-w) \circ \|y_t - G(x,z)\|_1\Big] \tag{4.10}$$

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. 4.3.4. We replaced the L_1 loss in Eq. (4.7) with our weighted L_1 loss in Eq. (4.10), and it eliminated the speckle patterns and specular artifacts just prior to the shallowest interface.

Tissue Interface Segmentation Network (TISN)

As mentioned in Sec. 4.3.3, 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. 5.21 and 5.22, the shallowest interface was extracted from this binary mask [1] and fitted with a curve [123].

4.3.4 Experiments

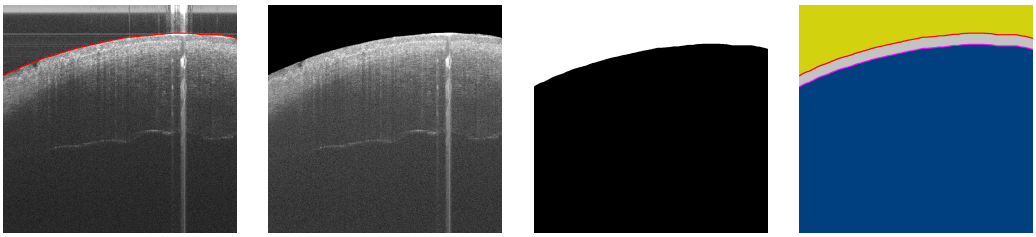
Data Preparation

From the 25 datasets, we had a total of 1250 corneal images images, and 7 datasets were randomly chosen for training the cGAN, and the remaining 18 were used for testing. These datasets were chosen such that they came from both eyes; the number of patients that were imaged could not be ascertained as the database contained deidentified datasets. From the total set, we chose the training set to comprise of a balanced number of corneal datasets (7 each) that exhibited different magnitudes of specular artifacts, shadowing, and speckle. The training set contained 350 corneal images, and the remaining were set aside in the testing set. Considering the varying dimensions of the OCT images acquired from two 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 [5,77], we sliced the input images width-wise to produce a set of images of dimensions 256×1024 pixels, and in this way, we preserved the OCT 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 grader is shown in Fig. 4.22(a). To generate the gold standard pre-segmentation images for training, we eliminated the speckle noise and specular artifacts by setting the region just 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. 4.22(b). The binary mask w that was used in the Eq. (4.10) is shown in Fig. 4.22(c). Using the image in Fig. 4.22(d) as reference, we detail the process of obtaining w. In Fig. 4.22(d), the original annotation of the tissue interface boundary by the grader 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.

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. 4.22(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 Fig. 4.21(b). In such a scenario, there is no large penalty applied to the erosion of pixels as detailed in Eq. (4.10). 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 (green + red + gray) to avoid the erosion of the tissue interface.



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

Figure 4.22: (a) Expert annotation of an original B-scan in a 6×6mm OCT volume acquired by Device 3, (b) Gold standard pre-segmentation image for training, (c) Binary mask w used in Eq. (4.10) for training the cGAN, (d) Label map detailing the process of generating w (see Sec. 4.3.4).

Data Augmentation

As our training datasets were smaller in number in contrast to those from datasets typically available in computer vision tasks, such as image recognition [126], we augmented our datasets to increase the variety of the images that were seen during the training. These augmentations [121] 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 [124]. For the cGAN, augmentation was done by simply flipping each input slice horizontally along the X-axis.

cGAN Training

Training of the cGAN commenced from scratch using the architecture shown in Fig. 5.22. 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 λ = 100, and α = 10 in final objective function, and optimized the network parameters using the ADAM optimizer [122]. 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.

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 [122], was used for training. In this work, we used MSE loss to be consistent with the original CorNet implementation [5], but the MSE loss can easily be substituted for the cross entropy loss [82] or the dice loss [127]. 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 (the boundary of the interface and those pixels below it), and the second channel corresponded to the background pixel segmentation (above the tissue interface). The predicted segmentation

does not provide a contour of the shallowest interface. Instead, it highlights the area of the pixels at the shallowest interface and those below it as belonging to the foreground. Finally, the predicted segmentation was fitted with a curve [123] after the tissue interface was identified using a fast GPU-based method [1]. We show our final results in Fig. 4.23.

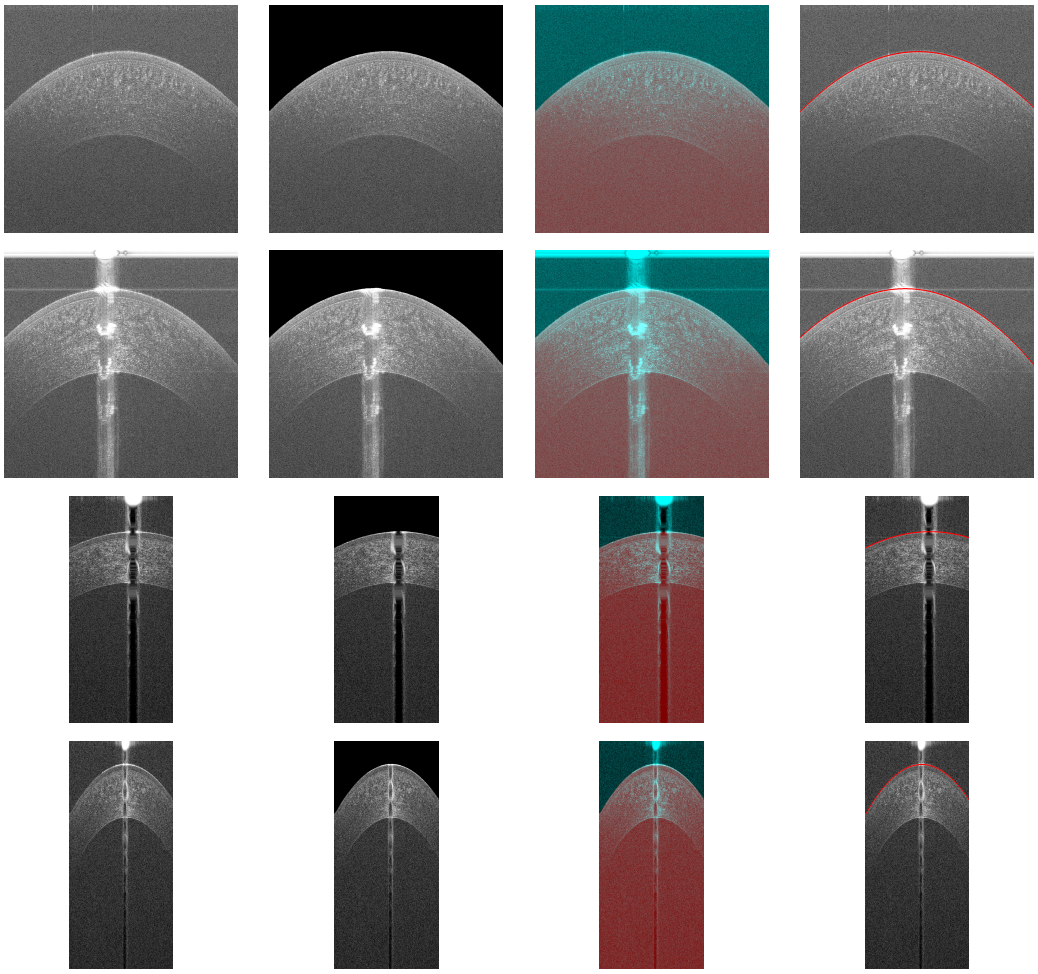


Figure 4.23: Corneal interface segmentation results for datasets acquired using Devices 1 and 2. Columns from left to right: (a) Original B-scans in corneal OCT datasets, (b) Pre-segmented OCT images from the cGAN with the specular artifact and speckle noise patterns removed just prior to the shallowest 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).

Comparison Against Other Algorithmic Approaches

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 [3] 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 [128], and segmented the tissue interface. We denote this approach 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 [3] 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 our algorithmic approach by the acronym: Traditional With Pre-Segmentation (TWPS).

Third, we trained a CorNet architecture [5] 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 combinations of algorithmic approaches were considered for performance evaluation:

1. TWOPS - A traditional image analysis-based algorithm [3] that directly segmented the tissue interface.

2. TWPS - The hybrid framework.

3. DLWOPS - A deep learning-based approach [5] that directly segmented the tissue interface.

4. DLWPS - The cascaded framework.

Annotation

Each corneal dataset was annotated by an expert grader (G1; Grader 1) and a trained grader (G2; Grader 2). 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 algorithmic approaches. 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.

4.3.5 Results

Metrics

In order to compare the segmentation accuracy across the different algorithmic approaches, we calculated the two 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. (5.4) and (5.5), 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 Euclidean distance of a point p in G to the closest point in S, and similarly for $d_G(p)$.

We chose MADLBP in Eq. (5.4) as one of our error metrics since it was used in [3], and we desired a direct comparison with prior work for 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. (5.5) captured the greatest of all Euclidean distances between the points in the segmentation and annotation. Therefore, it quantitatively describes the worst segmentation error in microns, which may be 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.

$$\text{MADLBP} = \frac{1}{X}\sum_{x=0}^{X-1} |y_G(w) - y_S(w)| \tag{4.11}$$

$$\text{HD} = \max\left(\max_{p\in G} d_S(p),\ \max_{p\in S} d_G(p)\right) \tag{4.12}$$

In Fig. 4.24, the HD error and the MADLBP error across all algorithmic approaches for the corneal datasets acquired from devices 1 and 2 were compared. In Fig. 4.25, the benefit of pre-segmenting the OCT image was verified by first grouping the approaches 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.

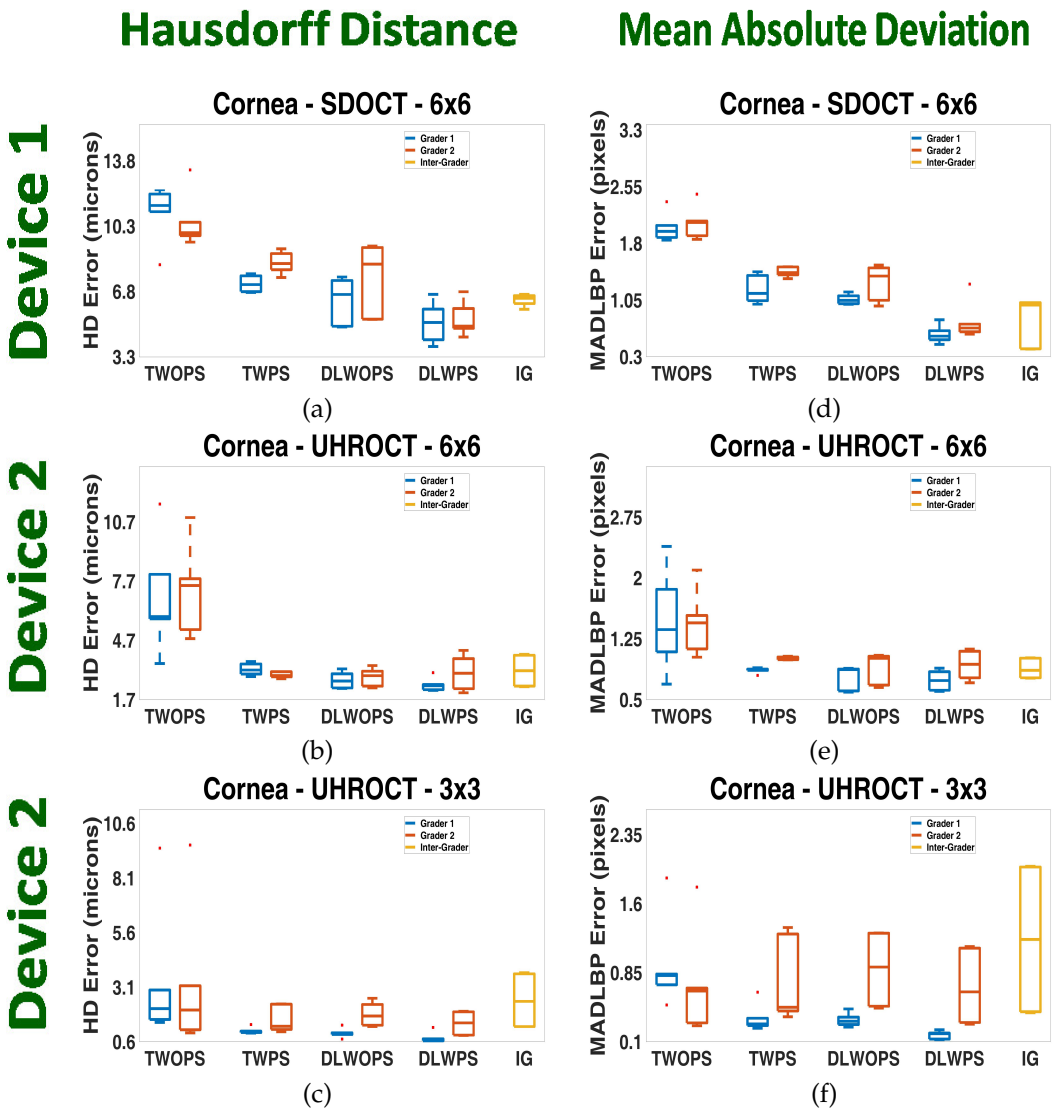


Figure 4.24: (a)-(c) HD error and (d)-(f) MADLBP error comparison for the corneal datasets acquired with Devices 1 and 2 respectively. In the boxplots, the segmentation results obtained for each algorithmic approach are contrasted against expert grader (blue) and trained grader (red) annotations, while the Inter-Grader (IG) variability is shown in yellow.

Discussion - Segmentation Accuracy of Corneal Interface

From the HD and MADLBP errors in Figs. 4.24, the error is worse for the TWOPS approach, where the traditional algorithm [3] used the original OCT image (without the pre-

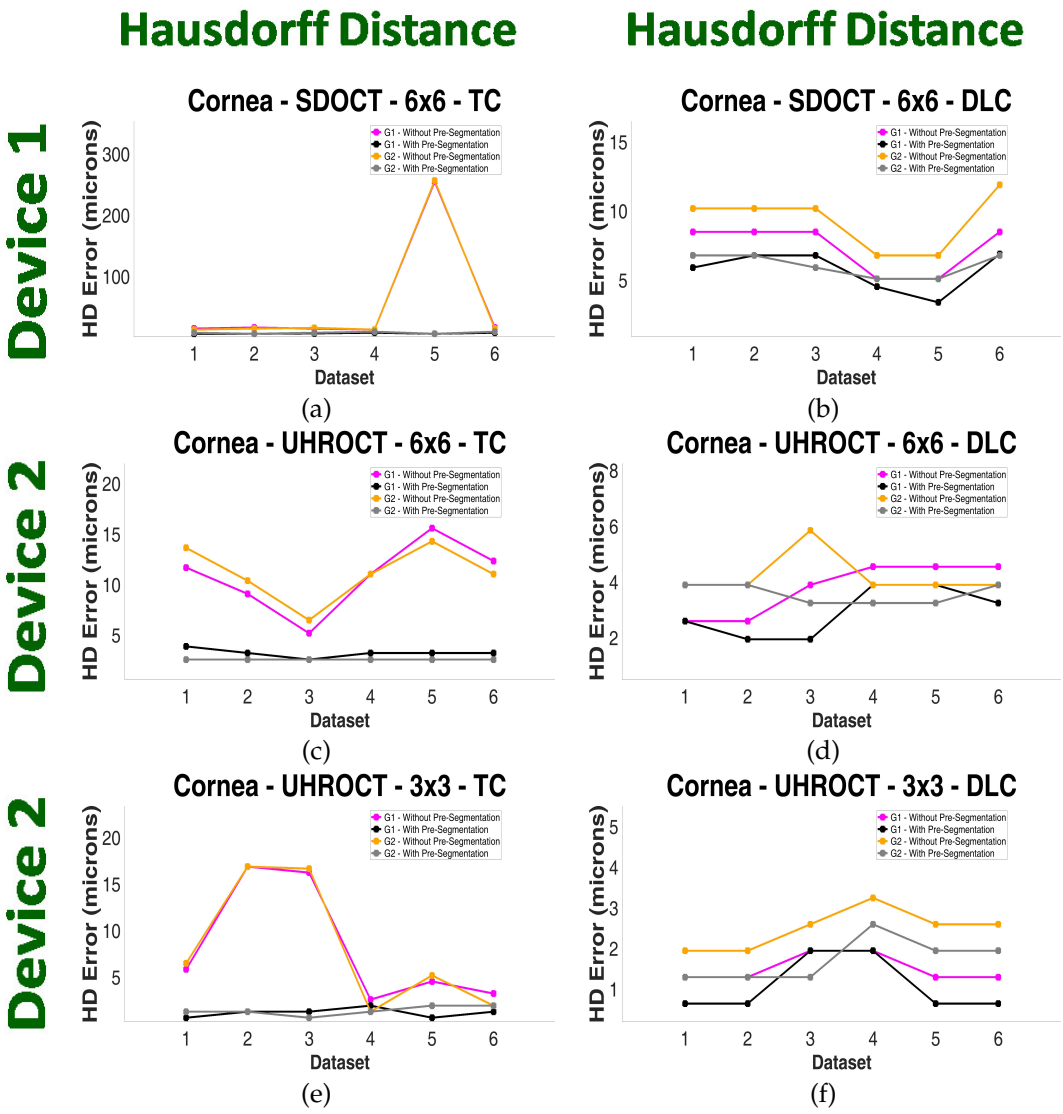


Figure 4.25: Quantitative estimation of the benefit of pre-segmenting the corneal OCT image. The algorithmic approaches were grouped into two categories: Traditional Comparison (TC; TWOPS vs TWPS), and Deep Learning Comparison (DLC; DLWOPS vs DLWPS). The first column corresponds to the former, and the 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 approaches respectively. The stored location indicies were then used to retrieve the corresponding HD errors from the TWPS and DLWPS algorithmic approaches 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).

segmentation) to directly segment the interface. The hand-crafted features in this algorithm failed to handle severe specular artifacts and noise patterns as seen in Fig. 4.15. In contrast, the TWPS approach (hybrid framework), which uses the pre-segmented image instead of the original OCT image, produced a lower segmentation error. To quantify these observations, a paired t-test between the TWOPS and TWPS algorithmic approaches 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 approach in Fig. 4.24 had lower HD and MADLBP errors than the TWPS approach 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. 4.24(c) and 4.24(f), due to the large inter-grader variability. On the other hand, our DLWPS approach, which used the pre-segmented image, fared better in contrast to the other three approaches. Again, we computed paired t-tests between the DLWPS approach and all other approaches to determine the improvement in segmentation accuracy for each error metric. From the p-values in Table. 4.7 and Fig. 4.24, the cascaded framework generated results that were an improvement upon the other approaches, and indicated statistically significant results across all corneal datasets ($p < 0.05$).

Table 4.7: Statistical significance between our cascaded framework (DLWPS) against each approach 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

To determine the improvement in segmentation accuracy on an per-image basis in each

of the corneal test datasets, we first grouped the algorithmic approaches 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 approaches respectively, and we plotted these maximum HD errors for each grader in Fig. 4.25 (purple and yellow colored curves). Then, we queried the errors for the same images (using the image indicies) in the TWPS and DLWPS approaches respectively, and plotted the corresponding HD errors for each grader in Fig. 4.25 (black and gray curves). From Fig. 4.25, we noted that the algorithmic approaches incorporating the pre-segemented OCT image performed better than one that did not include the pre-segmentation. 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 quantitatively demonstrates the potential benefit of utilizing the pre-segmented OCT image as part of a segmentation framework.

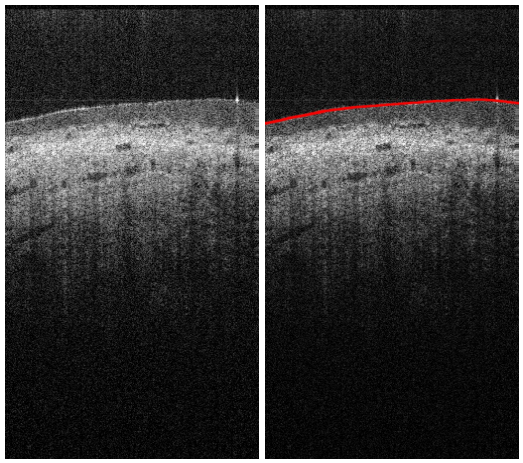
Chapter 5

Visualizing The Palisades of Vogt

5.1 Classical Approach

5.1.1 Problem Statement

Given a SD-OCT image I in a volume, the problem can be defined as finding the set of pixels $\mathcal{F}$ that lie on the boundary of the limbal tissue interface. Every pixel in $\mathcal{F}$ takes a specific label $L = 1$ if it lies on the tissue interface boundary, or it takes a label $L = 0$ if it does not lie on the boundary (see Fig. 5.1). The segmented limbal tissue interface should then be used to guide the registration of all the images in a volume to a reference image in the volume. The reference image is usually an image in the middle of the volume. Registration should lead to the generation of an aligned 3D volume, which can be reconstructed to visualize the imaged limbal region.



(a) (b)

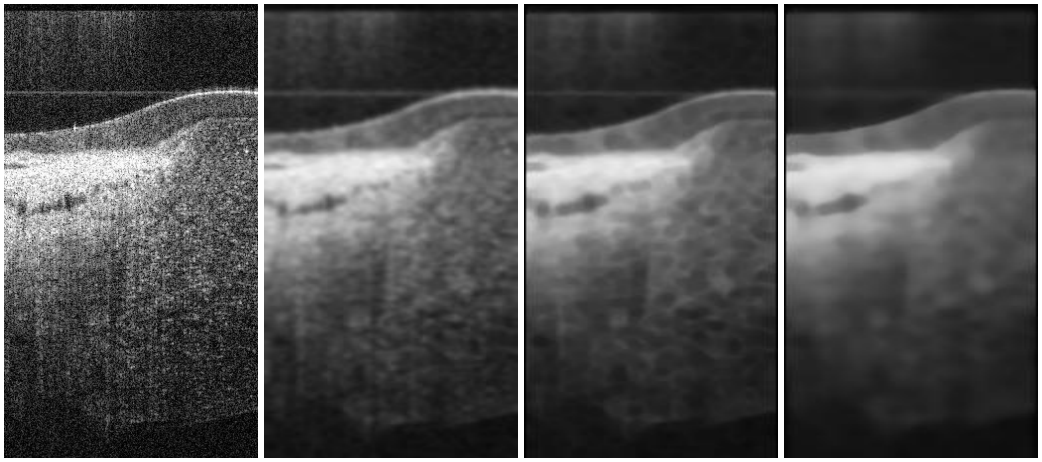
Figure 5.1: (a) shows the 1^{st} B-scan imaging the limbal area in a SD-OCT volume, and the expertly annotated tissue interface is shown in (b).

5.1.2 Data Acquisition

Previously acquired data from an existing research database was de-identified for this pilot work. 16 raster scanned volumes from both eyes of five subjects were scanned with a high-speed ultra-high resolution OCT (hsUHR-OCT) scanning system [70]. With the hsUHR-OCT system, two regions of different sizes were scanned: a 4×4mm region, and a 3×3mm region. The 4×4mm region was scanned to yield volumes, which contained 300 B-scans of dimensions 1024×300 pixels each. The 3×3mm region was scanned to yield volumes, which contained 250 B-scans of dimensions 1024×250 pixels each. These volumes were cropped or padded to 1024×256 pixels (H×W), where the POV was seen prominently, with the resulting volumes consisting of a variable number of B-scans per volume (150-300) after padding/cropping. The axial pixel pitch and resolution of 1.3μm was maintained.

5.1.3 Noise Reduction

In this section, properties of the noise affecting the limbal surface segmentation are discussed, and our noise suppression approach is presented. In addition to speckle noise, two types of artifacts affect the segmentation; horizontal and vertical artifacts (see Fig. 5.2(a)). These artifacts are due to specular reflection and system noise during imaging [57]. To minimize these effects, the B-scans were first downsampled by a factor of 4 in each dimension, and then upsampled by a factor of 2 in each dimension (see Fig. 5.2(b)). It was then filtered with a 20^{th} Percentile filter [1] of size 5×5 pixels to preserve edges while reducing the noise in the image (see Fig. 5.2(c)).



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

Figure 5.2: (a) shows the 195^{th} B-scan in a volume; (b) the down-sampled image in (a); (c) shows the result of percentile filtering the downsampled B-scan; (d) show the result of further bilateral filtering applied to the percentile-filtered B-scan.

Following percentile filtering, small amplitude residual noise was left in the image. It was empirically observed that applying the percentile filter to the image a second time diminished their presence, but did not conserve the edges with low intensity. These poorly visible boundaries are crucial elements of our segmentation, and need to be retained. To

that end, a bilateral filter [129] of size 5×5 pixels was used to preserve the edges in the image while smoothing the small amplitude noise regions further (see Fig. 5.2(d)). As seen in Figs. 5.2(a) and 5.2(d), the key objective was to retain visible tissue structure, while diminishing the speckle and artifact noise. Although our approach did not remove the artifacts entirely, their effects were mitigated before segmentation.

5.1.4 Segmentation

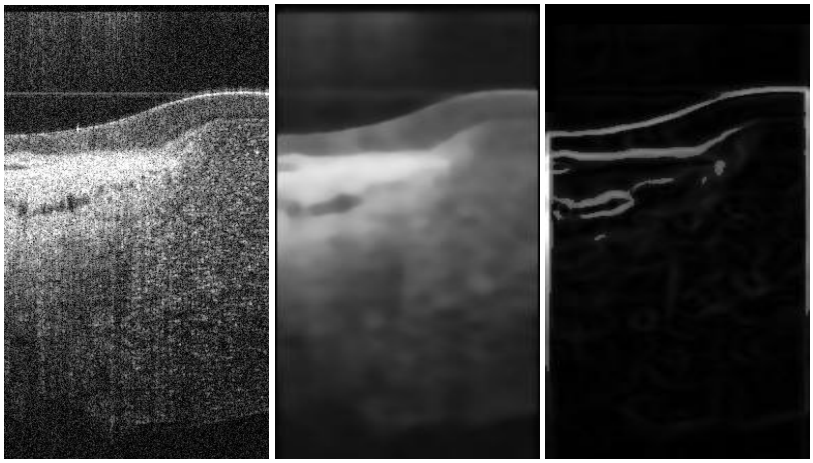
In this section, a multi-scale edge detection scheme highlights and segments the limbal surface in the bilateral filtered image.

Edge Detection

Similar to Sec. 6.1.4, the Cauchy filter was used for edge detection. The parameter values used were $w_o = [5, 10, 15]$, $a = 1$, and $s = 1/w_o$. The image B was filtered with the Cauchy filter at three different scales, and the two odd components of the monogenic signal [130,131], $[f_{o1}, f_{o2}]$, were obtained for each scale. Next, the two odd parts were combined to produce the magnitude for each scale as defined in Eq. (5.1). Finally, the magnitudes estimated for each scale were averaged together using Eq. (5.2) (see Fig. 5.3(c)).

$$f_o = \sqrt{f_{o1}^2 + f_{o2}^2} \tag{5.1}$$

$$f_{oc} = \frac{f_o^5 + f_o^{10} + f_o^{15}}{3} \tag{5.2}$$



(a) (b) (c)

Figure 5.3: (a) shows the 195^{th} B-scan in a volume (same image in Fig. 5.2(a)); (b) shows the bilateral filtered image; (c) shows the odd magnitude component f_{oc} of the monogenic signal (scaled for visualization here).

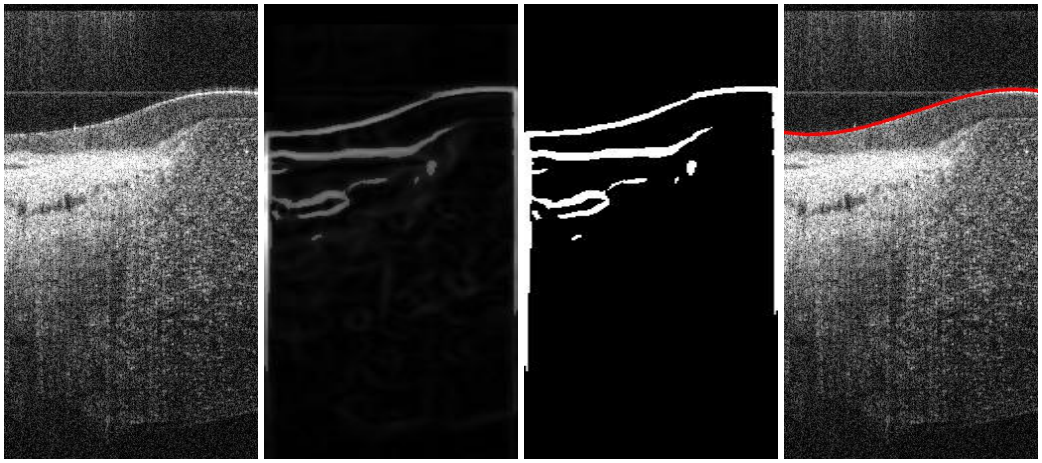
Surface Segmentation

At this point, the edges are highlighted in f_{oc}, but they have not yet been segmented. To this end, an estimate of the foreground intensity along each A-scan is needed. It can be obtained from the range of the intensity differences as we proceed axially along each A-scan in the image. This range was estimated from the vertical gradient G of f_{oc}, which was obtained by convolving f_{oc} with a kernel such as $[-1, 0, 1]^T$. Next, the maximum value of the gradients for each A-scan was calculated, yielding a set of gradient values across all A-scans. Finally, the mean value of this range was computed, and the threshold $thresh_M$ for segmenting f_{oc} was set to be half the mean value as defined in Eq. (5.3). Intuitively, the list of maximum values across all A-scans represents the gradient values at the transition point from dark-to-bright regions, which in the ideal case, would correspond to the exact location of a shallow tissue surface. However, as tissue interfaces are often fuzzy in OCT images, a band of pixels in an A-scan (and indeed, across the image) in f_{oc} will represent a

boundary. For this reason, 50% of the mean value of this range captures most of the tissue boundary as shown in Fig. 5.4(c).

$$thresh_M = \frac{1}{2W} \sum_{w=1}^{W} \max_{h \in H}(G_{w,h}) \tag{5.3}$$

The edge points corresponding to the limbal surface were extracted from the binary segmentation map using the fast A-scan based approach that was published in [1]. Briefly, the approach in [1] used a fast parallel GPU-based search to accurately localize the tissue structure in each A-scan of the image. Once, the edge points were detected, a second-, third-, or fourth-order polynomial curve was fit [104] to the detected points as shown in Fig. 5.4(d).



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

Figure 5.4: (a) shows the 195^{th} B-scan in a volume (same image in Fig. 5.2; (b) shows the odd magnitude component f_{oc} of the monogenic signal (scaled for visualization here); (c) shows segmentation result after selecting a suitable threshold $thresh_M$; (d) shows the final result after detecting the shallowest surface and fitting a 3^{rd} order curve to the detected edge points. Note the true tissue surface detected in (d) even in the presence of noise artifacts.

5.1.5 Registration

In this section, the shallowest segmented surface in each B-scan was registered to the surface segmented in a reference B-scan. First, a B-scan was "flattened" [57] by finding the shallowest A-scan position on the segmented curve. Relative to this position, the offset in position of the segmented curve in all other A-scans was computed. Based on the estimated offsets, A-scans were circularly shifted up to match the shallowest position (Figs. 5.5(c) and 5.5(d)). This process was done for all B-scans in a volume. Next, a rigid transformation was computed between the flattened landmarks of a reference B-scan (usually the middle of the volume), and the flattened landmarks of each individual B-scan. By following this approach, individual B-scans were aligned, resulting in a flattened volume.

At this point, horizontal cross-sections at different depths were extracted from the flattened volume, as seen in Fig. 5.6, by incrementing the flattened and registered segmentation points for each B-scan by a constant value. Finally, each registered B-scan was "unflattened" [57] by circularly shifting rows down to match the detected curvature of the reference B-scan. This approach aligned the shallowest surface detected in each B-scan with the surface detected in the reference B-scan (see Figs. 5.5(e) and 5.5(f)), and a registered volume was reconstructed as shown in Fig. 5.5(g).

5.1.6 Results

In this section, the segmentation accuracy and registration accuracy (as a function of cross-section extraction) were evaluated. Segmentation accuracy was measured by comparing the automatically segmented results against an experts' manual annotations. Images in all datasets that were used were manually annotated by the expert. The expert manual annotations were aligned using the same procedure described in Sec. 5.1.5, and cross-sections

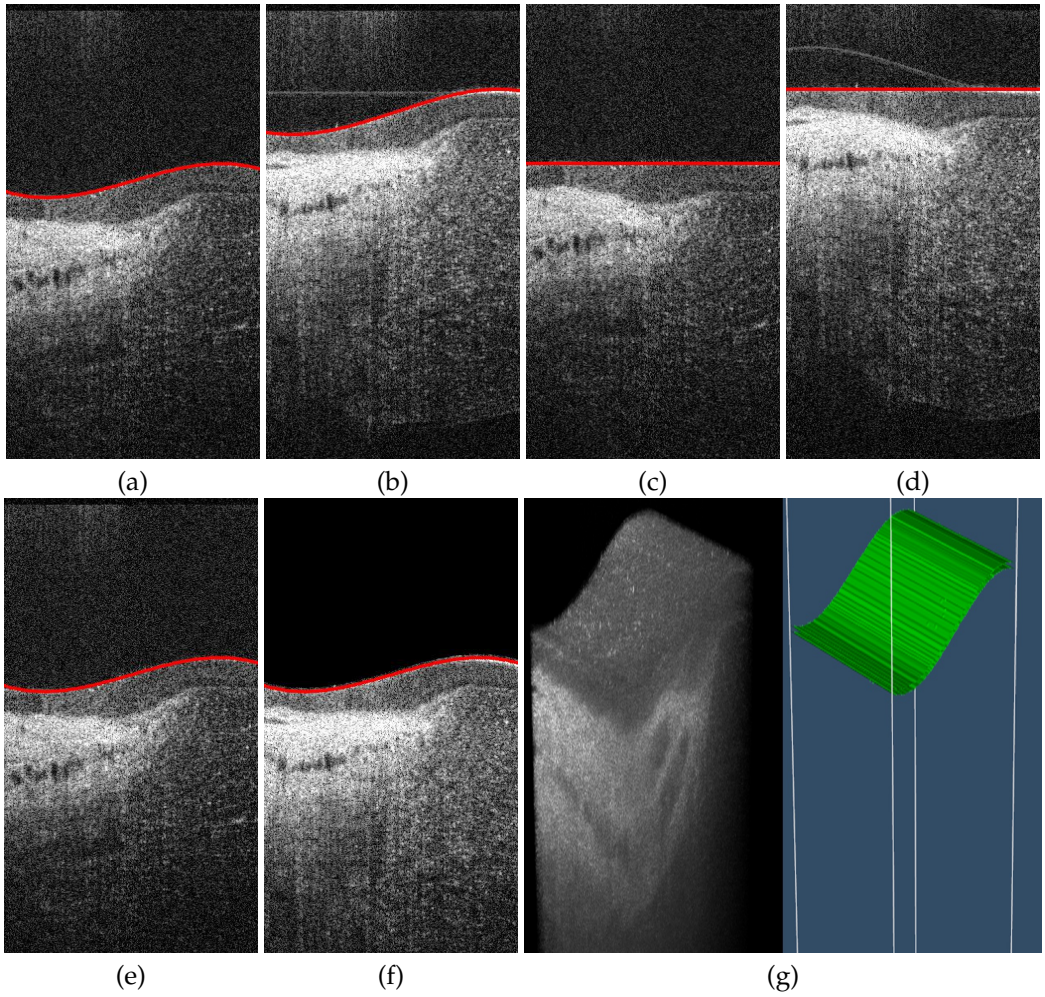


Figure 5.5: Registration procedure followed by our algorithm. (a)-(b) the 100^{th} (reference) and 195^{th} (target) B-scans in a volume; (c)-(d) the "flattening" of the images using the segmentation derived by our algorithm; (e)-(f) the "unflattened" images obtained after rigid registration of the flattened images in (c) and (d); (g) a side view profile of the reconstructed volume and meshed outer surface following the registration. Note the clear visualization of the limbal region, and the associated connective tissues appearing white and vessels seen as black ducts in the left volumetric rendering of (g).

were extracted. The Structural Similarity (SSIM) image metric [132] was used to determine the correlation between cross-sections that are automatically extracted and cross-sections that were extracted after the alignment of the expert manual annotations. Registration effectiveness was assessed by the cross-sectional SSIM value. The validation protocol in [57]

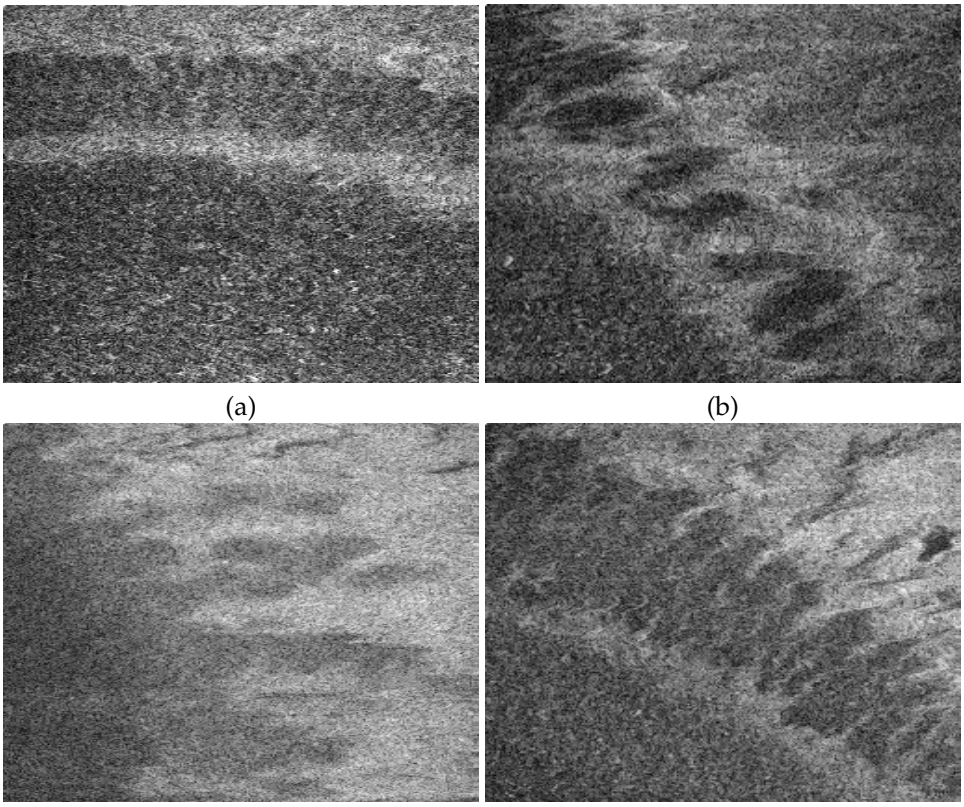


Figure 5.6: (a)-(d) show cross-sectional visualizations of the unique POV structural configuration.

was followed. For fair comparison, the expert annotation was fitted with a curve, and compared against the automatic segmentations. The following metrics were calculated: 1) *Mean Absolute Distance* (MAD) - the mean residual distance in boundary position between the annotations (fitted with a curve) and the automatic segmentation (see Eq. (6.22)), and 2) *SSIM* between cross-sections [132]. Fig. 5.7 and Table 5.1 summarize the segmentation accuracy and registration accuracy, respectively, over all datasets.

From Fig. $5.7(a)$, the surface segmentation error is minimal when MAD values are close to 0. Associated higher SSIM values (which can take a maximum value of 1) in Fig. $5.7(b)$ indicate a similarity between the extracted cross-sections. From Table 5.1 and Fig. $5.7(a)$,

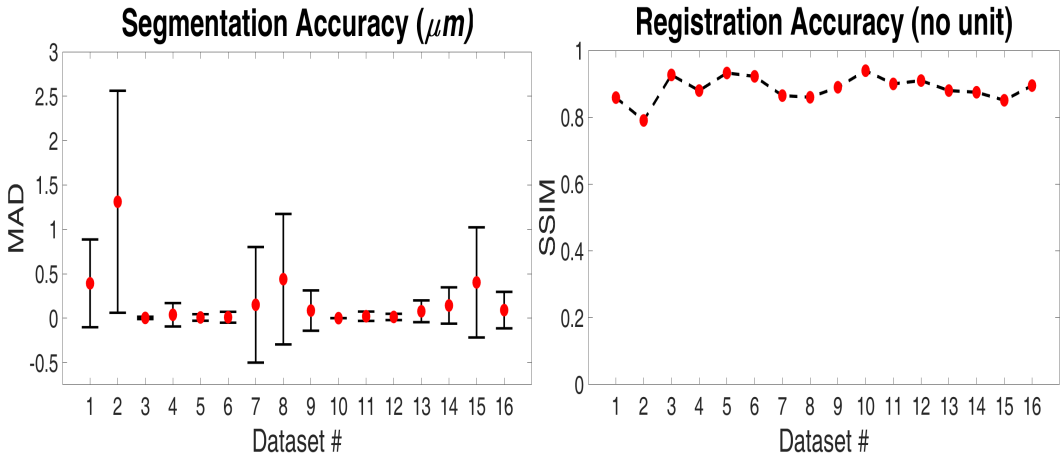


Figure 5.7: (a) Mean (red circle) and standard deviation error (black error bars) between automated segmentation and manual annotation. (b) SSIM values of cross-section comparison.

Table 5.1: Average quantitative error metrics across all 16 datasets. The pixel resolution in an A-scan is 1.3µm.

	Grader 1
MAD	0.19 ± 0.30 µm
SSIM	0.89

on average, the MAD error plus standard deviation was less than $0.5\mu\text{m}$. From the graphs, only dataset $\#2$ had a high MAD error and a corresponding lower SSIM value. Upon closer examination of the MAD error, it was due to lower signal-to-noise ratios at the tissue interfaces near the edges of the image, especially when they were corrupted by speckle noise. While the expert mentally extrapolated the missing regions in the image during annotation, the algorithm was slightly unconstrained at these poorly defined boundaries, and it had to extrapolate the curve to fit this area. Furthermore, fitting a curve to the manual annotations induces a small degree of error, which is unavoidable, during comparison. Nevertheless, the maximum error obtained in our work was less than $3\mu\text{m}$ (or ~ 2 pixels), which is acceptable error for the purposes of visualization of the Palisades of Vogt. The per-frame

runtime of the algorithm was ~ 351 milliseconds on a NVIDIA Quadro K6000 GPU with C++, CUDA, OpenCV, and VTK.

Through the consistency between the automatic and manual alignment results, we have demonstrated that our adopted methodology removes the subjectivity of manual segmentation. Since it is also faster than manual alignment, we have shown that it constitutes transferable knowledge to the OCT imaging community. However, the registration algorithm proposed in this work - rigid registration guided by limbal surface segmentation - fails to account for out-of-plane tissue motion. This is a challenging registration problem to analyze without constraints, and our future work is directed towards utilizing deep neural networks to segment the limbal interface and intelligently determine the tissue structure motion from one frame to another.

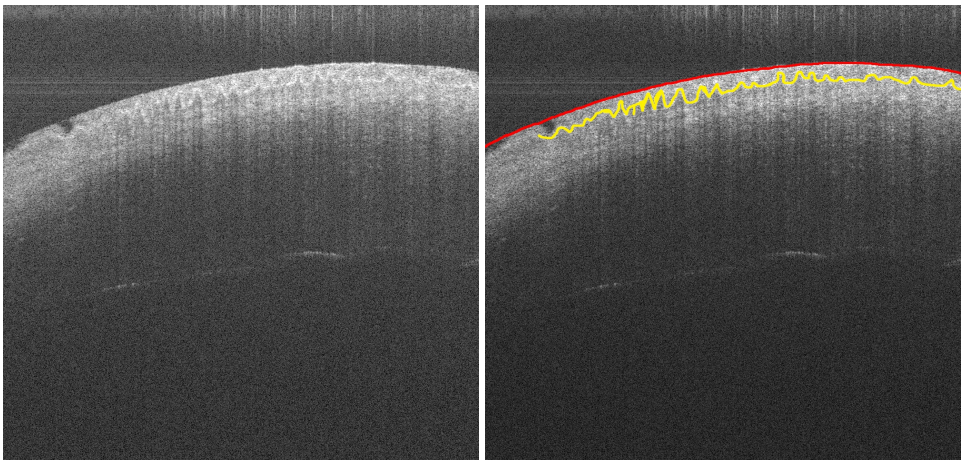
5.2 Learning-Based Approach

5.2.1 Problem Statement

For an anterior segment SD-OCT image I of the limbal region of the eye in a volume of images, the problem can be considered as assigning every pixel in the image a specific label L. As shown in Fig. 5.8, we can consider the segmentation task as assigning every pixel in the image I as being on the boundary of the limbal tissue interface ($L = 1$) or not ($L = 0$). Since each B-scan is acquired independently in a SD-OCT volume, an external tracking mechanism is unavailable, and a guaranteed global alignment is not available as well. The segmented limbal tissue interface will then be used to create an aligned 3D volume by guiding the registration of all the images in a volume to a reference image in the volume. The reference image is usually an image in the middle of the volume. The registered 3D volume should be reconstructed to visualize the limbal region.

5.2.2 Data Acquisition

Similar to Sec. 5.1.2, data from an existing research database was de-identified. 36 raster scanned volumes from both eyes of many subjects were scanned with multiple SD-OCT scanning systems. In addition to the datasets in Sec. 5.1.2, volumes from a different SD-OCT system (Bioptigen SD-OCT scanner [133]) were acquired as shown in Fig. 5.9. Together, these volumes comprised the 36 volumes, for which manual annotations were available for all images in all 36 volumes. The volumes included images showing severe artifacts, pathology, and excess corneal surface degradation. Some of the datasets in this pool of 36 volumes were used to train the network. We elaborate on how the available data was split into training, validation, and testing phases in the next section. The expert grader



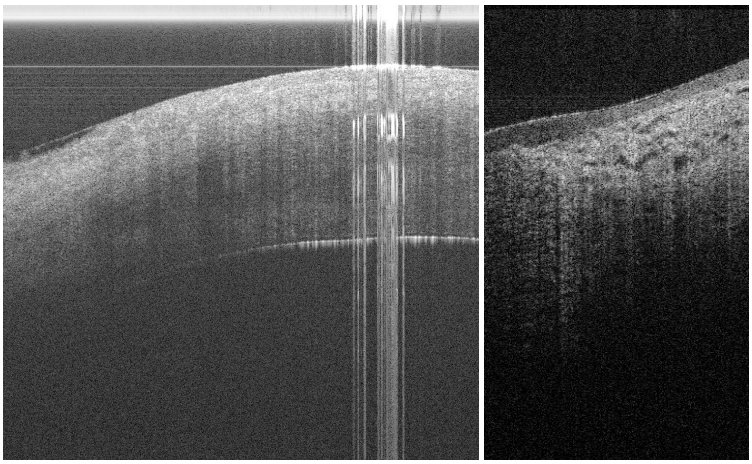
(a) (b)

Figure 5.8: (a) Pathological SD-OCT image of a patient acquired with a Bioptigen SD-OCT scanner [133]. (b) The limbal tissue interface (red curve) and the Basal Cell Layer (BCL) where the POV (yellow) are present is highlighted. The images were manually annotated by an expert. Notice the lack of clear and visible structure of the BCL and the POV in (a) as they appear wavy and are difficult to pinpoint on the left corner inside the tissue in (b). Also notice the pathological changes in the tissue structure on the left side of the image signifying a degradation of the tissue structure.

was allowed to annotate the images using a 3-pixel band. They were asked to stay within 1 pixel of the true tissue boundary location. This was mostly done in order to provide a slight leeway in annotations as tissue boundaries can appear diffuse and make it difficult to annotate.

The dimensions of the volumes varied; the minimum dimensions of images in the volumes were 1024×250 pixels, while the maximum dimensions were 1024×1000 pixels. The total number of images per volume varied in the range of 150-320. The dimensions of the images fed into the network was 1024×256 pixels (H×W). All images in the volumes were padded to 1024×256 pixels if the column count was lower than the required input size. If they were larger than the required network dimensions, then the images in volumes were sliced width-wise to produce an input to the network of size 1024×256 pixels. There was

overlap in the images that were sliced and fed to the network if the dimensions were not a multiple of 256. The predicted output from the network for each slice was then used to recreate the original dimension of the full sized image.



(a) (b)

Figure 5.9: (a) Example SD-OCT image acquired by the Bioptigen system [133]. (b) Example SD-OCT image acquired by the hsUHROCT system [34]. In contrast to (b), note the increased intensity of the tissue structure and imaging artifacts in the image in (a).

5.2.3 CorNet Architecture

The architecture of the convolutional neural network (CNN) used in this work was the CorNet, previously described in Sec. 4.2.2, and is shown in Fig. 5.10. The CorNet architecture was originally designed to segment multiple corneal tissue interfaces. Since the morphology of the limbal region is similar to the cornea, we extended our work to segment the shallowest limbal tissue interface. The CorNet architecture closely resembles the model proposed in [79]. In this network design, the input image is convolved and downsampled from one layer to another with the number of filters used to detect objects in the image exponentially increasing until a predefined depth is reached. This contracting path [82] is

called the descending branch. Following the contracting path, the downsampled images are subsequently upsampled and convolved to reach the original dimensions of the image. This expanding path [82] is called the ascending branch. Skip connections [134] allow the high resolution features from the descending/contracting path to be combined with the corresponding layer in the ascending branch to allow the localization of tissue structures and allow successive layers to learn precise positions of the tissue interface based on this additional information.

The U-Net architecture [82] is shown in Fig. 5.11. In a conventional U-Net architecture [82], there are several issues that occur when segmenting the tissue boundary of interest. As the largest possible object in an image that can be segmented depends on the receptive field of the network, we have observed:

- With the U-Net, a 3×3 pixels kernel that is used for a network depth of 5 layers exhibits holes [79] when segmenting tissue objects containing a large discontinuity, such as when there is shadowing or pathology, that is greater than $3 * 2^5 = 96$ pixels. This has been experimentally observed in our experiments, and [79] have seen this in their experiments as well.

- As the network grows in depth, the convergence rate of the network decreases, and this is due to the vanishing gradient problem [135]. The gradients of the network output with respect to the parameters (weights) in each layer that are computed using back-propagation in a CNN become smaller and smaller. The shrinking value of the gradients occurs as the gradients pass through a deep network, with the effect dependent upon the choice of an activation function, such as tanh or sigmoid activation functions. Even though the U-Net uses a ReLU activation function, it still undergoes the problem of vanishing gradients due to the depth of the network architecture.

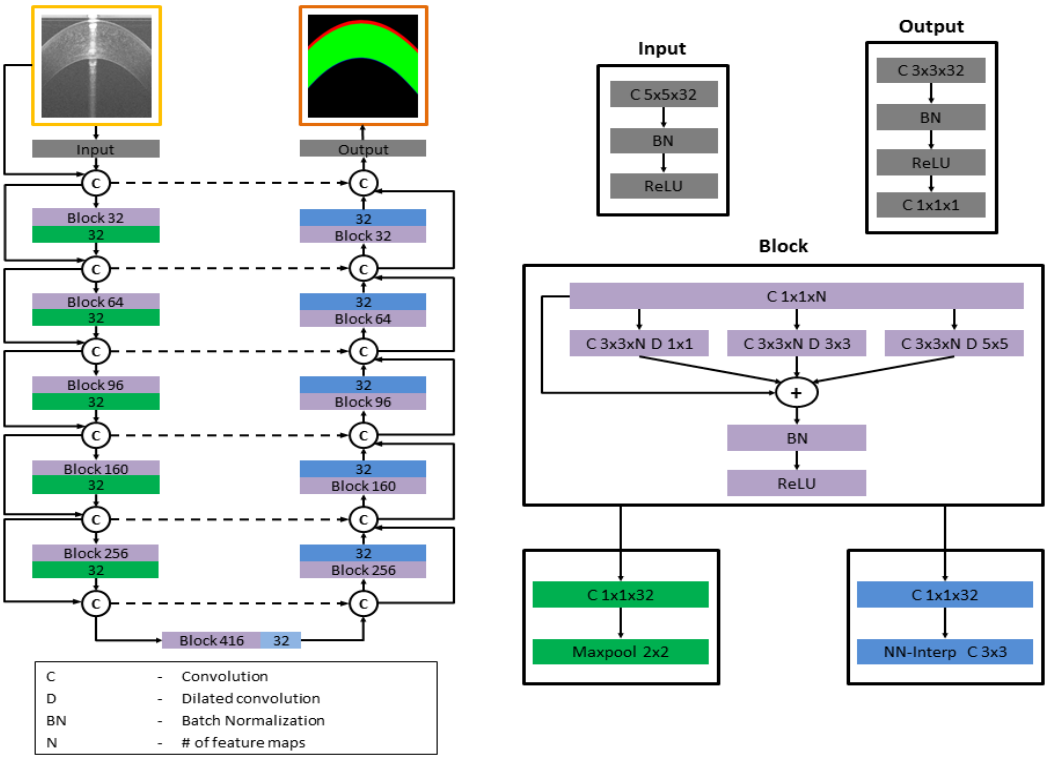


Figure 5.10: Our network architecture comprises of contracting and expanding branches. The dark green and blue blocks represent downsampling and upsampling computations respectively. Our network makes efficient use of residual and dense connections to generate the limbal interface segmentation in the final image, where each pixel is assigned the label of the tissue it belongs to. The input image is split width-wise into a set of slices of dimensions 256×1024 pixels, the network predicts an output for each slice, and the slices are aligned to recreate the original input dimension. Dense connections concatenate feature maps from previous layers. The light blue block at the bottom of the "U" (on the right end of the rectangle) does not perform upsampling, but as with the other blue blocks it functions as a bottleneck for the layer and generates feature maps of the same dimensions as the output feature maps from the previous layer.

- A deep network architecture with 5 layers [82] or 7 layers [79] utilized lots of parameters (44- and 176- million parameters), exhausting the computation power available in typical research settings.

To tackle the above issues, we propose the following, which is similar to [79]:

1. Our CorNet architecture is modular as we have a building block that consists of di-

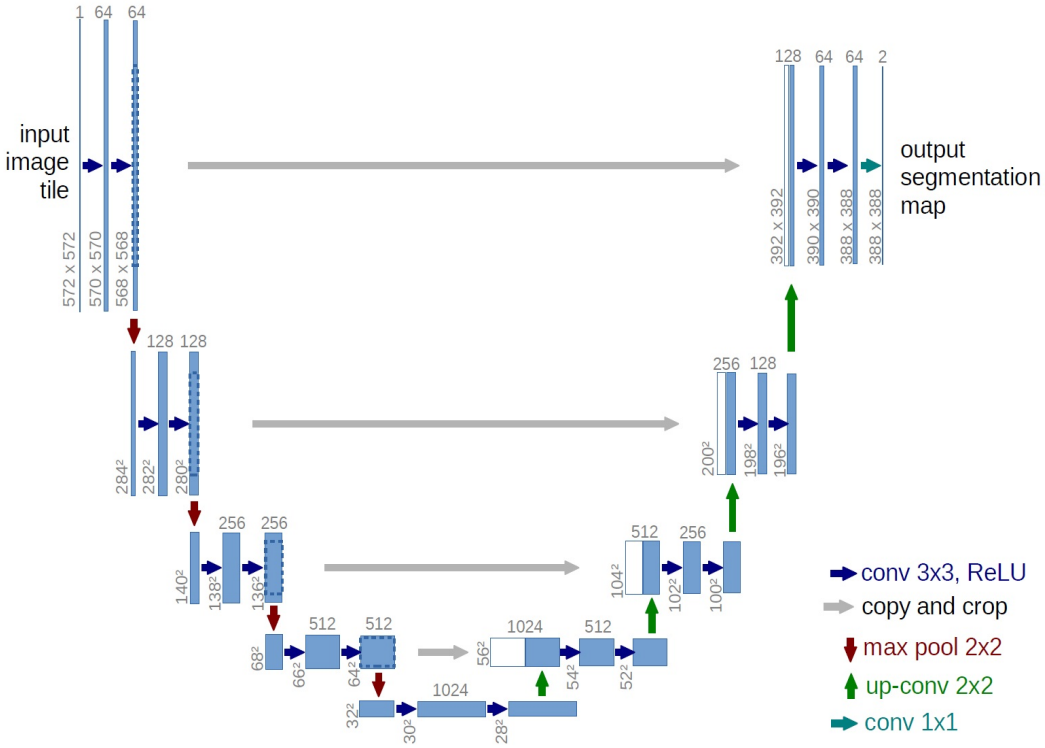


Figure 5.11: The U-Net network architecture proposed in [82]. Figure used with permission - Springer LNCS, Elsevier, license 4337120750575 [82].

lated convolutions with dilation rates of {1,3,5}. This building block is shown in magenta titled 'Block' in Fig. 5.10. The dilated convolutions increase the receptive field of the network, allowing each layer to see a larger context of the original input image. It also has the added benefit of not increasing the number of trainable parameters.

2. Each modular block in Fig. 5.10 is enhanced through the employment of residual connections [114, 136] and batch normalization [113]. With residual connections, the input of a layer is summed with the output of a layer, with no transformation being applied on the input. It enables a layer to learn new representations that are different from the input that was passed to it. Batch normalization allows a hidden layer in the network to be more robust to changes in the underlying distribution of the in-

put images that are passed to the network by normalizing the input batch of images to contain zero mean and a variance of a defined value. By reducing this internal covariate shift [113], it also allows for faster training, and higher accuracy.

3. The number of feature maps at the output of each modular block follows a Fibonnaci sequence. It was shown that this achieves a good trade-off between segmentation performance, network depth, and trainable parameters [79].

4. Bottleneck connections [112, 137, 138] are added between the modular blocks so that the total number of parameters that need to be trained is lower.

5. Appropriately downsampled input images are concatenated with the output of each modular block in both the descending and ascending paths.

Crucially, we have modified the network architecture in [79] to not only include the appropriately downsampled input image to a layer (the branch residual), but to also concatenate the output of each modular block to successive blocks (the deep residual). This important modification is shown in the main architecture of Fig. 5.10. Similar to the Residual Network [114], the incorporation of the branch and deep residual from previous layers as the input to new layers was done in order to force the network to encode information that is different from something that it has already learned. This is particularly useful in dealing with shadowing and large drops in SNR along the tissue boundary. This modification also allows the CorNet to learn to encode subtle changes in the tissue structure.

Throughout the architecture of the CorNet, we employed a 3×3 pixel convolutional kernel with n feature maps with n increasing according to a Fibonacci sequence defined by {64,96,160,256,416}. The network depth was limited to 416 feature maps. This prohibits the number of trainable parameters from reaching the level in traditional U-Net type archi-

tectures. The total number of parameters in the U-Net that was trained was ∼44 million parameters, while the number of trainable parameters was ∼29 million in the CorNet, in spite of the extra connections. This combination of hyper-parameters for training the CorNet allowed for an increase in segmentation accuracy and a reduction in the number of trainable parameters.

5.2.4 Curve Fitting and Registration

The output of the network is a prediction for each B-scan in the input OCT sequence. The sliced output prediction was used to recreate the original input dimensions as shown in Fig. 5.12. Next, the predicted boundaries are fitted with a second or third order polynomial, as seen in Figs. 5.12(b) and 5.12(d), following the same curve fitting [104] approach in Sec. 5.1.4. Due to the presence of severe pathology, it can be quite challenging even for a human annotator to accurately determine the tissue interface position. As expert detection of a pathological boundary is difficult, an automated algorithm would not be expected to beat the expert especially when the position of the boundary is ambiguous. To this end, we fit curves to both the expert annotations and to the automated segmentations, and then extract C-mode sections as shown in Fig. 5.14 following the same approach in Sec. 5.1.5.

5.2.5 Experiments

Training

The output of the CorNet is an image of the same dimensions as the input image, namely 1024×256 pixels. Every pixel in the output image took a value in the range [0,1], which corresponds to the value of the limbal tissue interface. The loss function used in this network architecture was the Mean Squared Error (MSE) loss. This loss function minimizes

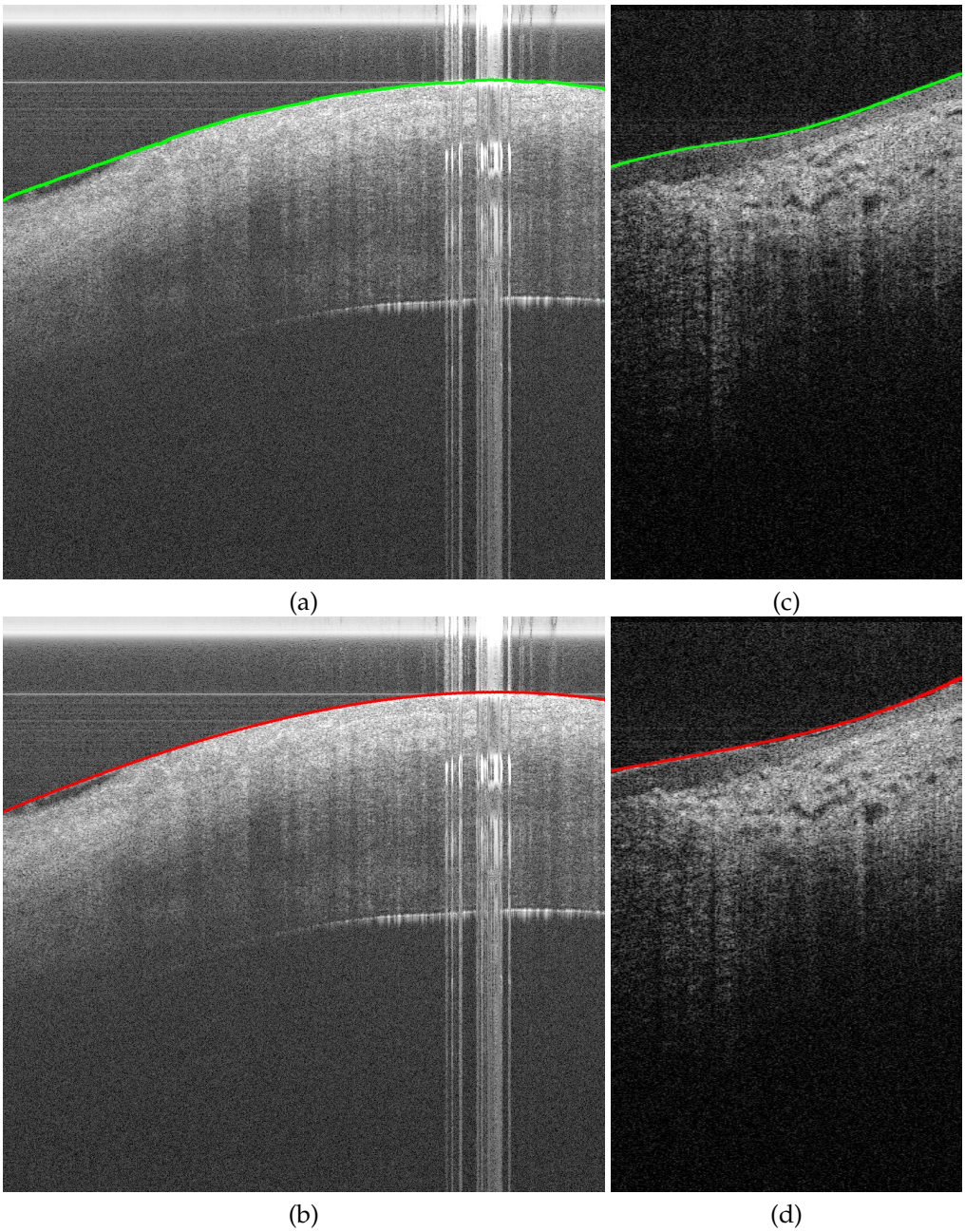
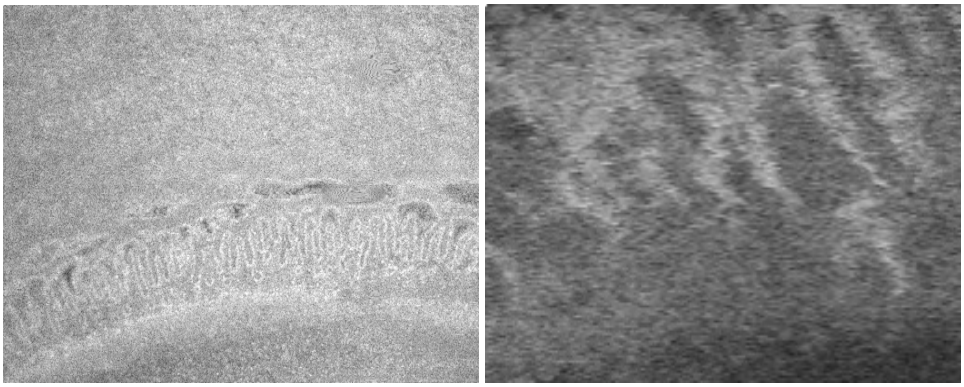
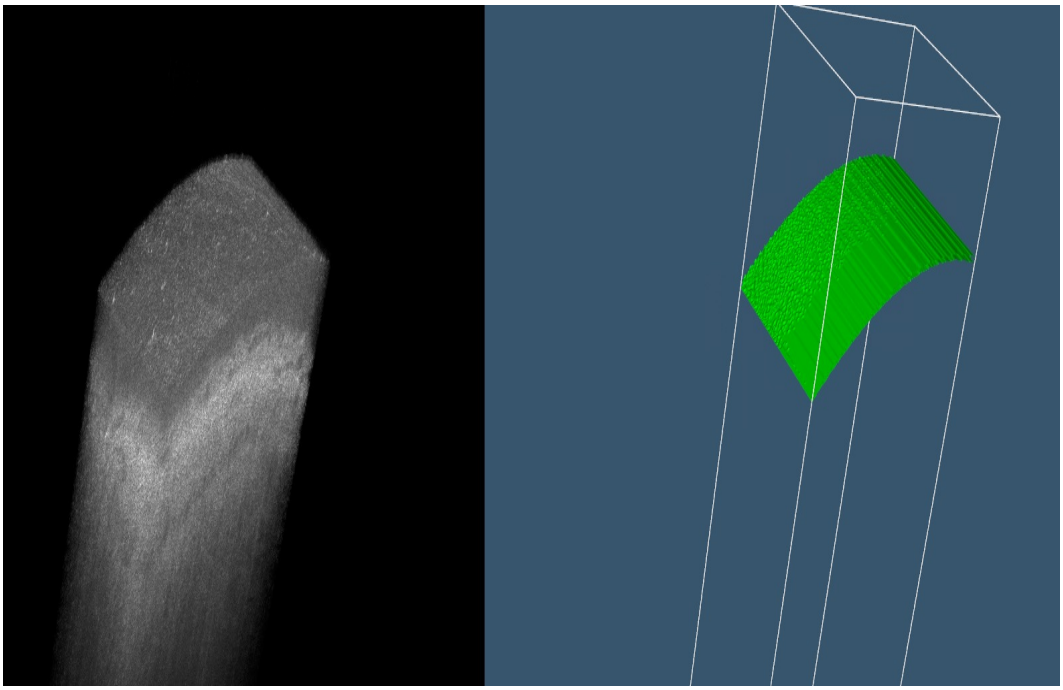


Figure 5.12: (a) and (c) show the tissue interface boundary predicted (green) by the CorNet. (b) and (d) show the result of fitting curves (red) to the predicted tissue interface boundary using the approach in [104].



(a) (b)

Figure 5.13: (a)-(b) show cross-sectional visualizations of the unique POV structural configuration. Notice the ridge- or finger-like cavities in the image, which represent the POV.



(a)

Figure 5.14: Side view of the reconstructed 3D volume of the limbal region. The volume was created by segmenting the visible tissue interface in every B-scan in the volume and aligning them, with the segmented interfaces being shown as the green surface on the right side of the rendering.

the pixel-wise error in the MSE between the predicted output and the ground truth annotation. The network parameters are updated using back-propagation and the Adam optimizer [139]. Out of the 36 annotated volumes that were available, 24 volumes were used as a training set to train the network for 150 epochs. The entire training dataset, without augmentation of data, comprised of 7612 images. The images were then augmented [121] by the addition of gaussian noise, gaussian filtering, flipping, bilateral filtering, and gamma adjustment with values 0.9, 1.2 and 1.5. The total augmented training dataset comprised of 60896 images, an $\sim 8\times$ increase in the original number of images. Among the remaining 12 datasets, a dedicated validation set of 4 volumes, that is distinct from the training dataset, was used compute the validation loss, and trigger early stopping if the validation loss did not improve.

An initial learning rate of 10^{-3} was used to commence training. This learning rate was reduced by a factor of 2 when the validation loss did not improve for 2 consecutive epochs. The lowest the learning rate can drop down to is 10^{-12}. The training of the network was also stopped early if the validation loss did not improve for 5 consecutive epochs. The training ends when the validation loss does not improve, and the network weights at the stage with the lowest validation loss are saved. The testing dataset comprised of 8 held-out volumes that were not seen during training or validation. During testing, the images were either padded or sliced width-wise to match the input dimensions of the network (1024×256 pixels), and are put back together to match the original image size after prediction.

5.2.6 Results

The segmentation accuracy and registration accuracy (as a function of cross-section extraction) were evaluated. Segmentation accuracy was measured by comparing the predicted re-

sult from the network after curve fitting against the experts' manual annotations. Images in all the 8 testing datasets that were used were manually annotated by the expert. The manual annotations were registered using the procedure described in Sec. 5.1.5, and cross-sections were extracted. The Structural Similarity (SSIM) image metric [132] was used to determine the correlation between cross-sections that are automatically extracted and cross-sections that were extracted after the alignment of the expert manual annotations. Registration effectiveness was assessed by the cross-sectional SSIM value. The validation protocol in [57] was followed, and the expert annotation was fitted with a curve, and compared against the curve fitted results. The following metrics were calculated: 1) *Chamfer Distance* [140], 2) *Hausdorff Distance* [141], 3) *Mean Absolute Distance* (MAD) [93], and 4) *SSIM* between cross-sections [132]. These metrics were chosen since traditional metrics, such as MAD, that rely on pixel based distances do not convey the sharp changes in the convexity/concavity of the curved surface, and are not well suited to validation. As the CorNet is not restricted to only convex shapes, the Chamfer and Hausdorff distance measures convey the changes in curvature to the ground truth annotation (fitted with a curve). For comparison, we also trained a U-Net and estimated the segmentation and registration accuracy results. The results for the U-Net are shown in Fig. 5.15 and Table 5.2, while the results for the CorNet are shown in Fig. 5.16 and Table 5.3. The surface segmentation error is minimal when error values are close to 0. As the segmentation error is closer to 0, SSIM values (which can take a maximum value of 1) are higher indicating a similarity between the extracted cross-sections.

At first glance, the CorNet seems to be performing only marginally better than the U-Net. However, this does not indicate the inferiority of the underlying predictions of the U-Net in several failure cases that the CorNet is able to handle. As some of the additional and most challenging test datasets that we have tested our algorithm on have not been annotated

by an expert, there are no metrics available yet in these cases to quantitatively measure the predictive power of the CorNet over the U-Net. Instead, we rely on the output images of the CorNet and U-Net to describe the improvement in performance shown by the CorNet over the U-Net. Below, we first detail the results for datasets that have been annotated, and compare the U-Net results against the CorNet results. Then, we elaborate on the failure cases when the U-Net performs worse than the CorNet, and explain its utility that allowed it to fare significantly better than U-Net. We provide output images from the intermediate stages of the CorNet that attest to the performance gains that are seen.

Table 5.2: U-Net - average quantitative error metrics across all 8 datasets. The pixel resolution in an A-scan is 1.3µm.

	Grader 1
Chamfer Distance	6.56 ± 0.51 pixels
Hausdorff Distance	4.8 ± 1.45 µm
MAD	0.59 ± 0.27 µm
SSIM	0.84

Table 5.3: CorNet - average quantitative error metrics across all 8 datasets. The pixel resolution in an A-scan is 1.3µm.

	Grader 1
Chamfer Distance	6.47 ± 0.51 pixels
Hausdorff Distance	4.58 ± 1.31 µm
MAD	0.55 ± 0.25 µm
SSIM	0.86

In Table 5.3 and Fig. 5.16(a), the MAD error, plus standard deviation, of the predicted output of the CorNet (fitted with a curve) is 0.80µm. This value is 7.5% better than the MAD error, plus standard deviation, of the predicted output of the U-Net (fitted with a curve), which was 0.86µm as seen in Table 5.2. This error is also lower than the hsUHR-OCT system pixel resolution. However, this error does not provide much information as it does not quantify the distance of the ground truth annotation (fitted with a curve) from the

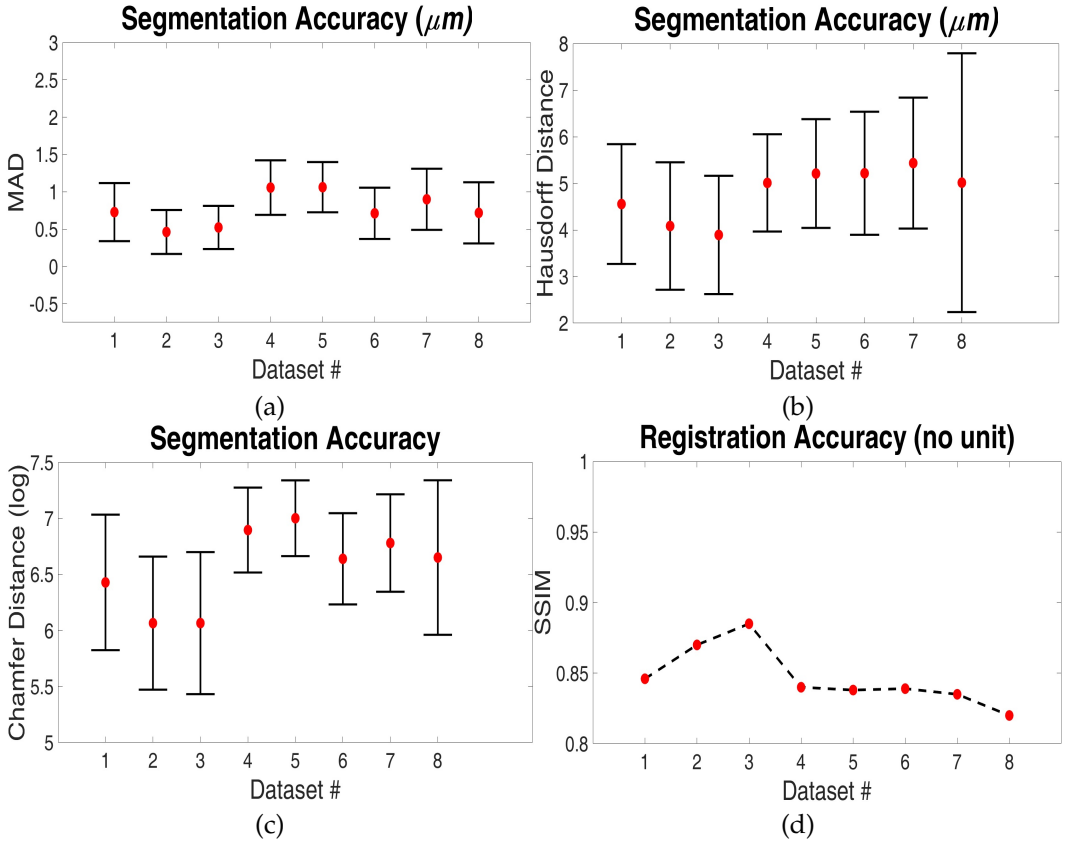


Figure 5.15: U-Net based segmentation and registration accuracy results. (a) Mean (red circle) and standard deviation error (black error bars) of the MAD error between automated segmentation and manual annotation; (b) Mean (red circle) and standard deviation error (black error bars) of the Hausdorff Distance error between automated segmentation and manual annotation; (c) Mean (red circle) and standard deviation error (black error bars) of the Chamfer Distance error between automated segmentation and manual annotation; (d) SSIM values of cross-section comparison.

prediction (fitted with a curve), nor does it show the changes in the ground truth annotation (fitted with a curve) that affects the predicted curves' distance to the annotation curve.

To this end, we computed the Hausdorff error and the Chamfer error. From Table 5.3 and Fig. 5.16(b), the average Hausdorff distance error, plus standard deviation, was less than 5.89µm or ∼4.53 pixels. This value was lower than the average Hausdorff distance error, plus standard deviation, of the U-Net, which was 6.25µm or ∼4.8 pixels as seen in Table 5.2.

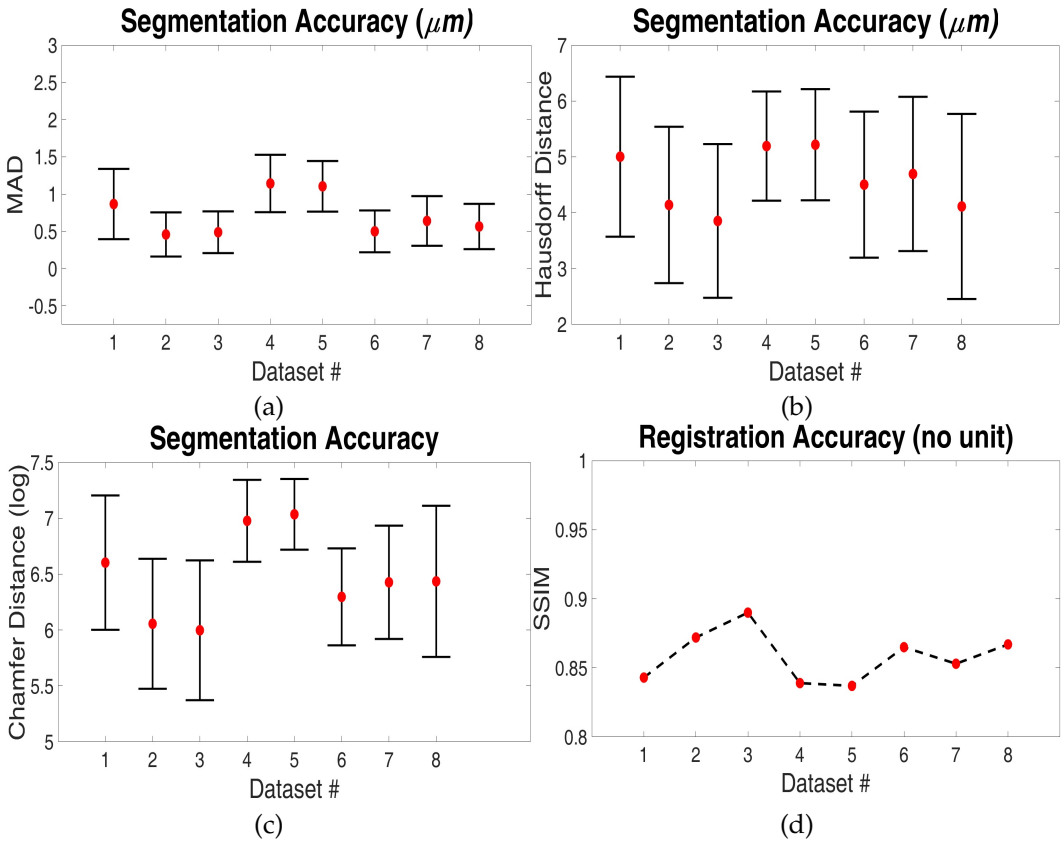


Figure 5.16: CorNet based segmentation and registration accuracy results. (a) Mean (red circle) and standard deviation error (black error bars) of the MAD error between automated segmentation and manual annotation; (b) Mean (red circle) and standard deviation error (black error bars) of the Hausdorff Distance error between automated segmentation and manual annotation; (c) Mean (red circle) and standard deviation error (black error bars) of the Chamfer Distance error between automated segmentation and manual annotation; (d) SSIM values of cross-section comparison.

The maximum distance between any point of the annotated curve to the predicted curve is defined by this value, thereby showing that the CorNet prediction (fitted with a curve) is closer to the ground truth annotation (fitted with a curve). It justifies the sensitivity to the position of each curve, and details the increase in segmentation error in microns of the U-Net prediction (fitted with a curve).

Similarly, the Chamfer error, plus standard deviation, of the predicted output of the Cor-

Net (fitted with a curve) was less than 6.98 pixels as described in Table 5.3 and Fig. 5.16(c). This value is also lower than the corresponding Chamfer error for the U-Net prediction (fitted with a curve), which was 7.07 pixels as shown in Table 5.2. This means that the distance from the ground truth annotation (fitted with a curve) to the CorNet predicted output (fitted with a curve) only differs by $\sim$7 pixels.

From the graphs, only datasets # 1, 4 and 5 had a higher MAD errors. Dataset #8 has a smaller MAD error, but the Hausdorff distance and Chamfer distance errors show a higher standard deviation. This is seen in the Chamfer and Hausdorff distances in Figs. 5.16(b)-(c) and Figs. 5.15(b)-(c). Corresponding lower SSIM values are also seen in Fig. 5.16(d). Closer examination of the errors revealed that slight changes in the numerical values of the estimated polynomial coefficients contributed to slight changes in the position of each curve, and thereby the error values were higher. We postulate that fitting a curve to the manual annotations induces a small degree of error, which is unavoidable. Nevertheless, the maximum error is acceptable as the similarity of the C-mode cross-sections that were extracted is higher than 0.86, and on average, this is the case as Table 5.3 shows that the average SSIM value is 0.86.

The network took $\sim$ 0.125 milliseconds to generate the predictions for a single image, and the fitting algorithm took $\sim$ 193 milliseconds per image to fit a curve and extract the C-mode cross-section. The network was trained on a NVIDIA Titan Xp GPU with Tensorflow and Keras, while the curve fitting and registration algorithm was implemented using C++, CUDA, OpenCV, and VTK. Consistent with our previous work in [3], we have shown that through deep learning and the results obtained from the network, it is possible to generate reproducible metrics that are consistent with expert annotations. As in our previous work [3], we have again demonstrated that our new methodology removes the subjectivity

of manual segmentation. Most importantly, it is generalizable to datasets that have been acquired with different SD-OCT systems, and this is a crucial drawback in [3]. Furthermore, the automated approach is also faster than manual alignment, and it constitutes transferable knowledge to the OCT imaging community. However, similar to the work in [3], the proposed deep learning based segmentation and registration algorithm does not take into account the out-of-plane tissue motion. Future work is directed towards using the segmentation and registration results to further infer the deformation field that is necessary to align the structures that are present in the B-scan to a reference B-scan.

Experimental Observations

As mentioned before, the error statistics for the CorNet and the U-Net are nearly similar. However, results were only computed on 8 annotated volumes. When tested on other unannotated datasets, there were significant differences in the prediction accuracy between the two networks. One significant difference was visualized when dealing with very low intensity SD-OCT images in volumes with the underlying tissue having undergone significant morphological changes in tissue structure due to pathology/disease as seen in Fig. 5.17. In Fig. 5.17, there are significant changes in the tissue structure directly underneath the shallowest tissue interface. If this tissue interface is not detected correctly, then the underlying pathology cannot be examined appropriately. Furthermore, it is quite difficult to accurately pinpoint the tissue boundary on the left side of the image in Fig. 5.17(a).

In the U-Net, the lack of appropriate context from the first few layers, as shown in Fig. 5.18(a), does not adequately inform the later layers in the network of the true tissue interface position. This localized information is lost in the deeper layers of the U-Net. On the other hand, the CorNet utilizes the contextual information from the first few layers, as shown

in Fig. 5.18(b), and is aided in localizing the tissue interface using through the use of the residual formulation [114, 136]. It is postulated that OCT image segmentation would benefit from an increased number of feature maps in the first layer of the network. It helps enhance the subtle tissue boundaries that may sometimes be lost due to corruption of the tissue boundary through speckle noise. In contrast to the U-Net, the last output layer of the CorNet architecture has accuratly defined and localized the tissue boundary as seen in Fig. 5.19. We attribute this to the skip connections that helps in training deeper models, while providing features from the encoder blocks to the decoder blocks for improved segmentation accuracy. The residual connections help the network learn to encode a different representation from the ones that have already been learned by the earlier layers of the network. This can be plainly seen in Fig. 5.19(b) when contrasted with the U-Net representation in Fig. 5.19(a).

At the lowest downsampling level in the contracting path of both the U-Net and the CorNet, lots of contextual information from the higher levels has been encoded. At this level, only gross depiction of the tissue boundary and surrounding structures are seen, while subtle changes in the features, such as edges, are not visualized. At the bottom of descending path in both architectures, the CorNet has learned to represent the desired tissue boundary better than the U-Net as shown in Fig. 5.20.

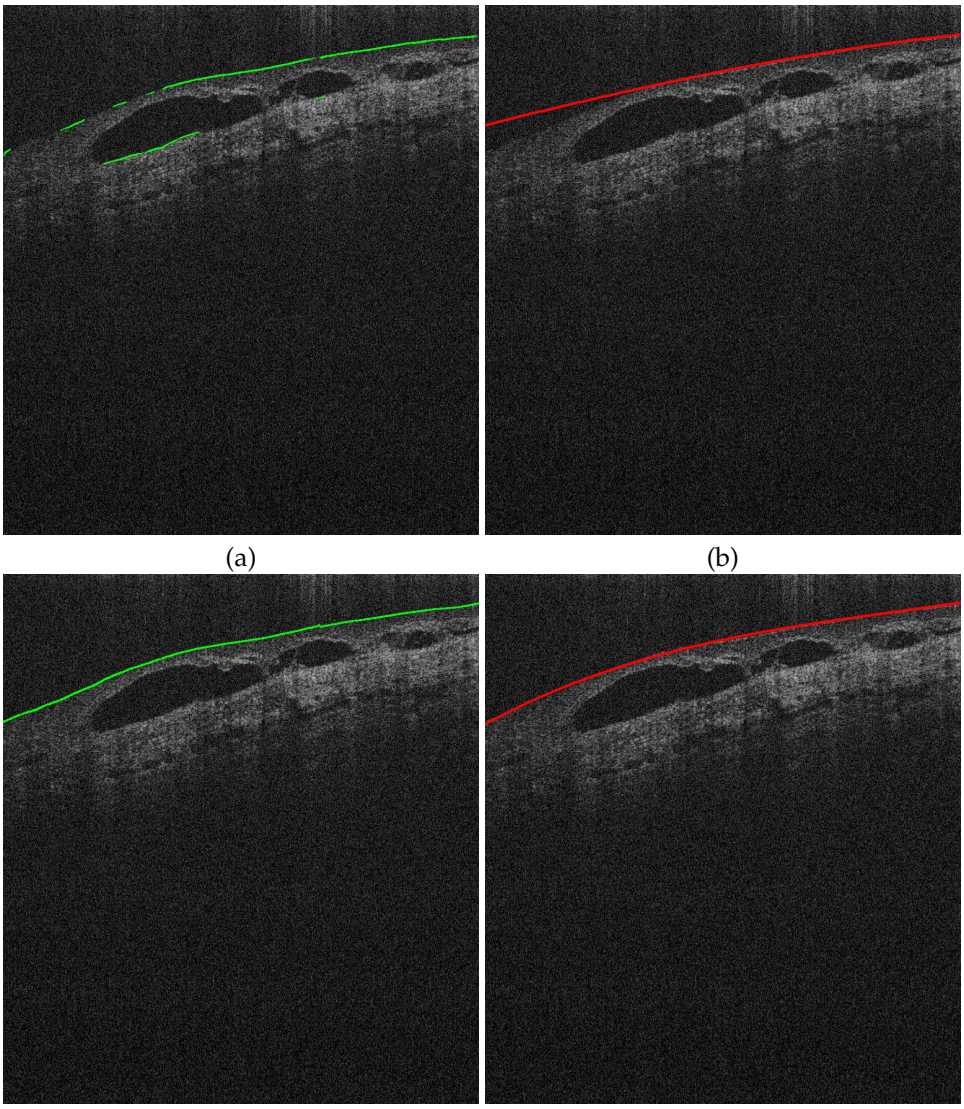
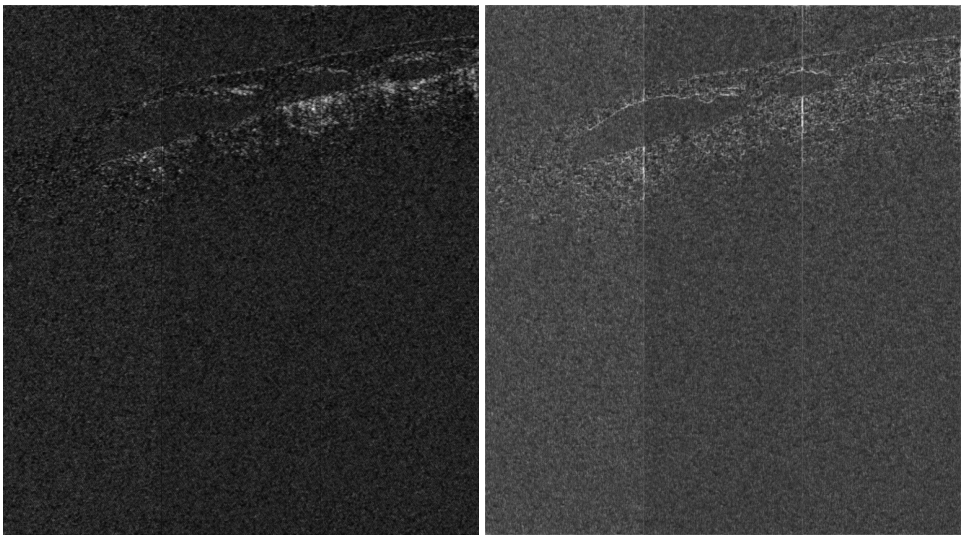
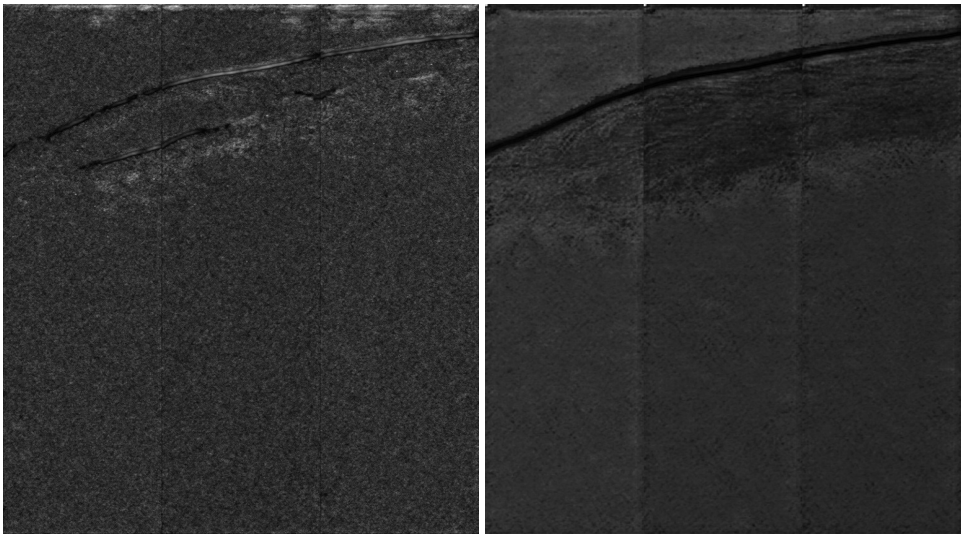


Figure 5.17: (a) shows the prediction (green) of the U-Net on the 58^{th} B-scan in a volume where the intensity of the tissue interface is significantly lower than a regular SD-OCT image of the limbal region; (b) shows the curve (red) fitted to the U-Net predicted output; (c) shows the prediction (green) of the CorNet on the same 58^{th} B-scan; (d) shows the curve (red) fitted to the CorNet predicted output. Notice the lack of a continuous predicted output in (a) as opposed to the predicted output in (c). In the absence of a prediction at the left side of the image in (a), the fitted curve in (b) is incorrect. A correctly predicted output (c) leads to a better curve fit (d), and thereby, accurate C-mode slice extraction.



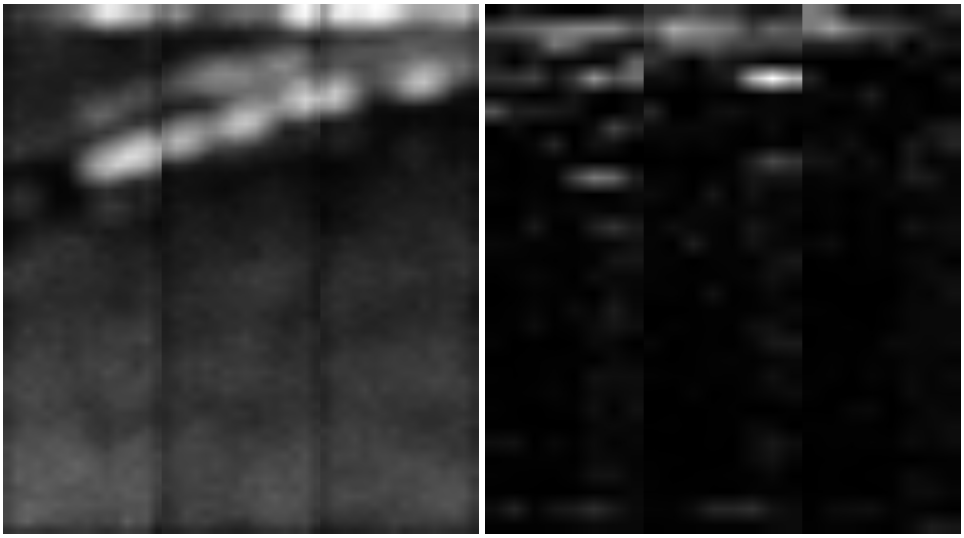
(a) (b)

Figure 5.18: (a) shows the network activation at the first input block of the U-Net. (b) shows the network activation at the first input block of the CorNet. The number of feature maps at the output of each layer was 128, and the images have been resized to match the original input image dimensions. The CorNet clearly seems to be learning tissue structure better. Even though the object boundary was not captured in the left side of the image, in the later layers of the CorNet, this object boundary was accurately examined.



(a) (b)

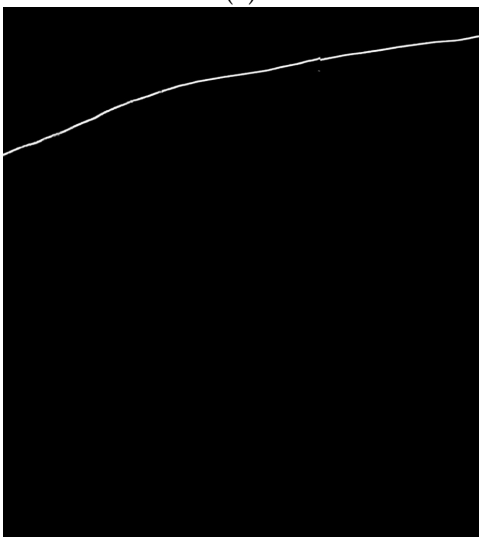
Figure 5.19: (a) shows the network activation at the last output block of the U-Net. (b) shows the network activation at the last output block of the CorNet. The number of feature maps at the output of each layer was 128, and the images have been resized to match the original input image dimensions. The CorNet has accurately determined the position of the limbal tissue interface.



(a)

(b)

(c)



(d)

Figure 5.20: (a) shows the network activation at the lowest downsampled output block of the U-Net with the number of feature maps being 1024. Each feature map has dimensions of 64×16 pixels. (b) shows the network activation at the lowest downsampled output block of the CorNet with the number of feature maps being 64 ([672, 64] block). Each feature map has dimensions of 32×8 pixels. The images have been resized to match the original input image dimensions. (c) and (d) show the final activations of the U-Net and the CorNet. The CorNet has clearly learned to represent the desired tissue interface more accurately.

5.3 Hybrid and Cascaded Approaches

5.3.1 Problem Statement

The problem statement for this work is the same as that provided in Sec. 4.3.1. In addition to generating accurate segmentations on corneal OCT images, we would like to also generate accuarate segmentations on limbal OCT images. The shallowest limbal interface in OCT image sequences are segmented using a cascaded approach and a hybrid approach. Our frameworks are pictorially shown in Fig. 5.21. We refer the reader to Sec. 4.3.1 for a brief overview of a similar problem statement applied to corneal images.

5.3.2 Data Acquisition

In addition to the corneal datasets chosen from the research database [5] in Sec. 4.3.2, limbal datasets were also added; 25 corneal datasets and 25 limbal datasets, totaling 50 datasets, were randomly selected from an existing research database [5]. 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 [109], a high-speed ultra-high resolution OCT (hsUHR-OCT) scanner (Device 2) [70], and a Leica (formerly Bioptigen) Envisu C2300 SD-OCT system (Device 3) [142]. Device 1 had a $3.4\mu m$ axial and $6\mu m$ 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.

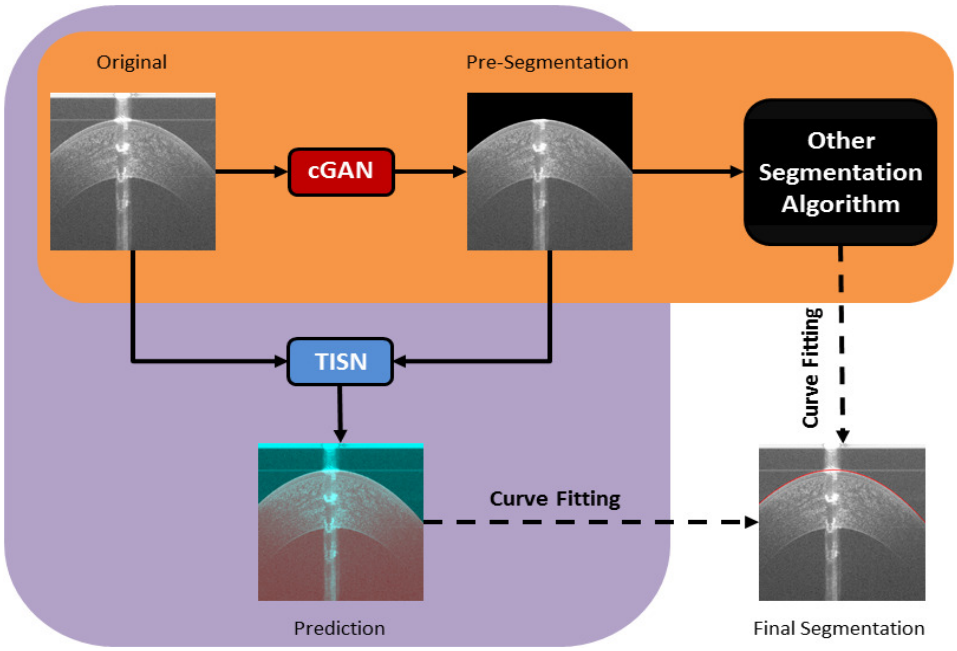


Figure 5.21: Our proposed approach contains two frameworks: a cascaded framework (purple) and a hybrid framework (orange). First, a conditional Generative Adversarial Network (cGAN) takes an input OCT image, and produces an intermediate pre-segmentation image. In the pre-segmentation, pixels just prior 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.

Devices 1 and 2 were solely used to scan the cornea, with the former producing datasets of dimensions $1024 \times 1000 \times 50$ pixels, and the latter generating datasets of dimensions $400 \times 1024 \times 50$ 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.

5.3.3 Adversarial Network Architecture

As described in Sec. 4.3.3, we utilized the same CorNet architecture design Sec. 4.2.2 in order to generate pre-segmentations of the limbal OCT images. The pre-segmentations were obtained from the conditional Generative Adversarial Network (cGAN) detailed in Sec. 4.3.3. Once the pre-segmentations are generated, they are utilized in a hybrid or cascaded framework by two distinct approachs. The cascaded framework uses the cGAN and the Tissue Interface Segmentation Network (TISN), which was described in Sec. 4.3.3, to segment the shallowest limbal interface in the limbal OCT sequence. The hybrid approach can utilize any prior segmentation algorithm [3,20,57,65,67–69,77,78,81,108,143–150] in order to estimate the limbal interface. In this work, we chose to use our previously described approach in Sec. 5.1 to segment the shallowest limbal tissue interface. We have described this approach in [3].

5.3.4 Experiments

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. These datasets were chosen such that they came from both eyes; the number of patients that were imaged could not be ascertained as the database contained deidentified datasets. 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.

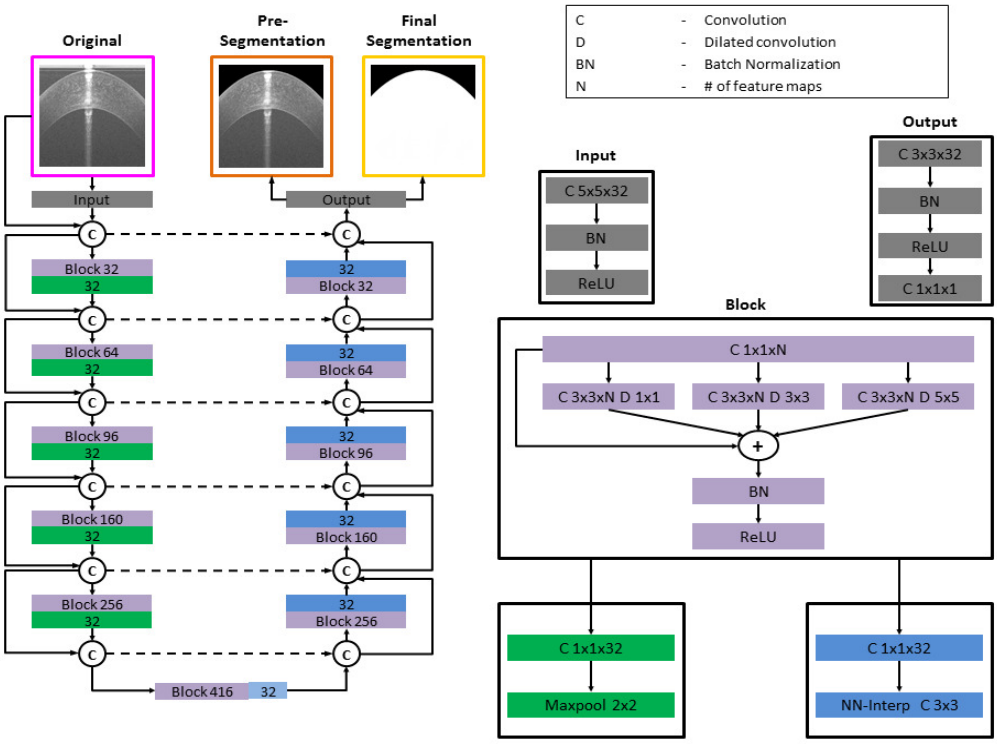


Figure 5.22: The CorNet model is the base architecture used for training both the cGAN and TISN. The input to the cGAN is a two-channel image, the input OCT image and binary mask w (see Sec. 4.3.4), and the output is a pre-segmented OCT image (orange box). The TISN gets a two-channel input (magenta and 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 block at the bottom of the "U" (on the right end of the rectangle) does not perform upsampling, but as with the other blue blocks it functions as a bottleneck for the layer and generates feature maps of the same dimensions as the output feature maps from the previous layer.

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 [5, 77], we sliced the input images width-wise to produce a set of images of dimensions 256×1024 pixels, and in this way, we preserved the OCT image resolution. We used the same datasets that were selected in the training set

for training both the cGAN and the TISN. We also refer the reader to Sec. 4.3.4 for related information regarding data preparation.

Data Augmentation

As our training datasets were smaller in number in contrast to those from datasets typically available in computer vision tasks, such as image recognition [126], we augmented our datasets to increase the variety of the images that were seen during the training. These augmentations [121] 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 [124]. For the cGAN, augmentation was done by simply flipping each input slice horizontally along the X-axis.

cGAN and TISN Training

Similar to Secs. 4.3.3 and 4.3.3, training of the cGAN and the TISN commenced from scratch using the architecture shown in Fig. 5.22. 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. During test time, the input OCT image is replicated to produce a two-channel input to the cGAN. 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 feature maps in the final layer

of the network are activated using the softmax function to produce a two-channel output. 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 [123] after the tissue interface was identified using a fast GPU-based method [1]. We show our final results in Fig. 5.23.

Comparison Against Other Algorithmic Approaches

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 [3] 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 [128], and segmented the tissue interface. We denote this approach 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 [3] 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 algorithmic approach by the acronym: Traditional With Pre-Segmentation (TWPS).

Third, we trained a CorNet architecture [5] to directly segment the foreground in the input OCT image, *without* including the cGAN pre-segmentation as an additional input chan-

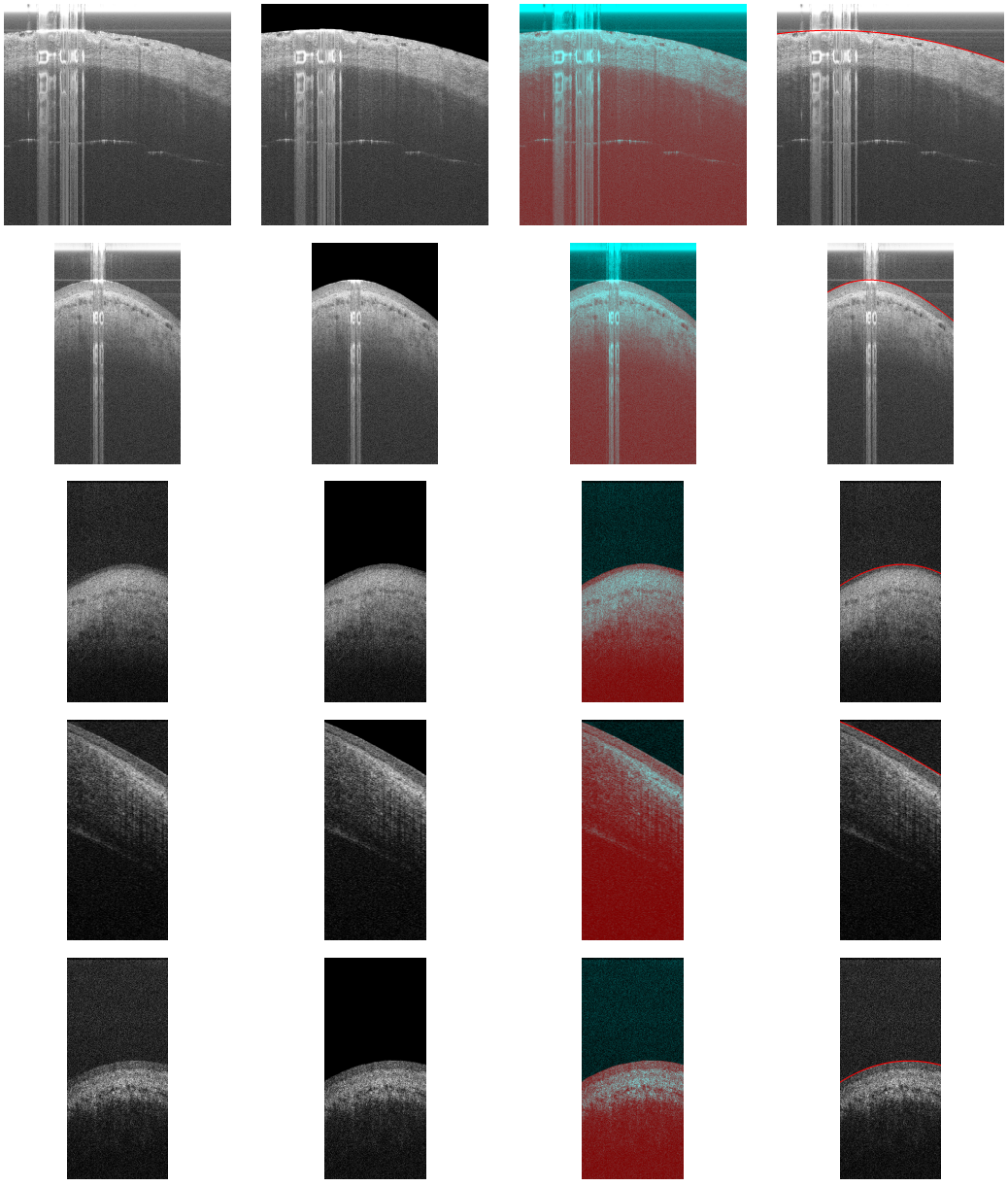


Figure 5.23: Limbal interface segmentation results for datasets acquired using Devices 2 and 3. Columns from left to right: (a) Original B-scans in the limbal OCT datasets, (b) Pre-segmented OCT images from the cGAN with the specular artifact and speckle noise patterns removed above the shallowest tissue 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).

nel. 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 algorithmic approaches were considered for performance evaluation:

1. TWOPS - A traditional image analysis-based algorithm [3] that directly segmented the tissue interface.

2. TWPS - The hybrid framework.

3. DLWOPS - A deep learning-based approach [5] that directly segmented the tissue interface.

4. DLWPS - The cascaded framework.

Annotation

Each limbal dataset was annotated by an expert grader (G1; Grader 1). The grader was asked to annotate the shallowest interface in all test datasets. For each dataset, the grader 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 approaches.

5.3.5 Results

Metrics

In order to compare the segmentation accuracy across the different algorithmic approaches, 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. (5.4) and (5.5), 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 Euclidean distance of a point p in G to the closest point in S, and similarly for $d_G(p)$.

We chose MADLBP in Eq. (5.4) as one of our error metrics since it was used in [3] 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. (5.5) 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.

$$\text{MADLBP} = \frac{1}{X}\sum_{x=0}^{X-1}|y_G(w) - y_S(w)| \tag{5.4}$$

$$\text{HD} = \max\left(\max_{p\in G} d_S(p),\ \max_{p\in S} d_G(p)\right) \tag{5.5}$$

We also determined the HD and MADLBP error across the limbal datasets in Figs. 5.24 and 5.26. Again in Fig. 5.25, we estimated the benefit of pre-segmenting limbal datasets by grouping methodologies into two categories- Traditional Comparison (TC; TWOPS vs TWPS) and Deep Learning Comparison (DLC; DLWOPS vs DLWPS) - 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. 5.25 (results after the red vertical line).

Discussion - Segmentation Accuracy of Limbal Interface

We plotted the segmentation error for the algorithmic approaches executed on limbal datasets in Figs. 5.24, 5.25 and 5.26. In Fig. 5.24, 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. 5.26, we plot the errors only for the successful instances of interface segmentation. From Figs. 5.24 and 5.26, the error for the TWOPS approach is the worst amongst all methodologies as it failed to handle strong specular artifacts and severe speckle noise. On the other hand, the TWPS approach fared better with lower errors than the TWOPS approach. 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. 5.25. From the errors (after the red vertical dashed line) in Figs. 5.25(a) and 5.25(c), the hy-

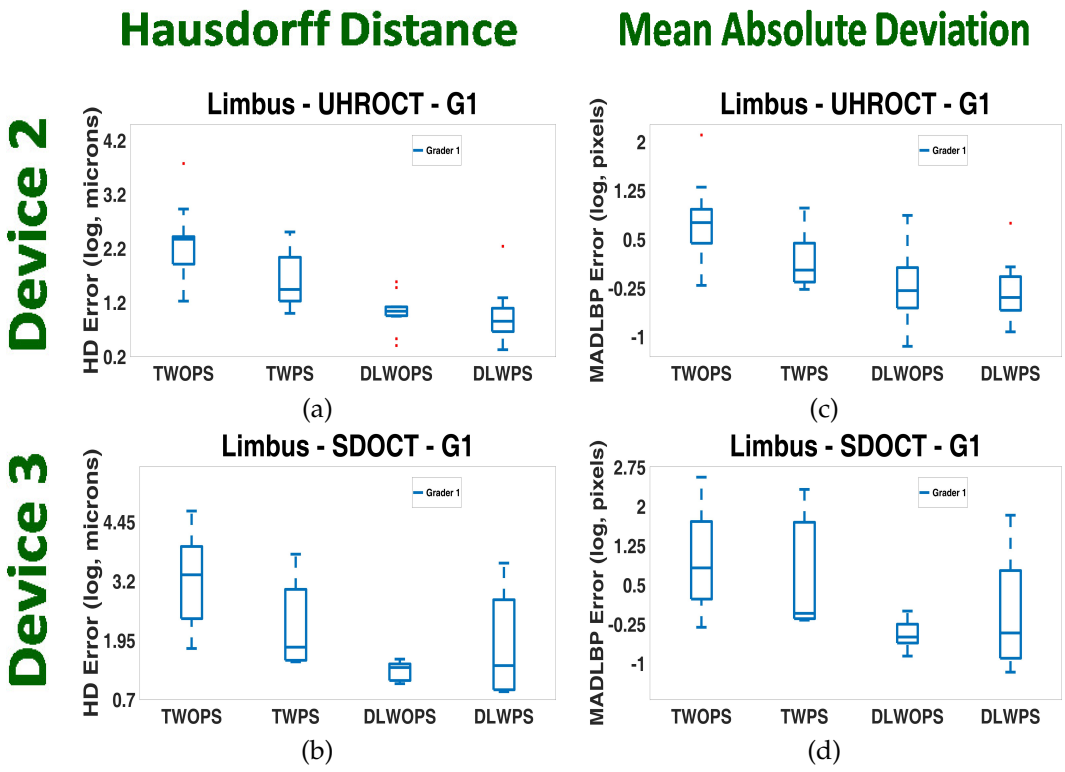


Figure 5.24: (a)-(b) HD error and (c)-(d) MADLBP error comparison for the limbal datasets acquired with Devices 2 and 3 respectively. For the limbal datasets, the segmentation results obtained for each algorithmic approach were contrasted exclusively against the expert annotations (G1). This graph plots the errors across all limbal datasets, including the failure cases. In contrast to Fig. 5.26, note the increased segmentation error in the DLWPS approach due to imprecise pre-segmentations.

brid framework (TWPS approach) was able to reduce the segmentation error even with an incorrect OCT image pre-segmentation. Therefore, the incorporation of the pre-segmented OCT image in the hybrid framework lead to lower errors for the traditional image analysis-based approach.

The DLWOPS approach had lower errors as shown in Figs. 5.24 and 5.26 as compared to the TWOPS and TWPS approaches. But, at an image level, it sometimes yielded higher segmentation errors as seen in Figs. 5.25(b) and 5.25(d). On the other hand, the DLWPS approach (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. 5.27, 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.

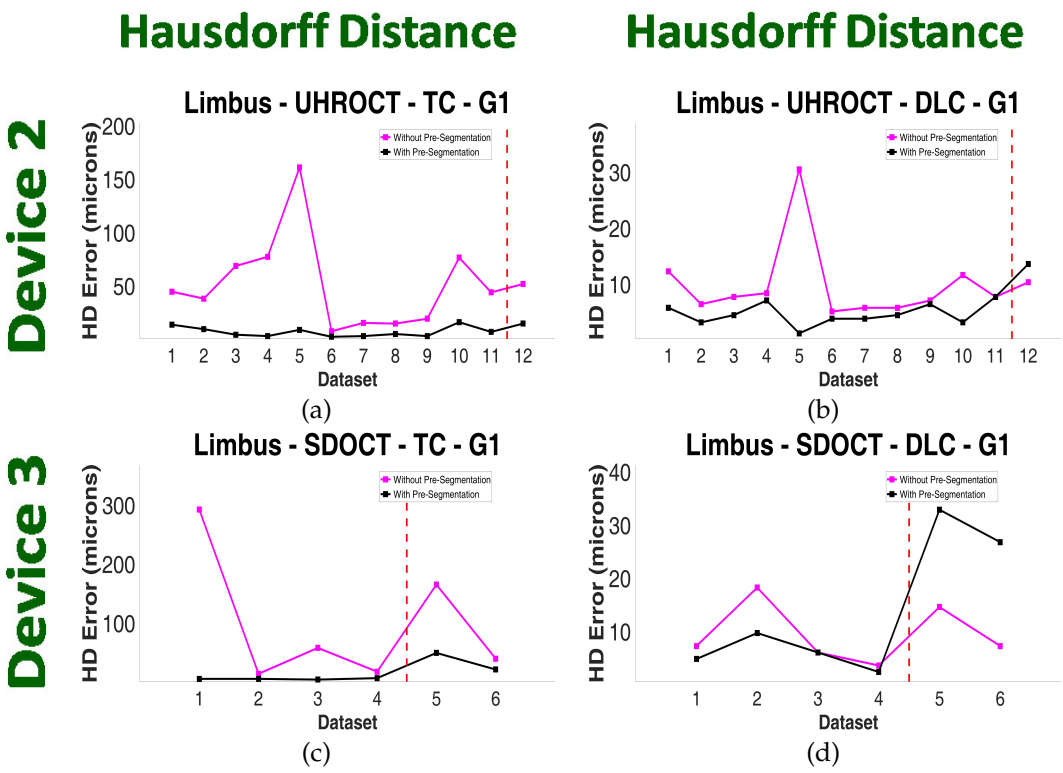


Figure 5.25: Quantitative estimation of the benefit of pre-segmenting the corneal OCT image. All the approaches were grouped into two categories: TC (TWOPS vs TWPS), and DLC (DLWOPS vs DLWPS). The first column corresponds to the former, and the second column corresponds to the latter. For each test dataset, 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 approaches respectively. The stored location indicies were then used to retrieve the corresponding HD errors from the TWPS and DLWPS approaches 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.

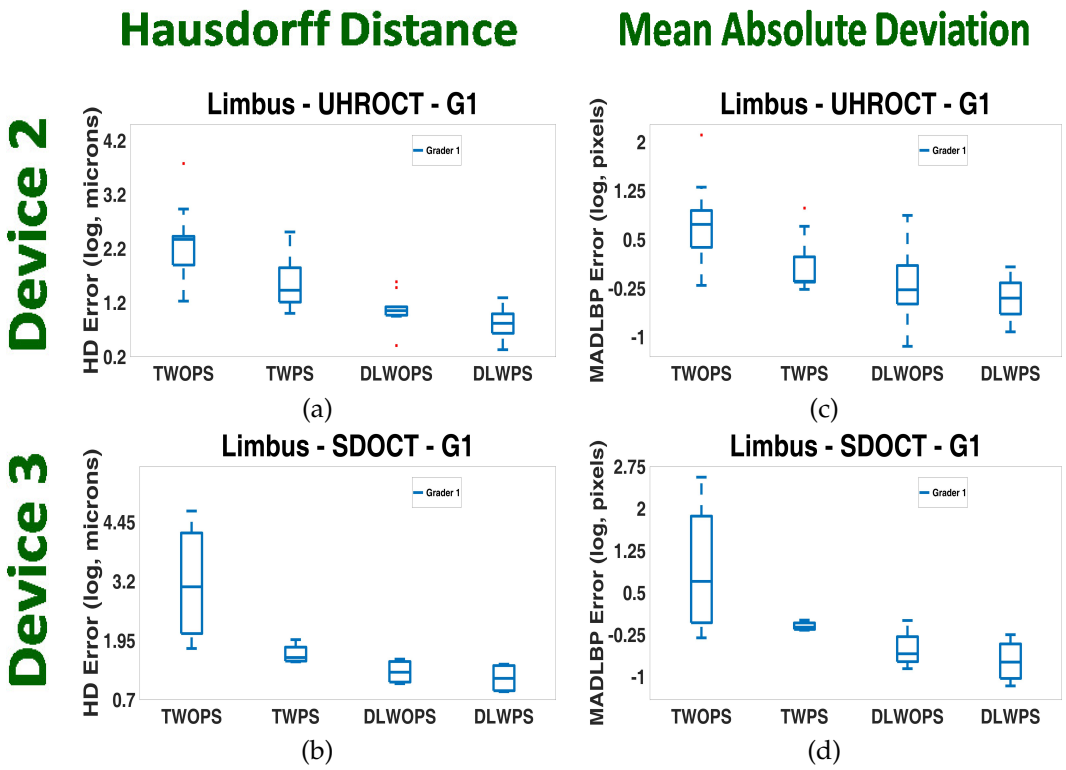


Figure 5.26: (a)-(b) HD error and (c)-(d) MADLBP error comparison for the limbal datasets acquired with Devices 2 and 3 respectively. For the limbal datasets, the segmentation results obtained for each algorithmic approach were contrasted exclusively against the expert annotations (G1). These graphs plot errors for the successful segmentation results on 15 limbal test datasets.

Consequently, the segmentation error of the TWPS (hybrid framework) and DLWPS (cascaded framework) approaches was increased. As seen in Fig. 5.25 (after the red vertical dashed line), the DLWOPS approach performed the best among all other approaches 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 [3, 5, 57] as shown in Fig. 5.27(a) (top two rows), or blend in with the background and specular artifacts of the same amplitude [57]

as seen in Fig. 5.27(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. 4.22) 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. 5.27(b). Hence, the performance of the TISN on such incorrectly pre-segmented OCT images is poor.

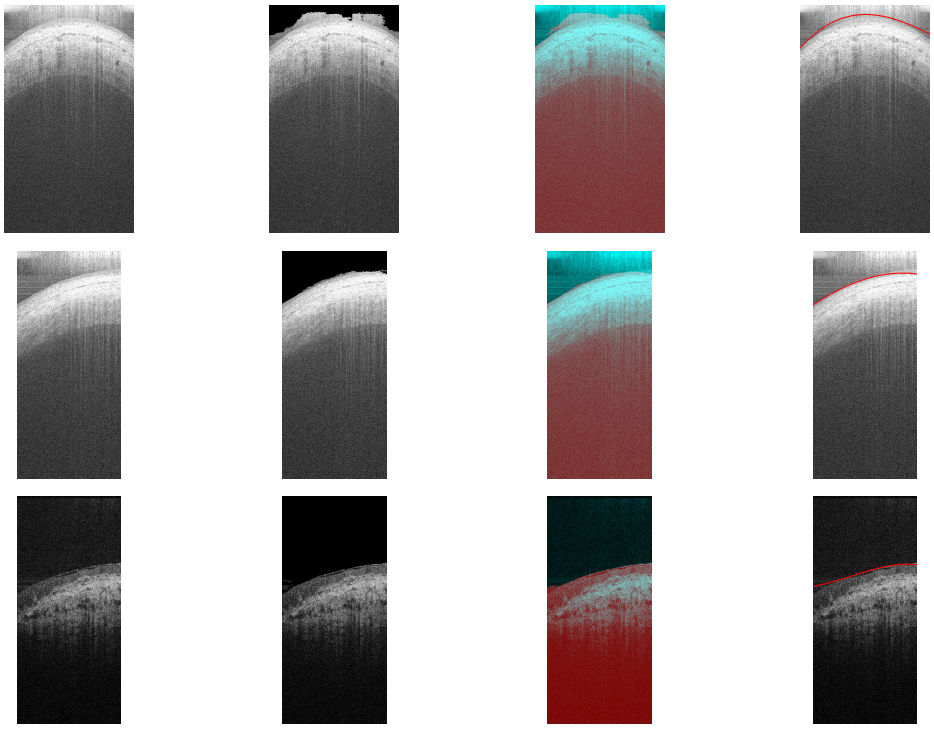
One way to address this issue is to re-train the framework with gold standard pre-segmentations that have corrupted boundaries. In this pilot work, we did not introduce any corruption to the gold standard pre-segmentation used during training as we wanted to directly measure the performance of the TISN 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 B-scans in the dataset through recurrent neural networks, which retain long-term information in memory in order to deal with such challenging datasets. 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 methodologies for each error metric,

and determined that our errors were statistically significant ($p_{\text{HD}} = 0.0471$, $p_{\text{MADLBP}} = 0.0313$). We also calculated paired t-tests between the DLWPS approach and all other approaches to determine the statistical significance of our results for each error metric. As seen in Table. 5.4, our DLWPS cascaded framework generated statistically significant results ($p < 0.05$).

Table 5.4: Statistical significance between our cascaded framework (DLWPS) against each approach 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



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

Figure 5.27: 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).

5.3.6 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. 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. 5.28, wherein the shallowest interface is identified in B-scans that partially overlap both the cornea and the limbus.

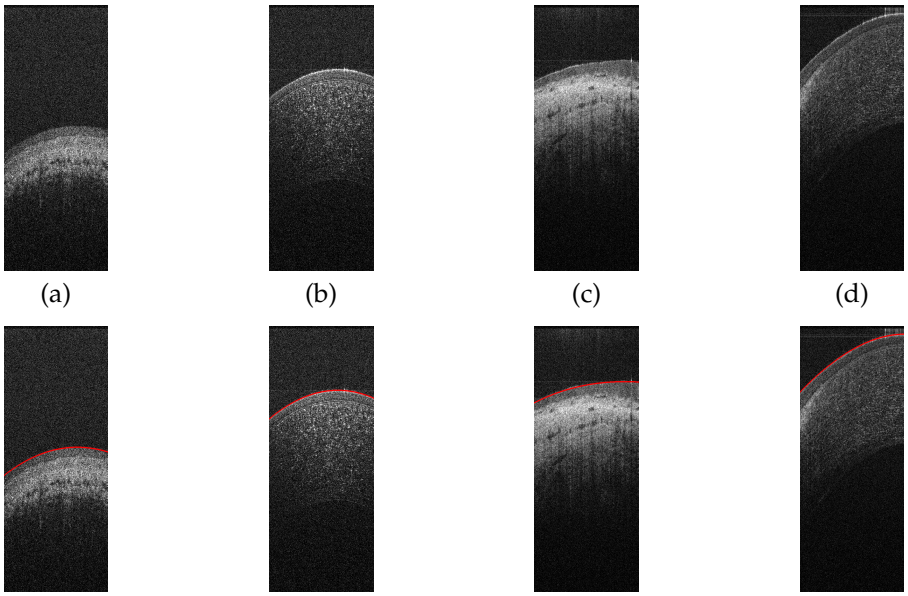


Figure 5.28: Segmenting the shallowest tissue interface in OCT datasets, wherein the OCT scanner commenced imaging from the limbus and crossed over into the cornea, thereby encompassing the limbal junction. (a),(b) B-scans #1 and #300 in an OCT dataset corresponding to the limbus and the cornea respectively. (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).

5.4 Longitudinal Study of Morphological Changes in the Palisades of Vogt

5.4.1 Problem Statement

A longitudinal study was conducted to determine the changes in structure of the Palisades of Vogt (POV). The goal is to extract C-Mode cross sections of the POV from 3D reconstructions of the limbal regions, and quantify the morphological changes in the POV by examining the cross-sections across each of the three days.

5.4.2 Data Acquisition

The cornea-scleral rims (intersection of the cornea and the sclera) of two cadaveric eyes (left and right eyes) were imaged using using an SD-OCT scanner [142]. We named this OCT scanner "Device 3". It had a $\sim$2.44μm axial and 12μm lateral spacing when fitted with the 18mm anterior imaging lens. The corneal rim containing the limbus can be divided into twelve (12) equal parts, similar to the hours shown on a clock. Datasets were obtained for each clock hour and the imaging was conducted for a total of 12 days.

However, the POV are most visible in the superior and anterior corneal rim section of the eye between the following clock hours: superior (10 o'clock - 2 o'clock) and inferior (4 o'clock - 8 o'clock). We only utilized the datasets in the superior and inferior corneal rim regions between the following clock hours: superior (11 o'clock - 1 o'clock) and inferior (5 o'clock - 7 o'clock). This corresponded to 3 datasets per region with 6 datasets per eye. Both eyes were imaged for 3 days, resulting in 12 datasets per day and 36 datasets in total. Due to manual positioning of the OCT scanner, the clock-hour locations were only approximate and are not consistent from day to day.

We compared the datasets across 3 days because the changes in the POV were not quantifiable by our approach after the 4^{th} day since the tissue had degraded to the point where curve fitting no longer captured the right regions of interest. Subsequent cross-sectional slices that were extracted were incorrect as there was misalignment between B-scans in the 3D reconstruction.

5.4.3 Approach

The shallowest limbal interface was identified from each OCT dataset as described in Secs. 5.2.3 and 5.2.4. Once these C-mode cross-sections were extracted from the registered limbal volumes, we manually compensated for day-to-day clock-hour misalignment by visually finding 2D correspondences between days. We matched visually similar POV regions across multiple days for each of the clock hours. Then, these corresponding POV structures were manually annotated by an expert and the area of the annotations was computed, allowing comparison of POV area across each day. If multiple annotated areas of correspondence were obtained from the expert, then the similarities across the largest annotated area was compared. An example of this process is shown in Figs. 5.29 and 5.30.

5.4.4 Results

The results of estimating the morphological change over each of the three days are shown in Table. 5.5. From our results, there is a trend of decreasing area measures for the annotated POV regions. This means that the size of the POV is diminishing across each day, and across each eye. The structure of the POV regions certainly appears to be stable across the first two days as shown in Figs. 5.29 and 5.30. However, there appears to be deterioration in the general shape of the annotated POV regions on the third day - see (c) and (f) in Figs. 5.29

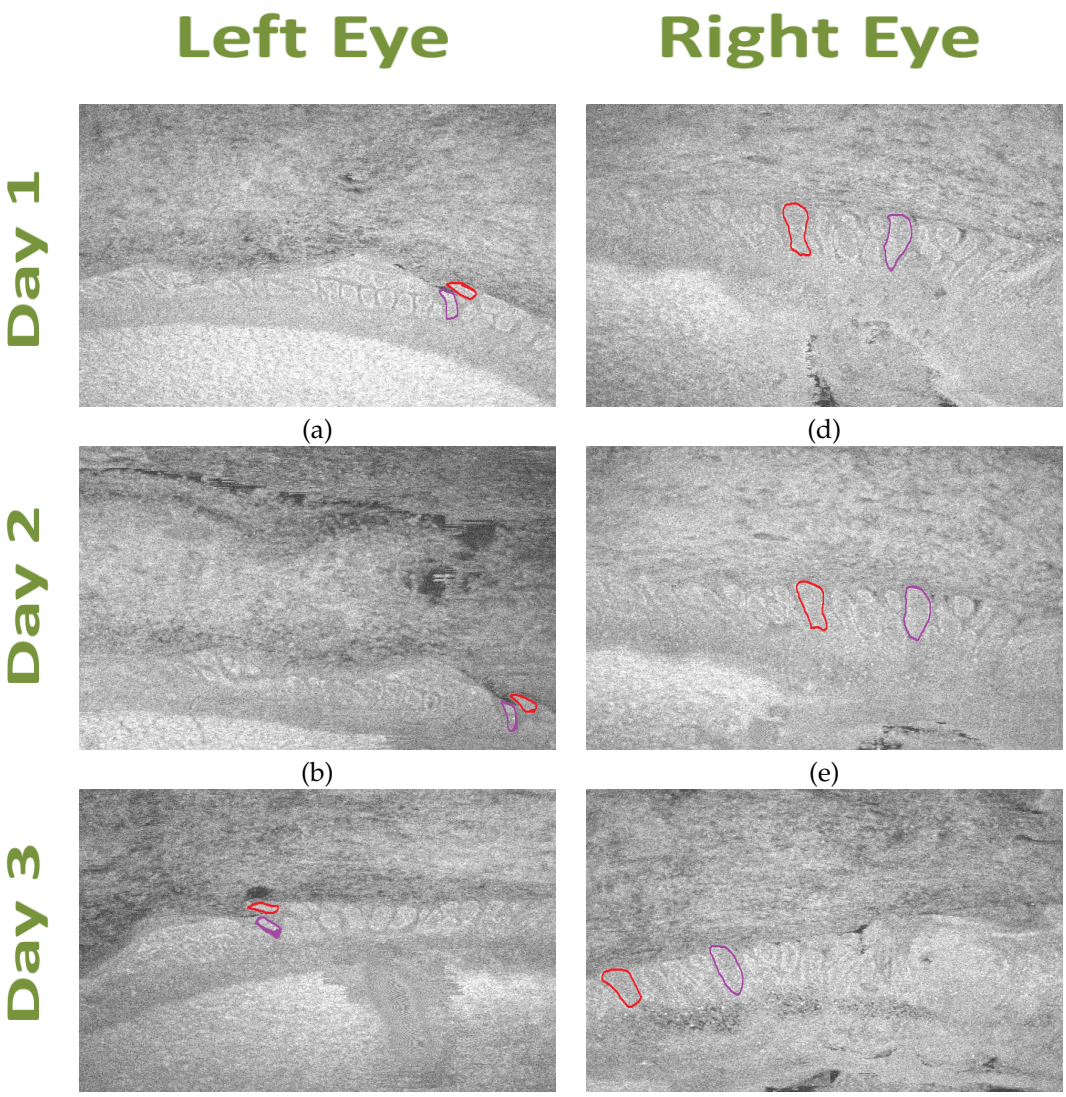


Figure 5.29: In this figure, cross-sections of the POV are shown for the 12 o'clock hour. C-mode cross-sections were extracted from the associated 3D limbal reconstructions for each eye across three days, and were visually compared for similarity. Next, two regions corresponding to the POV structures were manually annotated as shown by the red and purple contours in each image. Finally, the area of each contour was computed, and the contour with the maximum area was compared across each day and for each eye. In general, there is a trend of a decrease in the POV area over each day. We postulate that this may be due to the degradation of the tissue structure over time.

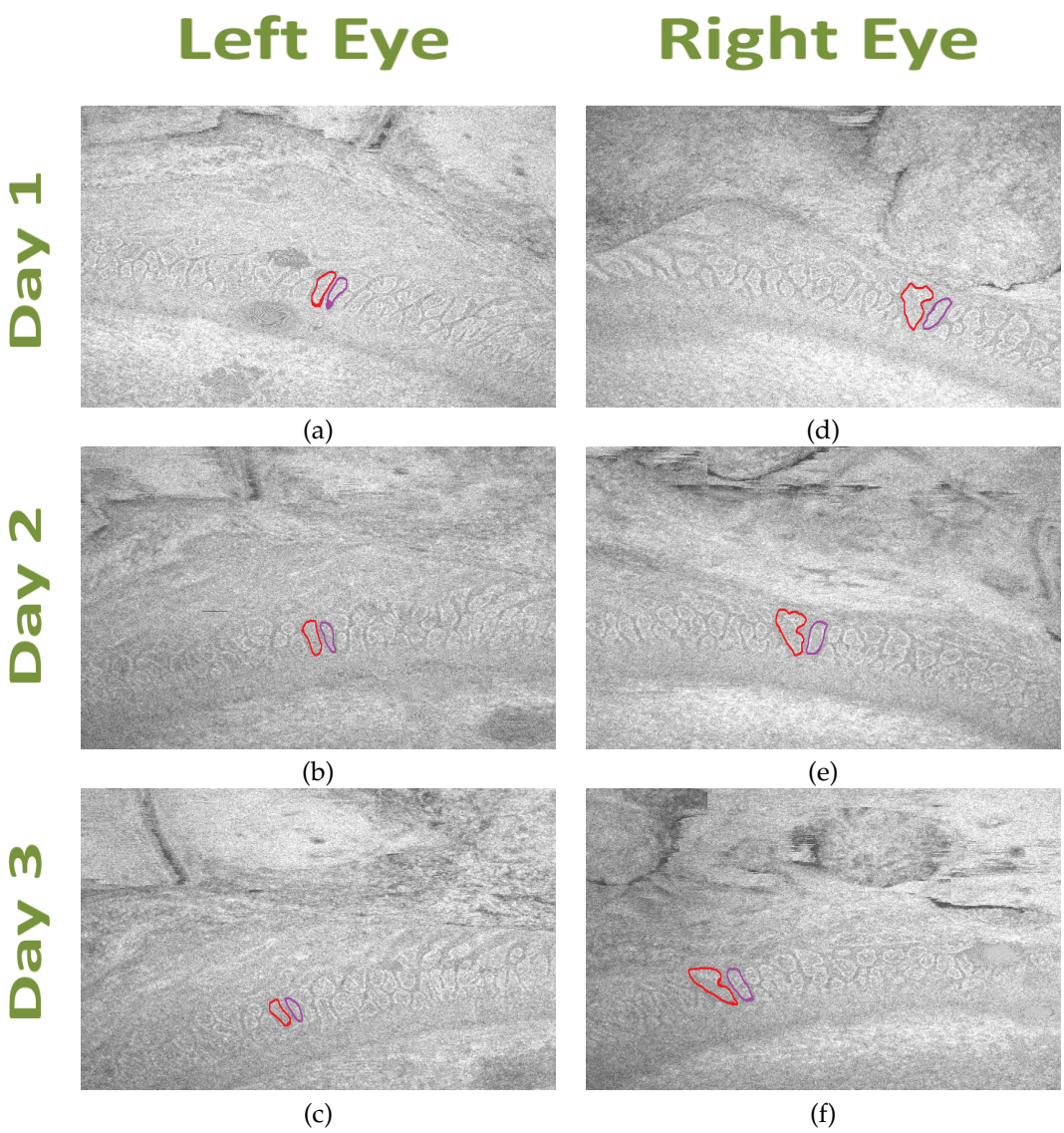


Figure 5.30: In this figure, cross-sections of the POV are shown for the 6 o'clock hour. C-mode cross-sections were extracted from the associated 3D limbal reconstructions for each eye across three days, and were visually compared for similarity. Next, two regions corresponding to the POV structures were manually annotated as shown by the red and purple contours in each image. Finally, the area of each contour was computed, and the contour with the maximum area was compared across each day and for each eye. In general, there is a trend of a decrease in the POV area over each day. We postulate that this may be due to the degradation of the tissue structure over time.

and 5.30. Furthermore, the scleral and limbal structures around the POV regions appear to degrade as well. Based on our empirical observations, we believe that this structural degradation will become worse past the point of the third day.

The cross-sections presented in this thesis, along with the visualization of the structural changes and quantification of certain annotated POV regions, are more representative of geometric changes in individual palisades as opposed to previous work [37]. In prior work, the Palisade Density (PD) and the Epithelial Rete Peg (ERP) [37] were calculated from a small region in a averaged B-scan obtained after OCT imaging of the limbal region. This B-scan is similar to the cross-sectional B-scan views presented in Figs. 5.8 and 5.9(a). However in this work, we provide longitudinal cross-sections (en-face) of the POV regions, which are enabled by our automated methods, and yield a better characterization of the length and extent of the POV at the corneo-scleral intersection. These reconstructed en-face views could enable new clinical practices for eye transplantation, including new donor-eye screening procedures, new time lines for presumed tissue viability, new monitoring of viability after transplant, and new science for future drugs to extend longevity of the corneal or limbal transplant before and after transplantation.

Table 5.5: Area of specific regions of the Palisades of Vogt (POV) that were identified and tracked in cross-sectional images across three days. Area measures are in pixels.

Day #	11 o'clock		12 o'clock		1 o'clock		5 o'clock		6 o'clock		7 o'clock	
	OS	OD	OS	OD	OS	OD	OS	OD	OS	OD	OS	OD
1	1044	2094	861	2764	1445	2767	889	2125	1291	2569	842	1410
2	1001	1707	788	2755	1408	2687	871	2010	1200	2557	826	1259
3	949	1170	707	2499	1392	2300	844	1922	900	2541	780	1203

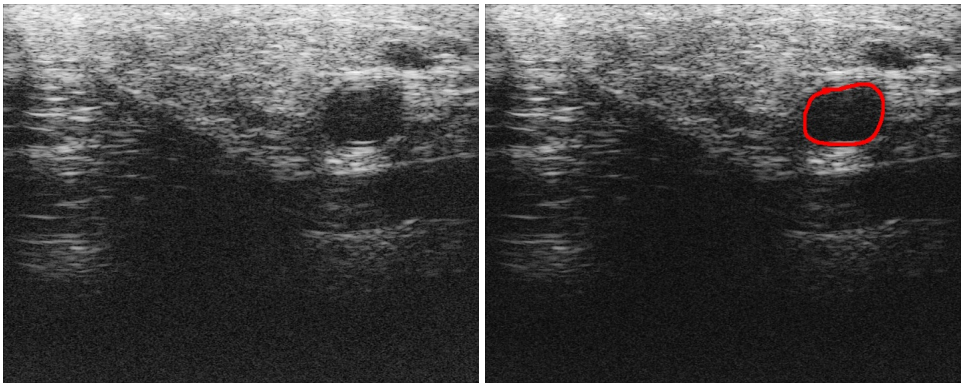
Chapter 6

Vessel Contour Segmentation in UHFUS and HFUS Sequences

6.1 Classical Approach

6.1.1 Problem Statement

Given an image I^t at time t in a UHFUS/HFUS image sequence and the initial vessel position in the sequence $s^{t=0}$ at time $t = 0$, the problem can be formulated as finding the set of pixels $\mathcal{D}$ that lie on the boundary of the vessel. As shown in Fig. 6.1, every pixel in $\mathcal{D}$ takes a specific label $L = 1$ if it lies on the boundary, or it takes a label $L = 0$ if it does not lie on the boundary. Furthermore, the vessel location s^t should also be tracked as the vessel undergoes deformation from one image to another in the image sequence.



(a) (b)

Figure 6.1: (a) shows the 30^{th} B-scan in an UHFUS image sequence, and the expertly annotated vessel boundary is shown in (b).

6.1.2 Data Acquisition

The Visualsonics Vevo 2100 UHFUS machine (Fujifilm, Canada) and a 50 MHz transducer (bandwidth extendable to 70 MHz) was used to acquire freehand ultrasound volumes. This UHFUS system has a physical resolution of 30µm, and the pixel pitch is 11.6µm between pixel centers. Validation was performed on 35 deidentified UHFUS sequences. The sequences were acquired over a wide range of gain values (40-70 dB), with the maximum gain value setting being 70 dB. The sequences contained a wide range of motions with the probe, such as longitudinal scanning, out-of-plane tissue deformation, beating vessel visualization, etc. Fig. 6.2(a) shows an example ultrasound image of the proper palmar digital artery acquired with the UHFUS system. Each sequence consisted of 100 2D B-scans with dimensions of 832×512 pixels.

To show the generality of our approach, 5 additional sequences were acquired from a traditional HFUS machine (Diasus, Dynamic Imaging, UK) using a 10-22 MHz transducer. The pixel resolution for the HFUS machine was 92.5µm, and each sequence consisted of 250 2D B-scans of dimensions 280×534 pixels.

6.1.3 Noise Reduction and Clustering

Noise Reduction

In contrast to traditional HFUS, speckle noise is greater in UHFUS as seen in Figs. 6.2. To mitigate the effects of speckle during segmentation, the UHFUS B-scans were first downsampled [151] by a factor of 4 in each dimension as shown in Fig. 6.3(a). Next, a bilateral filter [129] of size 5×5 pixels was applied to the downsampled image to smooth the small amplitude noise as shown in Fig. 6.3(b), while preserving vessel boundaries that are crucial to our segmentation. The bilateral filtered image is represented by I_B.

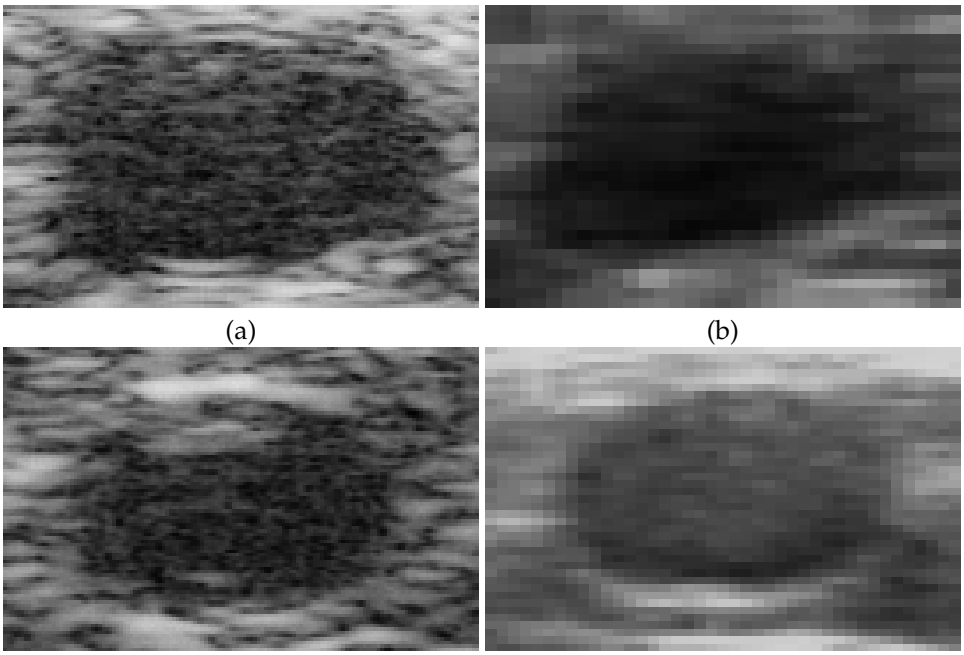
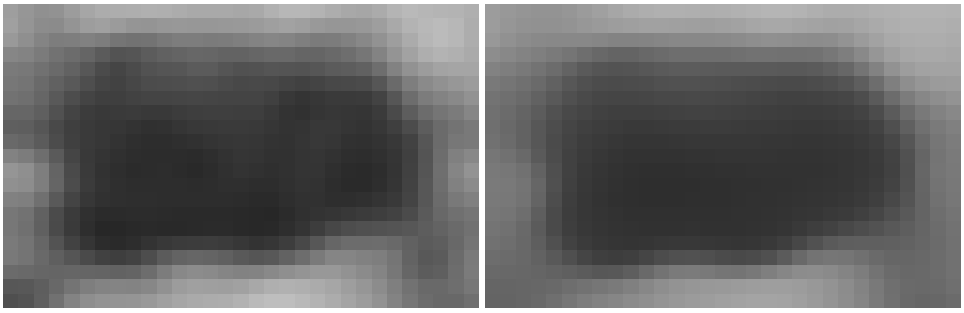


Figure 6.2: Left columns show vessels imaged by UHFUS, while the right columns show vessels imaged by HFUS.

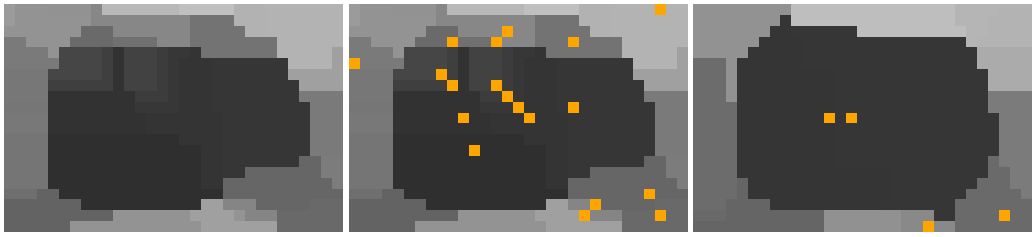


(a) (b)

Figure 6.3: (a) Downsampling the original ultrasound image, shown in Fig. 6.2(a), by a factor of 4 in each dimension. (b) Filtering the downsampled image with a bilateral filter of size 5×5 pixels.

Clustering

The approach published in [152], which has also shown applicability to MRI images, was used to produce an image I_C, where the pixels in I_B were clustered into homogeneous patches. The result of this clustering process is shown in Fig. 6.4. Each pixel in I_C can be represented by two elements: the mean intensity of the patch that it belongs to, and the associated cluster/patch center (root) to which it belongs. For each pixel in I_B, the mean intensity and variance is found in a circular neighborhood, whose size varies depending on the size of the vessel. For small vessels in UHFUS images ($\leq$70 pixel diameter or 0.81mm), the neighborhood size was 3×3 pixels, while it was 7×7 pixels for larger vessels (>70 pixels). Each patch root in I_C has the lowest local variance amongst all the members of the same patch [152]. Roots in I_C were used solely as seeds to track vessels over B-scans in an ultrasound sequence. As seen in Figs. 6.4(b) and 6.4(c), increasing the neighborhood size reduces the number of roots that can be tracked, which can cause tracking failure when large motion occurs. This phenomenon influenced the choice of the neighborhood size.



(a) (b) (c)

Figure 6.4: (a) Result of clustering pixels in I_{B} into patches in I_{C} based on the method presented in [152]; (b) With a kernel of size 3×3 pixels, the pixels in I_{B} are clustered into homogeneous patches in I_{C}, each with its own cluster center (root) (orange points); (c) I_{C} generated with a 7×7 pixels kernel. Notice the lower number of cluster centers as the size of the clustering neighborhood is increased.

6.1.4 Local Phase Analysis

Vessel boundaries in I_{B} were highlighted using a Cauchy filter, which has been shown to be better than a Log-Gabor filter at detecting edges in ultrasound [153]. We denote the spatial intensity value at a location $\mathbf{x}=[x \;\; y]^T$ in the image I_{B} by $I_{\mathrm{B}}(\mathbf{x})$. After applying a 2D Fourier transform, the corresponding 2D frequency domain value is $F(\mathbf{w})$, where $\mathbf{w} = [w_1 \;\; w_2]^T$. The Cauchy filter $C(\mathbf{w})$ applied to $F(\mathbf{w})$ is represented as:

$$C(\mathbf{w}) = \|\mathbf{w}\|_2^u \exp\left(-w_o \|\mathbf{w}\|_2\right), \qquad u \geq 1 \tag{6.1}$$

where u is a scaling parameter, and w_o is the center frequency. We chose the same optimal parameter values suggested in [153]: w_o=10, and u=1. Filtering $F(\mathbf{w})$ with $C(\mathbf{w})$ yielded the monogenic signal [130,131], from which the feature asymmetry map (I_{FA}) [130,131,153] was obtained (see Fig. 6.5). Pixel values in I_{FA} range between [0, 1]. The reader is referred to [154] for an excellent summary of the well-known mathematical theory behind local phase analysis.

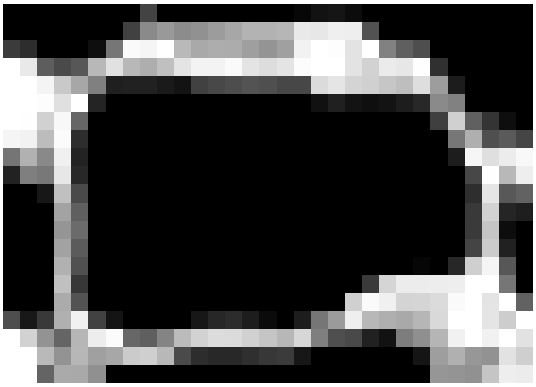


Figure 6.5: Feature Asymmetry map (I_{FA}) for image in Fig. 6.3(b).

6.1.5 Vessel Segmentation and Tracking

Initialization

As in [91,92], we manually initialize our system by clicking a point inside the vessel lumen in the first B-scan of a sequence. This pixel location is stored as a seed, denoted by $\mathbf{s}^0$ at time t=0, to segment the vessel boundary in the first B-scan, and initialize the vessel lumen tracking in subsequent B-scans.

Initial Boundary Segmentation

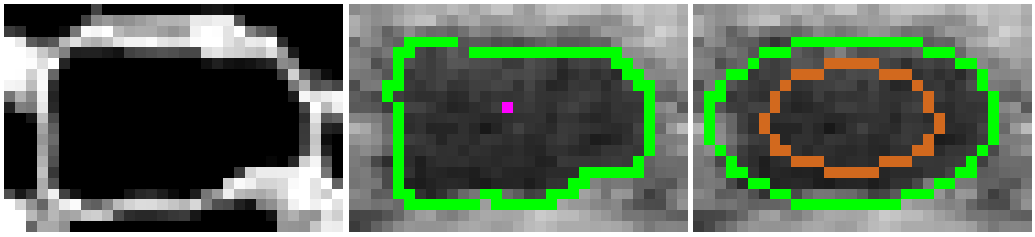
N = 360 radial lines of maximum search length M = 100, which corresponds to the largest observed vessel diameter, stem out from the seed location $\mathbf{s}^0$ to find the vessel boundaries in I_{FA}. The first local maximum on each radial line is included in a set $\mathcal{I}$ as an initial boundary point as shown in Fig. 6.6(b).

Segmentation Refinement

A rough estimate of the semi-major and semi-minor vessel axes was determined by fitting an ellipse [155] to the initial boundary locations in $\mathcal{I}$. Next, the estimated values were

shrunk by 75%, and used to initialize an elliptical binary level set function (LSF) ϕ_o defined by Eq. (6.2), as seen in Fig. 6.6(c), in a narrowband distance regularized level set evolution (DRLSE) [156] framework.

$$\phi_o(x) = \begin{cases} -c_o, & \text{if } x \geq R_o \\ c_o, & \text{otherwise} \end{cases} \tag{6.2}$$



(a) (b) (c)

Figure 6.6: (a) Feature Asymmetry map (I_{FA}); (b) Initial boundary locations (green points) estimated from I_{FA} using the tracked point $\mathbf{s}^t$ (magenta). (c) Ellipse (green) fitted to green points in (i), and then shrunk by 75% (brown ellipse) to initialize the level set evolution.

As the LSF initialization is close to the true boundaries, the DRLSE formulation allows quick propagation of LSF to the desired vessel locations $\mathcal{D}$ (see Fig. 6.7) with a large timestep $\Delta\tau$ [156]. For a LSF $\phi : \Omega \rightarrow \mathcal{R}$, defined on a domain Ω, the DRLSE framework minimizes an energy functional $\mathcal{E}(\phi)$ [156] written in Eq. (6.3):

$$\mathcal{E}(\phi) = \mu\mathcal{R}_p(\phi) + \mathcal{E}_{ext}(\phi) \tag{6.3}$$

$\mathcal{R}_p(\phi)$ is the level set regularization term [156] defined below:

$$\mathcal{R}_p(\phi) = \int_{\Omega} p(|\nabla\phi|)dx \tag{6.4}$$

where p is the double-well potential function:

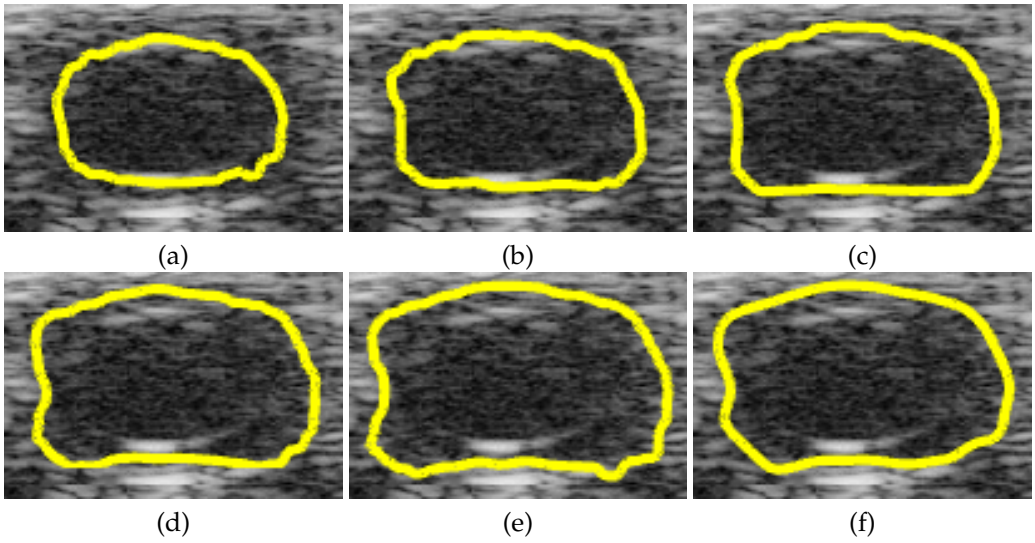


Figure 6.7: (a) LSF evolution (closed yellow contour) at 5^{th} iteration; (b) LSF evolution at 10^{th} iteration; (c) LSF evolution at 15^{th} iteration; (d) LSF evolution at 20^{th} iteration; (e) LSF evolution at 25^{th} iteration; (f) Refinement of LSF at the end of the 25^{th} iteration for another 10 iterations yields the final LSF position.

$$p_2(s) = \begin{cases} \frac{1}{(2\pi^2)}(1 - cos(2\pi s)), & \text{if } s \leq 1 \\ \frac{1}{2}(s-1)^2, & \text{if } s \geq 1 \end{cases} \tag{6.5}$$

$\mathcal{E}_{ext}(\phi)$ is the external energy term for the level set that depends on the external data, such as the image to be segmented. The $\mathcal{E}_{ext}(\phi)$ term can be written as the sum of an area term and an edge term:

$$\mathcal{E}_{ext}(\phi) = \lambda \int_{\Omega} g\delta_{\epsilon}(\phi)|\nabla\phi|dx \quad + \quad \alpha \int_{\Omega} gH_{\epsilon}(-\phi)dx \tag{6.6}$$

μ, λ, ϵ, and α are constants. g is an edge indicator function, which is defined as follows:

$$g = \frac{1}{1 + |\nabla G_{\sigma} * I|} \tag{6.7}$$

G_σ is a Gaussian kernel with a standard deviation σ, and it is used to blur the image I to reduce the noise. This function g has smaller values at the boundaries of the object, and higher values everywhere else. H_ϵ and δ_ϵ, defined in Eqs. (6.8) and (6.9), are the Heaviside function and first order derivative of the Heaviside function respectively:

$$H_\epsilon(x) = \begin{cases} \frac{1}{2}\Big(1 + \frac{x}{\epsilon} + \frac{1}{\pi} sin(\frac{\pi x}{\epsilon})\Big), & |x| \leq \epsilon \\ 1, & x > \epsilon \\ 0, & x < -\epsilon \end{cases} \tag{6.8}$$

$$\delta_\epsilon(x) = \begin{cases} \frac{1}{2\epsilon}\Big[1 + cos(\frac{\pi x}{\epsilon})\Big], & |x| \leq \epsilon \\ 0, & |x| > \epsilon \end{cases} \tag{6.9}$$

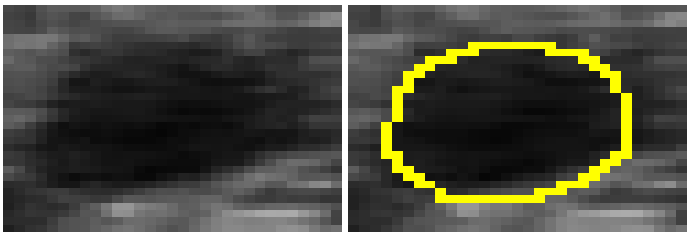
The energy function in Eq. (6.3) can now be minimized using the gradient in Eq. (6.10):

$$\frac{\partial \phi}{\partial \tau} = \mu \text{div}(d_p(|\nabla \phi|)\nabla \phi) \; + \; \lambda \delta_\epsilon(\phi) \text{div}\left(g \frac{\nabla \phi}{|\nabla \phi|}\right) \; + \; \alpha g \delta_\epsilon(\phi) \tag{6.10}$$

where d_p is derived from the first derivative $p_2'(s)$, written in Eq. (6.12), of the double-well potential $p_2(s)$ in Eq. (6.5). d_p is written below as:

$$d_p(s) \triangleq \frac{p_2'(s)}{s} \tag{6.11}$$

$$p_2'(s) = \begin{cases} \frac{1}{(2\pi^2)}(1 - cos(2\pi s)), & \text{if } s \leq 1 \\ \frac{1}{2}(s-1)^2, & \text{if } s \geq 1 \end{cases} \tag{6.12}$$



(a) (a)

Figure 6.8: Segmentation result in the HFUS image.

Vessel Tracking

To update the vessel lumen position $\mathbf{s}^t$ at time t to $\mathbf{s}^{t+1}$ at time $t+1$, two new potential seeds are found, from which one is chosen. The first seed is found using an EKF [93, 157], and the second seed, which is found using I_C, is needed in case the EKF fails to track the vessel lumen due to abrupt motion. The EKF tracks a state vector defined by: $\mathbf{x}^t = [c_x^t, c_y^t, a^t, b^t]$, where $\mathbf{s}_{\text{ekf}}^t$=$[c_x^t, c_y^t]$ is the EKF-tracked vessel lumen location and $[a^t, b^t]$ are the tracked semi-major and semi-minor vessel axes respectively. Instead of tracking all locations in $\mathcal{D}$, it is computationally efficient to track $\mathbf{x}^t$, whose elements are estimated by fitting an ellipse once again to the locations in $\mathcal{D}$ as shown in Fig. 6.9(a).

The EKF projects the current state $\mathbf{x}^t$ at time t to the next state $\mathbf{x}^{t+1}$ at time t+1 using the motion model in [93], which uses two state transition matrices A_1, A_2, the covariance error matrix P, and the process-noise covariance matrix Q. These matrices are initialized in Eqs. (6.13)-(6.16). Consistent with [93], the velocity of $\mathbf{x}^{t+1}$ was damped by 0.5 to avoid abrupt motion as the vessel can suddenly deform.

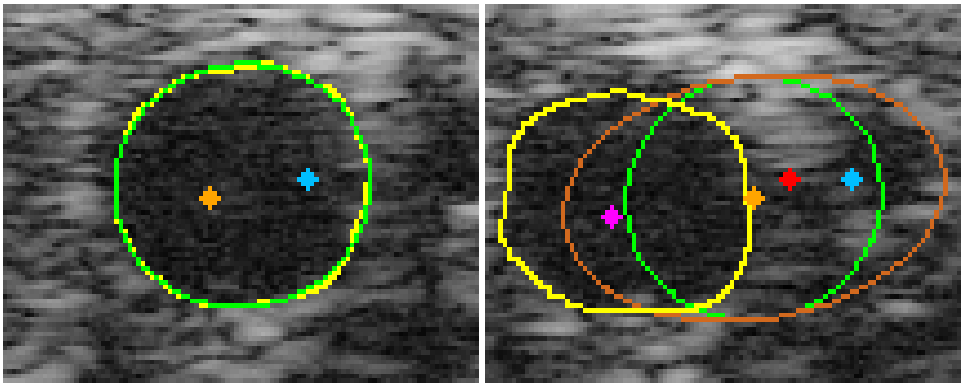
$$A_1 = diag([1.5, 1.5, 1.5, 1.5]) \tag{6.13}$$

$$A_2 = diag([-0.5, -0.5, -0.5, -0.5]) \tag{6.14}$$

$$P = diag([1000, 1000, 1000, 1000]) \tag{6.15}$$

$$Q = diag([0.001, 0.001, 0.001, 0.001]) \tag{6.16}$$

The second seed was found using the clustering result. At the location of the previous seed $\mathbf{s}^t$ in the clustering result I_C^{t+1} at time t+1, the EKF tracked ellipse axes $[a^{t+1}, b^{t+1}]$ were used to find the neighboring roots of $\mathbf{s}^t$ in an elliptical region of size $[1.5a^{t+1}, b^{t+1}]$ pixels. Amongst these roots, the root $\mathbf{s}_c^{t+1}$, which has the lowest mean pixel intensity representing a patch in the vessel lumen, is chosen as shown in Fig. 6.9(b). The elliptical region is robust to vessel compression, which enlarges the vessel horizontally.



(a) (b)

Figure 6.9: Tracking under large motion - (a) In frame 87, $\mathbf{s}_{ekf}^{t+1}$ (blue point) chosen over $\mathbf{s}_c^{t+1}$ (orange point) to segment vessel (yellow contour), which is then fitted with an ellipse (green); (b) In frame 88, EKF prediction (red) from its previous location (blue) is ignored as Eq. (6.17) is not satisfied. Instead, $\mathbf{s}_c^{t+1}$ (magenta point) is chosen as it falls under the elliptical neighborhood (brown) of $\mathbf{s}_c^t$ (orange).

The EKF prediction is sufficient for tracking during slow longitudinal scanning or still imaging as $\mathbf{s}_{\text{ekf}}^{t+1}$ and $\mathbf{s}_{\text{c}}^{t+1}$ lie close to each other. However, when large motion was encountered, the EKF incorrectly predicted the vessel location (see Fig 6.9(b)) as it corrected abrupt vessel motion, thereby leading to tracking failure. To mitigate tracking failure during large vessel motion, $\mathbf{s}_{\text{ekf}}^{t+1}$ was ignored, and $\mathbf{s}_{\text{c}}^{t+1}$ was updated as the new tracking seed according to the proposed rule in Eq. (6.17):

$$\mathbf{s}^{t+1} = \begin{cases} \mathbf{s}_{\text{c}}^{t+1} & if \quad \|\mathbf{s}_{\text{ekf}}^{t+1} - \mathbf{s}_{\text{c}}^{t+1}\|_2 > a^{t+1} \\ \mathbf{s}_{\text{ekf}}^{t+1} & otherwise \end{cases} \tag{6.17}$$

6.1.6 Results

Metrics

Segmentation accuracy of the proposed approach was evaluated by comparing the contour segmentations against the annotations of two graders. All images in all datasets were annotated by two graders. Tracking was deemed successful if the vessel was segmented in all B-scans of a sequence. Considering the set of ground truth contour points as G and the segmented contour points as S, the following metrics were calculated: 1) *Dice Similarity Coefficient* (DSC) [158] in Eq. (6.18), 2) *Hausdorff Distance* (HD) [141] in millimeters in Eq. (6.19), 3) *Definite False Positive Distance* (DFPD), and *Definite False Negative Distance* (DFND) in Eqs. (6.20) and (6.21). The latter represent weighted distances of false positives and negatives to the true annotation.

Let I_G and I_S be binary images containing 1 on and inside the area covered by G and S respectively, and 0 elsewhere. The Euclidean Distance Transform (EDT) is computed for I_G

and its inverse I_G^{Inv} [159]. DFPD and DFND are estimated from the element-wise product of I_S with EDT(I_G) and EDT(I_G^{Inv}) respectively ad defined in Eqs. (6.20) and (6.21). $d(i, G, S)$ is the distance from contour point i in G to the closest point in S. Inter-grader annotation variability was also measured. To compare against the results in [88] for HFUS sequences, the Mean Absolute Deviation (MAD) error defined in Eq. (6.22) was also computed.

$$\text{DSC} = \frac{2|G \cap S|}{|G| + |S|} \tag{6.18}$$

$$\text{HD} = max\Big(\max_{i \in [1,|G|]} d(i, G, S), \max_{i \in [1,|S|]} d(j, S, G)\Big) \tag{6.19}$$

$$\text{DFPD} = log\Big(\|\text{EDT}(I_G) \circ I_S\|_1\Big) \tag{6.20}$$

$$\text{DFND} = log\Big(\|\text{EDT}(I_G^{\text{Inv}}) \circ I_S\|_1\Big) \tag{6.21}$$

$$\text{MAD} = \frac{1}{2}\Big(\frac{1}{|G|}\sum_{i=1}^{|G|} d(i, G, S) + \frac{1}{|S|}\sum_{i=1}^{|S|} d(i, S, G)\Big) \tag{6.22}$$

UHFUS Results

We ran our algorithm on 35 UHFUS sequences (100 images each), and the corresponding results are shown in Figs. 6.10. The two graders varied in their estimation of the vessel boundary locations in UHFUS images due to the speckle noise obscuring the precise location of the vessel edges, as shown in the inter-grader Dice score in Fig. 6.10(a), inter-grader Hausdorff distance in Fig. 6.10(b), and inter-grader variation in Figs. 6.10(c) and 6.10(d). Grader 2 tended to under-segment the vessel (G1vG2, low DFPD and high DFND scores), while grader 1 tended to over-segment (G2vG1, high DFPD and low DFND scores). As desired, our segmentation tended to be within the region of uncertainty between the two graders (see Figs. 6.10(c) and 6.10(d)). Accordingly, the mean Dice score and mean Haus-

dorff distance of our algorithm against grader 1 (0.917±0.019, 0.097±0.019mm) and grader 2 (0.905±0.018, 0.091±0.019mm) were better than the inter-grader scores of (0.892±0.019, 0.105±0.02mm). *The largest observed Hausdorff distance error of 0.135mm is 6 times smaller than the smallest observed vessel diameter of 0.81mm. Similarly, the mean Hausdorff distance error of 0.094±0.019mm is ∼7 times smaller than smallest observed vessel diameter. This satisfies our goal of sub-mm vessel contour localization.* Tracking was successful as the vessel contours in all images in all sequences were segmented.

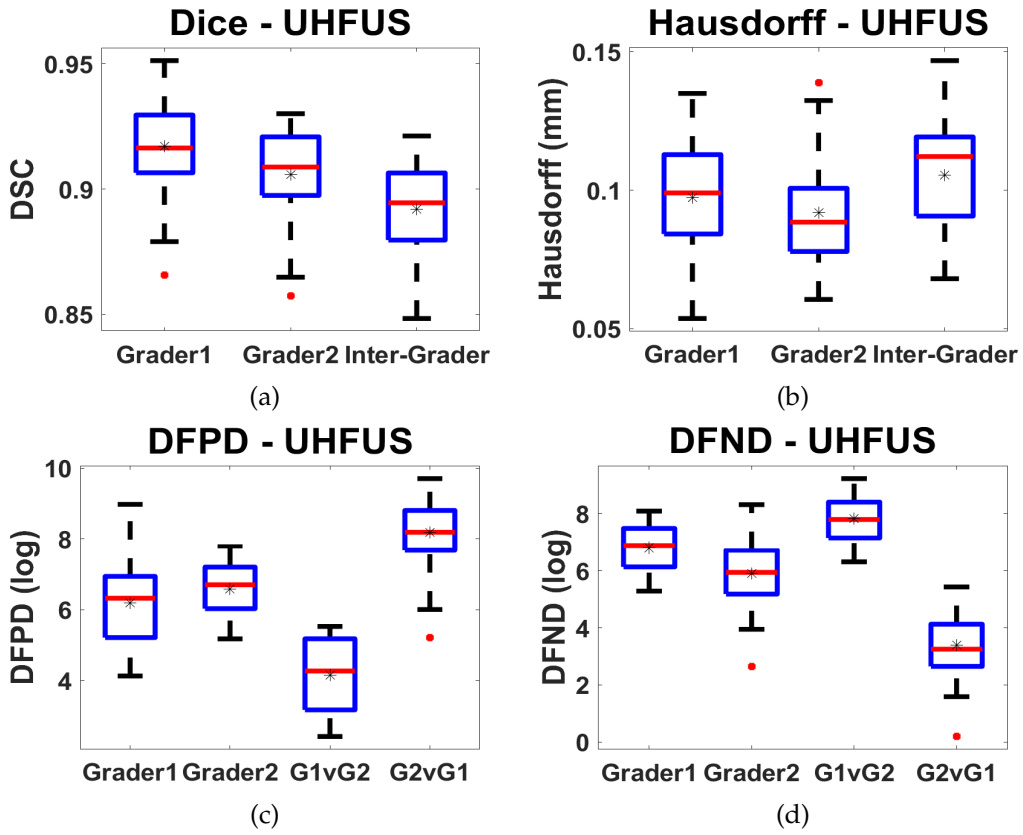


Figure 6.10: Quantitative segmentation and tracking accuracy metrics for 35 UHFUS sequences. The black * in each box plot represents the mean value of the metric. The terms 'G1vG2' and 'G2vG1' in Figs. 6.10(c) and 6.10(d) represent the inter-grader annotation variability when grader 1 annotation was considered the ground truth, and vice versa.

HFUS Results

To show the generality of our approach to HFUS, we ran our algorithm on 5 HFUS sequences (250 images each), and the corresponding results are shown in Fig. 6.8(b) and Figs. 6.11. As opposed to UHFUS, lower DFPD and DFND scores were seen with HFUS, meaning a greater consensus in annotation amongst the graders (see Figs. 6.11(c) and 6.11(d)). Notably, our algorithm still demonstrated the desirable property of final segmentations that lay in the uncertain region of annotation between the two graders. This is supported by comparing the mean Dice score and mean Hausdorff distance of our algorithm against grader 1 (0.915$\pm$0.008, 0.292$\pm$0.023mm) and grader 2 (0.912$\pm$0.021, 0.281$\pm$0.065mm), with the inter-grader scores (0.915$\pm$0.02, 0.273$\pm$0.04mm).

To compare against the 0.1mm Mean Absolute Deviation (MAD) error in [88], we also computed the MAD error for HFUS sequences (not shown in Fig. 6.11). The MAD error of our algorithm against grader 1 was 0.059$\pm$0.021mm, 0.057$\pm$0.024mm against grader 2, and 0.011$\pm$0.003mm between the graders. *Despite the lower pixel resolution* ($92.5\mu m$) *of the HFUS machine used in this work, our MAD errors were* $\sim 2\times$ *lower than the 0.1mm MAD error in [88].*

Performance

The average run-time of our algorithm on a desktop with an Intel i7 processor, 16GB RAM, and an entry-level NVIDIA GeForce GTX 760 GPU was 19.15 millisecond per B-scan and 1.915 seconds per sequence, thus achieving a potential real-time frame rate of 52 frames per second. The proposed approach is significantly faster than the real-time approaches proposed in [91–93]. Efficient utilization of the GPU unified memory was made to implement the DRLSE and EKF algorithms in parallel.

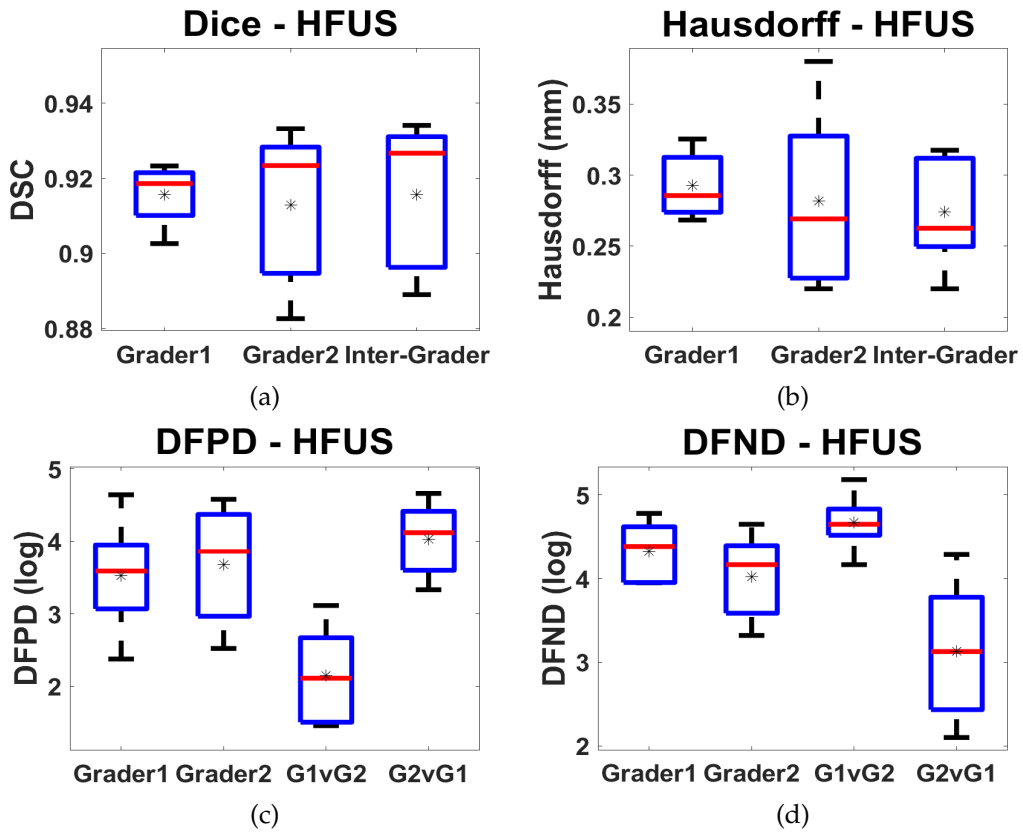


Figure 6.11: Quantitative segmentation and tracking accuracy metrics for 5 HFUS sequences. The black * in each box plot represents the mean value of the metric. The terms 'G1vG2' and 'G2vG1' in Figs. 6.11(c) and 6.11(d) represent the inter-grader annotation variability when grader 1 annotation was considered the ground truth, and vice versa.

Parameter Selection and Optimization

For all UHFUS datasets, the parameters used in all datasets were: $\Delta\tau = 10, \mu = 0.2, \lambda = 1, \alpha = -1, \epsilon = 1$ for a total of 15 iterations. Only minor changes in the parameters of the algorithm were required to transfer the methodology to HFUS sequences; namely, the bilateral filter size was 3×3 pixels, w_o=5, and $\Delta\tau$=8. No other changes were made to the level set parameters.

6.2 Deep Learning Approach

6.2.1 Problem Statement

Given an UHFUS/HFUS image $\mathcal{I}$, the task is to find a function $\mathcal{F} : \mathcal{I} \rightarrow \mathcal{L}$ that maps every pixel in $\mathcal{I}$ to a label $\mathcal{L} \in \{0, 1\}$. We would like to divide the ultrasound image into segmented regions which are: (1) Foreground vessel (filled) contour, and (0) being the background.

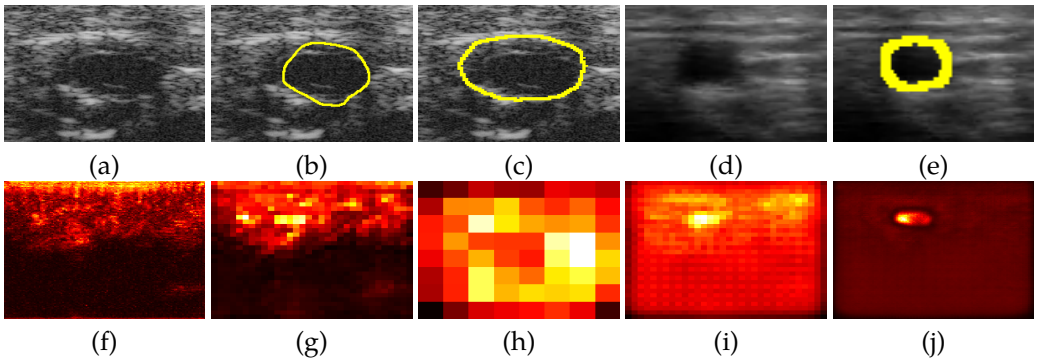


Figure 6.12: (a) Still frame capturing a pulsating vessel acquired using UHFUS; (b) Segmentation (yellow contour) from a level set method bleeds into the tissue region due to poor boundary contrast; (c) Final segmentation from the proposed USVS-Net; (d) Frame acquired using HFUS (zoomed), and (e) its associated final vessel segmentation; Activations of the network for the vessel imaged in (a) at different network depths: (f) downsampling level 1; (g) downsampling level 3; (h) downsampling level 5; (i) upsampling level 3; (j) upsampling level 1.

6.2.2 Data Acquisition

Previously acquired (free-hand) deidentified video sequences from an existing research database [4, 41] were used in this work, and they came from two scanners: a Visualsonics Vevo 2100 UHFUS machine (Fujifilm, Canada), and a Diasus HFUS scanner (Dynamic Imaging, UK). The UHFUS scanner provided a 50 MHz transducer with physical resolution of 30μm and a pixel spacing of 11.6μm. 58 UHFUS sequences were used, each containing 100 2D B-scans with dimensions of 832×512 pixels. The HFUS scanner had a 10-22 MHz

transducer with a pixel spacing of 92.5μm. 26 HFUS sequences were used, each containing a variable number of 2D B-scans (50-250) with dimensions of 280$\times$534 pixels. All the sequences contained arteries of the hand (eg. superficial palmar arch) with a wide range of adjustable gain settings (40-70 dB). Extensive probe motions were also acquired, such as longitudinal scanning, beating vessels, out-of-plane vessel deformation etc. An expert grader annotated all the 84 UHFUS and HFUS sequences.

6.2.3 Ultrasound Vessel Segmentation Neural Network (USVS-Net)

In this work, we propose to utilize a Convolutional Long Short Term Memory (LSTM) model to segment vessel cross-sections in UHFUS and HFUS video sequences. LSTM networks [160–165] intelligently combine multi-scale features by retaining relevant features over video time steps, and only update the features when required. Some of these approaches have shown applicability to different tasks [160, 161]; we tested the performance of these popular ConvLSTM methods on the challenging task of segmenting highly deformable and pulsating vessel contours in UHFUS and HFUS sequences. But, these approaches did not accurately segment vessel cross-sections, which inspired the development of our novel ConvLSTM-based architecture for vessel segmentation called the USVS-Net. This network was influenced by methods designed for different anatomies (retina [79], cornea [5], microscopy [160], X-Ray [161]). Validation of our method was conducted on 38 UHFUS and 6 HFUS sequences respectively. We also show the general applicability of our method to the Montgomery County Chest X-Ray dataset [166] with comparable results to the state-of-the-art.

As seen in Fig. 6.13, the proposed USVS-Net [167] is a Convolutional LSTM-based vessel segmentation architecture comprised of encoder and decoder sections. From prior work

[160–165], the memory mechanism in such network designs enables the retention of vessel appearance over multiple scales for dense pixel-wise output predictions.

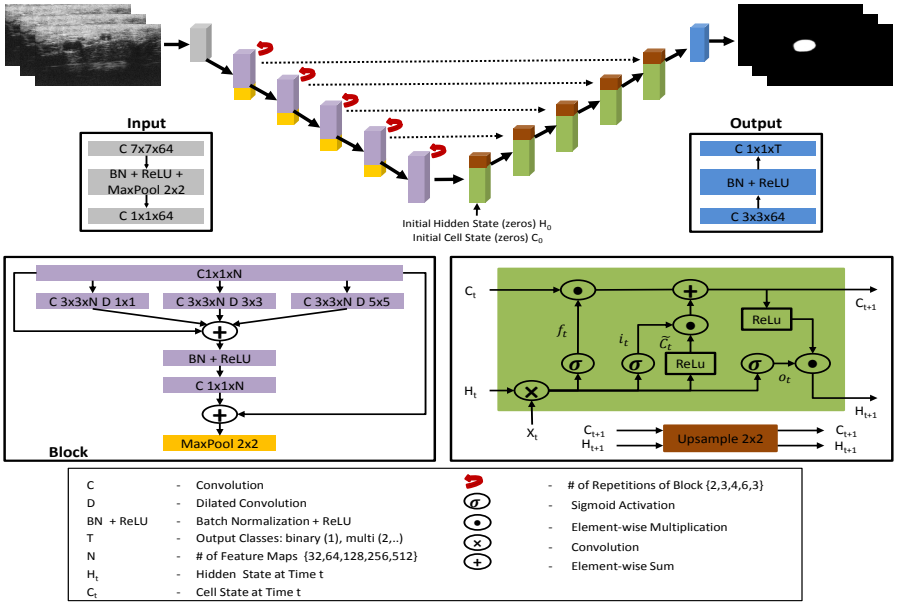


Figure 6.13: The USVS-Net architecture contains encoding (purple) and decoding (green) sections. The encoder uses of residual connections and dilated convolutions to extract features, while the decoder uses structured ConvLSTM blocks to retain vessel shape attributes and segment the vessel.

Encoder

The encoder structure is inspired by the approaches in [5,79], which have shown applicability to retina and cornea tissue interface segmentation. The blocks in the encoder section pull out meaningful representations of the vessel appearance over multiple scales using dilated convolutions [110] and residual connections [114]. These blocks can discern faintly defined boundaries better [5], thereby avoiding holes in the final segmentation. As shown in Fig. 6.12, the feature maps characterized at the first few layers of the encoder depict finely defined properties (edges, corners etc.), which are low-level attributes, and are limited due to their smaller receptive field. At the deeper layers of the network, coarse, but complex

attributes are seen with poorly defined contours. At this level, more of the image is seen on a global scale due to the larger receptive field. Yet, this hierarchical representation is not enough on its own to model the dynamics of vessel movement in a video sequence. By forwarding the feature maps extracted at different scales to the convolutional LSTM cells, which can retain relevant features of interest in memory, they can be integrated to produce segmentations of better quality and precision [160, 161].

Decoder

Every encoder block forwards its output feature maps to a LSTM unit in the decoder section. In this work, we incorporate the structured LSTM proposed in [162]. These LSTM cells consider the output of each encoder block X_t as a single time step, and implement a memory mechanism wherein the features extracted at multiple scales are integrated in a coarse-to-fine manner. This is done by gating structures that regulate the removal or addition of new information to the cell state C_t. In this manner, global contextual information from the deepest encoder layer is observed by the LSTM unit first, and as the receptive fields are reduced, finer details about the vessel contour are added.

From Fig. 6.13, each LSTM unit uses three feature sets (input, hidden, and cell), and outputs information using three gates: forget, input, and output. As seen in Eq. (6.23), the forget gate f_t removes information from the cell state. The current input X_t to the cell and the hidden state H_t is convolved and passed through a sigmoid function, which outputs a value in the range [0,1] for every element in the cell state C_t. 1 represents that a cell state value is kept entirely, while 0 represents that it is forgotten. W_f and b_f represent the weights and biases for the forget gate.

$$f_t = \sigma\big(W_f \cdot [X_t, H_t] + b_f\big) \tag{6.23}$$

The input gate determines the new information that will be incorporated into the cell state. In this step as written in Eq. (6.24), X_t and H_t are convolved and passed through a sigmoid function to determine the values to be updated. Then, a set of potential values that could be added to the cell state are obtained by passing the convolved output through a ReLu activation function as shown in Eq. (6.25). The next step is to update the current state C_t of the LSTM cell as in Eq. (6.26) to incorporate the new inputs, and convert it to a new cell state C_{t+1}. This is done by first multiplying the old state C_t with f_t to forget certain specific elements, and then adding it with $i_t \cdot \widetilde{C}_t$ to replace the elements removed with new values of interest.

$$i_t = \sigma\big(W_i \cdot [X_t, H_t] + b_i\big) \tag{6.24}$$

$$\widetilde{C}_t = ReLU\big(W_{\widetilde{C}} \cdot [X_t, H_t] + b_{\widetilde{C}}\big) \tag{6.25}$$

$$C_{t+1} = f_t * C_t + i_t * \widetilde{C}_t \tag{6.26}$$

Finally, the output of the LSTM unit H_{t+1} is regulated by the output gate. It is based on the new cell state C_{t+1} of the LSTM unit. The covolved output of X_t and H_t is first passed through a sigmoid layer to determine the values o_t that need to be output. Then, the new cell state C_{t+1} is passed through a ReLu function, and multiplied with o_t to estimate the final hidden state output H_{t+1} that needs to be obtained from the LSTM module.

$$o_t = \sigma\big(W_o \cdot [X_t, H_t] + b_o\big) \tag{6.27}$$

$$H_{t+1} = o_t * ReLU\big(C_{t+1}\big) \tag{6.28}$$

Contrary to [162], bi-directional LSTMs were not used in this work as our video sequences can be of arbitrary length with non-smooth vessel motion between consecutive frames, c.f. Fig. 6.14, making their implementation impractical. We employed convolution in the structured LSTM unit, and replaced the tanh operation with a ReLU as we empirically observed an improved segmentation accuracy. Similar to [161], the initial hidden and cell states were set to zero, and the hidden and cell states of the other LSTM units were upsampled from the LSTM unit below (see Fig. 6.13).

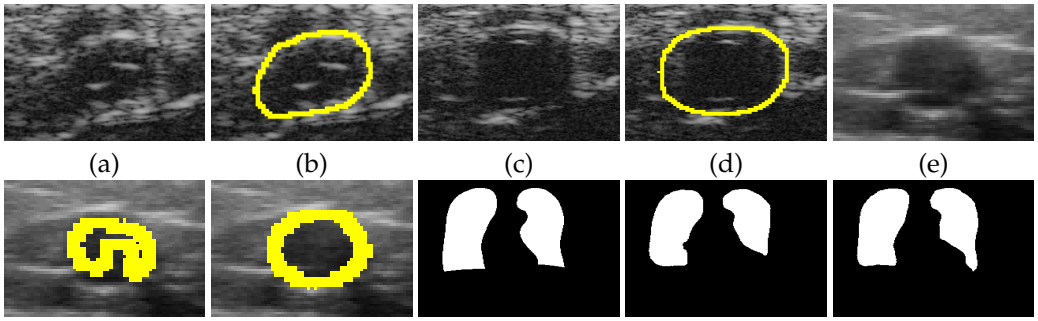


Figure 6.14: (a) Frame 152 in a UHFUS sequence showing a completely contracted vessel, and (b) its associated segmentation; (c) Next frame 153 in the same sequence showing a patent vessel, and (d) its segmentation; (e) Zoomed view of a HFUS B-scan (gain set to maximum); (f) Segmentation by the CFCM34 [161] and (g) our segmentation result; (h) Ground truth lung segmentation from the CXR dataset; (i) Result from CFCM34, and (j) our result (note the improved segmentation due to better contextual information).

6.2.4 Experiments

Setup

Of the 58 UHFUS sequences, 20 were chosen for training and the remaining 38 were used for testing. Similarly, from the 26 HFUS sequences, 20 were chosen for training and the remaining 6 were used for testing. We ran a 3-fold cross-validation for the vessel segmentation task. To simulate a clinical application, an ensemble of the two best models with the lowest validation loss (from a single fold) were used for testing. Similar to [161], we also ran a 3-fold cross validation for the lung segmentation task in the CXR dataset.

Training

Our sequences contained variable image sizes and training a ConvLSTM with full-sized images is limited by GPU RAM. We trained our USVS-Net by scaling each B-scan to 256×256 pixels. Data augmentation (elastic deformation, blurring etc.) was done to increase the training set to ∼120,000 images. To compare against [161], we used the generalized dice coefficient [161] loss with the ADAM optimizer [122], and set the batch size to 16 with a learning rate of 0.00001 for 30 epochs. The final pixel level probabilities were classified using the softmax function, and the connected component in the foreground class was considered the segmentation. For the DecLSTM [160], we used RMSProp optimizer [160], weighted cross-entropy loss [82], and a learning rate of 0.0001 for 30 epochs.

Baseline Comparisons

For the vessel segmentation task, we compared our errors against those from a level set-based method [4], and two LSTM-based segmentation approaches: DecLSTM [160] and CFCM34 [161]. For the lung segmentation task, we compared our results against the CFCM34

model [161].

6.2.5 Results

Metrics

We compared the results of each approach against the expert annotations. The following metrics were calculated to quantify errors: 1) Dice Similarity Coefficient (DSC) [4], 2) Hausdorff Distance (HD) in millimeters [4], 3) Mean Absolute Deviation (MAD) in millimeters [161], 4) Definite False Positive and Negative Distances (DFPD, DFND) [4], 5) Precision (Prec.) and 6) Recall (Rec.) [161].

UHFUS Results

From Table 6.1 (top), the traditional level set approach only succeeded in segmenting vessels in 33 of 38 sequences, while the LSTM-based methods successfully segmented vessels in all sequences. The proposed USVS-Net matched the expert annotations with the highest DSC, and lowest HD and MAD errors among all baselines. We estimated the statistical significance of our results using paired t-tests for every baseline, and determined that our results were statistically significant ($p < 0.05$) for all metrics except DFPD. Our largest HD error of 0.14mm was $\sim 15\times$ lower than the largest observed vessel diameter of 2.17mm. Similarly, the average HD error was $\sim 10\times$ lower than the smallest observed vessel diameter of 1.1mm. Although our method slightly over-segmented the boundaries (outer adventitia) as evidenced by the highest DFPD score, the low clinically relevant measures of HD and MAD were acceptable. Our primary intention for the USVS-Net was to segment vessels in UHFUS sequences, and through our results, we satisfactorily hit our target of sub-mm vessel localization in UHFUS sequences presenting with increased speckle, and large vessel

motion and deformation.

Table 6.1: Segmentation error comparison for the UHFUS (top) and HFUS (bottom) sequences. (* 33/38 sequences successful)

Method	DSC	HD (mm)	MAD (mm)	DFPD	DFND	Prec	Rec
UHFUS Results (Traditional scores exclude failure cases)							
Traditional* [4]	81.13 ± 3.72	0.21 ± 0.05	0.06 ± 0.02	*3.08 ± 1.68*	8.71 ± 0.55	*96.44 ± 2.56*	72.03 ± 4.9
DecLSTM [160]	88.83 ± 3.74	0.15 ± 0.06	0.04 ± 0.03	6.76 ± 1.05	**5.35 ± 1.4**	87.54 ± 4.45	92.46 ± 3.93
CFCM34 [161]	88.45 ± 3.97	0.15 ± 0.07	0.04 ± 0.04	**6.41 ± 1.21**	5.51 ± 1.39	88.07 ± 4.83	91.31 ± 3.87
USVS-Net	**92.15 ± 2.29**	**0.11 ± 0.03**	**0.03 ± 0.01**	6.83 ± 1.13	6.33 ± 1.36	**91.76 ± 3.78**	**93.2 ± 3.34**
HFUS Results							
Traditional [4]	83.6 ± 5.47	0.47 ± 0.13	0.08 ± 0.04	**2.08 ± 2.01**	6.02 ± 0.51	**95.13 ± 4.8**	75.42 ± 7.49
DecLSTM [160]	88.34 ± 5.21	0.39 ± 0.1	0.05 ± 0.3	4.23 ± 0.97	5.61 ± 0.78	87.21 ± 3.15	83.94 ± 7.61
CFCM34 [161]	89.44 ± 3.34	0.36 ± 0.09	0.05 ± 0.02	3.74 ± 1.04	5.23 ± 0.62	94.21 ± 3.48	85.74 ± 5.51
USVS-Net	**89.74 ± 3.05**	**0.36 ± 0.08**	**0.04 ± 0.02**	4.98 ± 0.86	**4.53 ± 1.03**	88.63 ± 0.05	**91.52 ± 0.05**

HFUS Results

As seen in Table 6.1 (bottom), the performance of the CFCM34 and the USVS-Net is comparable. The USVS-Net edges out the CFCM34 with a higher DSC score, along with lower HD, MAD, and DFND errors, and a higher recall rate. We postulate that this is due to lower speckle and clearer contrast along the vessel boundaries. Again, we conducted paired t-tests to assess the statistical significance of our results, and report that the results were statistically significant ($p < 0.05$) for all metrics except DFPD and Precision. The largest HD error of 0.45mm was $\sim$9.5$\times$ smaller than the largest observed vessel diameter of 4.35mm, while the average HD error was $\sim$8$\times$ lower in contrast to the smallest vessel diameter of 2.9mm. We note that the CFCM34 can be a useful alternative for clinical use in HFUS images, for which CFCM34 and USVS-Net could both be run (and results compared) for improved segmentation.

Chest X-Ray Results

To show the generalizability of our approach, we took our network designed for UHFUS vessel segmentation, retrained, and validated it on CXR images. As seen in Table 6.2, the errors between the CFCM34 and the USVS-Net are comparable. CFCM34 has a higher DSC, lower DFND, and a higher recall rate, while we achieve slightly lower HD and MAD errors, lower DFPD and higher precision. As seen in Fig. 6.14, the dilated convolutions in the USVS-Net provide the utility of incorporating regions excluded by the CFCM34 in the final segmentation. The increased contextual information available at the deepest layers of the network allowed it to segment the lung regions better.

Table 6.2: Segmentation error comparison (pixels) for the Montgomery County Chest X-Ray dataset.

Method	DSC	HD	MAD	DFPD	DFND	Prec	Rec
CFCM34 [161]	**97.01 ± 1.82**	11.05 ± 10.78	0.13 ± 0.31	6.67 ± 0.97	**6.39 ± 0.98**	96.93 ± 2.42	**97.25 ± 2.67**
USVS-Net	96.89 ± 1.80	**10.29 ± 8.26**	**0.10 ± 0.19**	**6.64 ± 0.89**	6.73 ± 1.04	**97.15 ± 1.65**	96.57 ± 2.97

Performance

The network training and testing was performed using Tensorflow on a desktop using a 3.5 GHz Intel i7 processor, 16 GB DDR3 RAM, and a NVIDIA Titan Xp GPU. The DecLSTM had 32.65 million parameters and a runtime of 5.83s (58.3ms per B-scan in 100 B-scan sequence). The CFCM34 had 49.16 million parameters and a runtime of 8.45s (84.5ms per B-scan in 100 B-scan sequence). The USVS-Net had 64.34 million parameters and a runtime of 9.95s (99.5ms per B-scan in 100 B-scan sequence). The level set method had a runtime of 2.03s (20.31ms per B-scan in 100 B-scan sequence), but yielded less accurate segmentations in contrast to the deep learning approaches. Testing was done with an ensemble of two models with the lowest validation loss.

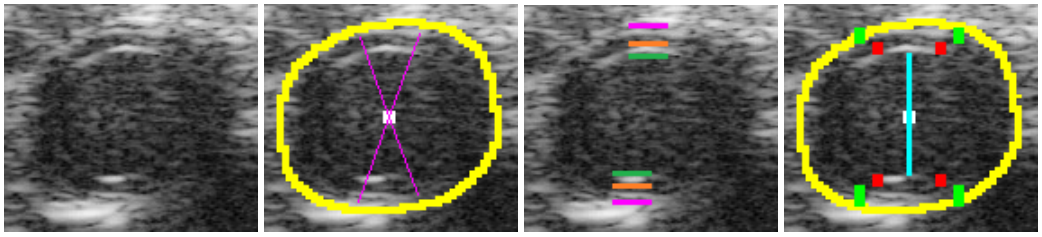
6.3 Vessel-Based Measurements

6.3.1 Problem Statement

An image $\mathcal{I}$ displaying a transverse cross-section of a vessel imaged using UHFUS is shown in Fig. 6.15(a). Once the vessel contour has been delineated by either the classical or deep learning-based approach, vessel-based measurements can be calculated. The predicted vessel contour by the proposed deep learning approach is shown in yellow in Fig. 6.15(b). These measurements include the Intimal Wall Thickness (IWT), Media Thickness (MT), and Intima-Medial Thickness (IMT). As shown in Fig. 6.15(c), we want to find the set of Y-coordinates that correspond to the inner (green lines) and outer edges (orange lines) of the intimal wall, and the inner side of the adventitia (magenta lines). With these locations, we can find the intimal wall thickness (pixels between the orange and green lines), media thickness (pixels between orange and magenta lines), and the intima-media thickness (pixels between magenta and green lines).

6.3.2 Algorithm

In this work, we have developed an automated algorithm that can determine these measurements once the vessel contour has been delineated in an UHFUS/HFUS image. Let $\mathcal{C}$ represent the vessel contour containing a set of points p_i where $i = 1, \ldots, N$ with N being the number of points in the contour. The automated algorithm first determines the geometric centroid c of the vessel contour. As the pixels on the contour represent the foreground values of the vessel boundary, the mean pixel intensity of the vessel contour m is calculated as seen in Eq. (6.29). Next, S radial spokes emanate from the centroid c in the upward and downward directions within a 45° zone. The length of each spoke L is bounded by the



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

Figure 6.15: (a) Example transverse cross-section of a vessel imaged using UHFUS; note the resolution with which the vessel walls at the top and bottom are identified. (b) Segmentation (yellow contour) of the adventitial wall from our deep learning approach. The centroid of the vessel is shown in white. The magenta lines correspond to the 45° sectors that represent the search for the vessel walls at the top and bottom. (c) The inner (green line) and outer (orange line) edges of the intimal wall, and the inner edge (magenta line) of the adventitial wall is depicted. (d) The final vessel-based measurements provided by the automated algorithm. The lumen diameter (cyan vertical line), the intimal wall thickness (red bars), and the medial thickness (green bars). The intima-media thickness is the sum of the red and green bars. Thicknesses are measured in terms of pixels, and are converted to microns after multiplication with the pixel resolution.

distance to the nearest point on the contour along that spoke.

$$m = \frac{1}{N}\sum_{i=1}^{N} I(p_i) \tag{6.29}$$

Theoretically, there can be spokes emanating from the centroid in all 360°. However as shown in Fig. 6.15(b), the intimal wall is most clearly seen at the top and bottom of the vessel due to higher axial resolution in ultrasound. There is also greater reflection of acoustic waves at interfaces that lie perpendicular to its direction of travel, which explains the increased brightness of the intimal walls. But, the intimal walls are poorly delineated in the lateral direction because the lateral resolution is not as high in ultrasound. Moreover, the walls are parallel to the direction of travel of the acoustic waves, thereby offering little in terms of acoustic reflection, and are poorly illuminated. Thus, in the upward direction,

the set of constituent angles were limited to a 45° sector (magenta lines) within the range of $s^\circ \in [250^\circ, 290^\circ]$. Similarly, the set of constituent angles in the downward direction were limited to a 45° sector in the range of $s^\circ \in [70^\circ, 110^\circ]$. Within these sectors, we would like to find the Y-coordinates of the intima edges (inner and outer) and inner adventitial side.

We denote by p^s the set of points on the spoke at angle s°. The set of Y-coordinates of the inner i and outer o edges of the intimal wall at the top t and bottom b of the vessel are given by the following equations:

$$Y_s^{ti} = \min\{y : (x,y) \in p^s \; and \; \mathcal{I}(x,y) \geq m\} \tag{6.30}$$

$$Y_s^{to} = \max\{y : (x,y) \in p^s \text{ and } \mathcal{I}(x,y) \geq m\} \tag{6.31}$$

$$Y_s^{bi} = \min\{y : (x,y) \in p^s \; and \; \mathcal{I}(x,y) \geq m\} \tag{6.32}$$

$$Y_s^{bo} = \max\{y : (x,y) \in p^s \text{ and } \mathcal{I}(x,y) \geq m\} \tag{6.33}$$

Once the set of Y-coordinates have been found, then the mean Y-coordinate of the inner and outer edges of the top and bottom of the intimal wall (IW) are computed as follows:

$$IW_y^{ti} = \frac{1}{45}\sum Y_s^{ti} \tag{6.34}$$

$$IW_y^{to} = \frac{1}{45}\sum Y_s^{to} \tag{6.35}$$

$$IW_y^{bi} = \frac{1}{45}\sum Y_s^{bi} \tag{6.36}$$

$$IW_y^{bo} = \frac{1}{45}\sum Y_s^{bo} \tag{6.37}$$

In similar fashion, the set of Y-coordinates of the inner edge of the adventitial wall for

the top and bottom of the vessel can be ascertained using the following equations:

$$AY_s^{ti} = \min\{y : (x,y) \in p^s \; and \; (x,y) \in \mathcal{C}\} \tag{6.38}$$

$$AY_s^{bi} = \min\{y : (x,y) \in p^s \; and \; (x,y) \in \mathcal{C}\} \tag{6.39}$$

The mean Y-coordinate of the inner edge of the adventitial wall at the top and bottom of the vessel can be found using the following equations:

$$AW_y^{ti} = \frac{1}{45}\sum AY_s^{ti} \tag{6.40}$$

$$AW_y^{bi} = \frac{1}{45}\sum AY_s^{bi} \tag{6.41}$$

Finally, the Medial Thickness (MT), the Intimal Wall Thickness (IWT), and the Intima-Media Thickness (IMT) at the top and bottom of the vessel can be measured as follows:

$$IWT^t = abs\big(IW_y^{to} - IW_y^{ti}\big) \tag{6.42}$$

$$IWT^b = abs\big(IW_y^{bo} - IW_y^{bi}\big) \tag{6.43}$$

$$MT^t = abs\big(IW_y^{to} - AW_y^{ti}\big) \tag{6.44}$$

$$MT^b = abs\big(AW_y^{bi} - IW_y^{bo}\big) \tag{6.45}$$

$$IMT^t = IWT^t + MT^t \tag{6.46}$$

$$IMT^b = IWT^b + MT^b \tag{6.47}$$

These measurements are depicted pictorially on a example image of a transverse vessel cross-section as seen in Fig. 6.15(d). In this figure, the white square represents the centroid of the vessel. The vertical line in cyan represents the vessel lumen diameter. The red bars represent the thickness of the IWT, and the green bars represent the medial thickness. The IMT is the sum of the thicknesses of the red and green bars. The yellow contour is the predicted vessel boundary.

6.3.3 Results

We ran this algorithm on UHFUS sequences acquired at the radial artery in the hand from 6 different subjects and show the variation in the measurements over 100 images in each sequences in Table. 6.3. We note that the measurements are consistent with those measurements seen [168, 169]. Most of the measurments were acquired on subjects who were healthy, and therefore there may not be representative estimates on subjects who have undergone hand transplants or have other associated cardiovascular risks.

Table 6.3: Mean and Standard Deviation of the thickness measures of the radial artery for various subjects. Intimal Wall Thickness (IWT), Media Thickness (MT), and Intima-Media Thickness (IMT). These measurements are in millimeters.

Subject #	IWT		MT		IMT	
	Top	Bottom	Top	Bottom	Top	Bottom
1	0.06 ± 0.009	0.06 ± 0.007	0.17 ± 0.015	0.14 ± 0.014	0.23 ± 0.013	0.2 ± 0.013
2	0.09 ± 0.038	0.07 ± 0.013	0.13 ± 0.034	0.11 ± 0.016	0.21 ± 0.025	0.18 ± 0.013
3	0.08 ± 0.026	0.06 ± 0.029	0.14 ± 0.029	0.1 ± 0.023	0.21 ± 0.018	0.15 ± 0.016
4	0.07 ± 0.015	0.06 ± 0.01	0.14 ± 0.018	0.13 ± 0.014	0.21 ± 0.008	0.18 ± 0.012
5	0.07 ± 0.028	0.07 ± 0.016	0.14 ± 0.031	0.13 ± 0.02	0.20 ± 0.02	0.2 ± 0.009
6	0.07 ± 0.012	0.07 ± 0.008	0.14 ± 0.018	0.13 ± 0.016	0.21 ± 0.014	0.19 ± 0.013

Chapter 7

Future Work

7.1 Exploiting Segmentations for Accurate Registration

In this thesis, we have proposed the following approaches for corneal interface segmentation: classical approach in Sec. 4.1, deep learning-based approach in Sec. 4.2, and a hybrid/cascaded approach in Sec. 4.3. All these approaches consider each OCT image in the sequence to be independent of each other. However, this is not the case as there is a lot of correlation across space and time between the OCT B-scans. It can be difficult to exploit this correlation without aligning the OCT B-scans first. But, if a set of B-scans are passed to the neural network architectures instead, then the corneal interface can be segmented with greater accuracy in scenarios when the SNR drop is quite large. Similarly, we believe that the above naturally extends to limbal interface segmentation as well.

Once the interfaces have been accurately segmented, it is possible to train a neural network architecture that can directly estimate the spatial transformation [170] between B-scans. In this way, B-scans in a sequence can be aligned to recreate the corneal or limbal

region in a 3D reconstruction. Furthermore, it will also enable the accurate measurement of corneal biometric parameters, such as corneal refractive index, and further enhance the quantification of morphological changes in the POV.

7.2 Uncertainty in Segmentation

It would be useful to know the level of trust to be placed in the output of a segmentation framework. This is especially true in one of our application areas: visualizing the POV. As mentioned in this thesis, datasets for the longitudinal study in Sec. 5.4 were only considered for the first three days of the study. Datasets from days 4 through 12 were not considered as the degradation in the tissue structure is significant, and it can be particularly difficult to determine the true location of the shallowest limbal interface. Furthermore, in order to be of clinical use, it is necessary to know the depletion in stem cells in the POV. This will help determine if corneas from donor eyes are truly viable for transplant into recipients.

Recently, there have been methods [171–174] proposed to determine the uncertainty of a neural network during the task of segmentation. These Bayesian uncertainty estimation methods [171–174] are highly useful to assess the performance of a framework designed for segmentation as it can provide a metric for the trust to be placed in the output of the framework. In the future, we plan to integrate these methods into our frameworks to provide confidence measures in our segmentation predictions.

7.3 ProbeSight

We plan to integrate our vessel segmentation and tracking approaches into an existing project in our lab called ProbeSight [175]. ProbeSight is a system designed to guide ul-

trasound transducers (along with other clinical tools) using one or more video camera(s) mounted directly on the probe (or tool) to a specified target position on the exterior of the patient. In this system, a prior high-resolution surface map of the patient's exterior (e.g. patient's arm) is correlated with the real-time video images of the patientâĂŹs exterior. This has direct utility in tracking anatomical changes, with particular interest in monitoring hand transplant rejection, as it is presumed necessary to precisely revisit the position, orientation, and compression of a previously acquired ultrasound image in order to monitor subtle changes in tissue structure and health.

Along with the real-time images of the patient's exterior, real-time images of the vascular structures can be obtained using ultrasound. The methods developed in this thesis - tracking of the vessels, segmenting their contours, and understanding the changes in vessel-based parameters (e.g. intima-medial thickness) - can be directly integrated with the design of ProbeSight in order to further enhance measurements of clinical markers of hand transplant rejection.

Bibliography

[1] Tejas Sudharshan Mathai, John Galeotti, Samantha Horvath, and George Stetten. Graphics processor unit (gpu) accelerated shallow transparent layer detection in optical coherence tomographic (oct) images for real-time corneal surgical guidance. In Cristian A. Linte, Ziv Yaniv, Pascal Fallavollita, Purang Abolmaesumi, and David R. Holmes, editors, *Augmented Environments for Computer-Assisted Interventions*, pages 1–13, Cham, 2014. Springer International Publishing.

[2] Chengqian Che, Tejas Sudharshan Mathai, and John Galeotti. Ultrasound registration: A review. *Methods*, 115:128 – 143, 2017. Image Processing for Biologists.

[3] 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)*, pages 1327–1331, April 2018.

[4] Tejas Sudharshan Mathai, Lingbo Jin, Vijay Gorantla, and John Galeotti. Fast vessel segmentation and tracking in ultra high-frequency ultrasound images. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 746–754, Cham, 2018. Springer International Publishing.

[5] Tejas Sudharshan Mathai, Kira L. Lathrop, and John M. Galeotti. Learning to segment corneal tissue interfaces in OCT images. *CoRR*, abs/1810.06612, 2018.

[6] Jiahong Ouyang, Tejas Sudharshan Mathai, Kira Lathrop, and John Galeotti. Accurate tissue interface segmentation via adversarial pre-segmentation of anterior segment oct images. volume abs/1905.02378, 2019.

[7] E. Bruce Goldstein. *Sensation and Perception*. Cengage, 2001.

[8] T. Olsen. On the calculation of power from curvature of the cornea. *British Journal of Ophthalmology*, 70(2):152–154, 1986.

[9] S. Klyce. Computer-assisted corneal topography: High-resolution graphic presentation and analysis of keratoscopy. *Investigative Ophthalmology and Visual Science*, 25(12):1426, 1984.

[10] J. Schwiegerling. Field guide to visual and ophthalmic optics. *SPIE Press*, 2004.

[11] A. Gullstrand. *Einfuhrung in die methoden der dioptrik des auges des menschen*. S. Hirzel, 1911.

[12] SA. Klein. Corneal topography: A review, new ansi standards and problems to solve. In *Vision Science and its Applications*, page NW8. Optical Society of America, 2000.

[13] NES. Norrby. Unfortunate discrepancies. *Journal of Cataract and Refractive Surgery*, 24:433, 1998.

[14] B. Seitz et al. Underestimation of intraocular lens power for cataract surgery after myopic photorefractive keratectomy. *Ophthalmology*, 106:693 – 702, 1999.

[15] American Society of Cataract and Refractive Surgery. Ascrs to participate in and co-fund study on post-lasik quality of life with u.s. food and drug administration. *ASCRS*, 2008.

[16] SC. Schallhorn, AA. Farjo, D. Huang, BS. Boxer Wachler, WB. Trattler, DJ. Tanzer, PA. Majmudar, and A. Sugar. Wavefront-guided lasik for the correction of primary myopia and astigmatism: A report by the american academy of ophthalmology. *Ophthalmology*, 115(7):1249 – 1261, 2008.

[17] G. Kymionis et al. Eleven-year follow-up of laser in situ keratomileusis. *Journal of Cataract and Refractive Surgery*, 33:191 – 196, 2007.

[18] Jorge L. Alio, Felipe Soria, Alessandro Abbouda, and Pablo Pena-Garcia. Laser in situ keratomileusis for âĹŠ6.00 to âĹŠ18.00 diopters of myopia and up to âĹŠ5.00 diopters of astigmatism: 15-year follow-up. *Journal of Cataract and Refractive Surgery*, 41(1):33 – 40, 2015.

[19] A. Williams, FA. Sloan, and PP. Lee. Longitudinal rates of cataract surgery. *Archives of Ophthalmology*, 124(9):1308–1314, 2006.

[20] Anthony N. Kuo, Ryan P. McNabb, Mingtao Zhao, Francesco LaRocca, Sandra S. Stinnett, Sina Farsiu, and Joseph A. Izatt. Corneal biometry from volumetric sdoct and comparison with existing clinical modalities. *Biomed. Opt. Express*, 3(6):1279–1290, Jun 2012.

[21] M. Zhao, AN Kuo, and JA Izatt. 3d refraction correction and extraction of clinical parameters from spectral domain optical coherence tomography of the cornea. *Optics Express*, 18(9):8923–8936, Apr 2010.

[22] RC. Lin, MA. Shure, AM. Rollins, JA. Izatt, and D. Huang. Group index of the human cornea at 1.3-um wavelength obtained in vitro by optical coherence domain reflectometry. *Opt. Lett.*, 29(1):83–85, Jan 2004.

[23] VADP. Sicam, M. Dubbelman, and RGL. van der Heijde. Spherical aberration of the anterior and posterior surfaces of the human cornea. *J. Opt. Soc. Am. A*, 23(3):544–549, Mar 2006.

[24] D. Malacara and Z. Malacara. Handbook of optical design. *CRC Press*, 2004.

[25] M. Dubbelman, HA. Weeber, RGL. Van Der Heijde, and HJ. VÃűlkerâĂŘDieben. Radius and asphericity of the posterior corneal surface determined by corrected scheimpflug photography. *Acta Ophthalmologica Scandinavica*, 80(4):379–383, 2002.

[26] MJ. Howcroft and JA. Parker. Aspheric curvatures for the human lens. *Vision Research*, 17(10):1217 – 1223, 1977.

[27] WM. Townsend. The limbal palisades of vogt. *Transactions of the American Ophthalmological Society*, 89:721–756, 1991.

[28] MF. Goldberg and AJ. Bron. Limbal palisades of vogt. *Transactions of the American Ophthalmological Society*, 80:155âĂŞ171, 1982.

[29] N. Di Girolamo TJ. Echevarria. Tissue-regenerating, vision-restoring corneal epithelial stem cells. *Stem Cell Rev*, 7:256 âĂŞ268, 2011.

[30] C. Sotozono et al. S. Kinoshita, W. Adachi. Characteristics of the human ocular surface epithelium. *Prog Retin Eye Res*, 20:639âĂŞ673, 2001.

[31] MA. Akinci JM. Wolosin, MT. Budak. Ocular surface epithelial and stem cell development. *Int J Dev Biol*, 48:981âĂŞ991, 2004.

[32] M. Zhou RM. Lavker MA. Akinci JM. Wolosin MT. Budak, OS. Alpdogan. Ocular surface epithelia contain abcg2-dependent side population cells exhibiting features associated with stem cells. *J Cell Sci*, 118(part 8):1715âĂŞ1724, 2005.

[33] J. Xu et al. J. Hong, T. Zheng. Assessment of limbus and central cornea in patients with keratolimbal allograft transplantation using in vivo laser scanning confocal microscopy: an observational study. *Graefes Arch Clin Exp Ophthalmol*, 249:701âĂŞ708, 2011.

[34] L. Kagemann JS. Schuman N. SundarRaj KL. Lathrop, D. Gupta. Optical coherence tomography as a rapid, accurate, noncontact method of visualizing the palisades of vogt. *Investigative Ophthalmology and Visual Science*, 53(3):1381–1387, 2012.

[35] Ammar Miri, Muhamed Al-Aqaba, Ahmad Muneer Otri, Usama Fares, Dalia G Said, Lana Akram Faraj, and Harminder S Dua. In vivo confocal microscopic features of normal limbus. *British Journal of Ophthalmology*, 96(4):530–536, 2012.

[36] 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, Sep 2017.

[37] Michel Haagdorens, Joséphine Behaegel, Jos Rozema, Veerle Van Gerwen, Sofie Michiels, Sorcha Ní Dhubhghaill, Marie-José Tassignon, and Nadia Zakaria. A method for quantifying limbal stem cell niches using oct imaging. *British Journal of Ophthalmology*, 101(9):1250–1255, 2017.

[38] Alex J. Shortt, Genevieve A. Secker, Peter M. Munro, Peng T. Khaw, Stephen J. Tuft, and Julie T. Daniels. Characterization of the limbal epithelial stem cell niche: Novel imaging techniques permit in vivo observation and targeted biopsy of limbal epithelial stem cells. *STEM CELLS*, 25(6):1402–1409, 2007.

[39] K. Bizheva, N. Hutchings, L. Sorbara, AA. 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, Jul 2011.

[40] AĐŘ. Voskresenskaya BV Gagloev NP. Pashtaev, NA. Pozdeyeva and AA. Shipunov. Comparative analysis of the value of information provided by anterior segment optical coherence tomography and in vivo confocal microscopy for identifying the palisades of vogt in normal limbus. *Russian Annals of Ophthalmology*, 1:60âĂŞ68, 2017.

[41] Vijay S. Gorantla and Anthony J. Demetris. Acute and chronic rejection in upper extremity transplantation: What have we learned? *Hand Clinics*, 27(4):481 – 493, 2011. Hand Transplantation.

[42] YC. Pan et al. WP. Lee, MJ. Yaremchuk. Relative antigenicity of components of a vascularized limb allograft. *Hand Clinics*, 87(3):401âĂŞ412, 1991. Plast Reconstr Surg.

[43] Jacqueline Saw, G.B. John Mancini, and Karin H. Humphries. Contemporary review on spontaneous coronary artery dissection. *Journal of the American College of Cardiology*, 68(3):297–312, 2016.

[44] V. Gorantla J. Keith. High resolution ultrasound for noninvasive monitoring of graft vasculopathy after human upper extremity transplantation. *Transplantation*, 94(10S):953, 2012.

[45] D. Huang, EA. Swanson, CP. Lin, JS. Schuman, WG. Stinson, W. Chang, MR. Hee, T. Flotte, K. Gregory, CA. Puliafito, and al. et. Optical coherence tomography. *Science*, 254(5035):1178–1181, 1991.

[46] D. Sampson and TR. Hillman. Lasers and current optical techniques in biology. *Lasers and Current Optical Techniques in Biology*, 4:481–572, 2004.

[47] JF. de Boer, B. Cense, BH. Park, MC. Pierce, GJJ. Tearney, and BE. Bouma. Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Optics Letters*, 28(21):2067–2069, Nov 2003.

[48] MA. Choma, MV. Sarunic, C. Yang, and JA. Izatt. Sensitivity advantage of swept source and fourier domain optical coherence tomography. *Optics Express*, 11(18):2183–2189, Sep 2003.

[49] R. Leitgeb, CK. Hitzenberger, and AF. Fercher. Performance of fourier domain vs. time domain optical coherence tomography. *Optics Express*, 11(8):889–894, Apr 2003.

[50] Joseph A Izatt and Michael A. Choma. *Theory of Optical Coherence Tomography*, pages 47–72. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

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

[52] A. Webb and GC. Kagadis. Introduction to biomedical imaging. *Medical Physics*, 30(8):2267–2267, 2003.

[53] I. Wendelhag, T. Gustavsson, M. SuurkÃijla, G. Berglund, and J. Wikstrand. Ultrasound measurement of wall thickness in the carotid artery: fundamental principles and description of a computerized analysing system. *Clinical Physiology*, 11(6):565–577, 1991.

[54] M. Avila D. Huang M. Tang, Y. Li. Measuring total corneal power before and after laser in situ keratomileusis with high-speed optical coherence tomography. *Journal of Cataract and Refractive Surgery*, 31(11), 2006.

[55] FW. Kremkau and KJ. Taylor. Artifacts in ultrasound imaging. *Journal of Ultrasound in Medicine*, 5(4):227–237, 1986.

[56] Jay Chhablani, Tandava Krishnan, Vaibhav Sethi, and Igor Kozak. Artifacts in optical coherence tomography. *Saudi Journal of Ophthalmology*, 28(2):81 – 87, 2014.

[57] Francesco LaRocca, Stephanie J. Chiu, Ryan P. McNabb, Anthony N. Kuo, Joseph A. Izatt, and Sina Farsiu. Robust automatic segmentation of corneal layer boundaries in sdoct images using graph theory and dynamic programming. *Biomed. Opt. Express*, 2(6):1524–1538, Jun 2011.

[58] Yan Li, Raj Shekhar, and David Huang. Segmentation of 830- and 1310-nm lasik corneal optical coherence tomography images. *Proceedings of SPIE - The International Society for Optical Engineering*, 4684, 05 2002.

[59] Kang Li, Xiaodong Wu, D. Z. Chen, and M. Sonka. Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):119–134, Jan 2006.

[60] 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 Transactions on Medical Imaging*, 29(1):159–168, Jan 2010.

[61] Michael D. Abramoff, Kyungmoo Lee, Meindert Niemeijer, Wallace L. M. Alward, Emily C. Greenlee, Mona K. Garvin, Milan Sonka, and Young H. Kwon. Automated segmentation of the cup and rim from spectral domain oct of the optic nerve head. *Investigative Ophthalmology and Visual Science*, 50(12):5778, 2009.

[62] Victor Robles, Bhavna Antony, Demelza Koehn, Michael Anderson, and Mona Garvin. 3d graph-based automated segmentation of corneal layers in anterior-segment optical coherence tomography images of mice. *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, 9038, 02 2014.

[63] Dominic Williams, Yalin Zheng, Fangjun Bao, and Ahmed Elsheikh. Fast segmentation of anterior segment optical coherence tomography images using graph cut. *Eye and Vision*, 2(1):1, Jan 2015.

[64] F. Graglia, J. L. Mari, G. Baikoff, and J. Sequeira. Contour detection of the cornea from oct radial images. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5612–5615, Aug 2007.

[65] Fangjun Bao Ahmed Elsheikh Dominic Williams, Yalin Zheng. Automatic segmentation of anterior segment optical coherence tomography images. *Journal of Biomedical Optics*, 18:18 – 18 – 8, 2013.

[66] Justin A. Eichel, Kostadinka K. Bizheva, David A. Clausi, and Paul W. Fieguth. Automated 3d reconstruction and segmentation from optical coherence tomography. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision – ECCV 2010*, pages 44–57, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[67] Kafieh R. Rabbani H. Dehnavi A. M. Peyman A. Hajizadeh F. Jahromi, M. K. and M. Ommani. An automatic algorithm for segmentation of the boundaries of corneal layers in optical coherence tomography images using gaussian mixture model. *Journal of Medical Signals and Sensors*, 4:171âĂŞ180, 2014.

[68] Hossein Rabbani, Rahele Kafieh, Mahdi Jahromi, Sahar Jorjandi, Alireza Dehnavi, Fedra Hajizadeh, and Alireza Peyman. Obtaining thickness maps of corneal layers using the optimal algorithm for intracorneal layer segmentation. *International Journal of Biomedical Imaging*, 2016:1 – 11, 2016.

[69] Meixiao; Tao Aizhu; Wang Jianhua; Dou Guopeng; Lu Fan Ge, Lili; Shen. Automatic segmentation of the central epithelium imaged with three optical coherence tomography devices. *Eye & Contact Lens*, 38:150âĂŞ157, 2012.

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

[71] Bo Wang, Larry Kagemann, Joel S. Schuman, Hiroshi Ishikawa, Richard A. Bilonick, Yun Ling, Ian A. Sigal, Zach Nadler, Andrew Francis, Michelle G. Sandrian, and Gadi Wollstein. Gold nanorods as a contrast agent for doppler optical coherence tomography. *PLOS ONE*, 9(3):1–7, 03 2014.

[72] Valentin Aranha dos Santos, Leopold Schmetterer, Hannes Stegmann, Martin Pfister, Alina Messner, Gerald Schmidinger, Gerhard Garhofer, and René M. Werkmeister. Corneanet: fast segmentation of cornea oct scans of healthy and keratoconic eyes using deep learning. *Biomed. Opt. Express*, 10(2):622–641, Feb 2019.

[73] Wajdene Ghouali, Rachid Tahiri Joutei Hassani, Zoubir Djerada, Hong Liang, Mohamed El Sanharawi, Antoine LabbÃĽ, and Christophe Baudouin. In vivo imaging of palisades of vogt in dry eye versus normal subjects using en-face spectral-domain optical coherence tomography. *PLOS ONE*, 12(11):1–13, 11 2017.

[74] P. Thevenaz, U. E. Ruttimann, and M. Unser. A pyramid approach to subpixel registration based on intensity. *IEEE Transactions on Image Processing*, 7(1):27–41, Jan 1998.

[75] P. Viola and W. M. Wells. Alignment by maximization of mutual information. In *Proceedings of IEEE International Conference on Computer Vision*, pages 16–23, Jun 1995.

[76] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. *In: Bizais*, 1995.

[77] Abhijit Guha Roy, Sailesh Conjeti, Sri Phani Krishna Karri, Debdoot Sheet, Amin Katouzian, Christian Wachinger, and Nassir Navab. Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomed. Opt. Express*, 8(8):3627–3642, Aug 2017.

[78] Leyuan Fang, David Cunefare, Chong Wang, Robyn H. Guymer, Shutao Li, and Sina 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, May 2017.

[79] Stefanos Apostolopoulos, Sandro De Zanet, Carlos Ciller, Sebastian Wolf, and Raphael Sznitman. Pathological oct retinal layer segmentation using branch residual u-shape networks. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne, editors, *Medical Image Computing and Computer-Assisted Intervention âĹŠ MICCAI 2017*, pages 294–301, Cham, 2017. Springer International Publishing.

[80] Yupeng Xu, Ke Yan, Jinman Kim, Xiuying Wang, Changyang Li, Li Su, Suqin Yu, Xun Xu, and Dagan David Feng. Dual-stage deep learning framework for pigment epithelium detachment segmentation in polypoidal choroidal vasculopathy. *Biomed. Opt. Express*, 8(9):4061–4076, Sep 2017.

[81] Cecilia S. Lee, Ariel J. Tyring, Nicolaas P. Deruyter, Yue Wu, Ariel Rokem, and Aaron Y. Lee. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed. Opt. Express*, 8(7):3440–3448, Jul 2017.

[82] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[83] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, June 2015.

[84] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[85] Fei Mao, Jeremy Gill, Donal Downey, and Aaron Fenster. Segmentation of carotid artery in ultrasound images: Method development and evaluation technique. *Medical Physics*, 27(8):1961–1970, 2000.

[86] C. Chen, T. L. Poepping, J. J. Beech-Brandt, S. J. Hammer, R. Baldock, B. Hill, P. Allan, W. J. Easson, and P. R. Hoskins. Segmentation of arterial geometry from ultrasound images using balloon models. In *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, pages 1319–1322 Vol. 2, April 2004.

[87] Ali K. Hamou, Said Osman, and Mahmoud R. El-Sakka. Carotid ultrasound segmentation using dp active contours. In Mohamed Kamel and Aurélio Campilho, editors, *Image Analysis and Recognition*, pages 961–971, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[88] J. Chaniot, B. Sciolla, P. Delachartre, T. Dambry, and B. Guibert. Vessel segmentation in high-frequency 2d/3d ultrasound images. In *2016 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, Sept 2016.

[89] P. Abolmaesumi, M. R. Sirouspour, and S. E. Salcudean. Real-time extraction of carotid artery contours from ultrasound images. In *Proceedings 13th IEEE Symposium on Computer-Based Medical Systems. CBMS 2000*, pages 181–186, 2000.

[90] P. Abolmaesumi, S. E. Salcudean, Wen-Hong Zhu, M. R. Sirouspour, and S. P. DiMaio. Image-guided control of a robot for medical ultrasound. *IEEE Transactions on Robotics and Automation*, 18(1):11–23, Feb 2002.

[91] J. Guerrero, S. E. Salcudean, J. A. McEwen, B. A. Masri, and S. Nicolaou. Real-time vessel segmentation and tracking for ultrasound imaging applications. *IEEE Transactions on Medical Imaging*, 26(8):1079–1090, Aug 2007.

[92] D. C. Wang, R. Klatzky, B. Wu, G. Weller, A. R. Sampson, and G. D. Stetten. Fully automated common carotid artery and internal jugular vein identification and tracking using b-mode ultrasound. *IEEE Transactions on Biomedical Engineering*, 56(6):1691–1699, June 2009.

[93] E. Smistad and F. Lindseth. Real-time automatic artery segmentation, reconstruction and registration for ultrasound-guided regional anaesthesia of the femoral nerve. *IEEE Transactions on Medical Imaging*, 35(3):752–761, March 2016.

[94] J. Stoitsis, S. Golemati, S. Kendros, and K. S. Nikita. Automated detection of the carotid artery wall in b-mode ultrasound images using active contours initialized by the hough transform. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3146–3149, Aug 2008.

[95] A. S. Jorgensen, S. E. Schmidt, N. H. Staalsen, and L. R. Ostergaard. Semi-automatic vessel tracking and segmentation using epicardial ultrasound in bypass surgery. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2331–2334, Aug 2012.

[96] Kun Qian, Takehiro Ando, Kensuke Nakamura, Hongen Liao, Etsuko Kobayashi, Naoki Yahagi, and Ichiro Sakuma. Ultrasound imaging method for internal jugular vein measurement and estimation of circulating blood volume. *International Journal of Computer Assisted Radiology and Surgery*, 9(2):231–239, Mar 2014.

[97] Lei Ma, Hidemichi Kiyomatsu, Keiichi Nakagawa, Junchen Wang, Etsuko Kobayashi, and Ichiro Sakuma. Accurate vessel segmentation in ultrasound images using a local-phase-based snake. *Biomedical Signal Processing and Control*, 43:236 – 243, 2018.

[98] Erik Smistad and Lasse Løvstakken. Vessel detection in ultrasound images using deep convolutional neural networks. In Gustavo Carneiro, Diana Mateus, Loïc Peter, Andrew Bradley, João Manuel R. S. Tavares, Vasileios Belagiannis, João Paulo Papa, Jacinto C. Nascimento, Marco Loog, Zhi Lu, Jaime S. Cardoso, and Julien Cornebise, editors, *Deep Learning and Data Labeling for Medical Applications*, pages 30–38, Cham, 2016. Springer International Publishing.

[99] Rosa-MarÃŋa MenchÃşn-Lara and JosÃĺ-Luis Sancho-GÃşmez. Fully automatic segmentation of ultrasound common carotid artery images based on machine learning. *Neurocomputing*, 151:161 – 167, 2015.

[100] J. Y. Shin, N. Tajbakhsh, R. T. Hurst, C. B. Kendall, and J. Liang. Automating carotid intima-media thickness video interpretation with convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2526–2535, June 2016.

[101] C. Azzopardi, Y. A. Hicks, and K. P. Camilleri. Automatic carotid ultrasound segmentation using deep convolutional neural networks and phase congruency maps. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 624–628, April 2017.

[102] Dan Z. Reinstein, Ronald H. Silverman, Mark J. Rondeau, and D. Jackson Coleman. Epithelial and corneal thickness measurements by high-frequency ultrasound digital signal processing. *Ophthalmology*, 101(1):140 – 146, 1994.

[103] Hong Fang Li, W. Matthew Petroll, Torben MÃÿller-Pedersen, James K. Maurer, H. Dwight Cavanagh, and James V. Jester. Epithelial and corneal thickness measurements by in vivo confocal microscopy through focusing (cmtf). *Current Eye Research*, 16(3):214–221, 1997.

[104] William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979.

[105] Zhengyou Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, Oct 1994.

[106] Martin F. Kraus, Benjamin Potsaid, Markus A. Mayer, Ruediger Bock, Bernhard Baumann, Jonathan J. Liu, Joachim Hornegger, and James G. Fujimoto. Motion correction in optical coherence tomography volumes on a per a-scan basis using orthogonal scan patterns. *Biomed. Opt. Express*, 3(6):1182–1199, Jun 2012.

[107] Kari V. Vienola, Boy Braaf, Christy K. Sheehy, Qiang Yang, Pavan Tiruveedhula, David W. Arathorn, Johannes F. de Boer, and Austin Roorda. Real-time eye motion compensation for oct imaging with tracking slo. *Biomed. Opt. Express*, 3(11):2950–2963, Nov 2012.

[108] 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.

[109] Bo Wang, Larry Kagemann, Joel S. Schuman, Hiroshi Ishikawa, Richard A. Bilonick, Yun Ling, Ian A. Sigal, Zach Nadler, Andrew Francis, Michelle G. Sandrian, and Gadi Wollstein. Gold nanorods as a contrast agent for doppler optical coherence tomography. *PLOS ONE*, 9(3):1–7, 03 2014.

[110] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In *ICLR*, 2016.

[111] Sripad Krishna Devalla, Prajwal K. Renukanand, Bharathwaj K. Sreedhar, Giridhar Subramanian, Liang Zhang, Shamira Perera, Jean-Martial Mari, Khai Sing Chin, Tin A. Tun, Nicholas G. Strouthidis, Tin Aung, Alexandre H. Thiéry, and Michaël 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, Jul 2018.

[112] C. Szegedy, Wei Liu, Yangqing 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)*, pages 1–9, June 2015.

[113] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 448–456. JMLR.org, 2015.

[114] 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)*, pages 770–778, June 2016.

[115] Suman Sedai, Bhavna Antony, Dwarikanath Mahapatra, and Rahil Garnavi. Joint segmentation and uncertainty visualization of retinal layers in optical coherence tomography images using bayesian deep learning. In Danail Stoyanov, Zeike Taylor, Francesco Ciompi, Yanwu Xu, Anne Martel, Lena Maier-Hein, Nasir Rajpoot, Jeroen van der Laak, Mitko Veta, Stephen McKenna, David Snead, Emanuele Trucco, Mona K. Garvin, Xin Jan Chen, and Hrvoje Bogunovic, editors, *Computational Pathology and Ophthalmic Medical Image Analysis*, pages 219–227, Cham, 2018. Springer International Publishing.

[116] 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)*, pages 2261–2269, July 2017.

[117] S. JÃľgou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1175–1183, July 2017.

[118] Naji Khosravan and Ulas Bagci. S4ND: single-shot single-scale lung nodule detection. *CoRR*, abs/1805.02279, 2018.

[119] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 2016.

[120] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1520–1528, Dec 2015.

[121] 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.*, pages 958–963, Aug 2003.

[122] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.

[123] WS Cleveland. Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician,*, 35:54, 1981.

[124] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.

[125] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[126] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, Dec 2015.

[127] 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)*, pages 565–571, Oct 2016.

[128] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, Dec 2001.

[129] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 839–846, Jan 1998.

[130] M. Felsberg and G. Sommer. The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. *Journal of Mathematical Imaging and Vision*, 21(1):5–26, Jul 2004.

[131] Peter Kovesi. Symmetry and asymmetry from local phase. In *Tenth Australian Joint Converence on Artificial Intelligence*, pages 2–4, 1997.

[132] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[133] Virender S. Sangwan Sayan Basu Kira L. Lathrop, Jessica F. Steele and Swapna Shanbhag. Ex-vivo remodeling of palisades of vogt in preserved tissue. In *ARVO Annual Meeting Abstract, June, 2015*, page 3465, 2016.

[134] Michal Drozdzal, Eugene Vorontsov, Gabriel Chartrand, Samuel Kadoury, and Chris Pal. The importance of skip connections in biomedical image segmentation. In Gustavo Carneiro, Diana Mateus, Loïc Peter, Andrew Bradley, João Manuel R. S. Tavares, Vasileios Belagiannis, João Paulo Papa, Jacinto C. Nascimento, Marco Loog, Zhi Lu, Jaime S. Cardoso, and Julien Cornebise, editors, *Deep Learning and Data Labeling for Medical Applications*, pages 179–187, Cham, 2016. Springer International Publishing.

[135] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, and JÃijrgen Schmidhuber. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.

[136] Zifeng Wu, Chunhua Shen, and Anton van den Hengel. Wider or deeper: Revisiting the resnet model for visual recognition. *CoRR*, abs/1611.10080, 2016.

[137] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016.

[138] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 4278–4284, 2017.

[139] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[140] M. Akmal Butt and P. Maragos. Optimum design of chamfer distance transforms. *IEEE Transactions on Image Processing*, 7(10):1477–1484, Oct 1998.

[141] Jeff Henrikson. Completeness and total boundedness of the hausdorff metric. *MIT Undergraduate Journal of Mathematics*, 1999.

[142] Leica. Leica envisu c2300 system specifications. *Leica Product Documentation*, 2019.

[143] Brenton Keller, Mark Draelos, Gao Tang, Sina Farsiu, Anthony N. Kuo, Kris Hauser, and Joseph A. Izatt. Real-time corneal segmentation and 3d needle tracking in intrasurgical oct. *Biomed. Opt. Express*, 9(6):2716–2732, Jun 2018.

[144] Yan Li, Raj Shekhar, and David Huang. Corneal pachymetry mapping with high-speed optical coherence tomography. *Ophthalmology*, 113:792–9.e2, 06 2006.

[145] Dominic Williams, Yalin Zheng, Pinakin Gunvant Davey, Fangjun Bao, Meixiao Shen, and Ahmed Elsheikh. Reconstruction of 3d surface maps from anterior segment optical coherence tomography images using graph theory and genetic algorithms. *Biomedical Signal Processing and Control*, 25:91 – 98, 2016.

[146] Tilman Schmoll, Angelika Unterhuber, Christoph Kolbitsch, Tuan Dzung Le, Andreas Dr. Stingl, and Rainer A. Leitgeb. Precise thickness measurements of bowman's layer, epithelium, and tear film. *Optometry and vision science : official publication of the American Academy of Optometry*, 89 5:E795–802, 2012.

[147] Min Chen, Jiancong Wang, Ipek Oguz, Brian L. VanderBeek, and James C. Gee. Automated segmentation of the choroid inÂăedi-oct images with retinal pathology using convolution neural networks. In M. Jorge Cardoso, Tal Arbel, Andrew Melbourne, Hrvoje Bogunovic, Pim Moeskops, Xinjian Chen, Ernst Schwartz, Mona Garvin, Emma Robinson, Emanuele Trucco, Michael Ebner, Yanwu Xu, Antonios Makropoulos, Adrien Desjardin, and Tom Vercauteren, editors, *Fetal, Infant and Ophthalmic Medical Image Analysis*, pages 177–184, Cham, 2017. Springer International Publishing.

[148] Xiaodan Sui, Yuanjie Zheng, Benzheng Wei, Hongsheng Bi, Jianfeng Wu, Xuemei Pan, Yilong Yin, and Shaoting Zhang. Choroid segmentation from opticalÂăcoherence tomography with graph-edge weights learned from deep convolutional neural networks. *Neurocomputing*, 237:332 – 341, 2017.

[149] Freerk G. Venhuizen, Bram van Ginneken, Bart Liefers, Mark J.J.P. van Grinsven, Sascha Fauser, Carel Hoyng, Thomas Theelen, and Clara I. Sánchez. Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks. *Biomed. Opt. Express*, 8(7):3292–3316, Jul 2017.

[150] Abhay Shah, Michael D. Abramoff, and Xiaodong Wu. Simultaneous multiple surface segmentation using deep learning. In M. Jorge Cardoso, Tal Arbel, Gustavo Carneiro, Tanveer Syeda-Mahmood, João Manuel R.S. Tavares, Mehdi Moradi, Andrew Bradley, Hayit Greenspan, João Paulo Papa, Anant Madabhushi, Jacinto C. Nascimento, Jaime S. Cardoso, Vasileios Belagiannis, and Zhi Lu, editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, Cham, 2017. Springer International Publishing.

[151] Peter J. Burt and Edward H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions ON Communications*, 31:532–540, 1983.

[152] G. Stetten, C. Wong, V. Shivaprabhu, A. Zhang, S. Horvath, J. Wang, J. Galeotti, V. Gorantla, and H. Aizenstein. Descending variance graphs for segmenting neurological structures. In *2013 International Workshop on Pattern Recognition in Neuroimaging*, pages 174–177, June 2013.

[153] A. Belaid, D. Boukerroui, Y. Maingourd, and J. F. Lerallut. Phase-based level set segmentation of ultrasound images. *IEEE Transactions on Information Technology in Biomedicine*, 15(1):138–147, Jan 2011.

[154] C.P. Bridge. Introduction to the monogenic signal. *arXiv*, 2017.

[155] Andrew W. Fitzgibbon and Robert B. Fisher. A buyer's guide to conic fitting. In *Proceedings of the 6th British Conference on Machine Vision (Vol. 2)*, BMVC '95, pages 513–522, Surrey, UK, UK, 1995. BMVA Press.

[156] C. Li, C. Xu, C. Gui, and M. D. Fox. Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing*, 19(12):3243–3254, Dec 2010.

[157] RE. Kalman. A new approach to linear filtering and prediction problems. *ASME. J. Basic Eng.*, 82(1):35–45, 1960.

[158] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[159] C. R. Maurer, Rensheng Qi, and V. Raghavan. A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):265–270, Feb 2003.

[160] Assaf Arbelle and Tammy Riklin Raviv. Microscopy cell segmentation via convolutional LSTM networks. *CoRR*, abs/1805.11247, 2018.

[161] Fausto Milletari, Nicola Rieke, Maximilian Baust, Marco Esposito, and Nassir Navab. CFCM: segmentation via coarse to fine context memory. *CoRR*, abs/1806.01413, 2018.

[162] Y. Gao, J. M. Phillips, Y. Zheng, R. Min, P. T. Fletcher, and G. Gerig. Fully convolutional structured lstm networks for joint 4d medical image segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1104–1108, April 2018.

[163] D. Zhang, I. Icke, B. Dogdas, S. Parimal, S. Sampath, J. Forbes, A. Bagchi, C. Chin, and A. Chen. A multi-level convolutional lstm model for the segmentation of left ventricle myocardium in infarcted porcine cine mr images. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 470–473, April 2018.

[164] Chaojie Zhao, Peng Zhang, Jian Zhu, Chengrui Wu, Huaimin Wang, and Kele Xu. Predicting tongue motion in unlabeled ultrasound videos using convolutional LSTM neural network. *CoRR*, abs/1902.06927, 2019.

[165] Nicolas Basty and Vicente Grau. Super resolution of cardiac cine mri sequences using deep learning. In Danail Stoyanov, Zeike Taylor, Bernhard Kainz, Gabriel Maicas, Reinhard R. Beichel, Anne Martel, Lena Maier-Hein, Kanwal Bhatia, Tom Vercauteren, Ozan Oktay, Gustavo Carneiro, Andrew P. Bradley, Jacinto Nascimento, Hang Min, Matthew S. Brown, Colin Jacobs, Bianca Lassen-Schmidt, Kensaku Mori, Jens Petersen, Raúl San José Estépar, Alexander Schmidt-Richberg, and Catarina Veiga, editors, *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 23–31, Cham, 2018. Springer International Publishing.

[166] S Jaeger, S Candemir, S Antani, Y WÃąng, P Lu, and G Thoma. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg.*, 4(6):475–477, 2014.

[167] Tejas Sudharshan Mathai, Vijay Gorantla, and John Galeotti. Segmentation of vessels in ultra high frequency ultrasound sequences using contextual memory. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, 2019.

[168] Susan M. Schultzb Lifeng Zhang Chandra M. Sehgal Emile R. Mohler III, Alexandra A. Sibleya. High-frequency ultrasound for evaluation of intimal thickness. *Journal of the American Society of Echocardiography*, 22(10):1129 – 1133, 2009.

[169] Charlotte Eklund, Peter Friberg, and Li-Ming Gan. High-resolution radial artery intima-media thickness and cardiovascular risk factors in patients with suspected coronary artery disease âĂŞ comparison with common carotid artery intima-media thickness. *Atherosclerosis*, 221(1):118 – 123, 2012.

[170] Hongming Li and Yong Fan. Non-rigid image registration using self-supervised fully convolutional networks without training data. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1075–1078, April 2018.

[171] Tanya Nair, Doina Precup, Douglas L. Arnold, and Tal Arbel. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *CoRR*, abs/1808.01200, 2018.

[172] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *CoRR*, abs/1703.04977, 2017.

[173] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. *CoRR*, abs/1703.02910, 2017.

[174] José Ignacio Orlando, Philipp Seeböck, Hrvoje Bogunovic, Sophie Klimscha, Christoph Grechenig, Sebastian M. Waldstein, Bianca S. Gerendas, and Ursula Schmidt-Erfurth. U2-net: A bayesian u-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological OCT scans. *CoRR*, abs/1901.07929, 2019.

[175] Jihang Wang, Vikas Shivaprabhu, John Galeotti, Samantha Horvath, Vijay Gorantla, and George Stetten. Towards video guidance for ultrasound, using a prior high-resolution 3d surface map of

the external anatomy. In Cristian A. Linte, Ziv Yaniv, Pascal Fallavollita, Purang Abolmaesumi, and David R. Holmes, editors, *Augmented Environments for Computer-Assisted Interventions*, pages 51–59, Cham, 2014. Springer International Publishing.