



HAL
open science

3D Modeling of the Lamina Cribrosa in OCT Data

Nan Ding

► **To cite this version:**

Nan Ding. 3D Modeling of the Lamina Cribrosa in OCT Data. Human health and pathology. Sorbonne Université, 2024. English. NNT : 2024SORUS148 . tel-04685862

HAL Id: tel-04685862

<https://theses.hal.science/tel-04685862v1>

Submitted on 3 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Sorbonne Université

École doctorale n°130 : Informatique, Télécommunications et Électronique

Laboratoire d'Informatique, Signal et Image, Electronique et Télécommunications

3D Modeling of the Lamina Cribrosa in OCT Data

Nan Ding

pour obtenir le grade de docteur de Sorbonne Université

Discipline: Informatique

Thèse dirigée par:

Mme. Florence Rossant	Professeure, ISEP
M. Michel Paques	Professeur, Hôpital des Quinze-Vingts, Sorbonne Université

Thèse encadrée par:

Mme. Hélène Urien	Professeure associée, ISEP
M. Jérémie Sublime	Professeur associé, ISEP

Rapporteurs:

M. Gwénoél Quellec	Directeur de recherche (HDR), INSERM
M. Pierre Charbonnier	Directeur de recherche (HDR), Cerema

*Examinatrice,
présidente du jury:*

Mme. Kate Grieve	Directrice de recherche (HDR), INSERM
------------------	---------------------------------------

Acknowledgements

I would like to express my sincerest gratitude to my research supervisors, Prof. Florence Rossant and Prof. Michel Paques. This work would not have been possible without their patient guidance and invaluable comments, despite their busy schedules as team leaders. I was greatly inspired by their enthusiasm for research, which motivated me to strive for improvement throughout the thesis. I would also like to express my thanks to H el ene Urien and J er emie Sublime for co-supervising my thesis. They have always been kind and supportive, I am genuinely grateful for their dedication.

I would like to extend my gratitude to my thesis committee, Prof. Gw enol e Quellec, Prof. Pierre Charbonnier, and Prof. Kate Grieve, for their constructive comments on my thesis. I am truly honored to have had them as jury members. I am deeply grateful for their insightful suggestions, the time they invested in reading my manuscript, and for coming to my defense.

I also would like to take this opportunity to thank our esteemed medical partners, the ophthalmologists at the Quinze-Vingt Hospital, the teams of Prof. Christophe Baudouin and Prof. Antoine Labb e, especially Paul Bastelica and H el ene Claudel, they have been invaluable in the process of acquiring and annotating the OCT images.

Thanks to my fellow PhD students in the lab, Abir, Guillaume, Arthur, Seoyoung, Shufan, Jade, Hongxiu, Kevin, Fahim, Dalhatu, Abdul, Tatty, Eduardo, Yaya, Xia, Dayu, Xuanbang, Xiaodong and Souha. I feel so lucky to have spent time with you all, especially the long chats we had about all sorts of things in L307 and L110. The sunsets on our walks to Porte de Versailles are such great memories!

Cheers to my friends, Zaichao, Xiaoyan, Weichao, Yingkai, Ning, Jiru, Qian, and Lu, you've been there for me through thick and thin, *la vie est belle avec vous*. A special thanks to Giuseppe, you have been such a great source of support and encouragement in all aspects of my life.

Most importantly, I would like to thank my family: my parents, my husband, and my brother for their unconditional love and support. Sheng, you were a great babysitter for Nina while I was working on my deadline. A special thanks to my uncle for pushing me to do better. Without your guidance, I would not have started my research career.

Abstract

The lamina cribrosa (LC) is a 3D collagenous mesh in the optic nerve head that plays a crucial role in the mechanisms and diagnosis of glaucoma, the second leading cause of blindness in the world. The LC is composed of so-called “pores”, namely axonal paths within the collagenous mesh, through which the axons pass to reach the brain. *In vivo* 3D observation of the LC pores is now possible thanks to advances in Optical Coherence Tomography (OCT) technology. In this study, we aim to automatically perform the 3D reconstruction of pore paths from OCT volumes, in order to study the remodeling of the lamina cribrosa during glaucoma and better understand this disease.

The limited axial resolution of conventional OCT as well as the low signal to noise ratio (SNR) poses challenges for the robust characterization of axonal paths with enough reliability, knowing that it is difficult even for experts to identify the pores in a single en-face image. To this end, our first contribution introduces an innovative method to register and fuse 2 orthogonal 3D OCT volumes in order to enhance the pores. This is, to our knowledge, the first time that orthogonal OCT volumes are jointly exploited to achieve better image quality. Experimental results demonstrate that our algorithm is robust and leads to accurate alignment.

Our second contribution presents a context-aware attention U-Net method, a deep learning approach using partial points annotation for the accurate pore segmentation in every 2D en-face image. This work is also, to the best of our knowledge, the first attempt to look into the LC pore reconstruction problem using deep learning methods. Through a comparative analysis with other state-of-the-art methods, we demonstrate the superior performance of the proposed approach.

Our robust and accurate pore registration and segmentation methods provide a solid foundation for 3D reconstruction of axonal pathways, our third contribution. We propose a pore tracking method based on a locally applied parametric active contour algorithm. Our model integrates the characteristics of low intensity and regularity of pores. Combined with the 2D segmentation maps, it enables us to reconstruct the axonal paths in 3D plane by plane. These results pave the way for the calculation of biomarkers characterizing the LC and facilitate medical interpretation.

Key words: *lamina cribrosa, OCT, 3D reconstruction, pore segmentation, 3D image registration, U-Net.*

Résumé

La lame criblée (LC), située dans la tête du nerf optique, joue un rôle crucial dans le diagnostic et l'étude du glaucome, la deuxième cause de cécité. Il s'agit d'un maillage collagénique 3D formé de pores, par lesquels les fibres nerveuses passent pour atteindre le cerveau. L'observation 3D *in vivo* des pores de la LC est désormais possible grâce aux progrès de l'imagerie de tomographie de cohérence optique (OCT). Dans cette étude, nous visons à réaliser automatiquement la reconstruction 3D des pores à partir de volumes OCT, afin d'étudier le remodelage de la LC au cours du glaucome.

La résolution limitée de l'OCT conventionnel ainsi que le faible rapport signal à bruit (SNR) posent des problèmes pour caractériser les chemins axonaux avec suffisamment de fiabilité et de précision, sachant qu'il est difficile, même pour des experts, d'identifier les pores dans une seule image en-face. Ainsi, notre première contribution est une méthode innovante de recalage et de fusion de deux volumes OCT 3D orthogonaux pour l'amélioration de la qualité d'image et le rehaussement des pores, ce qui, à notre connaissance, n'avait jamais été réalisé. Les résultats expérimentaux démontrent que notre algorithme est robuste et conduit à un alignement précis.

Notre deuxième contribution est la conception d'un réseau de neurones profond, de type attention U-net, pour segmenter les pores de la LC dans les images 2D en-face. Il s'agit de la première tentative de résolution de ce problème par apprentissage profond, les défis posés relevant de l'incomplétude des annotations pour l'apprentissage, du faible contraste et de la mauvaise résolution des pores. L'analyse comparative avec d'autres méthodes montre que notre approche conduit aux meilleurs résultats.

La fusion des volumes OCT et la segmentation des pores dans les images en-face constituent les deux étapes préliminaires à la reconstruction 3D des trajets axonaux, notre troisième contribution. Nous proposons une méthode de suivi des pores fondée sur un algorithme de contour actif paramétrique appliqué localement. Notre modèle intègre les caractéristiques de faible intensité et de régularité des pores. Combiné aux cartes de segmentation 2D, il nous permet de reconstituer plan par plan les chemins axonaux en 3D. Ces résultats ouvrent la voie au calcul de biomarqueurs et facilitent l'interprétation médicale.

Mots clés: *lame criblée, OCT, reconstruction 3D, recalage d'images 3D, segmentation des pores, U-Net, modules d'attention, annotations partielles.*

Contents

1	Introduction	1
1.1	Medical Context	2
1.1.1	Anatomy of the Eye	2
1.1.2	Lamina Cribrosa	3
1.2	Optical Coherence Tomography	4
1.2.1	Imaging Modalities in Ophthalmology	4
1.2.2	Optical Coherence Tomography	7
1.3	Research Objectives	9
1.4	Challenges and Contributions	10
1.5	Thesis Outline	11
1.6	Conclusion	12
2	Related work on OCT image analysis	13
2.1	3D Reconstruction of the Lamina Cribrosa	14
2.2	OCT Image Enhancement	15
2.2.1	Noise Reduction	16
2.2.2	Feature Enhancement	17
2.3	OCT Image Registration and Fusion	19
2.3.1	Transformation Based Registration	20
2.3.2	Feature Based Registration	21
2.4	ONH Segmentation in OCT Images	22
2.4.1	ONH Segmentation with Traditional Methods	22
2.4.2	ONH Segmentation with Deep Learning	23
2.5	Conclusion	25

3	Dataset and Preprocessing	27
3.1	3D SD-OCT Dataset	27
3.1.1	Patient Recruitment	27
3.1.2	Image Acquisition	28
3.1.3	Artifacts	30
3.1.4	Dataset	31
3.2	Pre-processing	31
3.2.1	Pore Feature Enhancement	32
3.2.2	Ground Truth Generation	34
3.2.3	Automatic Region-of-Interest Selection	36
3.2.4	Volume-of-Interest Selection	38
3.3	Conclusion	40
4	3D Orthogonal SD-OCT Volumes Registration for the Enhancement of Pores	41
4.1	Coarse Registration by Translation	42
4.1.1	2D Reference En-face Images Registration by Translation	43
4.1.2	3D Registration by Translation	45
4.2	3D Affine Transform for the Refined Registration	46
4.3	Image Fusion	48
4.4	Experiments and Results	49
4.4.1	Ground Truth Construction	49
4.4.2	Evaluation of the Registration Performance	49
4.4.3	Evaluation of the Fusion Performance	51
4.5	Conclusion and Discussion	53
5	Segmentation of Pores in OCT Volumes Using Partial Points Annotation	55
5.1	Baseline U-Net	56
5.1.1	Key Components in U-Nets	56
5.2	Proposed Method	63
5.2.1	Network Architecture	63
5.2.2	Training Dataset	66
5.3	Experiment Setup	66
5.3.1	Implementation Details	67

5.3.2	Evaluation Metrics	67
5.4	Experiment Results	68
5.4.1	Comparison of Different Dataset	68
5.4.2	Comparison of Existing Networks	69
5.5	Conclusion and Discussion	71
6	3D Reconstruction of Pore Pathways	73
6.1	Parametric Active Contours	73
6.1.1	Energy Terms	74
6.1.2	Energy Minimization	75
6.1.3	Discrete Approximation	75
6.2	Proposed Method	76
6.2.1	Energy Function Selection	76
6.2.2	Proposed Algorithm	77
6.3	Experiments and Results	79
6.3.1	Evaluation Metrics	80
6.3.2	Quantitative Evaluation	81
6.4	Conclusion	82
7	Conclusion and perspectives	83
7.1	Conclusions	83
7.1.1	Contributions	83
7.2	Perspectives	84
7.2.1	Perspectives on the Methodology	84
7.2.2	Combination with Other Modalities	85
7.2.3	Towards Clinical Applications	85
8	Résumé en Français	87
8.1	Introduction	87
8.1.1	Contexte	87
8.1.2	Défis et contributions	89
8.2	État de l'Art sur les méthodes de Reconstruction des Pores de la LC en 3D	90
8.3	Recalage des Volumes OCT 3D Orthogonaux pour le Rehaussement des Pores de la LC	91

8.4	Attention U-Net pour la Segmentation des Pores de la LC	93
8.5	Suivi en 3D des Trajets Axonaux dans la LC	95
8.6	Conclusion et perspectives	97
References		98

List of Figures

1.1	Illustration of a human eye.	2
1.2	Glaucoma vision.	4
1.3	Optic nerve imaged by different modalities.	5
1.4	Schematic of different OCT techniques.	6
1.5	3D OCT scanning system schematic centering on the optic nerve head.	8
1.6	Three successive OCT slices of the ONH.	9
2.1	State-of-the-art methods for the 3D LC pore reconstruction.	14
2.2	Main techniques for the ONH segmentation in OCT slices/volumes.	22
3.1	Image acquisition protocol.	28
3.2	An example of the horizontal and vertical images of the same eye.	29
3.3	3D OCT data and en-face image extraction.	30
3.4	Main artifact and appearance variability of en-face images of the LC.	31
3.5	Pipeline of the pre-processing step.	32
3.6	Evaluation of the proposed method to enhance pores in the en-face image.	33
3.7	Manual annotation of a pore pathway by an expert.	34
3.8	Region growing for the incomplete ground truth generation for two different images (a) and (b).	36
3.9	Automatic region of interest detection.	37
3.10	Example of the proposed VOI selection method.	39
4.1	Top-level block diagram showing the stages of the registration.	41
4.2	Modules of the coarse registration of the orthogonal horizontal and vertical volumes.	43
4.3	Reference en-face image registration.	44
4.4	Coarse translation registration results on the enhanced en-face images.	45

4.5	3D affine registration results of different X-planes.	46
4.6	Registration results of the X-plane images at each stage.	47
4.7	Evaluation of the proposed image fusion method.	48
4.8	Cross-correlation scores of 2 subjects.	51
4.9	Registration results of the reference image at each stage.	52
4.10	Fused pores using our method and the <i>GT</i>	52
4.11	Blank region produced by the 3D affine transform.	53
5.1	Baseline U-Net architecture.	57
5.2	Common activation functions.	58
5.3	Proposed context-aware attention U-Net architecture.	64
5.4	Attention gate.	65
5.5	Evaluation using the fusion method proposed in chapter 4.	68
5.6	Comparison of the segmentation results with other state-of-the-art methods.	70
6.1	Schema showing how our snake model works for pore tracking.	78
6.2	Overview of the all reconstructed pathways on the enhanced fused image.	78
6.3	Reconstruction result using our method in a high contrast case.	79
6.4	Example of pathway interruption.	81
7.1	High pore visibility in adaptive optics (Zwillinger et al., 2016).	85
8.1	Données OCT 3D pour un balayage H ou V.	88
8.2	Principales étapes du recalage 3D.	91
8.3	Exemple de recalage et de fusion sur les images rehaussées.	93
8.4	L'architecture d'attention U-Net.	94
8.5	Reconstruction 3D des trajets axonaux.	96

List of Tables

3.1	OCT volume sizes in horizontal and vertical directions.	29
3.2	Manual annotation overview*.	35
3.3	Evaluation of the proposed sub-volume selection method *.	40
4.1	Evaluation of our proposed registration method on the VOI.	50
4.2	Evaluation of our proposed registration method on different X-planes .	51
5.1	Segmentation results on the test dataset.	69
5.2	Comparison with other state-of-the-art methods on the test dataset. . .	71
5.3	<i>A posteriori</i> evaluation by the expert (10 images).	71
6.1	Evaluation of the path length using our proposed method*.	80
6.2	Evaluation of the proposed 3D reconstruction method.	80
8.1	Evaluation du recalage proposé.	92
8.2	Résultats de la segmentation sur les données de test.	95

Chapter 1

Introduction

Contents

1.1 Medical Context	2
1.1.1 Anatomy of the Eye	2
1.1.2 Lamina Cribrosa	3
1.2 Optical Coherence Tomography	4
1.2.1 Imaging Modalities in Ophthalmology	4
1.2.2 Optical Coherence Tomography	7
1.3 Research Objectives	9
1.4 Challenges and Contributions	10
1.5 Thesis Outline	11
1.6 Conclusion	12

The eye is an organ that captures and processes light waves allowing us to interact with the surroundings. In order to better understand the principle of vision, we need to know how different structures of the eye cooperate to produce a sharp image. One of the main motivations of the ophthalmology research is to figure out the mechanisms responsible for the degeneration in some pathology, in order to limit their progression.

In this chapter, we first describe the medical context of this research in section 1.1. Then, we present the basic principles of the optical coherence tomography (OCT) in sections 1.2, an imaging technique useful to observe the detailed structure of the lamina cribrosa, which is the research topic of this thesis. Section 1.3 describes the research objectives, and section 1.4 presents the challenges and contributions of this thesis. Finally, we list the thesis outline in section 1.5 and conclude this chapter.

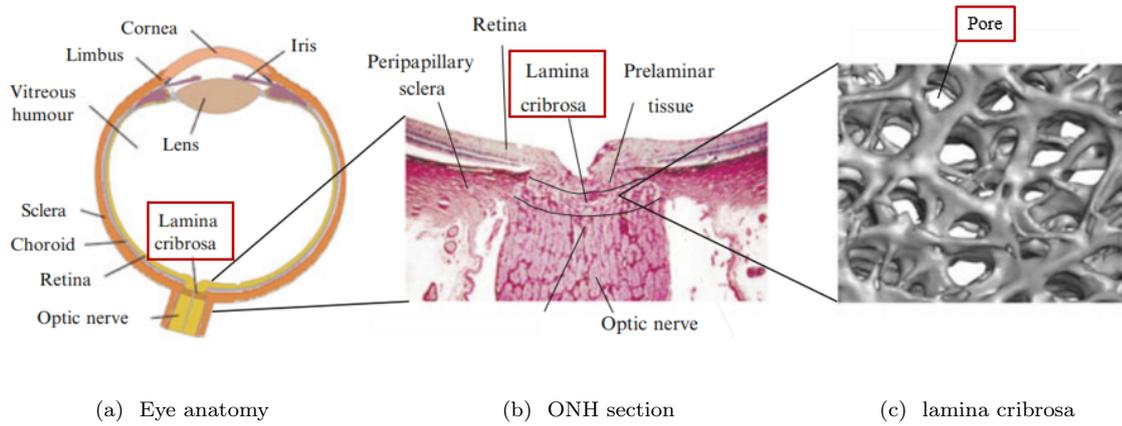


Figure 1.1: Illustration of a human eye. (a) Anatomy of the human eye. (b) Histologic section through the optic nerve head region illustrating the lamina cribrosa (LC). (c) 3D reconstruction of the beams and pores of the LC. The LC is presumed to be the principal site of damage in glaucoma. (Adapted from (Grytz et al., 2016)).

1.1 Medical Context

1.1.1 Anatomy of the Eye

Human eyeball is a cystic structure located within a protective bony cavity whose transverse diameter is up to 27 mm. Although small in size, the eye provides us with the most important of the five senses - vision.

Different structures work together step-forwardly to make a sharp vision, as shown in Figure 1.1 for the eye anatomy. Specifically, vision comes when light enters the eye via the pupil, and then the light is directed towards the lens with the help of other eye structures. Then acting as the lens in a camera, human lens refracts the light onto the retina. The retina, which is a light-sensitive structure, transforms the light into electrical energy that is sent to the optic disk, where the electrical energy will be conveyed into electrical impulses along the optic nerve to be processed by the brain.

The optic nerve may be divided into four topographic areas: **optic nerve head** (~ 1 mm), intraorbital (~ 30 mm), intracanalicular (~ 7 mm), and intracranial (~ 10 mm). The anatomy in Figure 1.1a only illustrates the optic nerve head and the intraorbital regions. In this thesis, we will focus on the optic nerve head which is the beginning of the optic nerve and the principal site of damage of many eye diseases.

1.1.2 Lamina Cribrosa

The **lamina cribrosa** (LC, as illustrated in Figure 1.1) is a mesh-like structure at the optic nerve head (ONH) that surrounds and supports the retinal ganglion cell (RGC) axons. In detail, the collagen structures of corneal scleral shell (Figure 1.1a) provide the structural integrity necessary to resist the intraocular pressure (IOP) load. At the posterior side of the scleral shell, the collagen structure is interrupted to form the scleral canal through which the RGC axons pass on their way from the retina to the brain. The tissues within in the scleral canal can be histologically divided into three regions (Figure 1.1b): in the prelaminar region, the RGC axons converge to form the optic nerve. Posterior to the prelaminar region is the LC. The LC is characterized by a porous collagenous structure (Figure 1.1c), where the RGC axons pass through the **pores** of the lamina cribrosa. The collagen architecture of the LC provides mechanical support to the RGC axons as they pass from the high pressure environment in the eye to the lower pressure environment in cerebrospinal fluid space. Posterior to the lamina cribrosa is the retrolaminar region, where the RGC axons become myelinated.

Glaucoma is an optic neuropathy that is the leading cause of irreversible, preventable blindness, affecting around 50 million people worldwide in 2010, going up to 80 millions in 2020 based on prevalence studies (Tham et al., 2014). The typical glaucomatous visual losses, shown in Figure 1.2, is called tunnel vision, where a person only sees what is directly in front of him. IOP is the most important treatable risk factor for glaucoma, but many other non-IOP risk factors have also been reported in contributing to glaucoma pathogenesis, including age (Leske et al., 2007), myopia (Suzuki et al., 2006), vascular risk factors (Grzybowski et al., 2020), and abnormalities in the LC (Faridi et al., 2014). Therefore, many researches suggest that glaucoma should be considered as a multifactorial disease. However, affected individuals are typically asymptomatic until later stages of the disease, so that a large proportion of them remain undiagnosed. Meanwhile, the glaucoma can be prevented with early diagnosis and treatment. As a consequence, the efforts to improve early diagnosis of glaucoma are essential to reduce the consequences of visual impairment and blindness.

Although the causes of glaucoma is still unclear, histological studies have identified the LC as the first site of the RGC axonal lesions in glaucoma (Quigley et al., 1981): The glaucomatous LC is subjected to compression/shear forces in response to increased IOP. This distorts the RGC axons within the LC pores. Since then, many

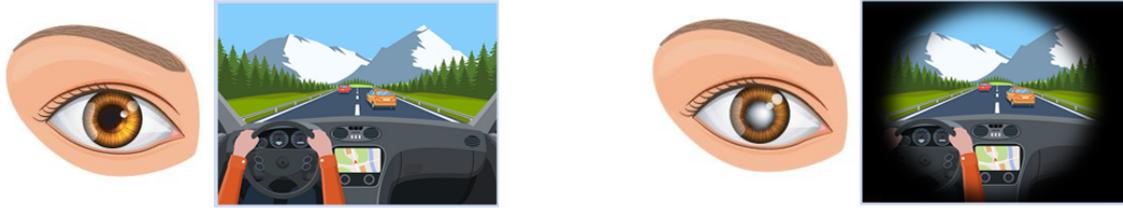


Figure 1.2: An example of normal vision (left) compared to what a person suffering from glaucoma (right) may be able to see (Wylee, 2023).

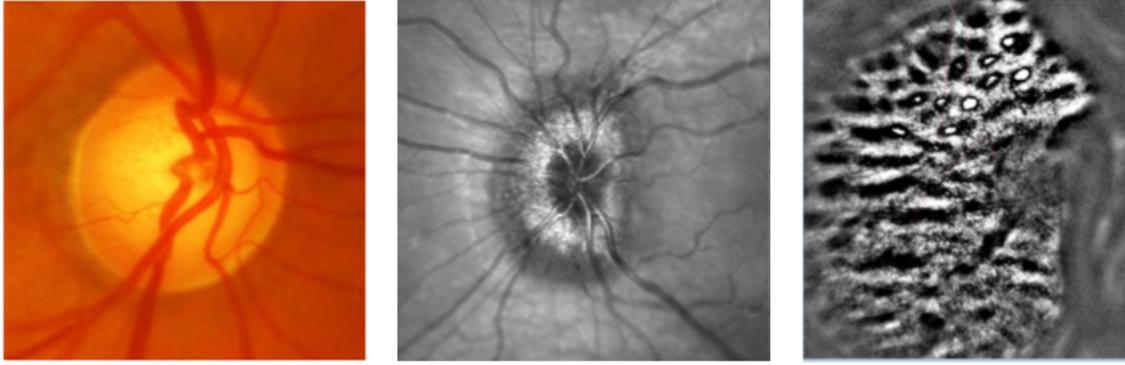
experimental studies have been conducted and demonstrated, on the one hand, the global deformations of the LC morphology such as the LC thickness (Omodaka et al., 2015), the LC curvature (Kim et al., 2016), and the anterior boundary of the LC (Tan et al., 2015). On the other hand, detailed LC pore changes like increase of area and elongation (Vilupuru et al., 2007), have also been observed. However, most of those researches are limited to analyze the LC only in 2D images of different modalities, without looking at the 3D pore paths through which the axons pass. In recent years, thanks to the progress in ophthalmic imaging technologies, especially the **optical coherence tomography** (OCT), it is now possible to image the LC pores *in vivo* in 3D, allowing the researchers to look into the deeper part of the optic nerve, and in particular the LC pores. Other recent researches also suggested that the LC pore damages may occur early, or even precede the first axonal lesions (Downs and Girkin, 2017). Therefore, it is of great interest to characterize the LC pores in 3D and assess changes over times, in order to further understand the pathogenesis of glaucoma, improve diagnosis and and develop new means of preventing or treating glaucoma.

To this end, our research work aims at segmenting the LC pores in 3D, in order to provide a 3D reconstruction of the main axonal pathways. Biomarkers will then be derived to model the LC and characterize deformations occurring over time.

1.2 Optical Coherence Tomography

1.2.1 Imaging Modalities in Ophthalmology

Ophthalmic imaging is important in clinical diagnosis and individualized treatment of eye diseases, and it has experienced progress in the past century (Kenry et al., 2018) to provide images of different eye structures with better quality, especially better resolution. Today, such imaging systems make it possible to monitor eye diseases such



(a) Fundus photography

(b) SLO

(c) AO fundus photography

Figure 1.3: Optic nerve imaged by (a) fundus camera, (b) scanning laser ophthalmoscope (SLO) and (c) Adaptive optics (AO). Adaptive optics offers the highest resolution to reveal details of the LC. (Bennett, 2020; Zwillinger et al., 2016)

as glaucoma and to assess the evolution of the pathology over time. They are of high interest for patient follow-up and clinical studies.

In practice, a number of different modalities, such as conventional fundus imaging, scanning laser ophthalmoscopy and optical coherent tomography, are used to image the human eyes. These modalities are complementary, each one has its advantages.

In this section, we present different ocular imaging techniques that have been designed to improve the visualization of ocular pathophysiology, and more specifically, the OCT imaging technique that allows us to visualize the LC pores in 3D, offering the opportunity of investigating into its morphological changes in glaucoma patients.

Fundus photography. Fundus camera is a specialized low power microscope with an attached camera introduced in the mid 19th century. It works by capturing a photography of the posterior chamber of the eye that could show the structures of the central and peripheral retina, retinal vessels, optic disc and macula (Figure 1.3a). Through the fundus photography, the appearances of retinal diseases could be visualized such as the damage to the optic nerve by glaucoma. With the technology development, fundus photography can also be performed with colored filters, or with specialized dyes.

Fundus photography provides a bird's view of the top most layer, the inner limiting membrane, as well as the vascular network. While optic nerve have different components (the LC, the prelaminar, etc.) that have specific functions in visual perception, different abnormalities often begin in a particular one before spreading into the others.

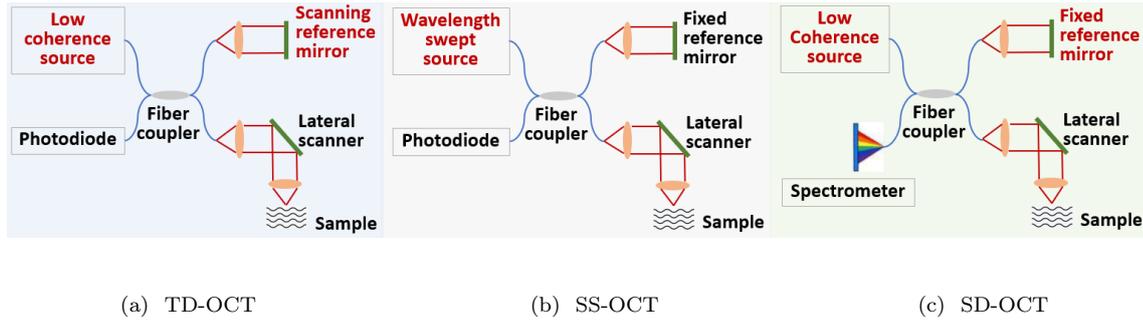


Figure 1.4: Schematic of different OCT techniques. Main differences are highlighted in red.

The fact that fundus camera could not offer specific depth examination of the optic nerve prevent it from offering early and accurate diagnosis of glaucoma.

Scanning Laser Ophthalmoscope (SLO). Scanning laser ophthalmoscope (SLO) is an eye examination method based on confocal laser scanning microscope for eye imaging. By scanning a point source rapidly across the retina with horizontal and vertical scanning mirrors, images of specific retina areas could be created, so that SLO could resolve optic nerve structure at microscopic level.

However, due to optical aberrations caused by the propagation of waves in the eye, using standard fundus photography or SLO to monitor individual cells is still problematic as the pixel resolution is too low. Therefore, to resolve cellular structure of the eye, adaptive optics (AO) techniques have been added to them to achieve better lateral resolution (Webb et al., 1987; Roorda and Williams, 1999). Theses systems are called AO-fundus photography and the AO-SLO, respectively (see Figure 1.3c). Further invention has also pushed the AO technique to achieve much higher lateral resolution (Roorda et al., 2002), larger field view (Ferguson et al., 2010), and the ability to visualize the distribution of cone photoreceptors around the fovea (Dubra et al., 2011).

Although AO techniques provide high lateral resolution with good axial resolution, the nature that the axial resolution is dependent on the numerical aperture of the eye still limits the axial sectioning ability, where only superficial layers of the LC could be assessed from a 3D reconstruction of a series of SLO images (Fitzke and Masters, 1993).

1.2.2 Optical Coherence Tomography

To overcome the mentioned disadvantages of fundus camera and SLO modalities for the imaging of the LC, another well-established imaging modality, called optical coherence tomography (OCT), is used in our research to obtain real-time 3D *in vivo* volumetric images. By measuring the magnitude and echo time delay of backscattered light, OCT devices provide micron-level resolution and millimeters of imaging depth. Compared with other imaging methods, its non-invasiveness, high resolution, high sensitivity, low cost and easy to use advantages make OCT a powerful imaging modality with applications across many clinical fields. In this section, we explained more details on OCT imaging technologies.

OCT techniques. Since the invention of OCT technique in the early 1990s (Huang et al., 1991), 2 main OCT principles have been developed to create images for the desired tissue depth range, namely the time-domain OCT (TD-OCT) and the Fourier-domain OCT (FD-OCT), as shown in Figure 1.4.

In a TD-OCT system, a reference mirror is moved mechanically to different positions, resulting in different flight time delays for the reference arm light. With the need for the scanning of reference mirror, the acquisition speed and signal-to-noise (SNR) with TD-OCT techniques is limited, that is why the FD-OCT techniques were developed with increased imaging speed and system sensitivity leading to higher SNR score.

In a FD-OCT system, axial line scans are acquired by the Fourier transform of the frequency encoded profiles principally using one of the two methods: the swept-source OCT (SS-OCT) and the spectral-domain OCT (SD-OCT). In both approaches, the backscattering signals are detected in frequency domain, which means the backscattered depth information at a given location could be collected without the movement of a reference mirror. The frequency information could be acquired with a broad-bandwidth light source in SD-OCT, or by sweeping a narrow-bandwidth light source through a broad range of frequencies in SS-OCT. A-scans (Figure 1.5) in both cases are obtained using a Fourier-transform of the detected frequencies, which facilitates rapid A-scan collection. In addition to improved speed, FD-OCT also offers the advantage of higher detection sensitivity, which means, higher SNR score.

3D OCT. The commercially available TD-OCT allows collecting of up to 400 A-scans (Figure 1.5) per second, which is quite slow and not suitable for 3D imaging,

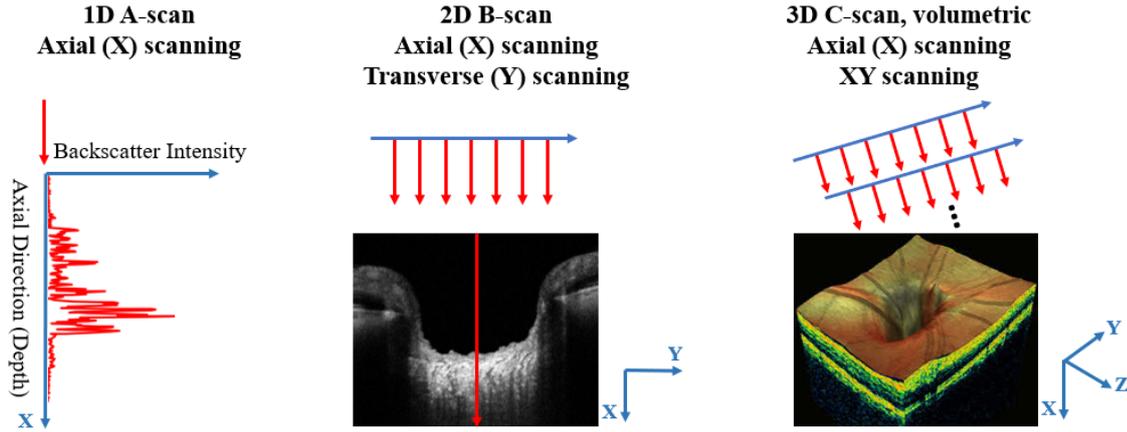


Figure 1.5: 3D OCT scanning system schematic centering on the optic nerve head.

because of motion artifacts, safety requirements limiting the amount of light that can be projected onto the retina, and patient comfort, 1–3 seconds per image or volume is essentially the limit of acceptance. However, with the improved speed of SD-OCT up to tens of thousands of A-scans per second, it is now possible to provide 2D cross-sectional (B-scan) and thus 3D volumetric internal sample structures (C-scan) by laterally scanning the optical beam and performing axial measurements of echo time delay (A-scan), as we can observe from Figure 1.5. Consequently, 3D OCT is now in wide clinical use and has become the standard of care.

EDI SD-OCT. Enhanced depth imaging SD-OCT (EDI SD-OCT) system places the objective lens of the SD-OCT closer to the eye, so that a stable inverted image is produced. The result of this practice is that the sensitivity is increased for imaging deeper layer structures. Compared with traditional OCT, this image can display the LC structure more clearly and helps to accurately reconstruct the LC in 3D (Girard et al., 2015).

Figure 1.6 shows an example of OCT slices forming the 3D OCT data. The pore pathways appear dark curves. These pathways are normally oriented vertically towards the depth of the ONH (larger X), allowing axons to pass to the brain. But one path might be less visible or seems interrupted in some slices due to the pore tortuosity, as pointed by blue arrows.

In summary, compared to other modalities, OCT allows to image the eye structures in 3D, which is indispensable for the analysis of LC pore morphological properties. Since LC is located deep in the eye and LC pores are small in size, high axial resolution

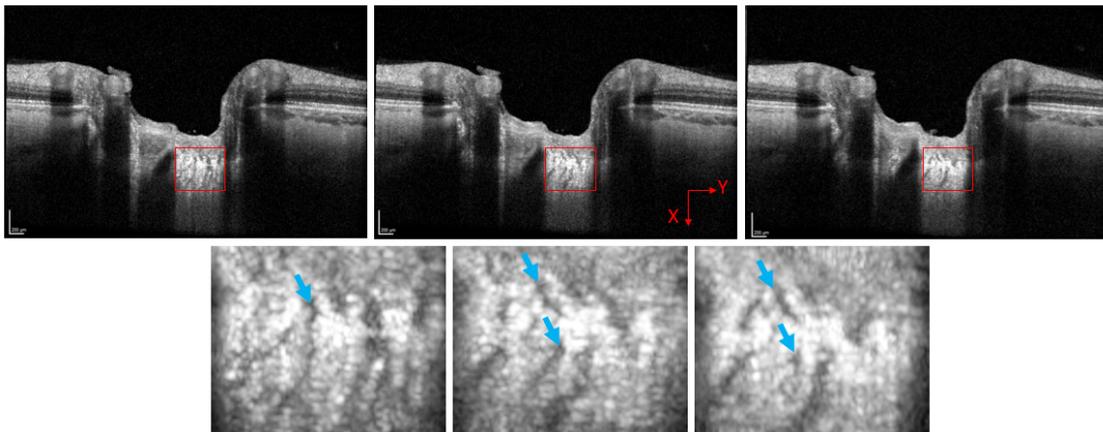


Figure 1.6: Three successive OCT slices (top row) using the Heidelberg Spectralis SD-OCT device under EDI mode. Each 3D volume consists of 131 such slices. The LC region is highlighted in a red box on the top row, and on the zoom view (bottom row). In the zoom view, dark curves correspond to parts of pore pathways. Blue arrows indicate the same pathway that might be less visible or interrupted in some slices. The axial resolution (X axis) and the lateral resolution (Y axis) are $4\mu\text{m}/\text{pixel}$ and $7\mu\text{m}/\text{pixel}$, respectively. The sampling step between two successive B-scan slice is around $15\mu\text{m}$. The X axis corresponds to the depth in the ONH.

is required for the OCT devices. To this end, the EDI SD-OCT techniques have been proposed to provide faster and more accurate scans.

1.3 Research Objectives

This research project is supported by ISEP and Institut de la Vision, an international research centre dedicated to eye diseases. At their clinical centre, several advanced SD-OCT devices have been provided to better image different eye structures, including the LC. Being able to conduct this study with the medical experts has granted us opportunities to gather insights from medical experts and finally validate our method for clinical usages.

The Heidelberg Spectralis SD-OCT device offers good LC visibility, and it is possible to scan the LC in two orthogonal directions: horizontal and vertical. The resulting (orthogonal) OCT volumes give us the idea to model the LC pores in detail, despite of the challenges such as the low signal to noise ratio (SNR) and the visibility of pores. The modeling of LC pores might allow us to calculate some biomarkers such as the tortuosity and the verticality, to eventually investigate into the pore morphological changes of glaucomatous patients.

In this thesis, we fully explore the pore information in SD-OCT C-scan volumes to reconstruct their pathways in 3D and to help the medical experts better understand the glaucoma. The goal is to design and implement an innovative 3D pore reconstruction method, capable of enhancing the pore features in a low SNR image, segmenting pores in en-face images despite partially annotated data and, finally, tracking the pore pathways based on the segmentation results.

1.4 Challenges and Contributions

The main challenge of 3D pore reconstruction is the low signal-to-noise ratio of OCT images and image defects, such as blood vessel shadowing or scanning artifacts. Also, considering the fact that the LC region is deeper than any retina layers, the attenuation in the LC region is more important than other structures. In additions, pores in the LC are small in size (compared to the voxel size) and weakly contrast to the surrounding tissue, which complicate the reconstruction problem.

In response to this, we proposed a registration algorithm to jointly exploit pore features in orthogonal OCT volumes acquired from the same examination with the same OCT device. Our experiment have shown good matches between counterpart pores in registered volumes, with high cross-correlation score. The registered orthogonal volumes are fused to get enhanced pores with better resolution.

Those contributions to the OCT volume registration method led to the following two publications:

- Nan Ding, Florence Rossant, H el ene Urien, J er emie Sublime, Michel Paques. 3D orthogonal SD-OCT volumes registration for the enhancement of pores in Lamina Cribrosa. 20th International Symposium on Biomedical Imaging (ISBI). IEEE, 2023. (Ding et al., 2023)
- Florence Rossant, Nan Ding, H el ene Urien, J er emie Sublime, Paul Bastelica, Christophe Baudouin, Michel Paques. Recalage des volumes OCT 3D orthogonaux pour le rehaussement des pores de la lame cribl ee. GRETSI, 2023. (Rossant et al., 2023)

The second challenge is to build a dataset with reliable manual annotations (gold standard) in OCT volumes. The boundary between the pores and the tissue is normally not clear, making the manual delineation difficult, especially in the borders of the optic disc, and the upper and lower limits of the LC. Moreover, pores are numerous in the en-

face images, approximating 227 ± 36 according to a *ex vivo* study (Jonas et al., 1991). As a result, due to the relatively low resolution of the OCT images, their noisy nature, and the small size and large amount of pores in the image, the manual annotation process is very challenging as even medical experts tend to disagree on what is a pore or not.

As a result, we proposed a context-aware attention U-Net to segment the pores using partial point annotation. By identifying the largest pores that allow most axons to pass, the medical experts are liberated from the heavy annotation task. Our proposed method is efficient in predicting pore candidates, thanks to the context-aware design and the attention gates integrated into the network.

Those contributions to the pore segmentation method in en-face images led to the following two publications:

- Nan Ding, H el ene Urien, Florence Rossant, J er emie Sublime, Michel Paques. Context-aware Attention U-Net for the segmentation of pores in Lamina Cribrosa using partial points annotation. 21st International Conference on Machine Learning and Applications (ICMLA). IEEE, 2022. (Ding et al., 2022b)
- Nan Ding, H el ene Urien, Florence Rossant, J er emie Sublime, Paul Bastelica, Michel Paques. Attention U-Net pour la Segmentation des Pores de la Lame Cribl ee. GRETSI, 2022. (Ding et al., 2022a)

Finally, with the robust segmentation result, we proposed a simple yet efficient active contour based tracking method to reconstruct pore paths. Our method explicitly models the main features of axonal pathways, namely the low intensity and the regularity.

Those contributions to the 3D pore reconstruction method led to the following publication:

- Nan Ding, Florence Rossant, H el ene Urien, J er emie Sublime, Paul Bastelica, Christophe Baudouin, Michel Paques. A complete method for the 3D reconstruction of axonal pathways from 2 orthogonal 3D OCT images of the Lamina Cribrosa. International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024. (Ding et al., 2024)

1.5 Thesis Outline

We organize the thesis as follows:

- Chapter 2 presents existing two works on the LC pore reconstruction. As they are not sufficient to provide an overview of the state of the art related to our work, we also review methods dedicated to OCT image analysis in general, and in 3D in particular.
- Chapter 3 first introduces the dataset used in our research project. Then it presents the proposed pre-processing method to enhance pore features and to select the optimal volume-of-interest (VOI) for further processing.
- Chapter 4 presents the proposed method to register two orthogonal VOI to further enhance pore features in a fused image.
- Chapter 5 presents the proposed context-aware attention U-Net to segment pores in every 2D en-face images, with partially annotated pores as ground truth.
- Chapter 6 presents the reconstruction of the pore pathways in 3D, with a tracking approach based on a dedicated active contour model.
- Chapter 7 concludes the thesis and offers perspectives for future work.

1.6 Conclusion

OCT is an non-invasive imaging technique that has developed rapidly over the last few decades, and now offers high resolution scans and high speed scanning rate. Thanks to recent advances in OCT imaging, such as the 3D-OCT based on the Fourier transform, it is now possible to scan the ONH and more specifically the LC in detail and in 3D.

However, the automatic reconstruction of the axonal pathways in 3D is very challenging for three main reasons: low image resolution given the pore size, poor signal to noise ratio and presence of artifacts, and the difficulty of relying on reliable and exhaustive annotated data. The proposed thesis work aims to solve these issues and propose a robust and accurate algorithm for reconstructing the main axonal pathways of the LC from two 3D OCT images acquired in two orthogonal scan directions.

Chapter 2

Related work on OCT image analysis

Contents

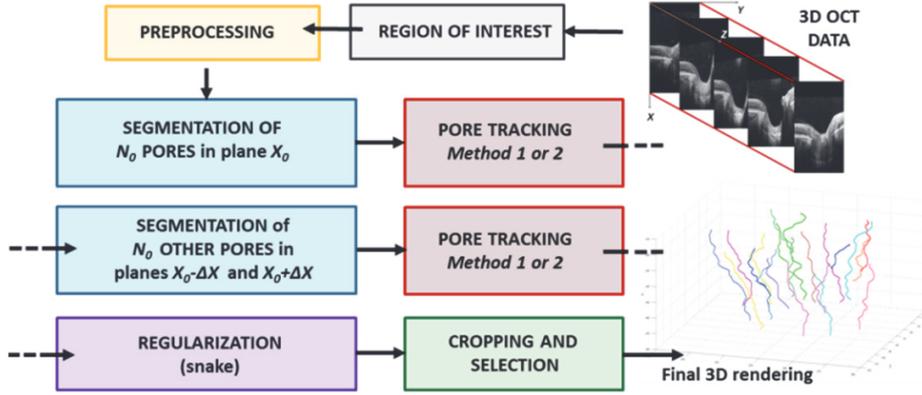
2.1	3D Reconstruction of the Lamina Cribrosa	14
2.2	OCT Image Enhancement	15
2.2.1	Noise Reduction	16
2.2.2	Feature Enhancement	17
2.3	OCT Image Registration and Fusion	19
2.3.1	Transformation Based Registration	20
2.3.2	Feature Based Registration	21
2.4	ONH Segmentation in OCT Images	22
2.4.1	ONH Segmentation with Traditional Methods	22
2.4.2	ONH Segmentation with Deep Learning	23
2.5	Conclusion	25

Since the use of OCT is relatively new in ophthalmic care compared to other modalities, processing OCT images using image analysis techniques has a shorter history. Nevertheless, it is rapidly growing especially as 3D volumetric scans of the ONH have been enabled with SD-OCT technology. As a result, it is important to develop advanced image analysis techniques for the in vivo reconstruction of the lamina cribrosa (LC) to better understand the glaucoma disease. However, few studies have addressed this reconstruction problem, mainly due to the limited visibility of the LC, not only because of the low SNR, but also the pore size of the LC being too small to be observed in some populations.

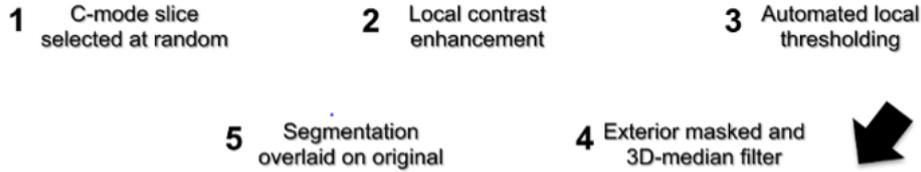
In this chapter, we first review the few existing works on the 3D reconstruction of LC pores. Then we present the state of the art methods for related but broader tasks in OCT image analysis especially for the ONH related tasks. We majorly review the

approaches related to enhancement, registration and segmentation, which can inspire us to better solve problems. Finally, we conclude this chapter by explaining the position of our work in relation to the existing literature.

2.1 3D Reconstruction of the Lamina Cribrosa



(a) (Rossant et al., 2017)



(b) (Wang et al., 2018). A particle tracking algorithm is applied after the segmentation step (5).

Figure 2.1: State-of-the-art methods for the 3D LC pore reconstruction.

The visualization of the structure of the LC with OCT has enabled researches on the global morphological modifications of the LC, such as the modified anterior surface features in 2D OCT slices (Tan et al., 2015; Kiumehr et al., 2012) and in 3D OCT volumes (Ren et al., 2014), or the decreased LC thickness in glaucomatous patients using 2D OCT images (Park et al., 2012; Devalla et al., 2018b) or 3D volumes (Omodaka et al., 2015), where the LC thickness is defined as the distance between the anterior and posterior borders of the highly reflective region. A number of groups have proposed methods for quantifying the LC pore structure alterations, such as larger and more elongated pores in glaucoma, with other image modalities, namely Adaptive Optics

(AO) (Vilupuru et al., 2007; Zwillinger et al., 2016) or standard fundus photography (Tezel et al., 2004). LC pore paths cannot be characterized with these methods that rely only on one 2D image per eye. Therefore, two groups, including ourselves, have proposed methods to model the pores of the LC in 3D and *in vivo* (Rossant et al., 2017; Wang et al., 2018) (see Figure 2.1), thanks to the progress in 3D OCT imaging techniques.

In detail, both methods could be divided into three steps: the pre-processing of 2D en-face images for pore enhancement and noise reduction, the pore segmentation step to extract pores in 2D en-face images, and finally the tracking step to link the pores in adjacent en-face images. For the pre-processing step, authors in (Wang et al., 2018) used the 3D Gaussian filter to reduce high frequency noise, and the contrast limited adaptive histogram equalization (CLAHE) (Niblack, 1985) algorithm to equalize local intensity differences. In (Rossant et al., 2017), authors proposed to use the alternating sequential filters (Gonzalez, 2009) to denoise and to enhance pore features.

As for the segmentation step, authors in (Wang et al., 2018) used the adaptive local thresholding (Niblack, 1985), as well as a 3D median filter with a kernel size of $1 \times 1 \times 3$ to take into account the continuity of pores in adjacent en-face images, while in (Rossant et al., 2017) the pores are selected as distant local minima of intensity. Finally, to individually track the segmented pores in 3D along different en-face images, authors in (Wang et al., 2018) used the particle tracking algorithm provided by MTrack2 software (Stuurman, 2003), while authors in (Rossant et al., 2017) used a Kalman filter (Welch et al., 1995) and a parametric snake model (Kass et al., 1988) to reconstruct pore pathways.

Besides these two methods, we want to explore more possibilities for the LC pore reconstruction task, especially deep learning based methods which have largely improved the performance of medical image analysis. To this end, we also review the state-of-the-art methods that might inspire us to better address the problem: pre-processing methods for OCT image enhancement and denoising, 3D-OCT volume registration and segmentation methods. The review is limited to OCT images, emphasizing the ONH analysis methods in which the LC is involved.

2.2 OCT Image Enhancement

Image enhancement is a typical pre-processing step in medical image analysis. Generic approaches for image enhancement have been employed for all eye imaging

modalities. Nonetheless, in order to address the problem pertinent only to OCT, many approaches have been proposed. Specifically, two main classes of methods for OCT image enhancement have been proposed, namely noise reduction and feature enhancement.

2.2.1 Noise Reduction

In OCT imaging, one of the broadly addressed noise is the speckle noise due to the retina reflectivity. OCT noise reduction techniques can be divided into two major groups: hardware approaches through fusion of multiple acquisitions and software techniques applied to post-acquisition data.

Hardware approaches require multiple scans of the same target which are combined to generate a fused image with improved image quality. Therefore, the compounding techniques have been widely investigated, including spatial compounding (Avanaki et al., 2013), angular compounding (Schmitt, 1997), polarization compounding (Kobayashi et al., 1991) and frequency compounding (Pircher et al., 2003) methods, mainly for speckle noise reduction in 2D OCT slices. The more images used for the compounding, the better the noise minimization and the tissue structure enhancement. Unfortunately, these methods require multiple scanning of the same targets, significantly prolonging the acquisition time, so they are not preferred for volumetric data.

Many methods of the second class have been proposed in the image processing community. The basic and commonly used methods are the convolution-based filters, and an example is the Gaussian filter that blurs and removes details. Authors in (Tan et al., 2008; Chiu et al., 2010) used the 2D Gaussian filter to denoise a single frame B-scan OCT image, while authors in (Nadler et al., 2013) chose the 3D Gaussian filter to reduce high frequency noise for C-scan OCT volumes. Due to the fact that the frequency response of most convolution-based filter are lowpass frequency filters, high frequencies in an image are attenuated after the filtering, thus the edges.

To better preserve the edges and the main structures, the well-known methods in the literature are based on the anisotropic diffusion technique (Perona and Malik, 1990). By applying Laplacian pyramid image decomposition, the frequency sub-bands of the images are obtained. In each decomposition layer, diffusion coefficients controlling the smoothness are computed separately, preserving the lines and structures. Since the invention, anisotropic diffusion technique has been widely applied for speckle noise

reduction in OCT images. For example, authors in (Wang, 2005) used the 2D image gradient to calculate the diffusion coefficient without considering the presented noise. On the contrary, authors in (Puvanathan and Bizheva, 2009) proposed to use Type II fuzzy sets to estimate the uncertainty in the diffusion process and thus to adjust the diffusion coefficient. The main problem with regular anisotropic diffusion algorithms is the large number of iterations before reaching a steady state, which implies bigger time consumption. Also the algorithm is not well performing in a low SNR image (Aja et al., 2001).

Another widely used method for denoising OCT images is wavelet thresholding (Adler et al., 2004; Hongwei et al., 2011). Spatial domain images are decomposed into wavelet coefficients by choosing an appropriate wavelet family such as the dual-tree complex wavelet transform (DT-CWT). As the noise is evenly distributed between different wavelet coefficients, while the informative content are mainly centered on several coefficients with high magnitude, the wavelet coefficients could be selected by defining an optimal threshold to reduce the noise. The resulting thresholded coefficients are finally used to reconstruct the spatial domain image.

For our dataset, the noise reduction algorithm, especially the speckle noise reduction, is not necessary because, firstly, the Heidelberg Spectralis SD-OCT device already integrated a speckle noise reduction algorithm in the latest software, secondly, if there is still too much noise in the images, we can already exclude it since in this case, the LC pores are too poorly contrasted to be identified. Moreover, if the above denoising algorithms were applied to our source dataset, it is very likely that we would damage the pore paths because of their small size and discontinuity in one single 2D B-scan image.

2.2.2 Feature Enhancement

Another image enhancement class focus on feature enhancement to get the targeted structure more contrasted to their surroundings. Authors in (Nadler et al., 2013; Xie et al., 2020) used the adaptive histogram equalization (AHE) technique to enhance the contrast in noisy OCT images, while the morphological filters have been used in (Rossant et al., 2017; Shehryar et al., 2020) to extract main features. We will go in details of the morphological filters that can denoise and enhance the features at the same time.

Gray-scale morphological filters

Mathematical morphology aims at quantitatively describing the geometrical structure of image objects (Maragos and Pessoa, 1999). They are suitable for shape analysis and play an important role in geometry-based enhancement.

We denote the gray-scale image $f(x, y)$ and the flat structuring element $b(x, y)$. The most fundamental morphological filters, namely the erosion (\ominus) and dilation (\oplus) at location (x, y) of the image f , are given as

$$[f \ominus b](x, y) = \min_{(s,t) \in b} f(x + s, y + t) \quad (2.1)$$

$$[f \oplus b](x, y) = \max_{(s,t) \in \hat{b}} f(x - s, y - t) \quad (2.2)$$

The erosion is determined by selecting the minimum value of f in the region coincident with b . Similarly, the dilation of f by b is defined as the maximum value of the image in the window spanned by \hat{b} , given that $\hat{b}(c, d) = b(-c, -d)$.

Based on the erosion and dilation operations, we can define two other important morphological filters: opening(\circ) and closing(\bullet).

$$[f \circ b] = (f \ominus b) \oplus b \quad (2.3)$$

$$[f \bullet b] = (f \oplus b) \ominus b \quad (2.4)$$

Cascading the opening and closing operations is called alternating sequential filters, where the opening–closing sequence starts with the original image, which smooth progressively from the smallest scale possible up to a maximum scale. Alternate sequential filters have a broad range of applications (Serra, 1982).

Another powerful morphological transformation is the morphological reconstruction, in which two gray-scale images and the structuring element b are involved. One image, the marker f , contains the starting points for reconstruction. The other image, the mask g , constrains the conditions for the reconstruction. Morphological reconstruction might be thought as the repeated dilations of the marker image until its contour fits under the mask image. Morphological reconstruction by dilation of a mask image g by a marker image f , denoted by $R_g^D(f)$, is defined as the geodesic dilation of f with respect to g , iterated until stability is reached.

$$R_g^D(f) = D_g^{(k)}(f), \text{ with } k \text{ such that } D_g^{(k)}(f) = D_g^{(k+1)}(f) \quad (2.5)$$

where the geodesic dilation of size k of f with respect to g is defined as

$$D_g^{(k)}(f) = D_g^{(1)}(D_g^{(k-1)}(f)), \text{ with } D_g^{(0)}(f) = f \quad (2.6)$$

and the geodesic dilation of size 1 is defined as

$$D_g^{(1)}(f) = (f \oplus b) \wedge g \quad (2.7)$$

where \wedge denotes the point-wise minimum operator, and the geodesic dilation of size 1 is obtained by computing the dilation of f by b , then selecting the minimum between the result and g at every point (x, y) . Similarly, the morphological reconstruction by erosion is defined as:

$$R_g^E(f) = E_g^{(k)}(f), \text{ with } k \text{ such that } E_g^{(k)}(f) = E_g^{(k+1)}(f) \quad (2.8)$$

$$E_g^{(k)}(f) = E_g^{(1)}(E_g^{(k-1)}(f)), \text{ with } E_g^{(0)}(f) = f \quad (2.9)$$

$$E_g^{(1)}(f) = (f \ominus b) \vee g \quad (2.10)$$

Based on those morphological, authors in (Rossant et al., 2017) used the bottom-hat filter (I_{BH}) to enhance the dark spots (pores) larger than the structuring element.

$$I_{BH} = 1 - ((I \bullet D_3) - I) \quad (2.11)$$

where D_3 is the structuring element with a disk of radius 3. They also defined a series of alternating sequential filters ($i = 1, 2, 3$) based on the reconstruction result at each step.

$$\begin{aligned} I_{open} &= R_{I_{fas}^{(i-1)}}^D(I_{fas}^{(i-1)} \circ D_i) \\ I_{fas}^i &= R_{I_{open}}^E(I_{open} \bullet D_i) \end{aligned} \quad (2.12)$$

This filter allows denoising and retaining main structures in the image larger than the structuring element (D_3), thus enhancing the pore features.

2.3 OCT Image Registration and Fusion

With the development of OCT imaging techniques, OCT image registration has become an important task. On the one hand, registration can be used for longitudinal studies by detecting the tiny but critical changes of the same structure, facilitating the disease monitoring and treatment evaluation. On the other hand, registration can be applied on images from the same examination, where images with significant overlap

could be fused into an image of better quality in term of resolution or SNR, enabling more accurate processing.

In this section, we present existing work on OCT image registration methods. These are few in number compared with other modalities such as CT or MRI, due to the fact that OCT images are more disturbed by noise, and fewer stable and precise landmarks can be identified, particularly in the case of lesions. We divide the OCT registration methods into two classes, namely the transformation based registration and the feature based registration. Both classes will be detailed in the following sections.

2.3.1 Transformation Based Registration

Transformation-based registration methods aim at maximizing the similarity between the fixed volume and the moving volume, both of which are normally obtained with the same imaging protocol, without the requirement for a specific anatomical model. Moreover, the similarity measures are generally the intensity-based metrics such as the cross correlation or the mutual information that are calculated and optimized over the volume.

Some transformation based registration methods have been implemented with unidirectional scan patterns. Authors in (Khansari et al., 2019) proposed a cross-subject 3D OCT volumes registration method to detect local retinal structural alterations due to Diabetic Retinopathy. They first used the rigid affine transformation for a global and coarse registration, refined by a B-spline non-rigid transformation to align the anatomical details. Both registration steps used the normalized mutual information as cost function, and the adaptive stochastic gradient descent for optimization. Authors in (Cheng et al., 2021) proposed a method to register successive 3D OCT volumes from the same examination for speckle noise reduction. They used a similar principle involving an affine transformation for coarse registration, followed by B-spline transformations. The registration is performed on lower resolution images by applying the Gaussian filter to speed up the registration process, and a penalty term is added on the cost function to avoid overfitting. Finally the registered volumes are fused by averaging. Both registration methods focus on the retina in OCT volumes, ignoring the details of the ONH, neither the LC.

In addition, orthogonal scan patterns have also been exploited by a team (Kraus et al., 2014; Ploner et al., 2021) for eye motion correction in 3D OCT volumes. For those methods, the registration is aiming at estimating the motion between the OCT device

and the object eye structure. To this end, the cost function is based on the residual intensity difference of neighbouring pixels, and the optimization is applied to find the displacement. To reduce the computational cost and avoid the local optimization, the source image is down-sampled by a Gaussian reduction. This process compensates for scanning artefacts and leads to a better image of the main structures. However, details in the ONH are ignored.

2.3.2 Feature Based Registration

Feature based registration methods use several important anatomical features to decide the transformation. Specifically, features could be landmark points, curves, surfaces or their combinations that are manually selected or automatically detected. For the automatic detection, accurate feature extraction algorithms have to be well designed. Feature based OCT volumes registration can be divided into 2 classes according to the nature of the features: control point based and layer based methods.

Multi-modal registration methods have been proposed to register 3D OCT images and 2D fundus images (Miri et al., 2016; Mokhtari et al., 2017; Kolar and Tasevsky, 2010), based on multiple control points, in order to achieve better vessel segmentation performance and more precise cup size estimation. In this case, OCT volumes are projected to get an en-face view. Features (control points) are derived from the retina vasculature, such as the vessel maps (Golabbakhsh et al., 2012; Kolar and Tasevsky, 2010), the branching points (Miri et al., 2016), or the center of fovea and ONH (Mokhtari et al., 2017).

In 3D OCT, alignment and flattening procedures of retinal layers have also been involved in layer segmentation methods to compensate for scanning artefacts and improve the segmentation result (Shi et al., 2014; Xiang et al., 2018). This is all the more important in pathological cases where the retinal layers may be severely structurally altered, necessitating to restore the surface smoothness to better highlight what is pathological disorganization.

Moreover, methods combining the feature based and transformation based registration have also been proposed (Pan et al., 2020, 2019) for a more robust registration.

However, according to the literature, few registration methods have investigated into the characteristics of the ONH, and to the best of our knowledge, no work has been carried out to register the orthogonal OCT volumes to enhance the LC in ONH.

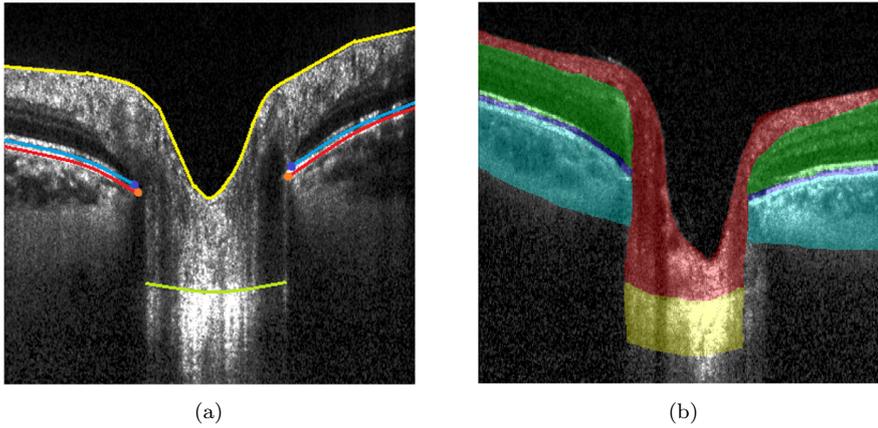


Figure 2.2: Two main techniques for the ONH segmentation in OCT slices/volumes (Marques et al., 2022). (a) Boundary-based method, where the LC anterior border is shown with a green line. (b) Region-based method, where the LC is shown in yellow.

2.4 ONH Segmentation in OCT Images

In the literature, LC segmentation is limited to the anterior or posterior borders (see Figure 2.2), thus to study the curvature or the thickness of the LC. In this chapter, we review those ONH segmentation methods to automatically segment the LC from other tissues, without investigating into the pores. Some methods require the manual delineation on B-scans (Ivers et al., 2015; Girkin et al., 2017), while the automatic methods are largely divided into two groups: the traditional methods and the deep learning algorithms.

2.4.1 ONH Segmentation with Traditional Methods

Traditional methods of the ONH segmentation techniques are divided into two groups: edge detection and active contour models.

Edge detection

Edge detection methods identify points in the image with discontinuities (edges) where the image brightness varies sharply. Authors in (Mao et al., 2019) proposed to use the customised 2D and 3D Canny edge detector to segment the anterior LC surface in 3D OCT volumes. They used 2D Canny edge detector to detect edges and, then, an additional 3D Canny edge detector helps to detect unnecessary edges from neighboring B-scan slices. To form the LC border, they select the point candidates by using the

minimum cost path approach. The cost map was accumulated by combining the edge map with the weighted intensity gradients, of which large values yield the detection of the LC border.

Another method for the LC anterior surface detection proposed by (Tan et al., 2015) operates on a single 2D OCT slice. They first isolate the vertical region of interest of the ONH, followed by an edge detector based on the local gradients of points on a seed curve. Point candidates are updated by iteration to find the best-fit curve that models the LC surface.

Active contours

In active contours methods, the targets are segmented by minimizing an energy function with contours guided by the surrounding pixels. The internal energy defines the continuity and smoothness of the contour, while the external energy is derived from the edge map. Different active contour based methods have been proposed to estimate the anterior border of the LC. Authors in (Sredar et al., 2013) modeled the 3D LC border by a plate smoothing spline using the kernel $f(r) = r^2 \log(r^2)$. The energy function to minimize is the sum of a regularization term and a bending energy term, and both terms depend on the kernel. In (Syga et al., 2018) the vertical regions of interest are extracted with Otsu thresholding (Otsu, 1979) and morphological operators. Then, the active contour is applied along three orthogonal directions of the B-scan data cuboid to perform the 3D LC surface estimation. Moreover, authors in (Belghith et al., 2015) proposed a shape-constrained surface evolution method to segment the LC anterior surface, by integrating the random Markov Field (MRF) and the non-local pixel similarity into the energy function.

2.4.2 ONH Segmentation with Deep Learning

Recently, deep learning methods, especially the U-Net network (Ronneberger et al., 2015) and its variants, have shown success in analyzing OCT images, from the segmentation of retina layers (Pekala et al., 2019) to the detection of the Bruch’s membrane opening (BMO) (Chen et al., 2019).

However, only several deep learning based segmentation methods have been proposed to investigate into the ONH, thus the LC. Authors in (Rahman et al., 2021) used 3840 2D B-scan slices acquired with Heidelberg Spectralis SD-OCT device. They proposed a 2-stage deep learning segmentation method to first detect the LC area

which is then refined by a segmentation module to segment the LC anterior surface. For the LC area detection step, YOLOv3-based Darknet-53 model was used. The fine segmentation step used the Attention U-Net model.

A major contribution of LC segmentation with deep learning is from the same author (Devalla et al., 2018a,b, 2020) by presenting different methods to segment the LC in 2D and 3D OCT images. In their initial work (Devalla et al., 2018a), the adaptive compensation algorithm is first applied on the images (Mari et al., 2013) to enhance the LC visibility, followed by an eight-layer CNN network (three convolution layers, three max-pooling layers and two fully connected layers) for the classification of different layers, where the LC is classified into the same class as the sclera. The model was trained on a small dataset (100 2D OCT images). Unsurprisingly, false predictions for the LC have been observed especially at the posterior border, which offered unreliable information on measuring the LC parameters.

This shortcoming was addressed in (Devalla et al., 2018b) with an architecture combining a 2D U-net and residual blocks. This time, the LC is separated from the other tissues of the ONH by the network. With this U-Net like network, both the contextual and local information were captured, and the residual connections help to improve the flow of the gradient information. Results showed that this new architecture performed better for the LC segmentation compared to the previous method in (Devalla et al., 2018a).

Finally, the latest work (Devalla et al., 2020) further improves the segmentation performance by introducing a deep learning network to enhance 3D OCT volumes and to harmonize image characteristics across OCT devices. A 3D U-Net based deep learning network to segment the ONH is proposed as well. The dataset for segmentation has also been enlarged with 60 volumes from 3 OCT devices. To address the LC segmentation problem, they used an ensembler to combine the predictions of three different networks for a more robust segmentation result.

It is also worth mentioning that deep learning has also shown its capacity to accurately segment small retinal fluid regions (Chen et al., 2020; Tennakoon et al., 2018) in OCT images with U-Net-based methods, making it promising that U-Net variants could perform well for the LC pore segmentation.

2.5 Conclusion

With the rapid development of OCT technologies, it is now possible to analyze the LC *in vivo*, but this remains a challenging task mainly due to the low SNR, as the LC is located deeper than any retina structure, and the small size of the LC substructures. Therefore, the existing state-of-the-art methods are mainly limited to identify LC borders, without investigating into the detailed structure of the LC.

However, analyzing the LC pore paths in detail is necessary to characterize the pore changes and to help the medical experts to better understand the glaucoma disease. To this end, more innovative and advanced algorithms should be designed to fully explore the information of pores in OCT volumes. From this literature review, we can presume that pre-processing will be necessary to improve the quality of OCT images while deep learning is promising in the segmentation task.

Chapter 3

Dataset and Preprocessing

Contents

3.1	3D SD-OCT Dataset	27
3.1.1	Patient Recruitment	27
3.1.2	Image Acquisition	28
3.1.3	Artifacts	30
3.1.4	Dataset	31
3.2	Pre-processing	31
3.2.1	Pore Feature Enhancement	32
3.2.2	Ground Truth Generation	34
3.2.3	Automatic Region-of-Interest Selection	36
3.2.4	Volume-of-Interest Selection	38
3.3	Conclusion	40

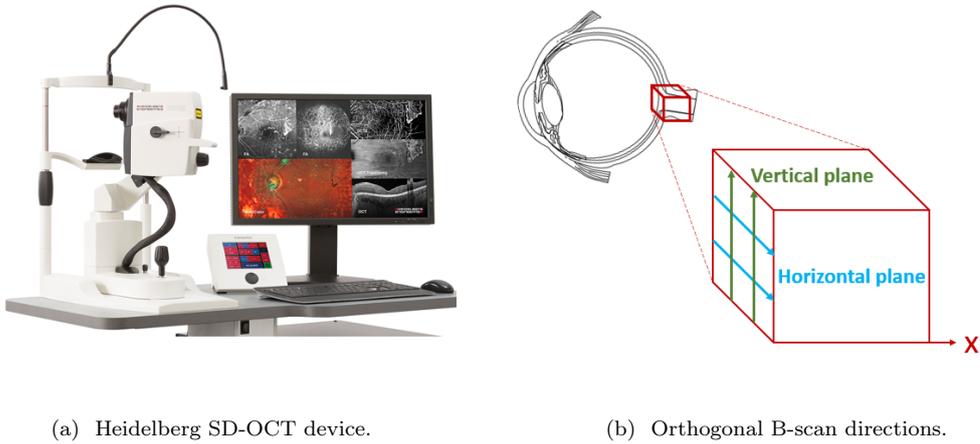
In this chapter, we first describe the dataset used in this research project, where there is no similar publicly available dataset. Then, we focus on the pre-processing steps to improve the image quality and to generate the incomplete ground truth segmentation map. Finally, we present the volume-of-interest selection algorithm, which aims at extracting the sub-volume of the lamina cribrosa (LC) containing exploitable information, to facilitate further processing steps.

3.1 3D SD-OCT Dataset

3.1.1 Patient Recruitment

This single-centre study was conducted at the Quinze-Vingts National Vision Hospital in Paris. Written consent for the re-use of each dataset was obtained from all

participants in the study. The protocol used in this project conformed to the principles of the Declaration of Helsinki. Only cooperative patients over 18 years of age with a sufficiently large optic cup to allow proper visualisation of the LC were recruited. Patients were excluded if the examination could not be performed under optimal conditions or if the quality of the OCT images was insufficient for visual localisation of the LC pores.



(a) Heidelberg SD-OCT device.

(b) Orthogonal B-scan directions.

Figure 3.1: Image acquisition protocol. Two volumetric OCT images are acquired for one eye during an examination with the Heidelberg SD-OCT device (a): the horizontal (blue) and the orthogonal vertical B-scan slices (green) centred on the ONH (red cube) in (b).

3.1.2 Image Acquisition

In this study, C-scan volumetric OCT data are acquired in both **horizontal** and **vertical** directions using the Heidelberg Spectralis SD-OCT device and the same scanning protocol in Enhanced Depth Imaging (EDI) mode (see Figure 3.1). This SD-OCT instrument has an A-scan rate of 85 kHz and a light source centred at 880 nm, allowing greater tissue penetration than conventional OCT instruments, ideal for imaging deeper ONH structures such as the LC. The axial (X axis) and lateral (Y axis) resolutions are $4\mu\text{m}/\text{pixel}$ and $7\mu\text{m}/\text{pixel}$ respectively (Figure 3.2). Thus, an acquired horizontal/vertical OCT B-scan slice of 496 (axial) \times 768 (lateral) pixels indicates that the scan has a depth of 2.0 mm and a lateral distance of 5.4 mm. Each examination centred on the ONH contains 131 such 2D B-scan slices, with a sampling step between two consecutive slices of approximately $15\ \mu\text{m}$ (Figure 3.3), giving the volume size at around $2.0 \times 5.4 \times 2.0\ \text{mm}^3$.

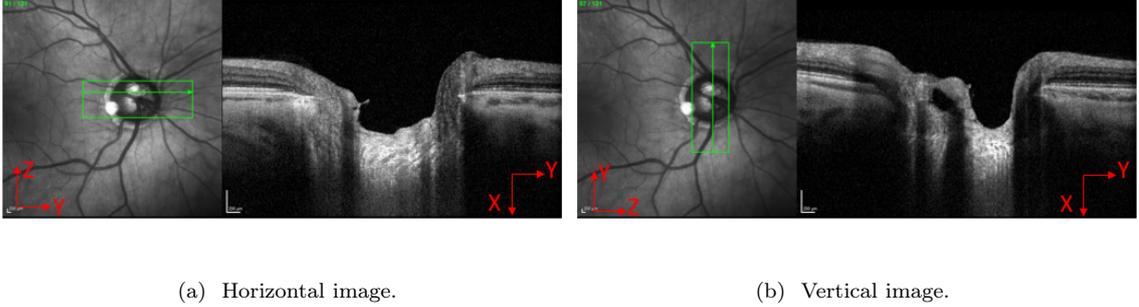


Figure 3.2: An example of the horizontal and vertical images of the same eye. The X, Y, Z coordinates are the depth, the lateral scan direction and the number of OCT slices, respectively.

Acquired volume sizes in both scan directions are shown in Table. 3.1. Resolutions in each direction are different, so that the appearance of a pore within horizontal and vertical volumes is different. In the following sections, we will refer to the horizontal and vertical B-scan slices as the horizontal and vertical volumes/images for the sake of simplicity.

Table 3.1: OCT volume sizes in horizontal and vertical directions.

Horizontal volume				Vertical volume			
Axis	Resolution	Pixels	Size	Axis	Resolution	Pixels	Size
X (axial)	4 μ m/pixel	496	2.0mm	X (axial)	4 μ m/pixel	496	2.0mm
Y (lateral)	7 μ m/pixel	768	5.4mm	Z (number)	15 μ m/scan	131	2.0mm
Z (number)	15 μ m/scan	131	2.0mm	Y (lateral)	7 μ m/pixel	768	5.4mm

A graphical interface has been developed by (Rossant et al., 2017) to navigate and visualise the volume at any given point (X,Y,Z) by displaying three orthogonal 2D planes with a given coordinate. The X, Y, Z abscissa represent the depth into the ONH, the lateral scan direction and the number of B-scans, respectively (see Figure 3.2). We are interested in the **en-face images** (also called X-plane images) extracted from the stacked B-scan slices (see Figure 3.3) because the dark spots in the en-face images correspond to pores of the LC through which the axons pass to deeper structures within the optic nerve (higher X) to join the brain.

Here the volume is resized by a factor of 2 along the z-axis to restore nearly square pixels (768 \times 262 pixels) in the en-face images (Figure 3.3c). The axonal pathways appear as black curves in the other two planes, namely the Y and Z planes, and are

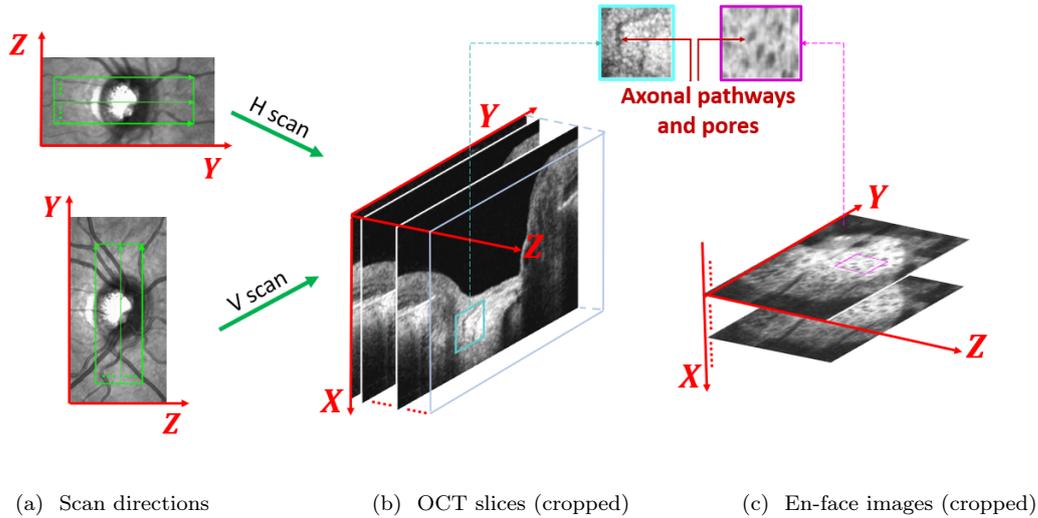


Figure 3.3: 3D OCT data and en-face image extraction. The OCT slices in (b) and the en-face image in (c) have been cropped for better visualisation of the pores. The volume size in (x,y,z) coordinates is about $1.9 \times 4.4 \times 2.0 \text{ mm}^3$.

usually interrupted in one Y or Z plane because of the tortuosity in 3D. Pores may also split or merge, but always with a tendency to higher X values.

3.1.3 Artifacts

Different types of OCT artifacts were observed in the en-face images, which could be due to errors in image acquisition and would affect the subsequent analysis. In our dataset, three main types of artifacts were identified, namely shadowing (Greig et al., 2020), poor signal (Anvari et al., 2021) and out-of-register (Chhablani et al., 2014).

The most common artifact is shadowing, which occurs when the OCT beam is blocked (mainly by the large vessels, Figure 3.4a) and cannot reach the outer LC components. This results in reduced contrast in the shadowed area as the signal reflectivity is reduced. Poor signal (Figure 3.4b) is another common type of artifact seen in OCT scans, where the images appear grainy or blurred, and the contrast between the pores and their surroundings is less clear. The poor signal can be due to patient-related factors such as dry eye, cataract and poor fixation (Anvari et al., 2021), but also to acquisition errors such as speckle noise (see section 2.2) produced by the OCT devices. Finally, the out-of-register artefact (Figure 3.4c) is also not negligible, where successive scans are shifted superiorly or inferiorly so that they cannot be fully aligned during stacking to extract the en-face images, resulting in the areas of displacement in the

rendered en-face images.

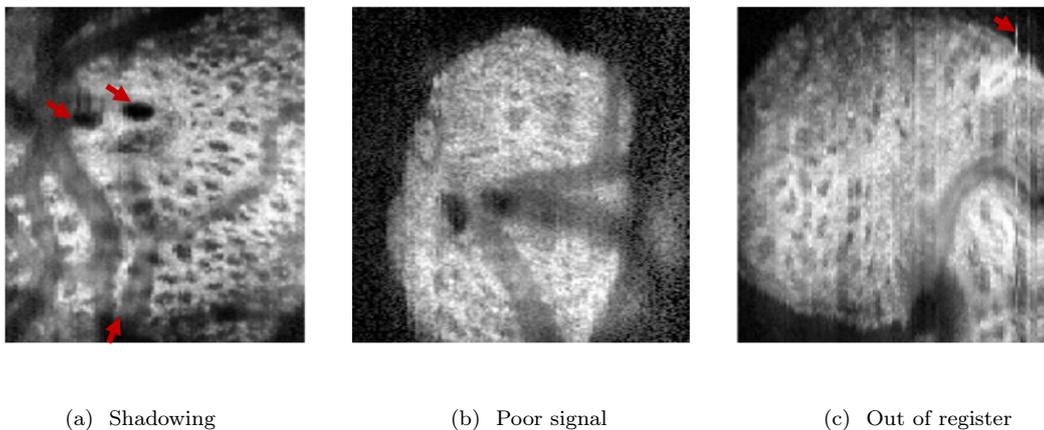


Figure 3.4: Main artifact and appearance variability of en-face images of the LC. Pores vary in shape, size and location, making them difficult to identify, especially together with various local artifacts shown as red arrows in (a) and (c), or the global artifact in (b).

3.1.4 Dataset

After exclusion of unusable volumes, our dataset consists of **158 OCT volumes**¹ in total, acquired in both horizontal and vertical directions, from 68 eyes of 42 subjects examined between 2021 and 2023 with the amount of 20,698 2D OCT slices. Multiple volumes may correspond to the same eye, either for longitudinal studies of healthy/glaucoma patients or for post-operative follow-ups.

In the following chapters we denote the grey levels of intensity (coded in the range $[0, 1]$) in horizontal and vertical images by $I_H(x, y, z)$ and $I_V(x, y, z)$, with $x \in [1, 496]$, $y \in [1, 768]$, $z \in [1, 262]$, and the corresponding en-face image extracted at coordinate X from I_H or I_V by $I_H^{(X)}(y, z)$ and $I_V^{(X)}(y, z)$ respectively. Both are also written as $I^{(X)}$ or $I^{(X)}(y, z)$ for simplicity.

3.2 Pre-processing

The pipeline of the pre-processing step is shown in Figure 3.5, where two main objectives are addressed: first, the ground truth segmentation map of each 2D en-face image is generated based on the partial annotation provided by the ophthalmologist; second, the main pore structures in the en-face images are enhanced to facilitate the

¹That's to say, 79 orthogonal horizontal and vertical volumes.

subsequent processing steps, and only the sub-volume containing the LC is retained. In the following sections of this chapter, each module of the pre-processing step will be described in detail.

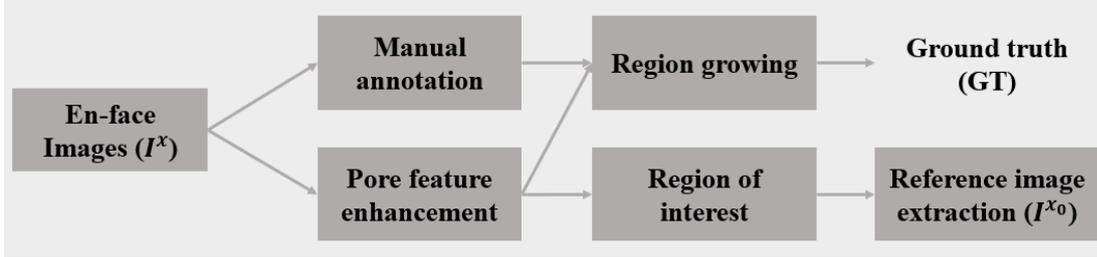


Figure 3.5: Pipeline of the pre-processing step.

3.2.1 Pore Feature Enhancement

Since pores in en-face images have an overall circular or ovoid shape and are weakly contrasted with the surrounding tissue, we applied the alternating sequential filter inspired by our team’s previous work in (Rossant et al., 2017) to enhance pore features. Similar to the definitions in section 2.2, the symbols \circ and \bullet denote the opening and closing operations, and both are constructed using a binary structuring element with a disc of radius r , denoted as D_r . $R_M^D(I)$, $R_M^E(I)$ are respectively the morphological reconstruction by dilation (\oplus) and by erosion (\ominus) of the image I in the mask M . Such a reconstruction repeats the dilation or erosion operation on the image I until the contour of I fits under the mask M .

The alternating sequential filter is denoted as $I_{ASF}^{(x)}$ (see Figure 3.6d), defined by a sequence of closings and openings with increasing size of the structuring element D_i (up to $i = 9$), with morphological reconstruction by dilation or erosion at each step i . This filter aims at denoising, while preserving the main dark structures (i.e., the pores).

$$I_{open}^{(i)(x)} = R_{I_{asf}^{(i-1)(x)}}^D(I_{asf}^{(i-1)(x)} \circ D_i), \text{ with initialization } I_{asf}^{(0)(x)} = I^{(x)} \quad (3.1)$$

$$I_{asf}^{(i)(x)} = R_{I_{open}^{(i)(x)}}^E(I_{open}^{(i)(x)} \bullet D_i) \quad (3.2)$$

$$I_{ASF}^{(x)} = \max(I_{asf}^{(9)(x)} - I^{(x)}, 0) \quad (3.3)$$

The pre-processed image is denoted as $I_{enh}(x)$ (Equation 3.4, see Figure 3.6e), where the transformation C allows to keep the mean and the standard deviation of the inten-

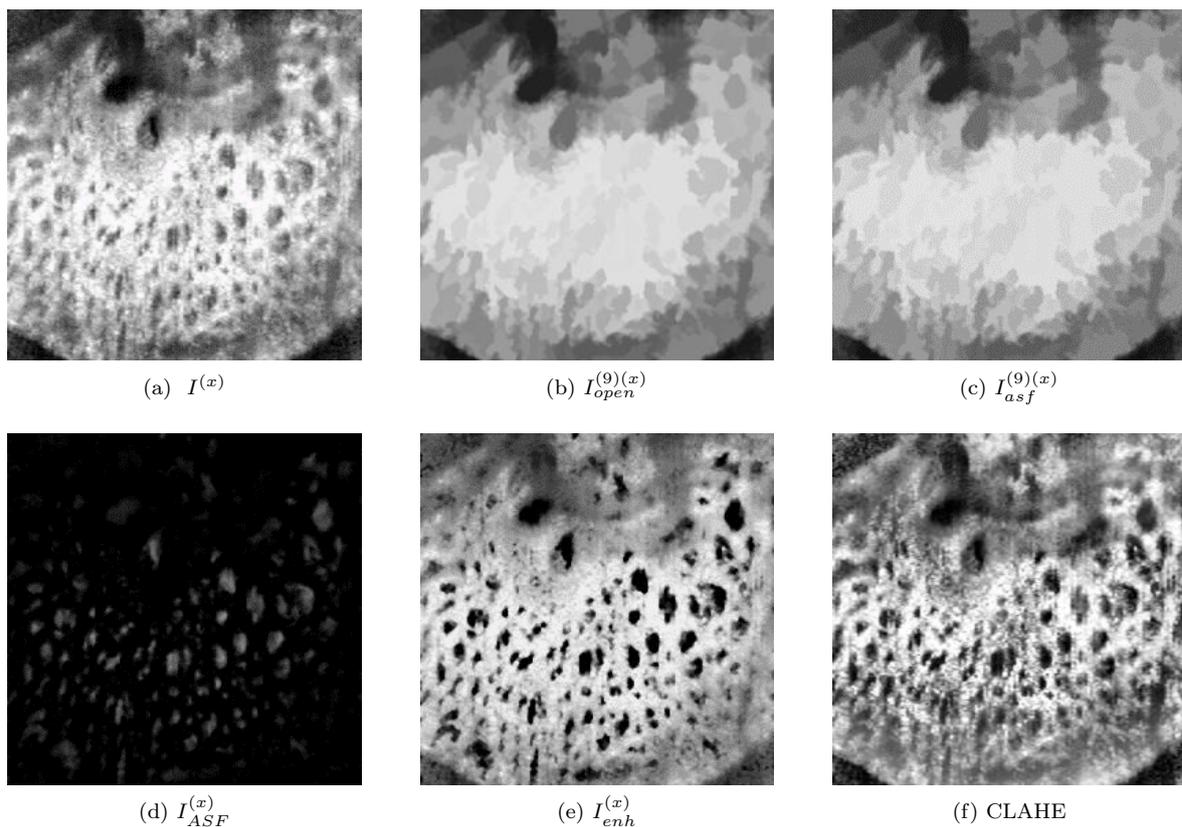


Figure 3.6: Evaluation of the proposed method to enhance pores in the en-face image (cropped for better visualisation). (a) The source OCT en-face image. (b-d) The intermediate results of the alternating sequential filter in equations (3.1), (3.2) and (3.3). (e) Enhancement result of our proposed method. (f) Enhancement result of the histogram equalisation based method (CLAHE).

sity levels in $I^{(x)}$, and the parameter $\alpha = 0.75$ allows to compromise between the two images to enhance the main pores and preserve the source image.

$$I_{enh}^{(x)} = C((1 - \alpha)I^{(x)} - \alpha I_{ASF}^{(x)}) \quad (3.4)$$

Evaluation. We subjectively compare the proposed pore enhancement algorithm with the Contrast Limited Adaptive Histogram Equalization (CLAHE) that was used in (Wang et al., 2018). Our method shows better performance compared to CLAHE, as shown in Figure 3.6. We also observe that our method preserves better the main pores without introducing additional noise compared to CLAHE, especially for the less contrasted areas.

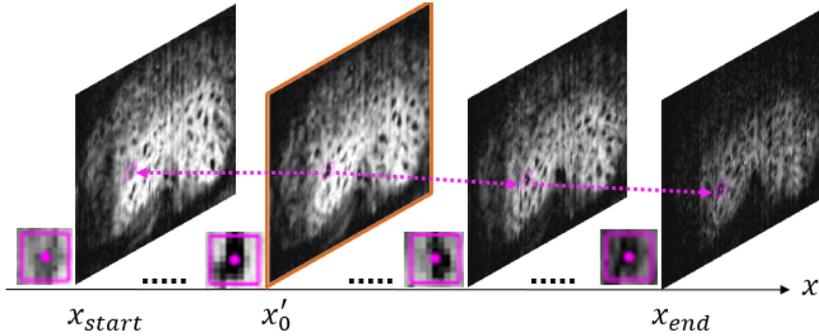


Figure 3.7: Manual annotation of a pore pathway by an expert, starting in the most contrasted plane ($I^{(x'_0)}$, in orange rectangle), and ending when the pore is no longer visible (x'_{start}, x'_{end}). The magenta rectangles (5×5 pixels) show a zoom of the annotated pores forming an axonal pathway.

3.2.2 Ground Truth Generation

As mentioned above, pores appear as dark spots in the X-plane images and also show the following regularity properties in successive X-plane images.

- Grey level intensity varies little from one X-plane image (x) to the previous/next ($x \pm 1$).
- The pore pathways are regular vertical curves, i.e. the Y and Z coordinates of a pore keep close values from one X-plane to the next.
- The axons will never stay in the same X-plane, they are globally oriented downwards into the brain (increasing X-values).

Thus, an ophthalmologist is asked to identify the main pores in the source OCT X-plane images using a semi-automated method proposed in (Rossant et al., 2017). The ophthalmologist retained only those pores that showed continuity in volume and high contrast (Figure 3.7), so the annotation is incomplete. Most pathways are initialised on a manually selected reference X-plane image (orange border image in Figure 3.7), denoted as $I^{(x'_0)}$, where the pores are most visible. The pathways continue in the upper and lower adjacent planes ($x \pm 1$), checking for continuity either with the pore centre positions suggested by a snake model, or with the subjective correction of the ophthalmologist. The pathways end when the pores can no longer be identified (x_{start} and x_{end} in Figure 3.7). Pathways could also be initialized in other X-planes if they do not pass through the reference plane.

The objective of such annotations is to identify the largest pores that allow most axons to pass through (Figure 3.8 left). Indeed, the exhaustive identification of all pores is unrealistic: they are far too numerous and most of them are poorly contrasted with the surrounding tissue, especially in regions with vascular shadows or scanning artefacts.

Finally, our dataset of 158 volumes consists of **8559** 2D en-face images containing at least one manually annotated pore, with an average of **6.73 \pm 6.03** pores identified in each en-face image, which is only a small fraction of the total number of existing pores, approximately 227 ± 36 according to a *ex vivo* study (Jonas et al., 1991). Table 3.2 shows the statistics on the annotation: we averaged the number of pore pathways for a given volume, and for all annotated pathways, we calculated the mean pathway length, and the mean number of pores in an en-face image if there is at least one. The annotation variability is mainly due to two reasons. On the one hand, the LC length and LC pore visibility are different from one subject to another, for example authors in (Dandona et al., 1990) found that blacks had a larger total lamina cribrosa area, and also a greater number of laminar pores than whites. On the other hand, the image quality may also be very different because of the artifacts, acquisition conditions, etc.

Table 3.2: Manual annotation overview*.

Mean number of paths in a volume	8.48 ± 4.92	Max	26
		Min	4
Mean path length	34.70 ± 12.03	Max	82
		Min	8
Mean number of pores in an en-face image	6.73 ± 6.03	Max	40
		Min	1
Number of annotated en-face images			8,559

*Unity: number of en-face planes.

The points marking the pore centroids could not be directly used for training since they represent only several pixels and there are too many false negatives in the image. To this end, a region growing algorithm was then applied in the enhanced en-face image $I_{enh}^{(x)}$ to obtain the binary masks by using the partial points annotated by the expert as seeds. The similarity criterion is the absolute value of the L1 distance (dst) between an unallocated pixel’s intensity value and the mean intensity of the current region.

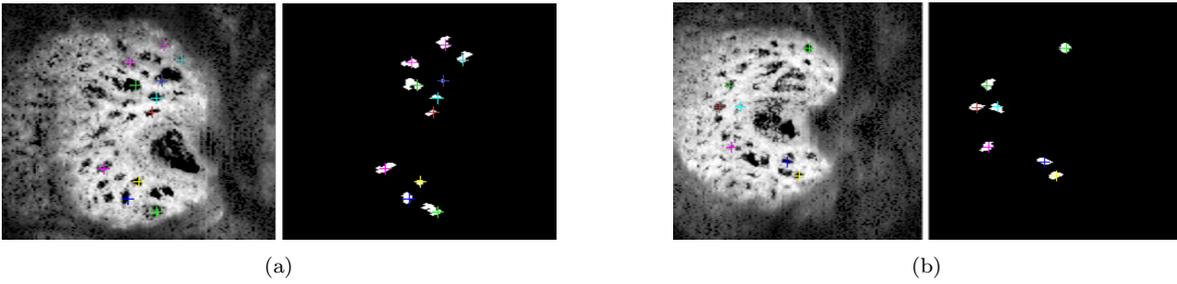


Figure 3.8: Region growing for the incomplete ground truth generation for two different images (a) and (b). (left) Manual annotated points on the source images, illustrated on the pre-processed en-face image. (right) Ground truth segmentation map used in this project by applying region growing on the pre-processed image.

Among the 8-connectivity neighbors, the pixel with the smallest dst is allocated to the region if dst is less than a given threshold $dst_{th} = 0.04$; the growing process stops when dst becomes larger than dst_{th} for all neighbors. A small threshold is applied to avoid over-segmentation since pore boundaries are barely identifiable in some cases. Examples of the region growing results are shown in Figure 3.8. The resulting binary segmentation map would serve as the ground truth for the following process.

3.2.3 Automatic Region-of-Interest Selection

The LC is observed in the central area of the ONH called the optic cup, the dark circle bounded by the green rectangle in Figure 3.2 and 3.3. On the top en-face images (small X coordinate), the cup appears as a dark connected component that can be automatically detected by thresholding. The Region-of-Interest (ROI) is defined as 262×262 pixels square bounding the cup area, and the detectable part of the LC is directly below this component, knowing that the size of an en-face image is 768×262 pixels. We denote by $ROI(I)$ the automatically detected ROI of the en-face images $I^{(x)}$ in the volume, and the detection method is described below.

First, we average the intensities of the half top sub-volume ($n = 248$) en-face images to get the dark cup area, and we denote the averaged image as $I_{avg}(y, z)$, $y \in [1, 768]$, and $z \in [1, 262]$ (Figure 3.9a). Then, the central point $P_0(y_0, 131)$ of the cup is defined where we get the minimum intensity among the points $I_{avg}(y, 131)$. Similarly, we can find the point $P(y_0, z_0)$ whose intensity is minimum along the Z-axis between the points

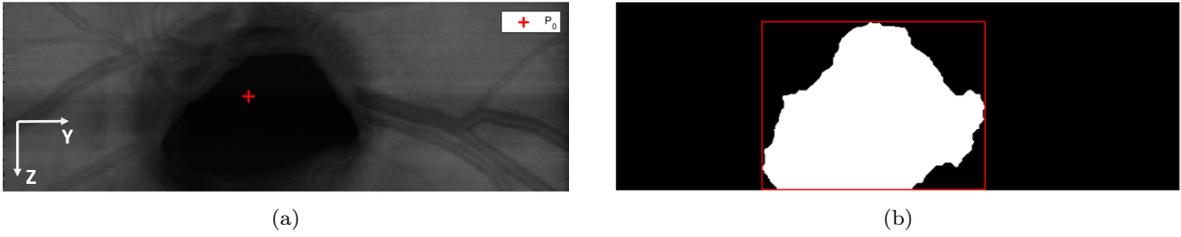


Figure 3.9: Automatic region of interest detection. (a) Averaged image of the top half sub-volume to identify the cup and the central point. (b) The automatically detected bounding box that extracts the LC area.

(y_0, z) .

$$I_{avg}(y, z) = \frac{1}{n} \sum_{x=1}^n I^{(x)}(y, z), n = \frac{496}{2} = 248$$

$$y_0 = \arg \min_{y \in [1, 768]} I_{avg}(y, 131)$$

$$z_0 = \arg \min_{z \in [1, 262]} I_{avg}(y_0, z)$$
(3.5)

We denote the intensity of the point $P(y_0, z_0)$ by s_0 , and a set of thresholds (s_i) is then defined as:

$$s_i = s_0 + \delta_i, \delta_i = 0.05 + 0.01(i - 1), i \in \{1, 2, \dots, 16\}$$
(3.6)

where s_i is used to threshold the image I_{avg} , and we denote the resulting binary image I_{s_i} .

$$I_{s_i}(y, z) = \begin{cases} 1 & \text{if } I_{avg}(y, z) < s_i \\ 0 & \text{otherwise} \end{cases}$$
(3.7)

Morphological reconstruction by dilation is then applied to a seed S_0 under the binary mask I_{s_i} , where the intensity of S_0 is zero everywhere except for the central point P_0 .

$$M_i = R_{I_{s_i}}^D(S_0)$$
(3.8)

We calculate the area (A_i) of the segmented component in M_i as well as the height (h_i) and the width (w_i) of its bounding box. A bounding box i is valid if both of the following conditions are satisfied:

$$0.75D < h_i, w_i < 1.25D, \text{ with } D = \min\{y_{max}, z_{max}\} = 262$$

$$\frac{D^2}{2} < A_i < D^2$$

We retain the index i that leads to the largest area while satisfying these conditions. The corresponding bounding box serves as the initialization of ROI(I).

An example is shown in Figure 3.9. Our algorithm is robust in identifying the region of interest, as the central point has been well selected on the average X-plane images to perform the morphological reconstruction, and this yields a single connected component. It is worth noticing that this step is an initialisation of the ROI(I), which will be further adjusted during the horizontal and vertical volume registration (Section 4).

3.2.4 Volume-of-Interest Selection

One volumetric OCT data consists of 496 en-face images, but the LC is only observable in an average of 54.17 en-face images (Table 3.3), according to the manual annotation in section 3.2.2. Therefore, an algorithm to extract the volume-of-interest (VOI) containing the LC is necessary for the optimisation of the volume size. This VOI selection method is divided into two steps: automatic reference plane selection and energy thresholding.

For the manual annotation in section 3.2.2, the reference plane was manually selected by the ophthalmologist as the en-face plane where the pores are the most contrasted. This reference plane is then used to initialize the pore pathways. Similarly, for the automatic detection, we also rely on the *a priori* knowledge of the size and shape of pores to extract the reference plane. To this end, we first apply a sliding window along the X axis and average the images $I_{ASF}^{(x)}$ on 5 consecutive en-face planes ($x \pm 2$) to get the $\bar{I}_{ASF}^{(x)}$ images with better SNR. Then, the OTSU algorithm (Otsu, 1979) is applied to get the binary image of potential pores.

$$I_{bin}^{(x)}(y, z) = \text{OTSU}(\bar{I}_{ASF}^{(x)}(y, z)) \quad (3.9)$$

Then morphological filters F_{morp} are applied to remove the connected components that may not be pores, those with an area smaller than a disk of radius 3 or larger than a disk of radius 9. The resulting image is multiplied by $I_{ASF}^{(x)}$, and finally all pixel intensity values in the current en-face image are summed up in $E(x)$ to calculate the global energy:

$$E(x) = \sum_{y,z} \{I_{ASF}^{(x)}(y, z) \cdot F_{morp}(I_{bin}^{(x)}(y, z))\} \quad (3.10)$$

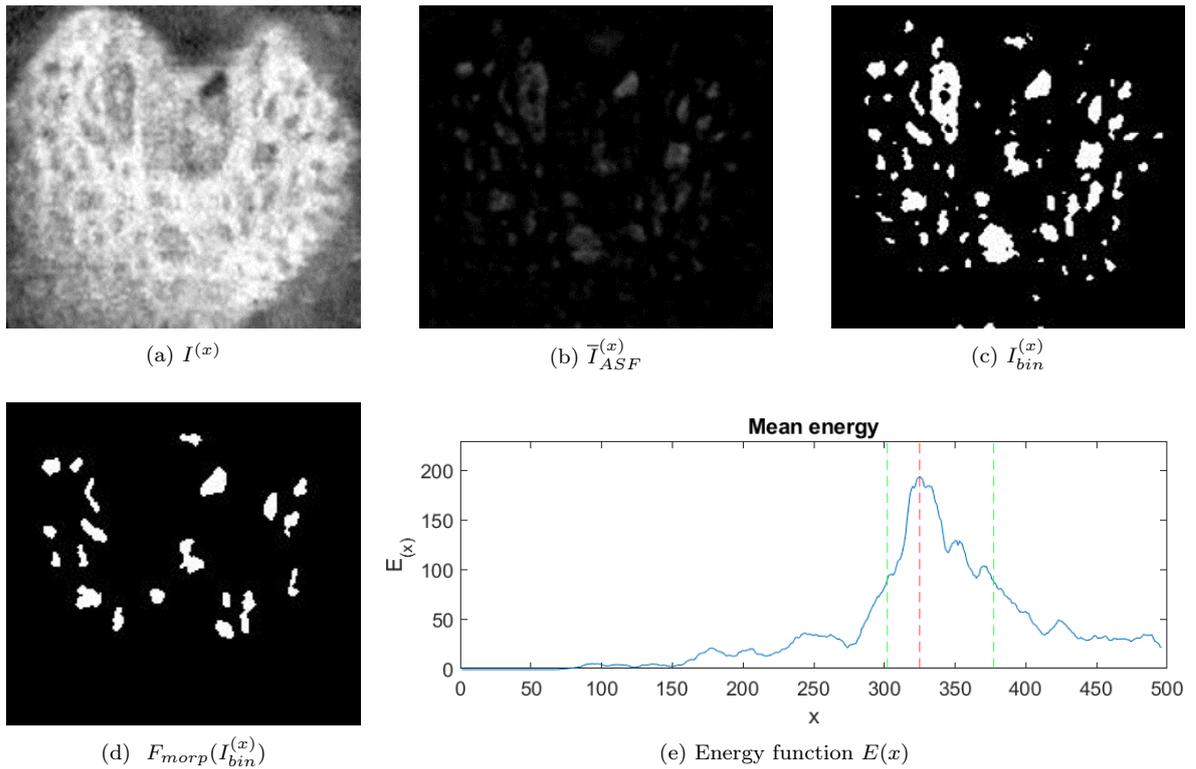


Figure 3.10: Example of the proposed VOI selection method. (a) Source en-face OCT image. (b) Averaged image $\bar{I}_{ASF}^{(x)}$ that reveals potential pores. (c) Binary image by thresholding image (b). (d) Morphological filters applied on (c) to denoise and retain main pores. (e) Energy function curve that allows to extract the VOI. The reference plane is in red, and the start/end planes in green, respectively.

The reference plane x_0 , namely x_0^H or x_0^V for the horizontal volume and vertical volume, is the one optimizing this cost function (red dotted line in Figure 3.10e). Thresholding $E(X)$ at $0.25E(x_0)$ enables us to define the X planes $[x_{start}, x_{end}]$ outside of which the pores are no longer visible (green dotted lines in Figure 3.10e).

To objectively evaluate the proposed method, the absolute difference (Diff) of the reference planes (x_0), the start (x_{start}) and the end (x_{end}) planes between the manual annotation (Manual) and the proposed method (Auto) are calculated. LC thickness ($x_{end}-x_{start}+1$) is also calculated for both methods. The result is shown in Table 3.3. Our method is accurate in predicting the reference plane (x_0), with low absolute difference and standard deviation (6.73 ± 5.94): our energy function (Figure 3.10e) correctly assesses where pores are contrasted and numerous. Our algorithm tends to select a deeper VOI than the ophthalmologist (23.98 ± 19.28). In any case, the upper and lower limits are highly ambiguous, as the pores are there hardly visible, and, at this

Table 3.3: Evaluation of the proposed sub-volume selection method *.

	Manual	Auto	Diff
x_0	316.91 ± 38.44	310.18 ± 49.80	6.93 ± 5.94
x_{start}	296.29 ± 41.82	263.15 ± 55.10	16.36 ± 15.11
x_{end}	349.47 ± 37.46	359.25 ± 61.33	17.27 ± 16.41
LC thickness	54.17 ± 13.39	76.62 ± 31.55	23.98 ± 19.28

*Unity: number of en-face planes.

stage, it is better to keep a VOI a little too deep than the opposite.

3.3 Conclusion

In this chapter, we present our dataset and the pre-processing applied to the 3D OCT images, in order to enhance the pores in the en-face images and define a volume of interest as well as a reference plane containing high-contrast pores. Image enhancement based on alternate sequential filters with morphological reconstructions proved highly effective compared with other approaches. This processing is essential for generating annotated data from pores manually pointed by the ophthalmologist, and will facilitate the following tasks. In addition, the ROI and subsequent VOI selection allow to focus on the most important area of the volume, thus speeding up the next processing.

Acquisitions of image comprise two orthogonal scan directions: horizontal or vertical. The two 3D volumes could be jointly exploited to further enhance the pore features, which is the objective of the next chapter.

Chapter 4

3D Orthogonal SD-OCT Volumes Registration for the Enhancement of Pores

Contents

4.1	Coarse Registration by Translation	42
4.1.1	2D Reference En-face Images Registration by Translation	43
4.1.2	3D Registration by Translation	45
4.2	3D Affine Transform for the Refined Registration	46
4.3	Image Fusion	48
4.4	Experiments and Results	49
4.4.1	Ground Truth Construction	49
4.4.2	Evaluation of the Registration Performance	49
4.4.3	Evaluation of the Fusion Performance	51
4.5	Conclusion and Discussion	53

In the previous chapter, we showed the pre-processing method to extract the reference planes ($I_H^{X_0}$, $I_V^{X_0}$) and to select the Volume-of-Interest (VOI) containing the LC from horizontal and orthogonal vertical volumes of the same examination. The proposed method is efficient in extracting the reference planes, close to the expert annotation.

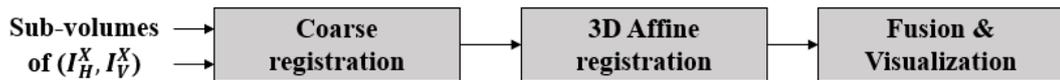


Figure 4.1: Top-level block diagram showing the stages of the registration.

In this chapter, we are interested in aligning the two orthogonal VOIs to enhance pore features in a fused image. The advantages of such fusion are twofold: on the one hand, to improve the SNR and, on the other, to achieve better pore resolution. Remember that the lateral resolution is $7\mu\text{m}/\text{pixel}$, while the sampling step between two slices is $15\mu\text{m}$ (Table 3.1). As a result, merging orthogonal slices will also improve the resolution of en-face images. However, registering the orthogonal volumes is very challenging: the SNR is very low, local shifts may occur because of scanning artefacts, the structures of interest are very small and the inter-correlation between neighbouring X-planes is high. Moreover, heterogeneous qualities of the orthogonal volumes and intensity variations increase the difficulty.

The pipeline for the registration process of the orthogonal OCT volumes is shown in Fig. 4.1. Since VOIs to be registered are acquired simultaneously with the same OCT device, linear geometric transforms, namely 2D and 3D affine transforms, have been fully exploited in this work as they are simple and commonly used with monomodal intensity-based registration problems (Zagorchev and Goshtasby, 2006). In the initial stage, the reliable information in the reference images is used to initialize the coarse registration process. Then the VOIs are further aligned by affine transformation for detailed registration of the pores. Finally the registered sub-volumes are fused to obtain an image of higher resolution and better SNR.

Given I_H , I_V the orthogonal volumes, the reference planes X_0^H and X_0^V , the ROIs in en-face images, and the vertical intervals $[X_{start}^H, X_{end}^H]$, $[X_{start}^V, X_{end}^V]$ (Section 3.2.4), we define the two sub-volumes containing the LC and we denote them by $VOI(I_H)$ and $VOI(I_V)$ ¹, respectively. Taking $VOI(I_H)$ as the fixed volume, we look for transforms that align $VOI(I_V)$ to $VOI(I_H)$.

4.1 Coarse Registration by Translation

For the initialization of the registration process, we use the **translation transformation** to get a coarse but robust alignment for the sub-volumes (VOI). Such coarse registration is divided into two steps, as shown in Figure 4.2. Horizontal and vertical reference en-face images ($I_H^{X_0}$, $I_V^{X_0}$) are first aligned in 2D, based on which the sub-volumes are then corrected and registered by translation in 3D for relocating the vertical reference X-plane image ($I_V^{X_0}$), thus $VOI(I_V)$, for better registration

¹ $VOI(I_H)$ and $VOI(I_V)$ can also be written as $VOI(ROI(I_H))$ and $VOI(ROI(I_V))$. Only the ROI of the en-face images is processed.

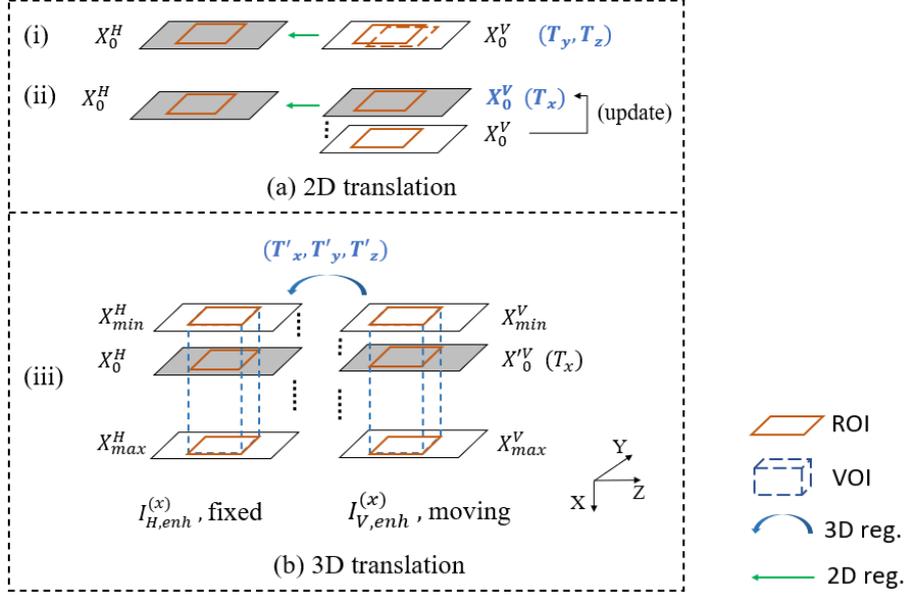


Figure 4.2: Modules of the coarse registration of the orthogonal horizontal and vertical volumes. (a) 2D translation transform to align the reference en-face images. (b) 3D translation transform to further register the sub-volumes. Main vectors to be optimized in both module are shown in blue.

performance. In the following sections, we will discuss the method in detail.

4.1.1 2D Reference En-face Images Registration by Translation

Since rich pore information could be found in reference images, both horizontal and vertical, the first step of our registration process is to align the orthogonal en-face images. As defined in chapter 3, $ROI(I_H^{X_0^H})$ and $ROI(I_V^{X_0^V})$ are the ROIs of the orthogonal reference images to be aligned.

To this end, we first average the ROI of the 3 consecutive en-face images ($X_0 \pm 1$) to get a less noisy image.

$$ROI(\bar{I}^{X_0}) = \sum_{i=-1}^{i=1} ROI(I^{X_0+i}) \quad (4.1)$$

Then the translation vector (T_y, T_z) in axes Y and Z is to be optimized by minimising the mean square error (MSE) between the two averaged images ($ROI(\bar{I}_H^{X_0^H}), ROI(\bar{I}_V^{X_0^V})$), and regular step gradient descent is used for the optimization.

Finally, X_0^V is updated by seeking the X-plane in $X_0^V \pm 50$ that maximizes the cross-correlation with $ROI(I_{H,enh}^{X_0^H})$, leading to a translation vector T_x . This two-step process, searching for the translation vector (T_y, T_z) and then for T_x , is iterated until

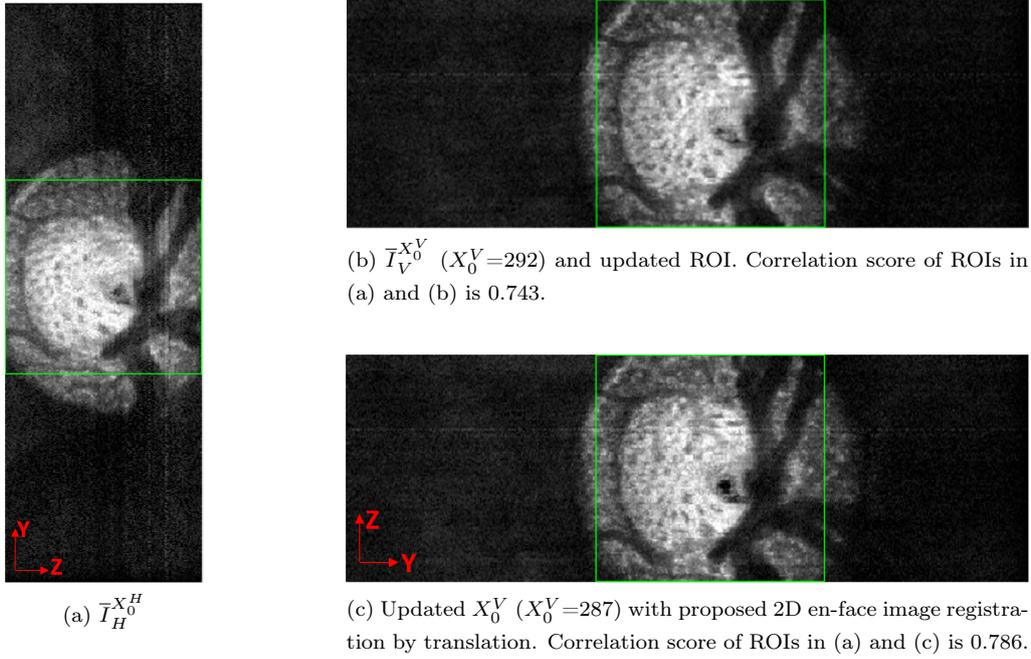


Figure 4.3: Reference en-face image registration.

the cross-correlation score cannot be improved anymore. At the end, we get a new reference plane X_0^V for I_V , and updated ROIs for I_H and I_V .

Our implementation does not only move the en-face vertical image ROI ($I_V^{X_0^V}$) in the search for the best (T_y, T_z) vector. Indeed, given the size of the horizontal and vertical en-face images, namely 768×262 and 262×768 pixels, it makes more sense to move the ROI of I_H along its own Y coordinate (Z coordinate for I_V) and the ROI of I_V along its Y coordinate. In Figure 4.3, just imagine that the green square is moved vertically in (a) and horizontally in (b). This way, the area for computing the cross-correlation is always a square of 262×262 pixels and there is no cropping.

The cross-correlation score between the ROIs of the reference planes is improved after this coarse 3D registration by translation of the reference en-face images (Figure 4.3). Based on this result, we want to further register the VOIs by translation, but this time, optimizing the cross-correlation on the VOIs, to get better global registration performance.

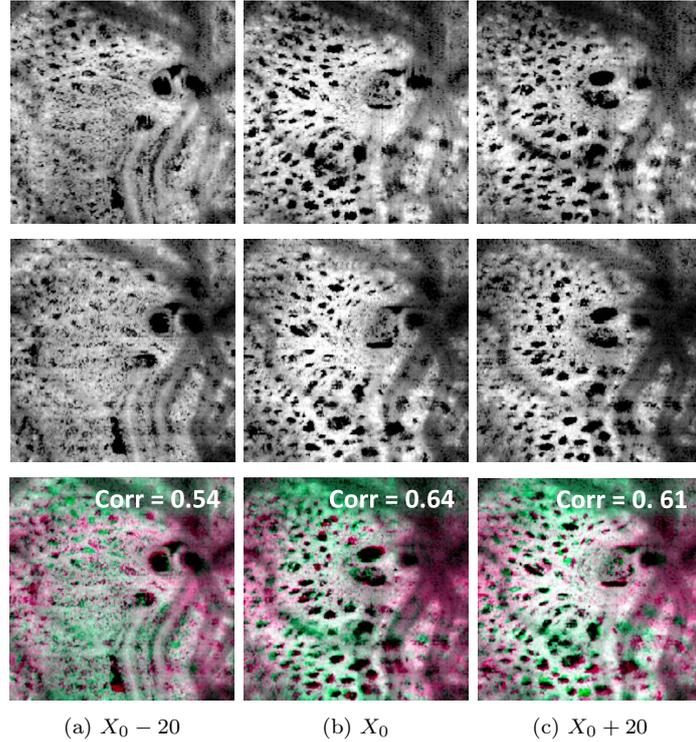


Figure 4.4: Coarse translation registration results on the enhanced en-face images. From top to bottom: ROI of the fixed horizontal planes, moving vertical planes, and the registration results.

4.1.2 3D Registration by Translation

After updating X_0^V , we can adjust the VOIs of the two volumes.

$$\delta_{up} = \min(X_0^H - X_{start}^H, X_0^V - X_{start}^V) \quad (4.2)$$

$$\delta_{dw} = \min(X_{end}^H - X_0^H, X_{end}^V - X_0^V) \quad (4.3)$$

In this way, the VOIs of the two volumes are rectangular cuboids defined vertically on $[X_0 - \delta_{up}, X_0 + \delta_{dw}]$. Such VOIs of enhanced en-face images are aligned by 3D translation to search for optimal solution on the entire VOI. We denote the three translation parameters by (T'_x, T'_y, T'_z) . We keep using the cross-correlation of pixel intensity as the similarity measure and the regular step gradient descent for the optimization for all of the following transforms, because of the statistical efficiency and computational efficiency for monomodal image registration problems (Kim and Fessler, 2004).

An example in Figure 4.4 gives the subjective evaluation of our method, where 3 planes ($X_0, X_0 \pm 20$) of the VOI before and after the coarse registration are shown.

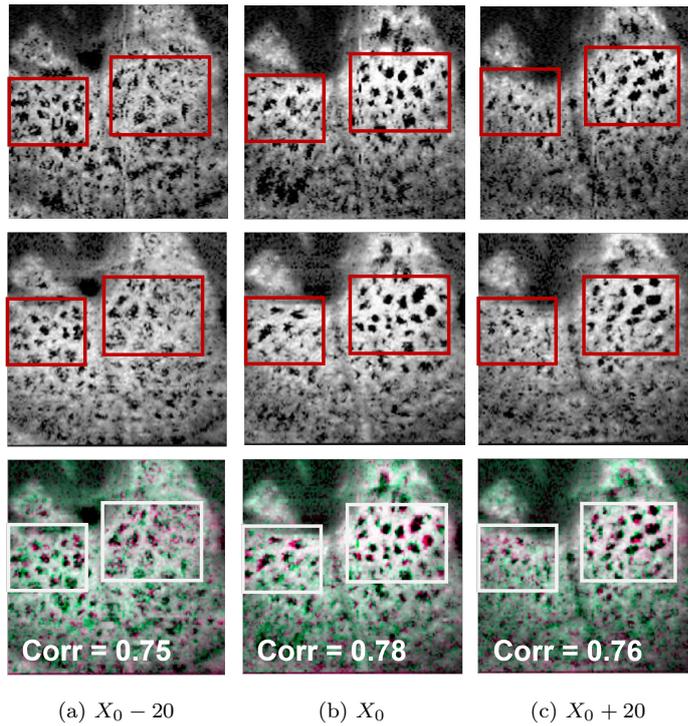


Figure 4.5: 3D affine registration results of different X-planes. From top to bottom: fixed horizontal planes, moving vertical planes, and the registration results. Cross-correlation scores after registration are shown for each plane.

We can observe that pores in high contrast images (X_0 , $X_0 + 20$) are better aligned, thanks to the pore enhancement algorithm. As X gets further away from X_0 , or the pixel reaching the border of the ROI, the registration performance is getting worse, mainly due to the fact that more noise is presented in both VOIs.

In summary, the proposed coarse registration method has proven to be robust: we first use the most reliable information, found in 2D reference images where the pores are numerous and contrasted, to approach the solution, and we refine the translation by considering a volume of interest. In this way, the problem of falling in a local minima of the cross-correlation function is overcome. Moreover, the pre-processing steps are crucial to enhance the useful pore information.

4.2 3D Affine Transform for the Refined Registration

The translation transform alone presented in former section is not enough since the optical axis may vary from one acquisition to another. In addition, pixel resolutions

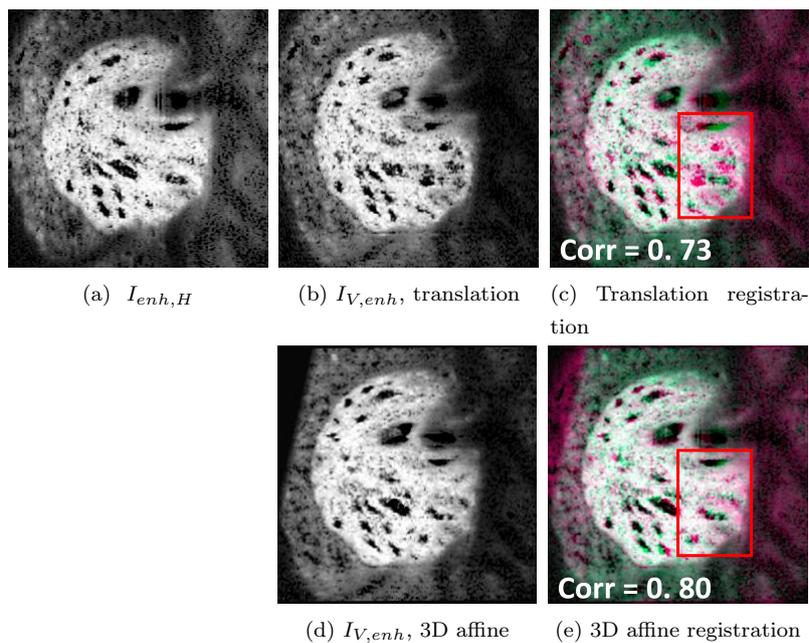


Figure 4.6: Registration results of the X-plane images at each stage. (a) Enhanced horizontal image. (b) Aligned vertical reference image by coarse registration. (c) Registration result of (b). (d) Aligned vertical reference image refined by 3D affine transform. (e) Registration result of (d). Registration performance is significantly improved after 3D affine transform, as can be observed in particular inside the red rectangle. Cross-correlation scores of X-plane images are shown as well.

are slightly different even we have applied the factor of 2 on the Z axis, and a rescaling factor has also to be applied. To address the possible rotation, scaling or other motion during the examination, a **3D affine transform**, namely the translation, scaling, rotation and shearing, is applied on the *VOI* to refine the registration. Regular step gradient descent is used for the optimization. An example in Figure 4.5 shows that in the high contrasted regions, pores are perfectly registered without any ambiguity, while in the noisy regions, there exist more doubt on whether pores exist.

This additional registration step is useful, confirmed by Figure 4.6. In this example, images are slightly deformed so that regions with ambiguity (red box) remain when limiting the transformation to a 3D translation. Such ambiguity is eliminated by the 3D affine transform. The improvement in performance can also be seen in the increase in the 2D cross-correlation score from 0.73 to 0.80.

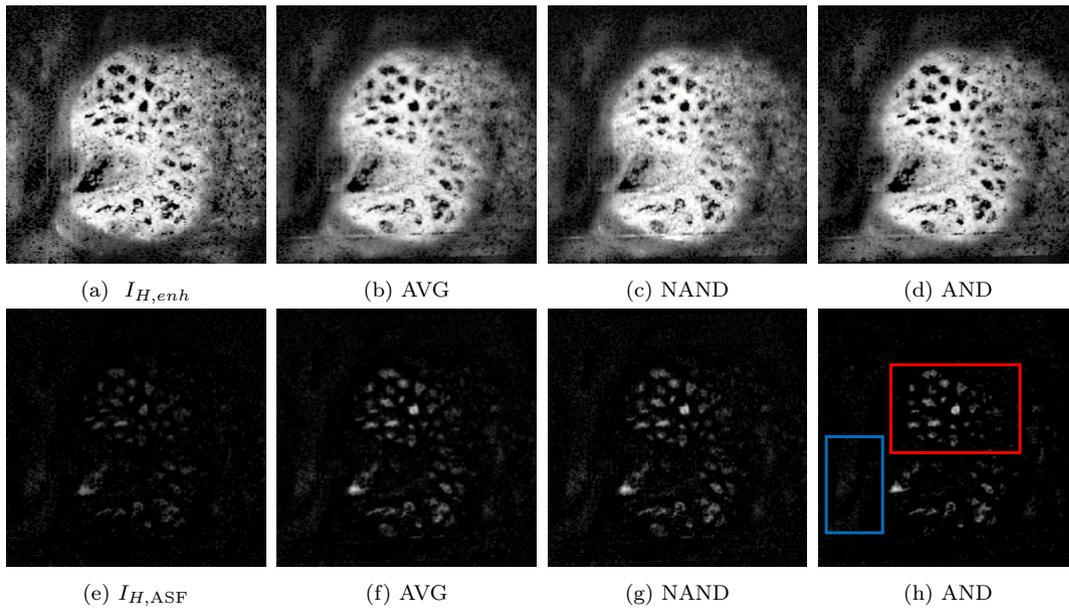


Figure 4.7: Evaluation of the proposed image fusion method. (a) Enhanced horizontal image. (b-d) Image fusion by average (AVG), not and (NAND), and and (AND), respectively. (e) Pore features extracted from (a). (f-h) Pore features with AVG, NAND, and AND, respectively. With the AND fusion method, we are able to better denoise (blue box) and retain main pores (red box) simultaneously.

4.3 Image Fusion

After 3D affine transform, pores are better aligned in the registered volumes. Now we look for an appropriate method to fuse the VOIs for better image quality. We propose to test three fundamental pixel-level fusion methods, namely the AVG, AND and NAND for the registered ROIs in two en-face images.

$$\text{AVG}(x, y) = \frac{1}{2}I_H(x, y) + \frac{1}{2}I_V(x, y) \quad (4.4)$$

$$\text{AND}(x, y) = \sqrt{I_H(x, y)I_V(x, y)} \quad (4.5)$$

$$\text{NAND}(x, y) = 1 - \sqrt{(1 - I_H(x, y))(1 - I_V(x, y))} \quad (4.6)$$

A subjective evaluation is shown in Figure 4.7. We can observe that all the three methods are efficient in isolating pores, as they are better contrasted, meanwhile less noise points are presented inside pores, with pore shapes more oval. In particular, with the AND fusion method, pores are better maintained without introducing additional noise (Figure 4.7h), so that we favor the AND fusion method for next processing steps.

Indeed, we prefer such subjective evaluation because on the one hand, the user perception is important but hard to model for the fusion performance, on the other hand, it is difficult to find a proper quantitative pore registration evaluation without ideal fused images, especially given the fact that our images are mostly of low SNR.

4.4 Experiments and Results

Experiments are carried out on Matlab, based on the image registration methods provided by the Image Processing Toolbox². In this section we detail the results obtained experimentally, and the protocol for validating our proposed method. The evaluation consists of the construction of the ground truth registration transform by aligning the manual annotation by the ophthalmologist, and the quantitative evaluation of our proposed registration and fusion method.

4.4.1 Ground Truth Construction

To be able to quantitatively assess the pore registration performance, we first construct the ground truth registration (“*GT*”) by aligning the paths that have been manually annotated by an ophthalmologist presented in section 3.2.2. In this case, the optimized criterion is the distance between the annotated paths in both volumes, rather than the cross-correlation (*Corr*) of the data. We selected 21 volumes with the most common annotated paths in both horizontal and vertical volumes to construct a robust ground truth. Finally we applied the found ground truth transform to the vertical enhanced volume, serving as registered ground truth image.

4.4.2 Evaluation of the Registration Performance

We applied the found geometric transform (“Ours”) to the annotated pores in vertical volumes and then, for each en-face image, we calculated the Euclidean distance (*MinDist*) between each pore in the horizontal volume and its counterpart in the vertical registered volume. The counterpart is defined as the nearest pore with the distance smaller than 10 if there exists. The smaller this distance, the better the alignment.

Table 4.1 displays the scores obtained on 21 volumes at each stage of the registration process (translation only, translation + affine transform). High cross-correlation scores between aligned sub-volumes are observed, with low standard-deviation (0.71 ± 0.09).

²<https://fr.mathworks.com/products/image-processing.html>

The distances between the aligned annotated pores are low (3.73 ± 0.49) compared to the typical pore sizes (from 5 to 20 pixels, average around 10), which validates the proposed method in terms of robustness and accuracy.

Table 4.1: Evaluation of our method with respect to the *GT* on the VOI with $X \in [X_{start}, X_{end}]$. “Trans.” and “Affine” refer to the performances after the translation and affine transform stages.

		<i>GT</i>		Ours	
		Trans.	Affine	Trans.	Affine
VOI	Corr	0.59 ± 0.09	0.70 ± 0.06	0.69 ± 0.11	0.71 ± 0.09
	MinDist*	3.59 ± 0.65	3.05 ± 0.73	4.82 ± 0.67	3.73 ± 0.49

*MinDist is expressed in pixels.

Then the same metrics are calculated on three planes ($X_0, X_0 \pm 20$), as shown in Table. 4.2. As expected, cross-correlation scores are the highest around the automatically selected reference images (X_0), where there is the most information with many contrasted pores. But we also obtain good matches in distant planes close to the X_{start} and X_{end} abscissas, in spite of the worse SNR. The distance between annotated pores in distant planes is higher and with higher standard-deviation, which is normal because the images are too noisy to identify the pores. However the metrics remain acceptable given the pore size, allowing for overlap in the fusion.

All results of the two tables confirm that the 3D affine transform leads to better alignment compared to the translation (lower *MinDist* and higher *Corr*), confirmed by Figure 4.8. Visually, Figure 4.9 illustrates these observations: the pores are well aligned in the fused image (see especially the areas in rectangles), better with the affine transform than only with the translation, and the green crosses are well inside the dark spots in the fused image.

Considering now the registration based on the annotated pores (“*GT*”), we also observe that the affine transform improves the registration on both criteria, cross-correlation and pore distance. The distance between the registered paths is higher with the proposed method (“Ours”) since the optimization relates to the cross-correlation of the image data. However, the averaged distances are not far from those of the ground truth (3.73 ± 0.49 against 3.05 ± 0.73) and remain below the typical pore diameter (Figure 4.10).

Table 4.2: Evaluation of our method on different X-planes.

		<i>GT</i>		Ours	
		Trans.	Affine	Trans.	Affine
X_{0-20}	Corr	0.56±0.09	0.68±0.10	0.69±0.07	0.71±0.06
	MinDist	3.99±1.03	3.20±1.18	5.08±1.43	4.22±0.77
X_0	Corr	0.63±0.06	0.71±0.07	0.72±0.08	0.78±0.06
	MinDist	3.41±0.18	2.53±0.54	3.95±1.09	3.16±0.70
X_{0+20}	Corr	0.54±0.13	0.68±0.10	0.69±0.08	0.70±0.08
	MinDist	4.13±0.82	3.63±0.99	6.42±1.51	5.12±1.23

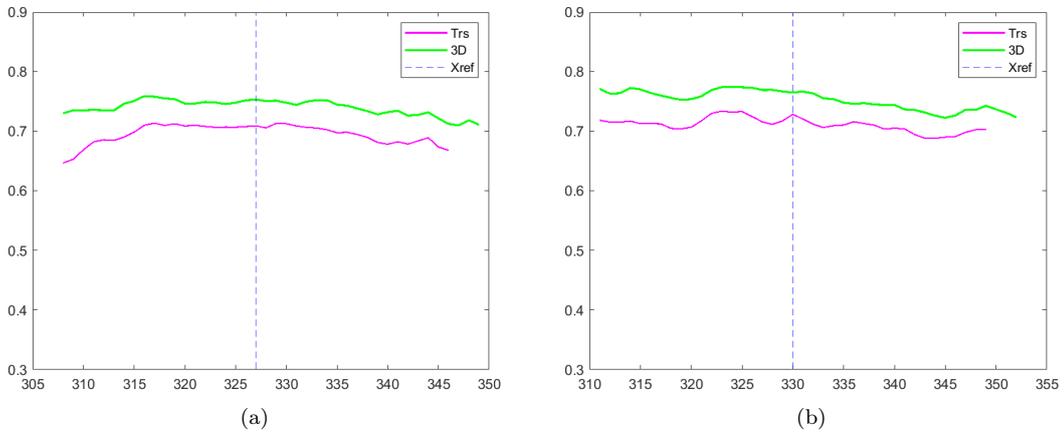


Figure 4.8: Cross-correlation scores of 2 subjects with respect to the number of X-plane images. Coarse translation registration results are shown in magenta, while in green is the final registration result.

4.4.3 Evaluation of the Fusion Performance

There are some metrics that can be used without ground-truth to evaluate the quality of the fused image. Cross-entropy (CE) is defined as the average of the relative entropies between each of the horizontal/vertical images (I_H, I_V) and the fused image (I_f).

$$CE(I_H, I_V; I_f) = \frac{D(h_{I_H} \| h_{I_f}) + D(h_{I_V} \| h_{I_f})}{2} \quad (4.7)$$

$$D(p \| q) = \sum_{i=0}^{L-1} p(i) \log_2 \frac{p(i)}{q(i)}$$

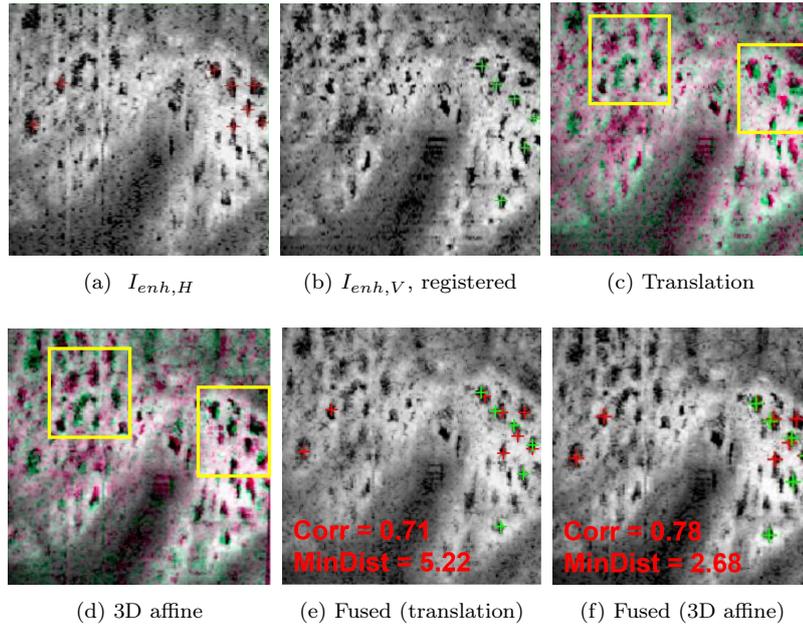


Figure 4.9: Registration results of the reference image (X_0) at each stage. (a) Enhanced horizontal reference image and with annotated pores (in red). (b) Enhanced vertical reference image with the annotated pores (in green). (c-d) Registration after the translation/the 3D affine transform. The registration performance is significantly improved after 3D affine transform, as can be observed in particular inside the yellow rectangle. (e-f) Fused image after the translation/the 3D affine transform, and the pore counterparts. Cross-correlation scores of en-face images and pore distances, are shown as well.

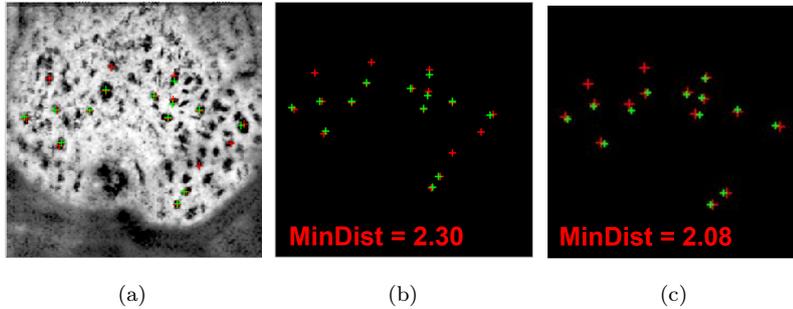


Figure 4.10: Comparison between our method and the *GT*. (a) Fused image using our method. (b) Aligned pores with our method. (c) Aligned pores with the transform found from the *GT*.

where h_{I_H} and h_{I_V} are the normalized histograms of the images I_H and I_V , respectively. The spatial frequency (SF) only operates on the fused image to get the frequency component by calculating the row frequency (RF) and column frequency (CF).

$$\begin{aligned}
SF &= \sqrt{RF^2 + CF^2} \\
RF &= \sqrt{\frac{1}{MN} \sum_{i=2}^M \sum_{j=2}^N (I_f(i, j) - I_f(i, j-1))^2} \\
CF &= \sqrt{\frac{1}{MN} \sum_{i=2}^M \sum_{j=2}^N (I_f(i, j) - I_f(i-1, j))^2}
\end{aligned} \tag{4.8}$$

Where I_f is the fused image; M, N are the width and height of the image. Low value of cross entropy means that the horizontal/vertical images contains similar information with the fused image, while high value of SF indicates the input images and fused image are similar.

We calculated the two metrics for the AND fusion method in section 4.3, on the reference planes (X_0). The averaged CE score is 0.07 ± 0.04 , and the averaged SF score is 0.28 ± 0.14 . The low CE score demonstrates that the fused image contains very similar information with the horizontal/vertical images.

4.5 Conclusion and Discussion

In this chapter, we proposed an innovative and efficient method to register the orthogonal OCT volumes, despite the difficulties of low SNR, resolution difference in three axes, and intensity inconsistencies between the orthogonal volumes. This method proved robust for pore enhancement, with high cross-correlation scores and short distance between matching pores, below the pore diameter. Better pore visibility and less point noise in the fused image offer the opportunities to segment pores with accuracy.

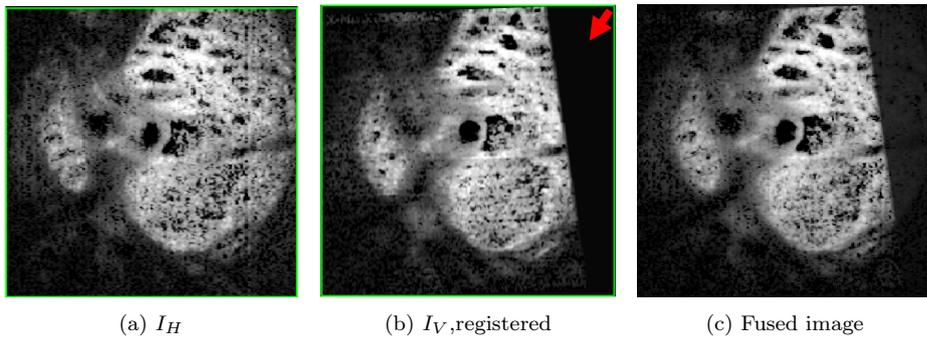


Figure 4.11: Blank region (red arrow) produced by the 3D affine transform.

One limitation of this registration method is that blank regions can appear in some cases, as illustrated in Figure 4.11. This phenomenon is caused by a significant variation in the optical axis between the two acquisitions, leading to large rotation and/or scaling factors in the final 3D affine transform. However, at current stage it is inevitable with our method, since the size of ROI square is already as large as possible (262×262 pixels). Clinically, it is possible to acquire more than 131 source OCT slices to capture a larger area around the LC, but at the cost of longer acquisition time, which might be unacceptable for some patients.

Chapter 5

Segmentation of Pores in OCT Volumes Using Partial Points Annotation

Contents

5.1	Baseline U-Net	56
5.1.1	Key Components in U-Nets	56
5.2	Proposed Method	63
5.2.1	Network Architecture	63
5.2.2	Training Dataset	66
5.3	Experiment Setup	66
5.3.1	Implementation Details	67
5.3.2	Evaluation Metrics	67
5.4	Experiment Results	68
5.4.1	Comparison of Different Dataset	68
5.4.2	Comparison of Existing Networks	69
5.5	Conclusion and Discussion	71

In this chapter, we aim to design a neural network architecture capable of accurately segmenting LC pores in OCT volumes, namely the horizontal/vertical and fused images. The network should address the problems of low image resolution, small pore size, and partial annotations in the ground truth. To the best of our knowledge, this task has never been investigated in a weakly supervised manner.

We will first give a brief literature review on U-Net and its variants (U-Nets), which have been widely adopted within the medical imaging community. We then present

the proposed weakly supervised method for pore segmentation. We describe the implementation details and finally evaluate the segmentation performance of the proposed method.

5.1 Baseline U-Net

Baseline U-Net (Ronneberger et al., 2015) is a convolutional neural network (CNN) that was designed for biomedical image segmentation using a small training dataset. Since its inception in 2015, U-Net has become the most popular network in medical imaging. Basically, U-Net (Figure 5.1) is a symmetric encoder-decoder architecture, where the encoder part (left path in Figure 5.1) is a typical fully convolutional network that aims to extract more context and complex information as the network goes deeper. The symmetric decoder part (right path in Figure 5.1), consists of up-convolutions and skip connections from high-resolution features in the encoder path, allowing a more precise localization of the target objects. The key components of U-Nets are presented in the following section.

5.1.1 Key Components in U-Nets

U-Nets share some common components with traditional CNN, such as convolution layers and pooling layers, especially at the encoder path. However, what makes them successful is the decoder part, where the skip connection is added for more precise segmentation. In this section, we will briefly present key component layers in U-Nets, which are also widely used with other neural networks.

Convolution

The convolution layers, both in the encoder and decoder paths, are the main building blocks of U-Nets. The input to this layer is a filter (kernel) and an array (2D array in U-Nets, input image or feature map from the previous layer). By applying successive convolutions between the kernel and the input array, the output is a collection of the convolution results, forming a feature map (also called an activation map) of $W \times H$.

Kernel weights are shared by all pixels in the array to reduce model complexity. They are learned during training iterations by minimizing a loss function. In general, the kernel size is smaller than the input array size, and 3 hyperparameters should be defined for the convolution operation: kernel size, stride and padding. Stride is defined as the number of rows and columns traversed per slide during the convolution, i.e.

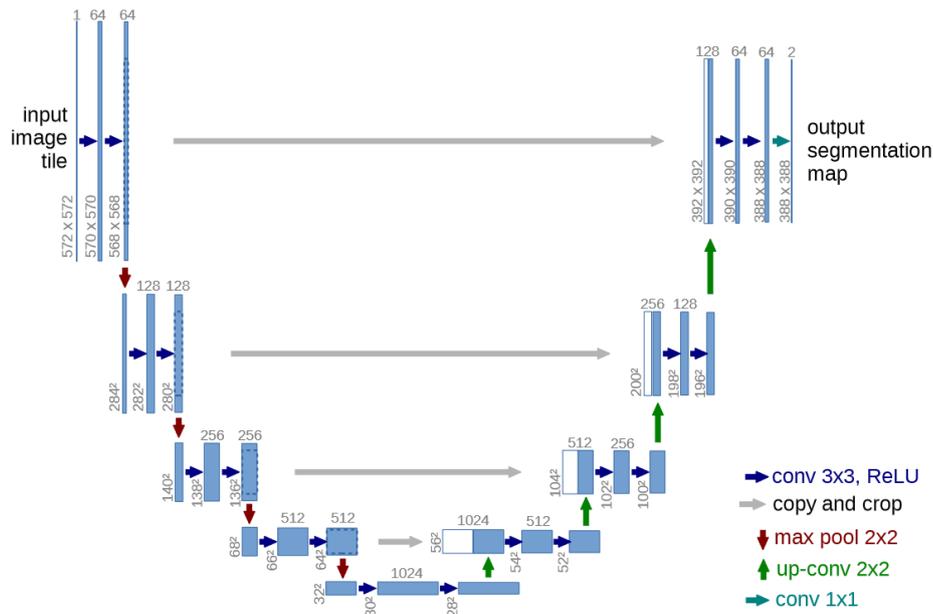


Figure 5.1: Baseline U-Net architecture (Ronneberger et al., 2015). Blue boxes represent the feature map at each layer, and white boxes correspond to the feature maps copied from the contracting path. Arrows represent different operations.

the convolution window could be moved for more than one element, skipping some intermediate locations. Padding is to add extra pixels around the boundary of the input array, thus increasing the effective size of the image. Typically, the extra pixel values are to set to zero.

One feature map of size $W \times H$ is not enough to describe the desired features, so a large number of feature maps (channel, denoted as C) are used to detect different features such as edges, textures, or higher-level features like shapes. The number of channels is also a hyperparameter: a larger number is used for the deeper layers to represent different abstract features.

Activation function

Convolutions are linear operations which are not enough to represent complex features. Therefore, activation functions are introduced to detect the nonlinear features and thus improve the CNN performance. Here, we present 3 commonly used activation functions, in the hidden layer and in the output layer.

First, the sigmoid function (Figure 5.2a) converts the input $x \in \mathbb{R}$ to an output on the interval $(0,1)$.

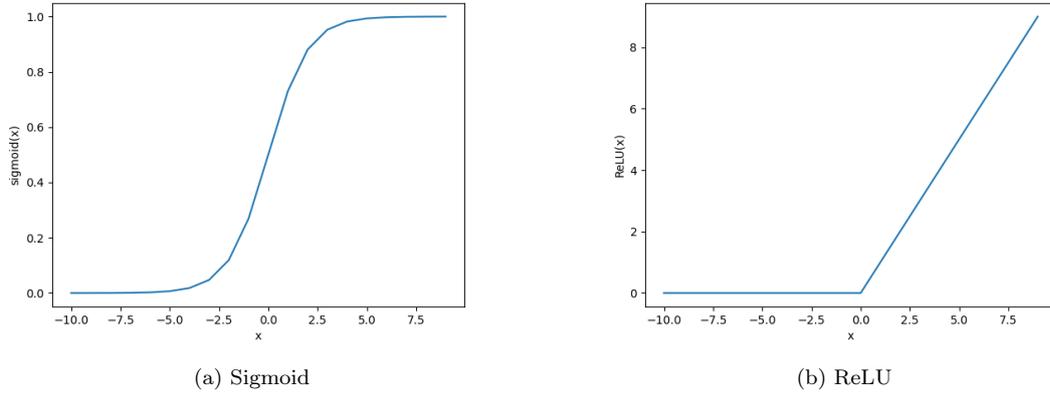


Figure 5.2: Common activation functions.

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (5.1)$$

where x is the input of the activation function. Sigmoid is a smooth, differentiable approximation to a thresholding unit. However, sigmoid poses challenges for optimization: the gradient vanishes for large positive and negative input.

Then, the most popular choice is the rectified linear unit (ReLU, see Figure 5.2b), due to its simplicity of implementation.

$$\text{ReLU}(x) = \max(x, 0) \quad (5.2)$$

The reason for using ReLU is that its derivatives are particularly well behaved: they either vanish or just let the argument through, which makes optimization better behaved and mitigates the problem of vanishing gradients.

Finally, softmax is often used for the output layer, with the advantage that the values of the output layers sum to 1.

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad (5.3)$$

Downsampling and upsampling

Let us take the baseline U-Net architecture as an example: after the convolutional and the activation function (ReLU) layers at the encoder path, the output feature map size is almost the same as the input. Therefore, in order to represent abstract features,

the spatial dimensions of the feature maps should be reduced for a larger reception field. Such an operation is called downsampling, and pooling is one of the most widely used downsampling methods.

Pooling layers divide the input array into small regions (reception fields) and perform a non-linear aggregation operation, such as the maximum (max pooling) or mean (mean pooling) value of the region within the reception field. The aggregation operations are predefined (no weights to train) and reduce the amount of data to learn for deeper layers. At the same time the reception field is enlarged because information from local regions is aggregated.

Symmetrically, in the decoder path, the operation to increase the spatial dimensions of intermediate feature maps is called upsampling. One of the most commonly used upsampling operations is the transposed convolution (up-conv in U-Net). Padding and stride parameters can also be defined for the transposed convolution, but unlike convolution, the numbers are applied to the output array.

Skip connection

One of the most successful innovations of U-Nets are the well-known skip connections (copy and crop in Figure 5.1). Although upsampling has increased the image dimensions, it is still difficult to recover the spatial information lost during downsampling, which is important in U-Nets for pixel-wise segmentation. Furthermore, skip connections provide an alternative way to ensure the reusability of features with the same dimensions from previous layers in the encoder path.

Loss function

Loss function compares the target and predicted output values. It is one of the most important elements of a CNN, as it directly measures how well the model fits the training data. We will go over the most commonly used loss functions under a binary classification formulation (background vs. foreground), as this represents image segmentation at the pixel level.

Consider now two segmentation maps: background and foreground. Let R be the target foreground segmentation (ground truth) of an image of N pixels, and the target value of pixel n in R be r_n . $r_n = 1$ (or 0) indicates that the pixel n is classified as foreground (or background). In the target background segmentation, the target background value of pixel n is $1 - r_n$. P is the predicted probability map for the

foreground label. $p_n = [0, 1]$ is the predicted probability for the foreground label of pixel n in P . The background class probability is $1 - p_n$.

The first family of loss functions is based on cross-entropy (CE), a measure of the difference between two probability distributions for random variables. CE-based loss functions are widely used for classification and segmentation tasks. The two most popular are the binary Cross-Entropy (BCE) and the weighted Cross-Entropy (WCE).

$$\text{BCE} = -\frac{1}{N} \sum_{n=1}^N r_n \log(p_n) + (1 - r_n) \log(1 - p_n) \quad (5.4)$$

$$\text{WCE} = -\frac{1}{N} \sum_{n=1}^N \omega r_n \log(p_n) + (1 - r_n) \log(1 - p_n) \quad (5.5)$$

$$\text{with } r_n = \begin{cases} 1 & \text{if } n \in \text{foreground pixels} \\ 0 & \text{otherwise} \end{cases} \quad (5.6)$$

where the weight (ω) applied to the foreground class may be defined as in (Sudre et al., 2017):

$$w = \frac{N - \sum_{n=1}^N p_n}{\sum_{n=1}^N p_n} \quad (5.7)$$

The only difference is the weight (ω) which allows us to penalise false positives or false negatives more severely. If $\omega > 1$ then the number of false negatives will be reduced, similarly $\omega < 1$ means false positives being decreased.

Another family of widely used loss functions is based on the dice coefficient, which measures the overlap between two images. The Dice coefficient has been adapted as a loss function known as Dice loss (DL).

$$\text{DL} = 1 - \frac{\sum_{n=1}^N p_n r_n + \epsilon}{\sum_{n=1}^N (p_n + r_n) + \epsilon} - \frac{\sum_{n=1}^N (1 - p_n)(1 - r_n) + \epsilon}{\sum_{n=1}^N (2 - p_n - r_n) + \epsilon} \quad (5.8)$$

where ϵ is used to ensure the stability by avoiding the problem of dividing by 0. The Generalized Dice Loss (GDL) (Sudre et al., 2017) by introducing the weight ω :

$$\text{GDL} = 1 - 2 \frac{\sum_{l=0}^1 \omega_l \sum_{n=1}^N p_{ln} r_{ln}}{\sum_{l=0}^1 \omega_l \sum_{n=1}^N (p_{ln} + r_{ln})} \quad (5.9)$$

$$r_{0n} = \begin{cases} 1 & n: \text{background pixel} \\ 0 & n: \text{foreground pixel} \end{cases}, \quad r_{1n} = \begin{cases} 1 & n: \text{foreground pixel} \\ 0 & n: \text{background pixel} \end{cases} \quad (5.10)$$

where $l \in \{0, 1\}$ is the background segmentation map ($l = 0$) or the foreground segmentation map ($l = 1$). As a result, $r_{ln} = 1$ indicates that the pixel n is classified as background ($l=0$) or foreground pixel ($l=1$) in the corresponding segmentation map (equation 5.10). Similarly, $p_{ln} = [0, 1]$ is the probability that the pixel n is a background pixel ($l=0$) or a foreground ($l = 1$) in the prediction. Invariance to different class set properties is obtained thanks to the weights w_l :

$$w_l = \frac{1}{(\sum_{n=1}^N r_{ln})^2} \quad (5.11)$$

Training a model with a given loss function is an optimization problem, which can be solved by gradient descent (Equation 5.12):

$$\theta_{t+1} = \theta_t - \eta g_t, \text{ with } g_t = \nabla_{\theta} L(\theta_t) \quad (5.12)$$

Where θ is the network parameter to be optimized, η is the learning rate, g_t is the gradient at time step t of the selected samples, and L is the loss function for model training.

Usually, the training data is split into subsets, which are called mini-batches. One epoch corresponds to the gradient descent of all mini-batches. So, for each training epoch, the model weights are updated by backpropagation based on the loss score obtained in the previous epoch. One of the most commonly used optimization methods is the stochastic gradient descent (SGD). The SGD replaces the actual gradient with an estimate of thereof that is calculated from randomly selected samples from the dataset, allowing for more efficient and computationally feasible optimization of deep models.

In order to ensure the convergence and accelerate the learning process, a well-known algorithm has been proposed: SGD with Momentum (SGD-M). The momentum contributes to faster convergence by adding a fraction of the previous update to the current update.

$$\begin{aligned} m_t &= \gamma m_{t-1} + \eta g_t \\ \theta_{t+1} &= \theta_t - m_t \end{aligned} \quad (5.13)$$

where m_t is the first-moment vector at time step t , and γ is the exponential decay rate (a hyperparameter, usually set to around 0.9) for the first moment estimates. m_{t-1} stores the moving average of the gradients. This smoothes the gradient and reduces oscillations around the minimum, thus speeding up convergence.

Both SGD and SGD-M update elements in θ with fixed learning rate (η). In practice, different learning rates are desired for different parameters: the gradient step size tend to decrease proportionally to the curvature of the stochastic objective function. Therefore, the Adam (Adaptive Moment Optimization) algorithm is proposed by introducing the second-moment vector:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ \nu_t &= \beta_2 \nu_{t-1} + (1 - \beta_2) g_t^2 \end{aligned} \quad (5.14)$$

where ν_t is the second-moment vector at time step t , and β_2 is the exponential decay rate for the second-moment estimation (hyperparameter, typically around 0.999). As m_0 and ν_0 are set to 0 for initialization, the bias is shown in equation 5.14. Therefore, the bias correction is necessary to ensure that the moving averages are representative, especially in the early stages.

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{\nu}_t = \frac{\nu_t}{1 - \beta_2^t} \quad (5.15)$$

where β^t denotes β to the power t . Finally the parameter vector θ can be updated:

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{\hat{\nu}_t} + \epsilon} \hat{m}_t \quad (5.16)$$

Where α is the learning rate, which is a crucial hyperparameter that determines the step size towards the minimum of the loss function, and ϵ prevents the division by zero. The amount by which each parameter is adjusted is influenced by both the first moment and the second moment. This adaptive fitting leads to efficient optimization, especially in deep models with a lot of parameters.

U-Net in Practice

U-Nets have proven successful in image segmentation for many medical imaging modalities. Naturally, many network architectures combining U-Nets with other techniques have been implemented by researchers, and a brief introduction to these well-known variants is described below.

- 3D U-Net (Çiçek et al., 2016). It is proposed for the 3D volumetric segmentation, by replacing all 2D operations with corresponding 3D operations.
- Attention U-Net (Oktay et al., 2018). It is proposed to focus on specific target objects while ignoring unnecessary areas by using the attention gate. The attention

gate allows estimation of potential areas of interest by removing feature activation in irrelevant regions, without the need to use explicit external ROI as supervision.

- Residual U-Net (He et al., 2016). This variant combines ResNet with U-Net to address the problems of training very deep neural networks. The skip connection in ResNet takes the feature map from one layer and adds it to another deeper layer, allowing better preservation of feature maps in deeper neural networks.
- nnU-Net (Isensee et al., 2021). It is proposed to address the problem of dataset diversity in medical images by automating the entire segmentation process, including dataset configuration, pre-processing, network training (2D/3D) and post-processing, without human intervention.

Each of these variants continues to be explored for different tasks, which means that U-nets still have a lot of potential to incorporate new ideas.

5.2 Proposed Method

For our dataset of 79 orthogonal OCT volumes (158 in total), the ophthalmologist only marked the main pores through which most axons pass. Therefore, pores are partially annotated in all volumes. In the previous chapter, the orthogonal OCT volumes are registered and fused to obtain en-face images with enhanced pores. Based on this, we explore further possibilities with the fused images and the partial point annotation to segment pores in each 2D en-face image.

5.2.1 Network Architecture

We use the U-Net architecture proposed in (Ronneberger et al., 2015) as a baseline method for the pore segmentation task. In our dataset, pore sizes are small, with a typical diameter of less than 10 pixels, given that the ROI size of an en-face image is 262×262 pixels. In order to improve the detection of such small areas, pore features generated by U-Net can be strengthened by integrating attention mechanisms (Vaswani et al., 2017) to help capture the regions of interest (ROIs).

A popular approach proposed in (Oktay et al., 2018) incorporates an Attention Gate (AG) module (Fig. 5.4) into the U-Net. The AG allows the estimation of potential areas where the pores are most likely to appear by removing feature activation in irrelevant regions, without the necessity of using explicit external ROIs as supervision.

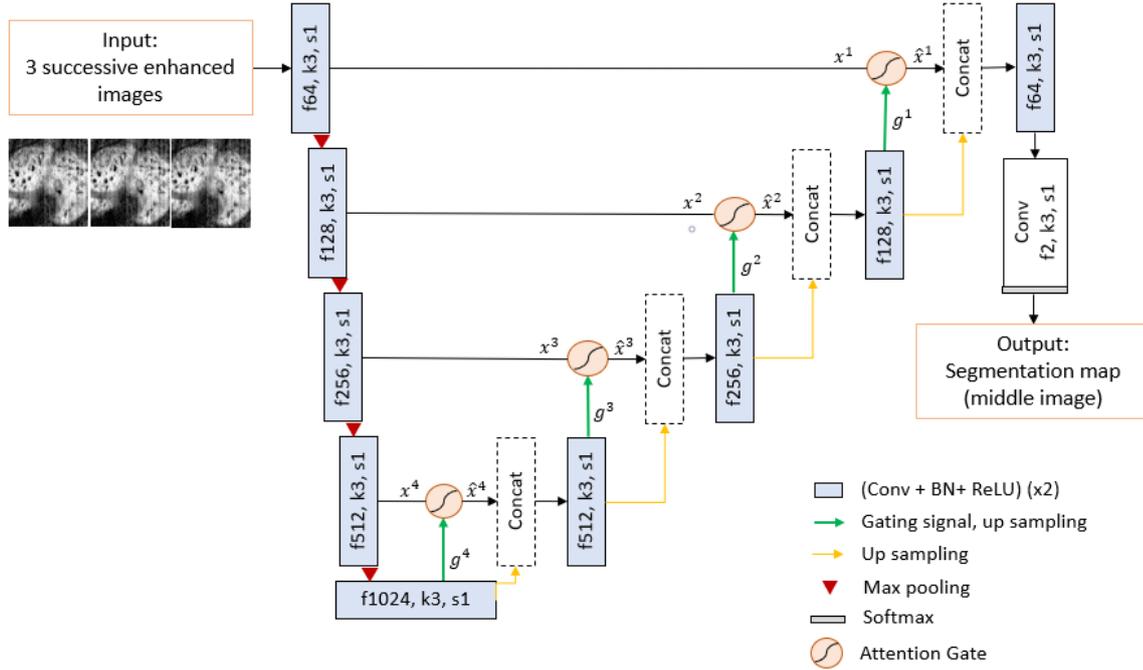


Figure 5.3: Proposed context-aware attention U-Net architecture. Input images from either horizontal/vertical volumes, or the registered and fused volumes in chapter 4. ‘f64,k3,s1’ indicates number of feature maps (64), kernel size (3) and stride (1).

Furthermore, axons passing through the ONH follow a fairly regular path: pore intensities are similar between successive en-face images, while their centroids and shapes show little spatial variation. Thus, a naive application of U-Net runs the risk of not fully exploiting these regularity properties. To this end, we design a context-aware network by inputting 3 successive en-face images, and outputting only a segmentation map for the middle image (Input and Output in Figure 5.3).

The proposed context-aware attention U-Net is shown in Figure 5.3. Three input images are progressively filtered by ($2\times$) convolution blocks and down-sampled in the encoder path. The convolution block, used in both the encoder and the decoder, is composed of a convolution layer (Conv), batch normalization (BN), and rectified linear unit (ReLU). In the decoder path, each layer has an attention gate through which features from the layer l in the encoder path must pass through before being concatenated to the up-sampled features in the coarser layer $(l + 1)$ of the decoder path. Finally a pixel-wise softmax is applied to generate probability maps to assign each pixel the corresponding class (pore or background).

The AG module is shown in Fig 5.4, where the architecture is adapted from (Oktay

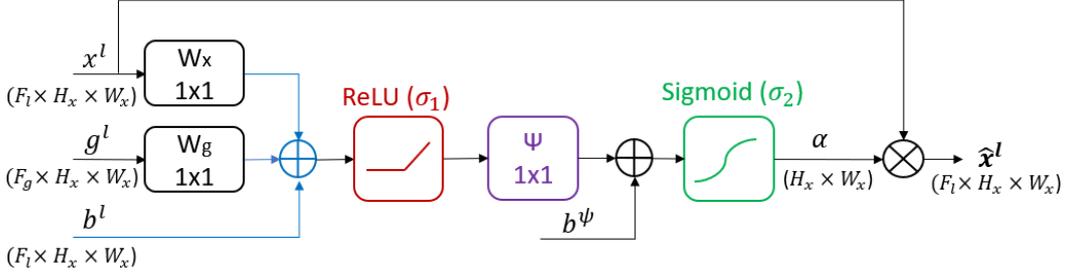


Figure 5.4: Attention gate. H_x , W_x refer to the height and width of the feature map x .

et al., 2018) for 2D pore segmentation. We define the feature map x at pixel $i \in \{1, \dots, N\}$ in layer $l \in \{1, \dots, L\}$ as $x_i^l \in \mathbb{R}^{F_l}$, where F_l refers to the number of feature maps in layer l . An attention coefficient $\alpha_i^l \in [0, 1]$ is calculated by the AG to identify the ROIs. The output of AG is an element-wise multiplication $\hat{x}^l = \alpha_i^l x_i^l$.

Feature maps are gradually down-sampled in the encoder to capture a large receptive field. Features on the coarse spatial grid level of layer $(l + 1)$ identify the location of target objects, and such coarse features may serve as gating signal $g^l \in \mathbb{R}^{F_g}$ to provide global information for x_i^l to disambiguate task-irrelevant feature content. Thus the additive attention coefficients α_i^l is calculated as:

$$\alpha_i^l = \sigma_2(q_{att}^l(x_i^l, g_i^l; \Theta_{att})) \quad (5.17)$$

$$q_{att}^l = \psi^T(\sigma_1(W_x^T x_i^l + W_g^T g_i^l + b_i^l)) + b_i^\psi \quad (5.18)$$

Where linear transforms $W_x \in \mathbb{R}^{F_l \times F_l}$, $W_g \in \mathbb{R}^{F_g \times F_l}$, $\psi \in \mathbb{R}^{F_l \times 1}$, and the bias $b_i^l \in \mathbb{R}^{F_l}$, $b_i^\psi \in \mathbb{R}$ together form the learnable parameter set $\Theta_{att} = \{W_x, W_g, \psi, b^l, b^\psi\}$ which characterize the AG. W_g and W_x ensure that the pixel-wise addition of x^l and g^l could be done to learn the salient regions. $\sigma_1(x)$ is the ReLU function for nonlinearity and $\sigma_2(x)$ is the sigmoid activation function for normalisation. Moreover, the linear function ψ allows to generate only one attention map α^l for all feature maps in layer l . In practice, linear transforms $\{W_x, W_g, \psi\}$ are implemented as 1×1 convolution layer.

The Generalized Dice Loss (*GDL*) presented in 5.1.1 is used to solve the unbalanced background/foreground problem, as only few pixels in our image are labeled.

$$GDL = 1 - 2 \frac{\sum_{l=0}^1 w_l \sum_{n=1}^N p_{ln} g_{ln}}{\sum_{l=0}^1 w_l \sum_{n=1}^N p_{ln} + g_{ln}}, w_l = \frac{1}{\sum_{n=1}^N g_{ln}}$$

Where $GT = \{g_{l1}, \dots, g_{lN} | g_{ln} \in \{0, 1\}\}$ is the *GT* background ($l = 0$) or foreground ($l = 1$) segmentation map of an image of N pixels, and $P = \{p_{l1}, \dots, p_{lN} | p_{ln} \in [0, 1]\}$

is the output segmentation map. Here, the weight ω is modified compared to the equation 5.11: as only a small portion of pores are segmented in the GT, many false negative pixels are presented in the GT. To this end, we want to penalize less the background pixels by choosing a lower ratio of ω_1/ω_0 . Finally, the probabilistic maps were thresholded at 0.5 to generate binary classification for each pixel (pore or not).

5.2.2 Training Dataset

Our dataset consists of 79 horizontal volumes and 79 vertical volumes, all annotated by the medical expert in section 3.2.2. In addition, the 79 fused volumes obtained in 4.3 can enrich our dataset. For these fused volumes, their ground truth segmentation maps are generated by applying the transform found in section 4.2 to fuse the ground truth of horizontal and vertical volumes.

Therefore, the input of our proposed context-aware network is three consecutive 2D images from different volumes: either the horizontal/vertical volumes ($I_H^{(x\pm 1)}, I_V^{(x\pm 1)}$), or the fused volumes ($I_{fus}^{(x\pm 1)}$). To evaluate the effectiveness of the fusion algorithm, we perform two experiments:

- In the first experiment (H+V), three 2D en-face images from horizontal ($I_H^{(x\pm 1)}$), or three 2D en-face images from vertical volumes ($I_V^{(x\pm 1)}$) are fed into the network.
- In the second experiment (H+V+Fusion), besides images from the horizontal and vertical volumes, three images from the fused volume ($I_{fus}^{(x\pm 1)}$) are also input to the network.

5.3 Experiment Setup

Experiments were carried out on the dataset described in section 3.1.4. In total, we have 158 OCT volumes (79 orthogonal scans) from 68 eyes of 42 subjects examined between 2021 and 2023. For the source horizontal or vertical volumes, after manual annotation by the medical expert, our dataset of 158 volumes consists of 8559 2D en-face images containing at least one manually annotated pore. On average, 6.73 ± 6.03 pores are identified in each en-face image. For the fused volumes, the manual annotation of orthogonal volumes is also fused to create a fused segmentation map. In this case, 79 fused volumes contain 3559 2D en-face images, with an average of 8.65 ± 7.80 pores identified in each 2D image.

5.3.1 Implementation Details

Hyperparameters. The same hyperparameters are set for all training processes. The learning rate is initialized as 3×10^{-4} , and would be reduced by a factor of 5 if the exponential moving average of training loss did not improve by at least 5×10^{-3} within the last 30 epochs. Adam optimizer is used with a weight decay of 3×10^{-5} , and a batch size of 32 for 1000 epochs on a TITAN RTX GPU.

Data augmentation. Data augmentation is performed by randomly combining horizontal/vertical flipping, rotating, brightness and contrast changing, Gaussian noise, and elastic deformation (Simard et al., 2003).

Cross-validation. Different volumes (longitudinal studies) from the same subject are in some cases very similar to each other, so subject-wise cross validation is used for our study. The dataset of 42 subjects is divided into 5 folds to perform the cross-validation: 3 folds (26 subjects) for training, 1 fold (8 subjects) for validation and 1 fold (8 subjects) for the test set. Given a test set (8 subjects), we perform the leave-one-out cross validation on the training (26 subjects) and validation sets (8 subjects). Therefore, for the test set, we have an ensemble of 4 trained models. The final prediction was obtained by averaging the 2 models that achieve the highest object-level recall and precision scores to increase the robustness of the model.

5.3.2 Evaluation Metrics

Different metrics have been used to evaluate the performance of the proposed pore segmentation method:

- The Pixel-level Dice similarity Coefficient (Dice) and Jaccard index (Jaccard) measure the pixel-wise segmentation accuracy between the GT and the output segmentation mask.

$$\text{Dice} = \frac{2TP}{2TP + FP + FN}, \quad \text{Jaccard} = \frac{TP}{TP + FP + FN} \quad (5.19)$$

where TP is the number of true positive pixels in an image that were correctly classified as pores; FP is the number of false positives that were incorrectly classified as pores; FN is the number of false negatives that were incorrectly classified as background. However, the ground truth is generated by the region growing algorithm using the expert annotations as seed points. As a result, the GT segmentation map is not accurate at the pixel level. For a more reliable evaluation,

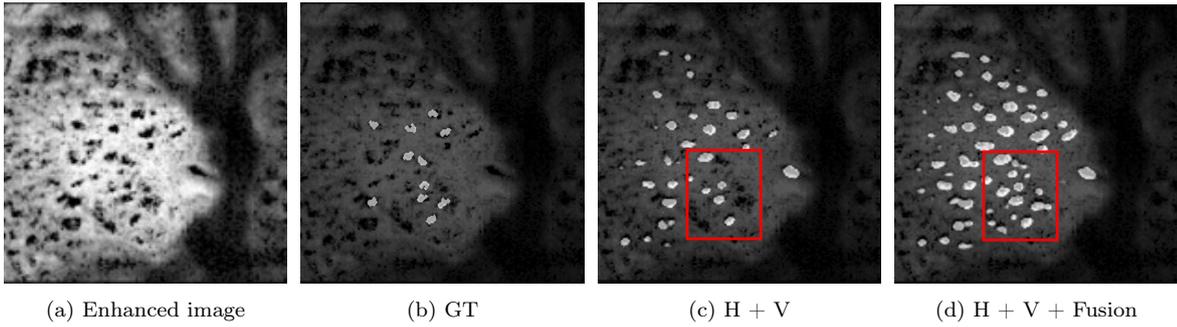


Figure 5.5: Evaluation using the fusion method proposed in chapter 4.

object-level pore detection metrics are also used.

- The object-level Recall (Recall) and Precision (Precision) metrics that measure the object-level accuracy of detection. Recall is used to assess the true positive rate of the proposed method and Precision is used to measure the positive predictive value.

$$\text{Recall} = \frac{TP_o}{TP_o + FN_o}, \quad \text{Precision} = \frac{TP_o}{TP_o + FP_o} \quad (5.20)$$

where TP_o is the number of pores in the ground truth segmentation that overlap with a pore candidate in the output segmentation (at least one pixel overlap in the 8-connectivity sense). FN_o is the number of missed pores in the output segmentation map compared with GT segmentation. FP_o is the number of misclassified pore candidates in the output segmentation map.

5.4 Experiment Results

In this section, we first compare the performance of different inputs, followed by a comparison of the proposed network with other state-of-the-art methods.

5.4.1 Comparison of Different Dataset

In order to evaluate the fusion algorithm proposed in the previous chapter, we apply our context-aware attention U-Net on two dataset: H+V (8559 2D images) and H+V+fusion (12118 2D images), as presented in section 5.2.2. Figure 5.5 and Table 5.1 show the performance of the proposed network on these two datasets:

- With additional fusion images as input, the network tends to predict more pore candidates, confirmed by the high recall score (0.746). Such result could be ex-

plained by the facts that on the one hand, pores are more contrasted in the fused images. Meanwhile, more pores are identified on the GT of the fused images thanks to the fusion process. On the other hand, the fused images could be considered as a data augmentation technique, which is powerful in small dataset regime like ours.

- The Dice, Jaccard and Precision are low for both models, which could be explained by the fact that only partial pores are annotated in the GT, so that false positive regions in the output segmentation map may correspond to pores that are not labeled in the ground truth.

Table 5.1: Segmentation results on the test dataset. The mean value of each metric is calculated over all pixels in the test dataset. We favor the Recall metric since the ground truth is incomplete.

Input images	Pixel-level		Object-level	
	Dice	Jaccard	Precision	Recall
H + V	0.261	0.169	0.282	0.715
H + V + Fusion	0.293	0.178	0.275	0.746

5.4.2 Comparison of Existing Networks

In this section, we used the dataset with fused images (H+V+Fusion) for all of the following methods. We compare the proposed method with region-based active contour approach proposed by Chan and Vese (Chan and Vese, 2001), the unsupervised method W-Net (Royer et al., 2021), as well as the supervised methods of baseline U-Net (Ronneberger et al., 2015) and Attention U-Net (Oktay et al., 2018). For the supervised methods, we used the same cross-validation principles.

Figure 5.6 shows the visual comparison of segmentation performances. We can observe that active contour and unsupervised W-Net approaches are sensitive to noise and artifacts, especially in low SNR regions like the LC borders. On contrary, U-Net based methods are more efficient in predicting the pore candidates. Our proposed method is able to predict more pore candidates in low contrast regions and in border regions, thanks to the context-aware design that is able to refer local consecutive potential pore areas. We also observe that pore sizes predicted by our method and the Attention U-Net are closer to their real sizes, meaning that the attention gate is efficient in highlighting the potential small regions in the image.

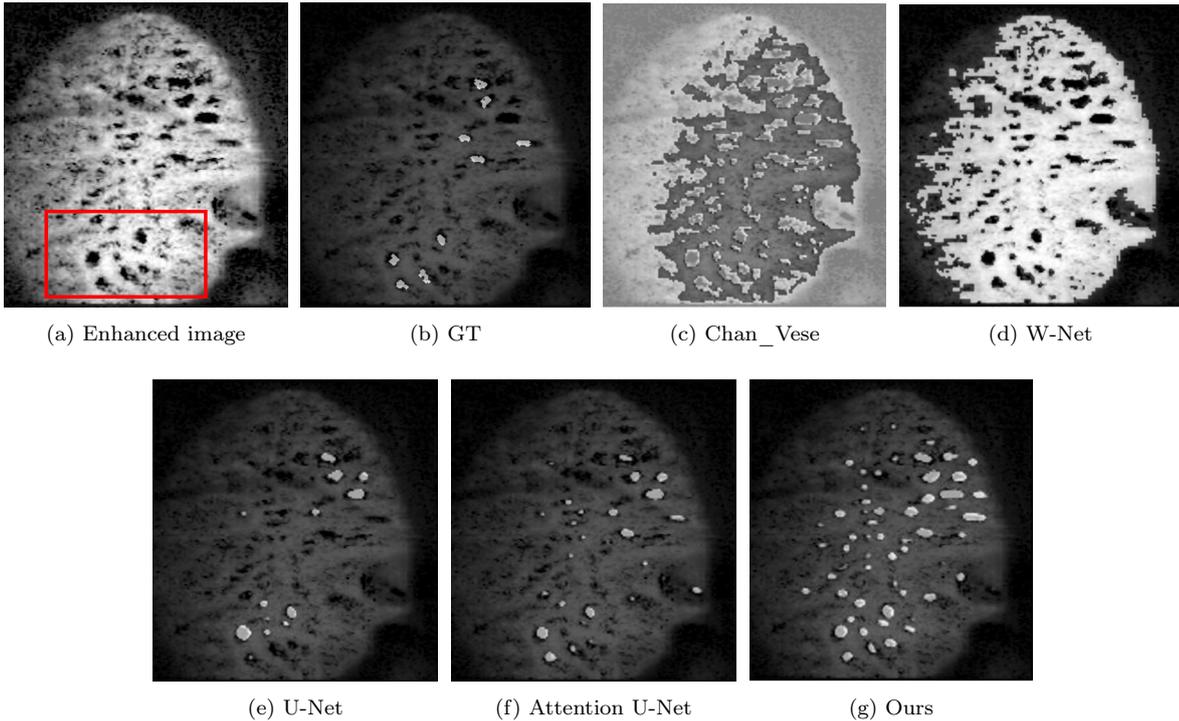


Figure 5.6: Comparison of the segmentation results with other state-of-the-art methods.

Our proposed method is robust in identifying true positive pores with the highest Recall value, as shown in Table 5.2. The Dice, Jaccard and Precision scores are not reliable to measure the performance with incomplete ground truth. To this end, we asked the expert to validate or not the pores suggested by the supervised methods and also to point out the missing pores. Using this method, we bypass the partial ground truth problem. Obviously this method is time-consuming, hence why we did it only on 10 random images of the test set.

Table 5.3 illustrates the *a posteriori* evaluation, instead of relying on the partial ground truth for score evaluation. We observe that our model predicts more TP pores thanks to the context-aware design, and meanwhile the attention gate helps to eliminate FP pores by gradually attenuating the activation of their surrounding background as the network goes to shallower layers in the decoder path, resulting in a more accurate prediction of pore location and pore sizes. Finally, the missing pores (FN) are mainly located at: (1) the start or the end of the pathway, where the image is of low SNR; (2) the artifact areas especially vessel shadows. In both cases it is hard even for experts to identify without referring to the continuity.

Table 5.2: Comparison with other state-of-the-art methods on the test dataset.

Model	Pixel-level		Object-level	
	Dice	Jaccard	Precision	Recall
Chan_Vese	0.084	0.056	0.126	0.677
W-Net	0.101	0.056	0.127	0.694
U-Net	0.265	0.169	0.262	0.656
Attention U-Net	0.269	0.173	0.283	0.697
Ours	0.293	0.178	0.275	0.746

Table 5.3: *A posteriori* evaluation by the expert (10 images).

Model	Object-level				
	TP	FP	FN	Recall	Precision
U-Net	30.4	2.3	7.1	0.81	0.93
Attention U-Net	36.7	2.4	4.4	0.89	0.94
Ours	38.2	3.5	2.8	0.93	0.92

5.5 Conclusion and Discussion

In this chapter, we proposed a simple but efficient context-aware Attention U-Net for pore segmentation using partial points annotation. The input of the proposed network could be the enhanced horizontal/vertical en-face images, or the fused images produced in the previous chapter. This is, to our knowledge, the first time a deep learning method is used for the *in vivo* LC pore segmentation task. Experimental results show that our method is robust in identifying true positive pores, despite the network being trained with an incomplete ground truth. Another advantage of the proposed network is that the manual delineation of peripheral masks is not required, thanks to the automatic ROI detection in section 4.1.1 and the supervised approach.

In our recent paper (Ding et al., 2024), the network architecture was slightly different: after the registration process, in addition to the three horizontal images, three registered vertical images are also fed into the network, resulting in 6 images as input to the network, while only 1 prediction map is output for the middle horizontal image. In this case, the manual annotation of the vertical volumes was ignored, which is not reasonable given the the small size of our dataset. Certainly, more training strategies are possible with the fused images, and we are also working on this for more robust

predictions.

We also tried object detection methods such as YOLO (Redmon et al., 2016) and Mask R-CNN (He et al., 2017), but none of them were successful. This is mainly due to the fact that pores are small (less than 10 pixels in a 262×262 pixel image) and have no particular appearance or characteristics, making it difficult for object detection networks to work perfectly.

Chapter 6

3D Reconstruction of Pore Pathways

Contents

6.1 Parametric Active Contours	73
6.1.1 Energy Terms	74
6.1.2 Energy Minimization	75
6.1.3 Discrete Approximation	75
6.2 Proposed Method	76
6.2.1 Energy Function Selection	76
6.2.2 Proposed Algorithm	77
6.3 Experiments and Results	79
6.3.1 Evaluation Metrics	80
6.3.2 Quantitative Evaluation	81
6.4 Conclusion	82

In the previous chapter, pores have been segmented in every 2D en-face image with the proposed algorithm, allowing for the 3D reconstruction of the axonal pathways in 3D. We take the same inspirations from our former work in (Rossant et al., 2017), where a parametric active contour algorithm was proposed to reconstruct pore pathways.

In this chapter, we propose an adapted active contour method for the tracking of pore pathways. Specifically, we first review the principles of active contours, followed by the details of the proposed tracking algorithm. Finally we evaluate the proposed reconstruction algorithm with other state-of-the-art methods.

6.1 Parametric Active Contours

Active contour models (also called deformable models or snakes), originally introduced by (Kass et al., 1988), have been used in a variety of different medical image

processing tasks, ranging from image segmentation to object tracking. The basic idea is that given an initial guess of the contour, the contour will be moved by forces to the boundaries of the desired object. This is solved by minimizing the energy function that is defined as the sum of different kind of energies along the length of the snake.

6.1.1 Energy Terms

Active contour is defined in the form of a closed or open parametric curve. It (dynamically) evolves the contour iteratively, from its initial position towards the edges of the object to be extracted. Active contours are essentially subject to two types of force: an internal force that regularises the curve and an image force that expresses the interaction of the contour with the image.

Let I be the image and C a contour of the image I that can be deformed spatially and temporally. The curve is parameterized by the curvilinear abscissa s and the time t . We denote $V(s, t)$ the position of a point on curve C at time t .

$$V(s, t) = (x(s, t), y(s, t)), s \in [0, 1] \quad (6.1)$$

The energy function associated with the contour ($E(V)$) consists of three terms:

$$E(V) = E_{int}(V) + E_{img}(V) + E_{ext}(V) \quad (6.2)$$

The internal energy (E_{int}) characterises the physical properties of the contour and imposes regularity properties on the curve, without looking into the image information.

$$E_{int}(V) = \int_0^1 (\alpha(s)|V'(s, t)|^2 + \beta(s)|V''(s, t)|^2) ds \quad (6.3)$$

where the first-order term represents the elasticity of the contour and is controlled by the parameter $\alpha(s)$. A large value of α will limit the stretching of the contour, while the discontinuity may be present if α is set to zero. The second-order term is the stiffness of the contour and is controlled by $\beta(s)$. A large value of β prevents the strong curvature, while the curve may be discontinuous at second order and contain corners if it is set to zero.

The image energy (E_{img}), also known as the context energy, depends on the image properties. This energy can be expressed as follows, so that the curve is attracted to areas with a strong gradient

$$E_{img}(V) = - \int_0^1 |\nabla I(V(s, t))|^2 ds \quad (6.4)$$

where $\nabla I(V(s, t))$ is the gradient of the image I at $V(s, t)$.

In most cases, the gradient of the smoothed image is calculated:

$$E_{img}(V) = - \int_0^1 |\nabla(g_\sigma * I)(V(s, t))|^2 ds \quad (6.5)$$

where g_σ is a Gaussian with mean zero and standard deviation σ .

Another way of defining image energy is simply modeling the pixel intensities in order to detect dark or light peaks, which is exactly what we are looking for the pore reconstruction task.

$$E_{img}(V) = \pm \int_0^1 I(V(s, t)) ds \quad (6.6)$$

The external energy (E_{ext}) are the constraints responsible for putting the snake near the desired local minimum, which may come from higher level interpretation, user interaction, etc.

6.1.2 Energy Minimization

The evolution of the active contour is obtained by minimising the total energy (equation 6.2). The weights α and β are considered constant to simplify the problem. Minimising Equation 6.2 is equivalent to solving the Euler-Lagrange equation

$$\min_V \int_0^1 E(s, V, V', V'') ds \quad \Leftrightarrow \quad \frac{\partial E}{\partial V} - \frac{\partial}{\partial s} \frac{\partial E}{\partial V'} + \frac{\partial^2}{\partial s^2} \frac{\partial E}{\partial V''} = 0 \quad (6.7)$$

And we have the following equalities

$$\frac{\partial E}{\partial V} = \nabla(P(V(s))), \quad \frac{\partial}{\partial s} \frac{\partial E}{\partial V'} = \alpha V''(s), \quad \text{and} \quad \frac{\partial^2}{\partial s^2} \frac{\partial E}{\partial V''} = \beta V''''(s) \quad (6.8)$$

The optimal condition is obtained by solving

$$\alpha V''(s) - \beta V''''(s) - \nabla(P(V(s))) = 0 \quad (6.9)$$

where $P(V(s))$ represents the image potential corresponding to the sum of the image and external energies. For the numerical computation of the active contours, the discretization of the curve V is considered in the next section.

6.1.3 Discrete Approximation

The N points that form path V are represented by the coordinates $[V_0, V_1, \dots, V_{N-1}]^T \in (\mathbb{R}^2)^N$. The derivatives involved in the Euler-Lagrange equation are conventionally es-

timated using the finite difference method.

$$\begin{aligned} V''(s) &\simeq V_{k+1} - 2V_k + V_{k-1} \\ V'''(s) &\simeq V_{k+2} - 4V_{k+1} + 6V_k - 4V_{k-1} + V_{k-2} \end{aligned} \quad (6.10)$$

Equation 6.9 can be written as

$$AV + \nabla(P(V)) = 0 \quad (6.11)$$

where A is a symmetric matrix of size $N \times N$. Then we have the following numerical scheme by introducing the time variable t .

$$AV_t + \nabla(P(V_{t-1})) + \gamma(V_t - V_{t-1}) = 0 \quad (6.12)$$

where γ is the step size of Euler's temporal discretization. Finally we can obtain V_t depending on V_{t-1}

$$V_t = (A + \gamma I_d)^{-1}(\gamma V_{t-1} - \nabla(P(V_{t-1}))) \quad (6.13)$$

At each iteration, the potential $\nabla(P(V_t))$ is calculated and the new position of each point on the curve is thus updated.

6.2 Proposed Method

The axonal pathways are reconstructed from the segmented pores on the fused X-plane image, based on a tracking process initialized at the reference X-plane image (X_0) that was selected in section 3.2.3. Every segmented pore in this plane initializes a pathway and the goal is to track it along the X axis, upwards and then downwards, at most to the start (x_{start}) and the end (x_{end}) of the LC.

6.2.1 Energy Function Selection

Let us consider one axon pathway l_x on the enhanced fused X-plane image $I_{fus}^{(x)}$. Given the gravity center location of the segmented pore $(x, y_x^{(l)}, z_x^{(l)})$, we want to determine the next position $(x+1, y_{x+1}^{(l)}, z_{x+1}^{(l)})$ in the downward case. To this end, we rely on a parametric active contour model described in section 6.1, which exploits the continuity of the dark intensities along any axonal pathway. Let us denote by $V_x^{(l)}$ this path indexed by l , initialized as a small vertical segment at the position $(y_x^{(l)}, z_x^{(l)})$. The energy functional to minimize is given by :

$$E(V_x^{(l)}) = \int_{x-\delta}^{x+\delta} \alpha |V_x^{(l)'}(s)|^2 + \beta |V_x^{(l)''}(s)|^2 + I_{fus}(V_x^{(l)}(s)) ds \quad (6.14)$$

where I_{fus} is the fused image and δ determines the length of the curve. The first two terms are the the internal energy, which regularizes and ensures that the path is smooth. The higher α and β the straighter the curve $V_x^{(l)}$. The last term is the image potential, which deforms the snake to reach the locations of lowest intensity, so to the pores. The minima of the energy function $E(V_x^{(l)})$ are found by solving the following Euler-Lagrange equations according to equation 6.9:

$$\begin{aligned} \alpha Y_x^{(l)''}(s) - \beta Y_x^{(l)''''}(s) - \frac{\partial I_{fus}(Y_x^{(l)}(s), Z_x^{(l)}(s))}{\partial Y} &= 0 \\ \alpha Z_x^{(l)''}(s) - \beta Z_x^{(l)''''}(s) - \frac{\partial I_{fus}(Y_x^{(l)}(s), Z_x^{(l)}(s))}{\partial Z} &= 0 \end{aligned} \quad (6.15)$$

where $Y_x^{(l)}(s)$ and $Z_x^{(l)}(s)$ are the coordinates of the curve $V_x^{(l)}$ in the plane s . In our case, the parameter s takes the discrete values $i = \{x - \delta, x - \delta + 1, \dots, x + \delta\}$, ensuring that each point of the discrete curve can only move along the Y and Z coordinates in its initial X-plane. The minimization is obtained through an iterative gradient descent algorithm and the solution in equation 6.13 become

$$\begin{aligned} Y_t &= (A + \gamma I_d)^{-1}(\gamma Y_{t-1} - \frac{\partial I_{fus}}{\partial Y}(Y_{t-1}, Z_{t-1})) \\ Z_t &= (A + \gamma I_d)^{-1}(\gamma Z_{t-1} - \frac{\partial I_{fus}}{\partial Z}(Y_{t-1}, Z_{t-1})) \end{aligned} \quad (6.16)$$

The parameters are set as $\alpha = 1$, $\beta = 0$, $\gamma = 4$ and $\delta = 3$, to ensure appropriate convergence speed without any oscillations around the local minimum reached by the algorithm.

6.2.2 Proposed Algorithm

The segmentation map of the 2D en-face images are thresholded using OTSU algorithm (Otsu, 1979), where the centroid of each connected component is extracted. The axonal pathways are initialized on the reference plane X_0 with the centroid of the path l being $(x_0, y_0^{(l)}, z_0^{(l)})$. Given a path at plane x , after the calculation of the minimization of the energy function, we get the predicted position $V_{x+1}^{(l)}$ in the plane $x+1$ for the next position of the axonal pathway (circles in Figure 6.1). If this position $(x+1, y_{x+1}^{(l)}, z_{x+1}^{(l)})$ falls inside a predicted pore candidate, then the pathway continues with the existing

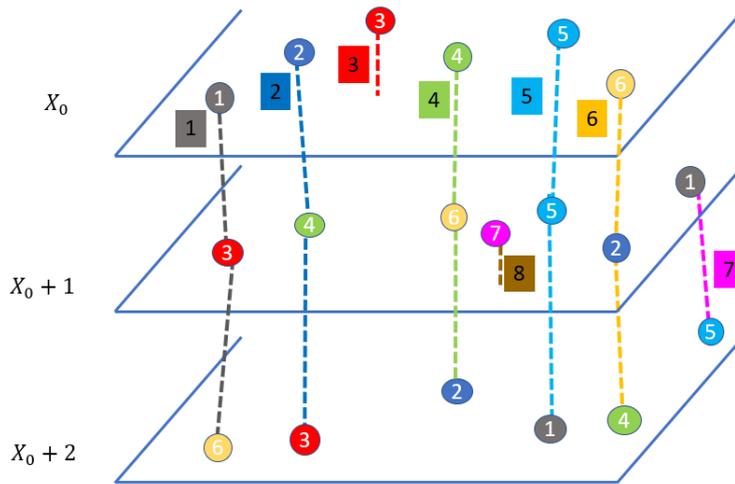


Figure 6.1: Schema showing how our snake model works. In circles are the pore candidates on the en-face planes, while in rectangles are the reconstructed pathways.

label, otherwise it ends. For pore candidates not linked to any existing pathway, they are initialized in the plane to a new path label, as they are first detected.

Finally, we only retain the paths that last for more than 5 planes since we are interested in identifying the main axonal pathways, which are clinically the most representative as well. Every axonal pathway is then refined and regularized by applying the same snake model along its entire length.

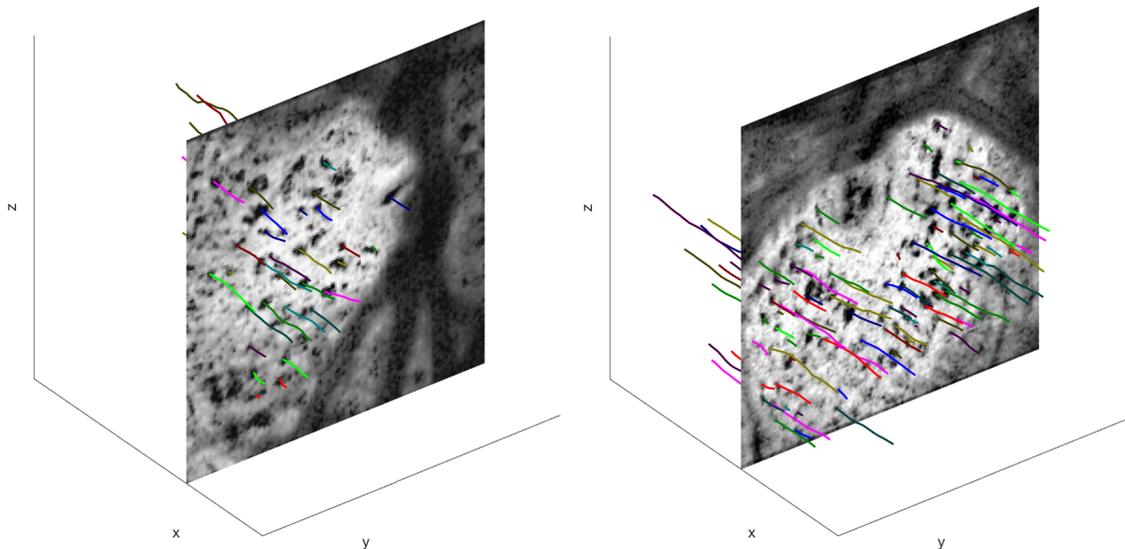


Figure 6.2: Overview of the all reconstructed pathways on the enhanced fused image.

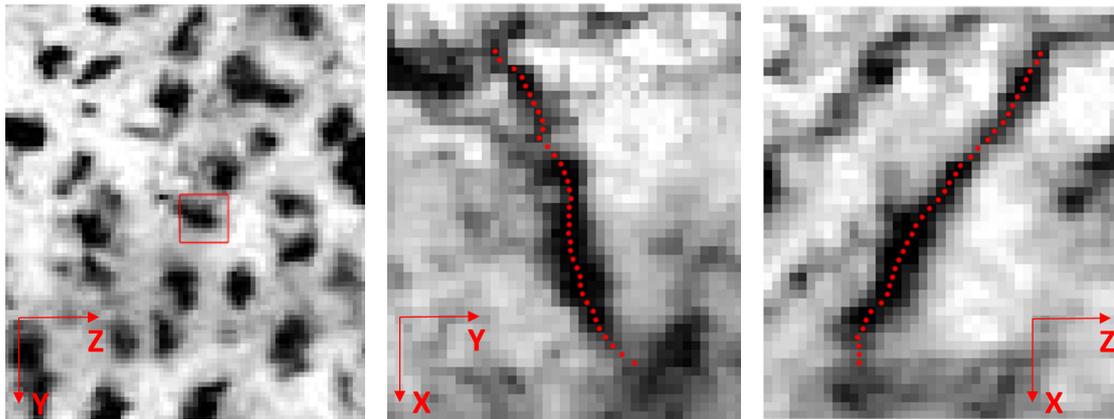


Figure 6.3: Reconstruction result using our method in a high contrast case (enhanced images, vertical slices reconstructed by interpolation)

This method is simple yet efficient for our task, since it explicitly models the main features of axonal pathways, that is their low intensity and their regularity. Two examples of the reconstructed pathways in Figure 6.2 shows that the proposed method is capable of reconstructing the main pathways through which most axons pass. This is due to the fact that the initialization of pore pathways in the reference X -plane X_0 is reliable because pores in this plane have good contrast. In addition, the initialization of the curve as a small vertical segment is accurate, since the axonal pathways are very regular at this scale, which prevents falling into a local maximum of the energy functional. Thus, our method makes it possible to track successive positions taking into account a small but significant local image context, unlike approaches that only consider the intensities in the adjacent plane. This property and the use of the binary segmentation map to validate the next position make our tracking approach both reliable and accurate in reconstructing the main axonal pathways. Finally, it is worth noting that our implementation allows to process all paths of a given x plane simultaneously, every path being small ($2\delta + 1 = 7$ points) and only a few iterations needed to converge the snakes. So the algorithm is not heavy. A 3D reconstructed path in Figure 6.3 shows that the path is determined without ambiguity in high contrast case.

6.3 Experiments and Results

Experiments were conducted on the test dataset in chapter 5 with 12 orthogonal OCT volumes (24 volumes in total) where pore candidates have been predicted with the

Table 6.1: Evaluation of the path length using our proposed method*.

Mean path length (GT)	32.66 ± 9.57	Max	56
		Min	11
Mean path length (Ours)	16.60 ± 13.25	Max	41
		Min	5

*Unity: number of crossed X-plane images.

Table 6.2: Evaluation of the proposed 3D reconstruction method.

	TP_{pth}	FN_{pth}	FP_{pth}
Ours	83(86.5%)	13(13.5%)	135(140.6%)
	TP_{pnt}	FN_{pnt}	FP_{pnt}
Ours	20.54 ± 8.17	15.20 ± 8.92	5.38 ± 5.04

proposed context-aware attention U-Net. For the manually annotated paths, we define the fused manual annotation as ground truth (GT), which means that the manual annotation on vertical volumes has been aligned and fused to the horizontal volumes by the found transform in chapter 4.

6.3.1 Evaluation Metrics

We used path detection metrics and path accuracy metrics defined in (Rossant et al., 2017) to evaluate the proposed method. At the path detection level, we calculated the number of True Positive pathways (TP_{pth} , number of paths present in both *GT* and our model), number of False Positives (FP_{pth} , extra paths of ours without equivalent in *GT*) and the number of False Negatives (FN_{pth} , paths of *GT* without equivalent in our model). Two paths are considered to be matched (TP_{pth}) if the distance between them is less than 2 pixels on at least 20% of the path of GT.

Then, the accuracy of the matching pathways (TP_{pth}) are also evaluated. Given two matching paths of our method (A^l_{ours}) and the ground truth (A^l_{GT}), we denote the X interval of overlapping by $I_{X,l}$. All points of A^l_{ours} inside $I_{X,l}$ are considered as true positive points (TP_{pnt}), whatever their distance to the GT. Similarly, the points of A^l_{ours} outside of $I_{X,l}$ are counted as false positives (FP_{pnt}), and those of A^l_{GT} outside of $I_{X,l}$ are the false negatives (FN_{pnt}). The root mean squared error (RMSE) between the matched points on the interval $I_{X,l}$ are also calculated to evaluate the accuracy of the

tracking performance. Finally, all metrics are averaged over the volumes and paths to have a statistical evaluation on the dataset. Mean path length (i.e. number of points) are also calculated to help interpretation.

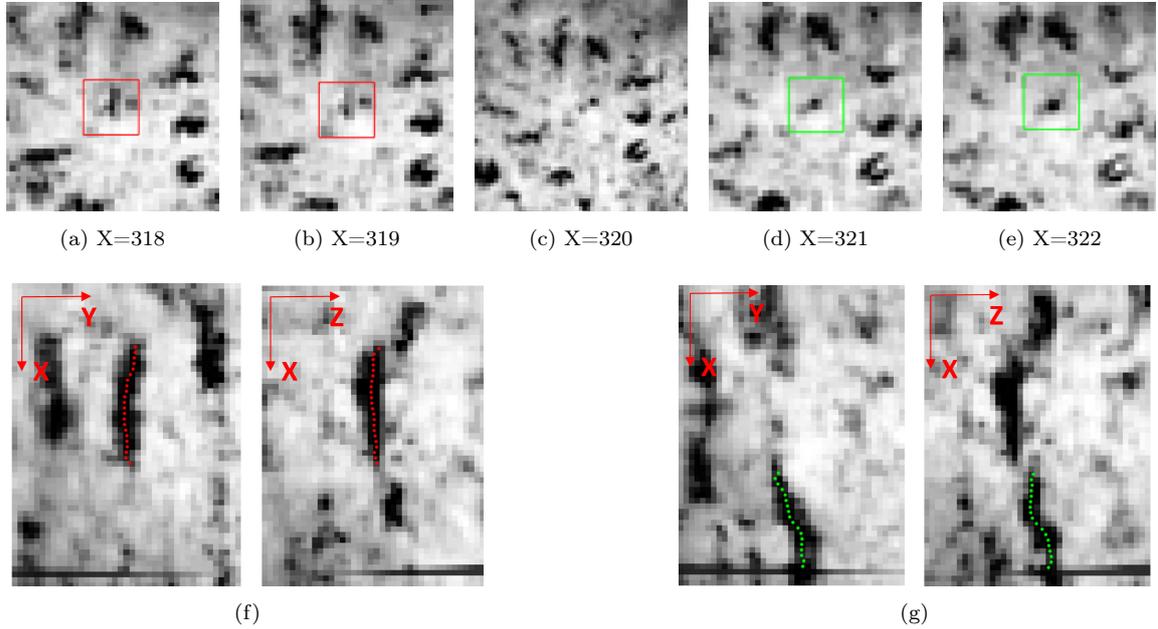


Figure 6.4: Example of pathway interruption. The first row shows the two reconstructed pathway with an interruption at $X=320$. The second row shows the the other two orthogonal views of the two pathways. The first pathway is shown in red, and the second in green, respectively.

6.3.2 Quantitative Evaluation

The mean path depth with our method (16.60 ± 13.25) is shorter than in the GT (32.66 ± 9.57) and with a higher standard-deviation, as shown in Table 6.1, which could be explained by the fact that the path could be interrupted for some X interval where the pore is less contrast. An example of such interruption is illustrated in Figure 6.4, where the first path stopped at $X=319$, while another started at $X=321$, without correspondence at $X=320$. By visual check the correspond pore at $X=320$ is observed, however, it is too small to be segmented by our algorithm, leading to the interruption of the path.

The reconstruction results are reported in Table. 6.2. Our method segmented more axonal pathways than the expert (218 vs. 96 paths), which is in agreement with the previous analysis and shows that the extra pores detected in en-face images correspond

to significant axonal pathways (but not labeled in the GT) that can be traced. We also observe that missing paths (FN_{pth}) are mainly due to the low SNR in some en-face images.

As for the accuracy of matching pathways, the FN_{pnt} and the FP_{pnt} are mainly observed at the start and the end of the pathways, where the en-face images are normally much less contrasted, leading to the disagreement even between the ophthalmologists.

Finally, the RMSE between the axonal pathways reconstructed with our method and the ground truth, on the area of overlap, is equal to 2.13 ± 1.20 pixels. This demonstrates that the 3D reconstruction is very accurate considering the typical pore diameter (~ 10 pixels).

6.4 Conclusion

In this chapter, we present a simple but efficient segmentation-based tracking algorithm to reconstruct the axonal pathways in 3D, given the enhanced fused volume and the binary segmentation maps of the pores in the en-face images. Our method relies on a parametric active contour model, which is initialized on the reference en-face image where pore information is the most reliable. It takes into account a local but significant context neighboring planes, which makes the tracking reliable and accurate. The predicted position is validated only if it falls into a segmented pore, so that only unambiguous segments are extracted. By qualitative and quantitative evaluation, the proposed method is robust in tracking the pore paths in adjacent en-face planes. At the same time the algorithm is light since the energy functional is minimized only on 7 en-face images and all pores simultaneously processed.

The efficiency of the algorithm is also owing to the proposed pre-processing and registration steps for pore enhancement, as well as the segmentation method for accurately isolating pores in each 2D en-face image. Improvements should essentially focus on managing pathway interruptions.

We hope that in the near future, the proposed method can be applied clinically to help ophthalmologists better model the LC, in particular pore morphology, in order to better understand glaucoma and therefore prevent it at an earlier stage.

Chapter 7

Conclusion and perspectives

Contents

7.1	Conclusions	83
7.1.1	Contributions	83
7.2	Perspectives	84
7.2.1	Perspectives on the Methodology	84
7.2.2	Combination with Other Modalities	85
7.2.3	Towards Clinical Applications	85

7.1 Conclusions

The objective of this thesis was to develop a complete and innovate method for the 3D reconstruction of pores in the lamina cribrosa (LC), which had been rarely investigated. In the course of this thesis, we have gradually collected and annotated our dataset containing horizontal and vertical 3D OCT volumes for each examination. Such dataset has allowed us to carry out the related tasks, which was challenging because of low signal-to-ratio (SNR) images and partially annotated pores on en-face images.

7.1.1 Contributions

In Chapter 3, we presented our dataset, from acquisition to expert annotation. We then introduced the pre-processing steps, which aims at enhancing pores in en-face images and extract the Volume-of-Interest. The enhancement results outperform the state-of-the-art method in denoising and enhancing main pores.

In Chapter 4, we proposed, for the first time, to register two orthogonal volumes to

enhance pore features. The registration consists of two steps: coarse registration by translation and 3D affine transform for refined registration. The proposed method is proven to be effective: (1) reliable *a priori* information in reference en-face images is fully exploited in the coarse registration process, (2) the issue of varying optical axis is better addressed with 3D affine transform. The experimental results show that the performance of the proposed network is comparable with the ground truth, with the distance between matching pores below the pore diameter.

In Chapter 5, we proposed a context-aware Attention U-Net network to segment pores in each 2D en-face image with partial points annotation. This method is shown to be robust due to the context-aware design to take into account for pore continuity in successive images, and the attention gate to highlight small regions-of-interest. Performance of the network is improved by inputting the fused images, proving conversely that our registration/fusion method is useful and efficient. The promising segmentation results allow us to track pores in successive en-face images with high reliability.

In Chapter 6, we proposed a segmentation-based tracking algorithm to reconstruct the axonal pathways in 3D, given the enhanced fused volume and the binary segmentation maps of pores in the 2D en-face images. Our method relies on a parametric active contour model that is initialized on the reference en-face image where pore information is most reliable. Local context of successive en-face images is considered, making the tracking algorithm accurate. The robust results are owing to the proposed pre-processing and registration steps for pore enhancement, and the accurate segmentation. All different processing components together make the proposed method promising for clinical applications.

7.2 Perspectives

7.2.1 Perspectives on the Methodology

Data. Currently, deep learning algorithms are widely applied to reconstruction tasks, meaning that a large training set and exhaustive annotation are required. Our segmentation result in section 5.3 used the *a posteriori* validation by pointing out all visible pores in ten 2D images. Similar pointing will also be possible on entire volumes to validate the tracking algorithm, but maybe limited to 5 volumes, cause such annotation is very time-consuming.

Advanced fusion method. For the registered H and V volumes, the pixel-based

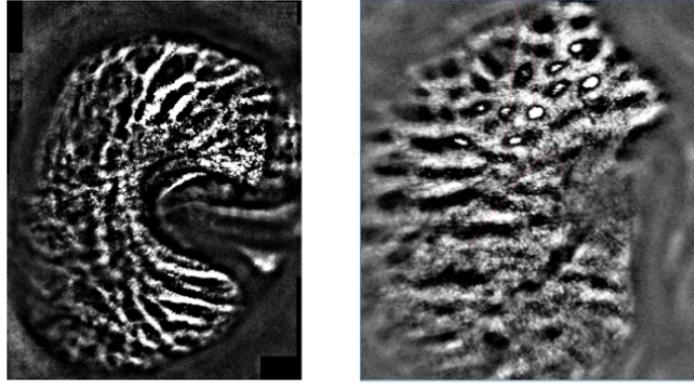


Figure 7.1: High pore visibility in adaptive optics (Zwillinger et al., 2016).

AND fusion method has been adopted. However, we can work on more advanced fusion schemes, in order to exploit the best of both images locally. These fusion schemes can be implemented using neural networks (Zhang et al., 2020; Liu et al., 2017) with more sophisticated methods than a simple AND operator. We can also test fusion schemes at the input or in the intermediate layers of a deep neural network.

Graph-based tracking algorithm. Our active contour based tracking method has one limitation: path interruption caused by a segmentation error is not considered. Future improvement may involve a graph-based tracking algorithm (Löffler et al., 2021) that could address the path interruption problem. Going further, a graph neural network (Weng et al., 2020) can also be an option in a tracking context.

7.2.2 Combination with Other Modalities

Multimodal analysis could also be a promising direction for this task, as pores are contrasted in high-resolution 2D images acquired using adaptive optics (AO, Figure 7.1). Such an AO image would allow us to select the most significant pores and extract additional information about their size and shape. In addition, multimodal image analysis using deep learning is also a trend in eye disease diagnosis (Jin et al., 2022).

7.2.3 Towards Clinical Applications

Ophthalmologists at the Quinze-Vingt Hospital have used the application developed by (Rossant et al., 2017) to carry out their studies (Bastelica, 2022). Their current method uses the semi-automatic method introduced in section 3.2.2 to track the pores. Our method is promising in helping them automate the reconstruction process.

The clinical applications of our research project could be: (1) axonal paths difference between healthy/glaucoma subjects; (2) axonal path modifications after surgery; (3) axonal path modifications for glaucoma suspect follow-ups. By analyzing pore paths, biomarkers characterising axonal paths can also be calculated for clinical studies, such as path tortuosity and verticality.

Chapter 8

Résumé en Français

Contents

8.1 Introduction	87
8.1.1 Contexte	87
8.1.2 Défis et contributions	89
8.2 État de l’Art sur les méthodes de Reconstruction des Pores de la LC en 3D	90
8.3 Recalage des Volumes OCT 3D Orthogonaux pour le Rehaussement des Pores de la LC	91
8.4 Attention U-Net pour la Segmentation des Pores de la LC	93
8.5 Suivi en 3D des Trajets Axonaux dans la LC	95
8.6 Conclusion et perspectives	97

8.1 Introduction

8.1.1 Contexte

Le glaucome est une neuropathie optique qui est la deuxième cause de cécité irréversible, touchant environ 80 millions de personnes dans le monde en 2020 sur la base d’études de prévalence (Quigley and Broman, 2006). Bien que les causes du glaucome ne soient pas encore claires, la lame criblée (LC) a été identifiée comme site principal des lésions dans le glaucome (Quigley et al., 1983). La LC est un maillage collagénique 3D dans la tête du nerf optique (ONH), composé de “pores” (i.e. trajets axonaux, Fig. 8.1) à travers lesquels les fibres nerveuses passent pour atteindre le cerveau.

L’observation in vivo en 3D des pores de la LC est désormais possible grâce aux

progrès de la tomographie de cohérence optique (OCT, Fig. 8.1). La modélisation des pores et de leur déformation au cours du glaucome à partir d'images OCT 3D permettrait de mieux comprendre cette pathologie cécitante.

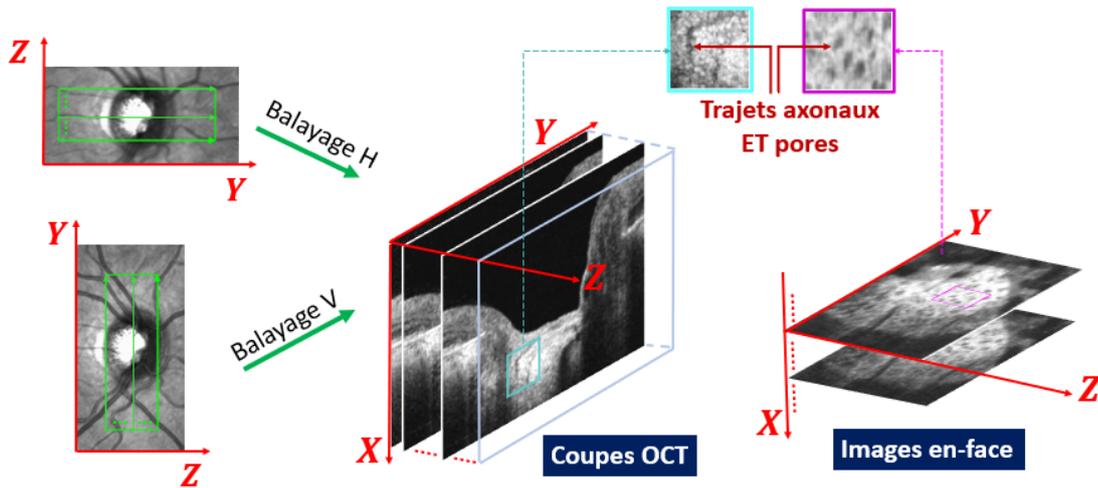


Figure 8.1: Données OCT 3D pour un balayage H ou V.

Dans le cadre de cette thèse, notre projet de recherche vise à caractériser les pores de la LC en cas de glaucome, ainsi que les modifications morphologiques survenant au cours de cette pathologie, par l'analyse automatique des images OCT 3D de la LC. Plus précisément, nous cherchons à concevoir une méthode de segmentation des trajets axonaux à partir des images OCT 3D. De cette segmentation, nous pourrions ensuite calculer des biomarqueurs caractérisant la LC et ses déformations au cours du glaucome. Pour cela, nous proposons des prétraitements permettant de rehausser les pores dans les images de faible rapport signal sur bruit, d'une part par le recalage de deux volumes OCT acquis dans les directions de balayage horizontale et verticale, et d'autre part par l'application de traitements de morphologie mathématique. Pour la segmentation, nous proposons une méthode en deux temps. Tout d'abord, nous avons conçu un réseau de neurones profond pour la segmentation des pores de la LC dans des images en-face, entraîné à partir de données partiellement annotées. Enfin nous reconstruisons les chemins axonaux en 3D par une procédure de suivi des pores segmentés.

8.1.2 Défis et contributions

Le principal défi est la qualité insuffisante des données (faible SNR, faible résolution transverse ($\sim 15\mu\text{m}$)), sachant qu’il est difficile, même pour des experts médicaux, d’identifier les pores dans une image (i.e. plan) en-face. En outre, les pores de la LC sont très petits et très peu contrastés, ce qui rend la segmentation automatique très difficile.

En réponse à cela, nous proposons d’exploiter le fait que les volumes OCT peuvent être acquis selon deux directions de balayage différentes, horizontale ou verticale, pour obtenir un volume fusionné de meilleure résolution et moins bruité. Pour ce faire, nous devons d’abord recalculer géométriquement les deux volumes orthogonaux. Notre méthode est basée sur l’intensité des pixels et est divisée en plusieurs étapes, afin de résoudre le problème d’optimisation de manière robuste et d’obtenir un recalage précis.

Le deuxième défi consiste à construire un jeu de données avec des annotations manuelles fiables dans les volumes OCT. En raison de la faible résolution des images OCT, de la petite taille et de la grande quantité de pores dans l’image, l’annotation manuelle est très difficile, car même les experts médicaux ont tendance à ne pas être d’accord sur ce qu’est un pore ou non.

Par conséquent, pour pouvoir détecter les pores les plus larges, ceux qui permettent à la plupart des axones de passer, nous avons proposé un réseau de neurones profond, un U-Net intégrant des modules d’attention (Oktay et al., 2018) et prenant en compte le contexte local par le traitement de plans adjacents. Cette architecture est bien adaptée à la segmentation précise des pores dans toutes les images en-face, malgré l’annotation incomplète des points. Elle nous permet de prédire avec robustesse et précision les pores candidats à la reconstruction 3D.

Enfin, nous avons proposé une méthode de reconstruction 3D des chemins axonaux à partir des pores segmentés dans les images en-face. Notre approche est fondée sur un modèle de contours actifs paramétriques dérivé du modèle de (Kass et al., 1988). Partant d’un plan de référence, nous appliquons ce modèle localement pour reconstruire pas à pas les trajectoires des pores. Notre méthode modélise explicitement les principales caractéristiques des chemins axonaux, à savoir leur faible intensité et leur régularité, ce qui conduit à un résultat robuste et précis.

8.2 État de l'Art sur les méthodes de Reconstruction des Pores de la LC en 3D

Les études récentes sur la LC se limitent principalement à l'analyse de l'épaisseur de la lame criblée (Devalla et al., 2018b) et de sa surface (Tan et al., 2015). Des changements morphologiques des pores, tels que leur allongement (Zwillinger et al., 2016), ont été observés dans des images en-face 2D d'optique adaptative (OA), donc en haute résolution, chez des patients glaucomateux.

A notre connaissance, seulement deux travaux, dont les nôtres, ont porté sur la reconstruction des pores de la LC en 3D et in-vivo (Rossant et al., 2017; Wang et al., 2018) grâce à l'OCT. Plus précisément, les deux méthodes peuvent être divisées en trois étapes principales : le pré-traitement pour la rehaussement des pores et la réduction du bruit, la segmentation des pores pour extraire les pores dans les images en-face en 2D, et enfin l'étape de reconstruction pour relier les pores dans les images en-face adjacentes.

Pour le pré-traitement, les auteurs de (Wang et al., 2018) ont appliqué un filtre Gaussien 3D pour réduire le bruit dans les hautes fréquences, ainsi que la méthode d'égalisation d'histogramme adaptatif limité par contraste (CLAHE) pour égaliser les différences d'intensité locales. Dans (Rossant et al., 2017), les auteurs ont proposé d'utiliser des filtres morphologiques, notamment les filtres alternés séquentiels et le filtre Bottom-Hat, pour débruiter et rehausser les caractéristiques des pores.

En ce qui concerne la segmentation des pores dans les images en-face, les auteurs de (Wang et al., 2018) ont utilisé le seuillage local adaptatif (Niblack, 1985), ainsi qu'un filtre médian 3D avec une taille de noyau de $1 \times 1 \times 3$ afin de tenir compte de la continuité des pores dans les images en-face adjacentes. Dans (Rossant et al., 2017), les pores ont été sélectionnés comme des minima locaux d'intensité les plus contrastés dans les images en-face. Enfin, pour reconstituer individuellement les trajets des pores segmentés, les auteurs de (Wang et al., 2018) ont utilisé l'algorithme de filtre particulière fourni par le logiciel MTrack2 (Stuurman, 2003), tandis que le filtre de Kalman (Welch et al., 1995) et le modèle de contour actif (Kass et al., 1988) ont été implémentés dans (Rossant et al., 2017).

8.3 Recalage des Volumes OCT 3D Orthogonaux pour le Rehaussement des Pores de la LC

La limite principale que nous avons identifiée dans notre état de l'art des méthodes de reconstruction des pores est que les trajets axonaux n'ont pas pu être caractérisés avec suffisamment de fiabilité, à cause de la qualité insuffisante des données. D'autre part, les deux méthodes citées précédemment nécessitent une délimitation manuelle des masques périphériques dans les images en-face, afin de ne traiter que les régions avec des pores identifiables. Nous proposons donc une méthode entièrement automatique pour déterminer le volume d'intérêt et pour aligner géométriquement les volumes OCT acquis dans les deux directions orthogonales de balayage (Figure 8.2). Cette étape permet d'obtenir une meilleure résolution des pores de la LC, sachant que, pour chaque volume, la distance entre deux coupes B-scan est de $15 \mu\text{m}$ tandis que la résolution latérale est de $5.7 \mu\text{m}/\text{pixel}$. Le rapport signal à bruit est aussi amélioré par la fusion des volumes recalés. A notre connaissance, c'est la première fois que l'étude est menée conjointement sur les volumes orthogonaux pour rehausser les pores.

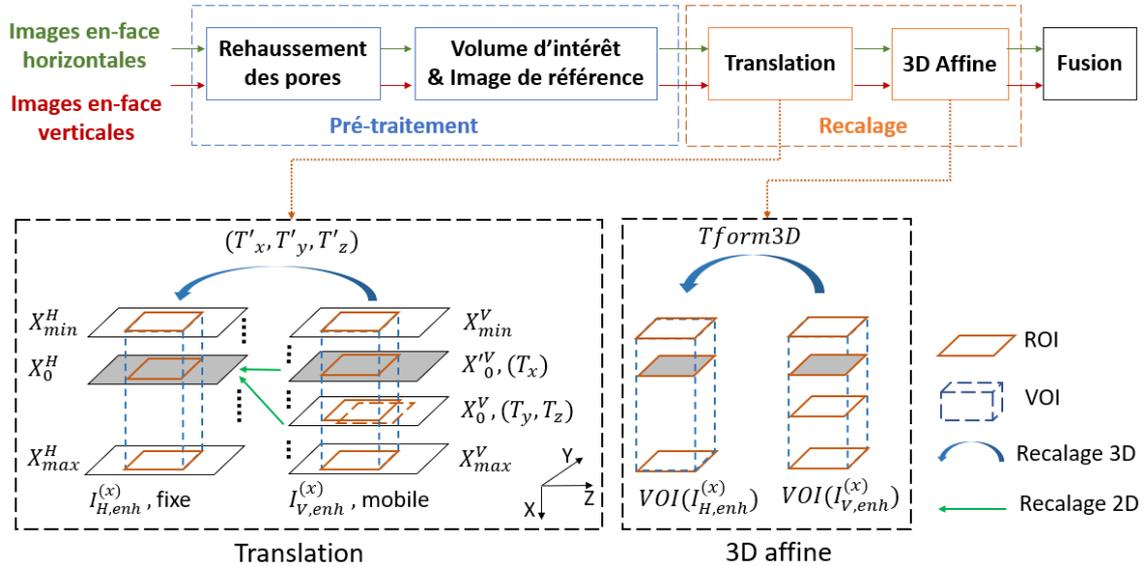


Figure 8.2: Principales étapes du recalage 3D, avec le pré-traitement pour rehausser des pores et sélectionner le volume d'intérêt.

Notre méthode prend en entrée les images en-face extraites des volumes H et V (Figure 8.2). Dans un premier temps, on rehausse les pores par morphologie mathématique pour pallier le faible SNR. On détermine pour chaque volume la région d'intérêt (ROI)

Table 8.1: Evaluation du recalage proposé sur le volume d'intérêt et sur 3 images en-face.

		<i>GT</i>		Notre méthode	
		Trans.	Affine	Trans.	Affine
VOI	Corr	0.59±0.09	0.70±0.06	0.69±0.11	0.71±0.09
	MinDist	3.59±0.65	3.05±0.73	4.82±0.67	3.73±0.49
X_{0-20}	Corr	0.56±0.09	0.68±0.10	0.69±0.07	0.71±0.06
	MinDist	3.99±1.03	3.20±1.18	5.08±1.43	4.22±0.77
X_0	Corr	0.63±0.06	0.71±0.07	0.72±0.08	0.78±0.06
	MinDist	3.41±0.18	2.53±0.54	3.95±1.09	3.16±0.70
X_{0+20}	Corr	0.54±0.13	0.68±0.10	0.69±0.08	0.70±0.08
	MinDist	4.13±0.82	3.63±0.99	6.42±1.51	5.12±1.23

dans les images en-face, puis un plan de référence, c'est-à-dire l'abscisse qui correspond à l'image en-face où les pores sont les plus nombreux et les mieux contrastés, grâce à l'optimisation d'une fonction de coût. Nous en déduisons également la zone d'intérêt dans la direction axiale, ce qui nous donne des volumes d'intérêt (VOI). Le recalage se fait par maximisation de l'intercorrélation et est divisé en plusieurs étapes, afin de résoudre le problème d'optimisation de manière robuste. Tout d'abord, nous utilisons les plans de référence dans lesquels l'information est la plus riche pour recalculer grossièrement le volume V sur le volume H par une translation 2D. Puis nous recherchons une translation verticale qui améliore le recalage, ce qui met à jour le plan de référence du volume V recalé. Ce processus est itéré tant que l'intercorrélation est améliorée. Ensuite, nous affinons ce résultat par une translation 3D sur les volumes d'intérêt puis par une transformation affine 3D. Finalement, les volumes alignés sont fusionnés pour produire un volume de meilleure résolution spatiale et de meilleur SNR.

Le tableau 1 présente les statistiques obtenues sur les 21 paires de volumes orthogonaux de notre base de données. Notre méthode de recalage conduit à un score de corrélation élevé sur le volume d'intérêt avec un faible écart-type (0.71 ± 0.09). Les scores de corrélation sont les plus élevés autour des images de référence sélectionnées automatiquement, là où il y a le plus d'information avec de nombreux pores contrastés. Les distances moyennes entre les pores annotés alignés sont faibles (3.73 ± 0.49 pixels) par rapport aux tailles typiques des pores (de 5 à 20 pixels, moyenne autour de 10); elles sont très proches de celles de la vérité terrain, alors que, dans ce cas, c'est le critère

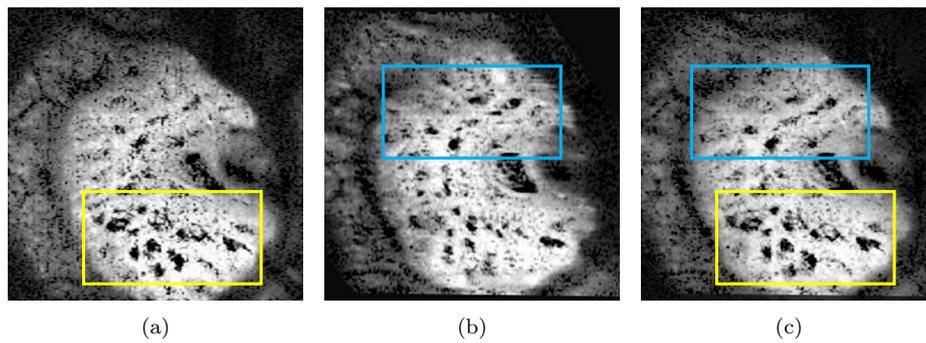


Figure 8.3: Exemple de recalage et de fusion sur les images rehaussées. (a) Image horizontale ; (b) Image verticale recalée ; (c) Image de fusion.

de distance qui est optimisé pour le recalage.

La transformée affine 3D apporte un gain significatif par rapport à la translation seule, que l'on recale sur les données (notre méthode) ou sur les pores annotés (GT), et ce en termes de distances et de corrélation. Tout ceci valide la méthode proposée, en ce qui concerne la robustesse et la précision des résultats.

Afin d'obtenir une meilleure reconstruction des trajets axonaux en 3D pour le suivi longitudinal des patients atteints de glaucome, nous proposons ensuite d'entraîner un réseau de neurones convolutionnel sur les images en-face H, V et fusionnées, afin d'obtenir de meilleurs résultats de segmentation.

8.4 Attention U-Net pour la Segmentation des Pores de la LC

Les pores dans les volumes recalés et fusionnés dans la section précédente sont mieux contrastés et mieux résolus par rapport aux volumes sources, ce qui permet de concevoir une méthode robuste pour les segmenter dans les images en-face.

Nous proposons donc une méthode de segmentation automatique des pores de la LC dans les images OCT en-face par apprentissage profond. C'est, à notre connaissance, la première fois que ce problème est abordé avec ce type d'approche, de plus sur les images H et V recalées. Les méthodes de segmentation faiblement supervisées ne permettent pas d'atteindre la fiabilité requise pour les études cliniques, car les difficultés sont trop importantes : très mauvais rapport SNR, faible résolution des images par rapport à la taille des pores, grande variabilité de forme et d'intensité de ces derniers. En revanche, les réseaux de neurones de type U-Net (Ronneberger et al., 2015) ont

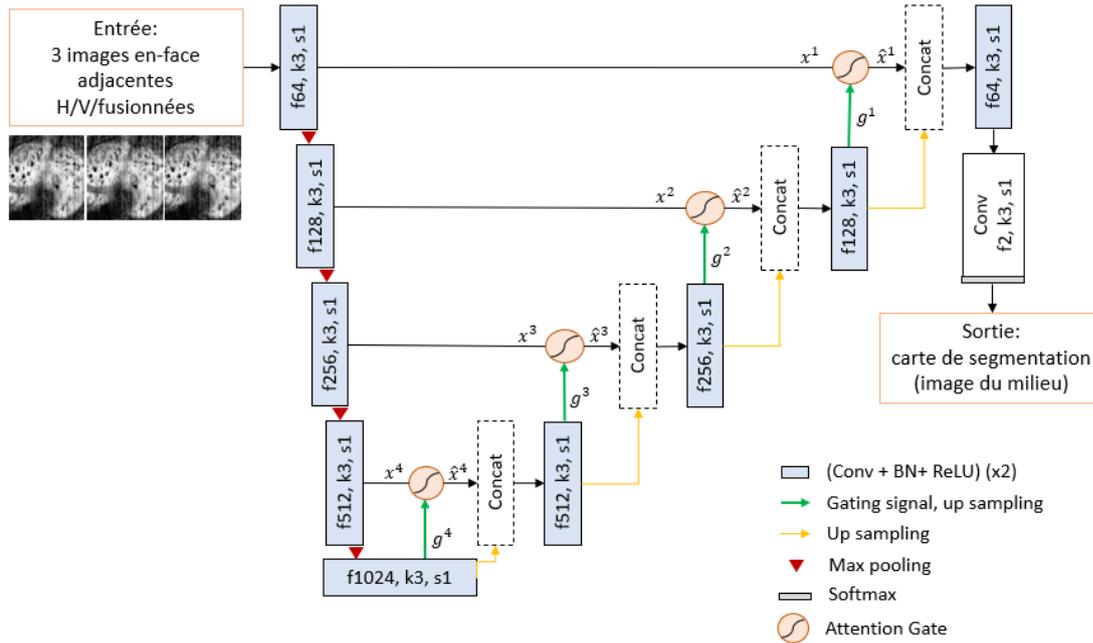


Figure 8.4: L'architecture d'attention U-Net adaptée de (Oktay et al. (2018)).

démonstré leur capacité à modéliser des problèmes complexes de segmentation par les données, y compris pour le traitement de structures ONH (Devalla et al., 2018b), même de petite taille (Chen et al., 2020). Notre objectif est d'obtenir, grâce à un réseau U-net adapté et optimisé, une segmentation précise des pores dans toutes les images en-face, suffisamment fiable pour permettre une future reconstruction 3D automatique des trajets axonaux, et ceci malgré les difficultés d'annotation manuelles des images pour l'apprentissage.

La méthode proposée avec l'attention U-Net est illustrée sur la Figure 8.4. Les *skip connexions* du U-Net aident à récupérer les informations spatiales perdues dans l'encodeur, permettant ainsi une localisation précise des objets d'intérêt. Pour notre application, les pores ont généralement une taille inférieure à 5×5 pixels pour des images d'entrée (zone d'intérêt) de 262×262 pixels. Afin d'améliorer la détection des petits objets, les *features* produites peuvent être améliorées en intégrant des mécanismes d'attention (*Attention Gate*, AG (Oktay et al., 2018)) dans l'U-Net, pour aider à capturer les régions d'intérêt. L'AG permet d'estimer les zones potentielles où les pores sont les plus susceptibles d'apparaître en supprimant l'activation des *features* dans les régions non pertinentes, le tout sans avoir besoin de supervision externe.

D'autre part, les trajets axonaux étant assez réguliers, les intensités des pores sont

similaires dans les images en-face adjacentes, tandis que leur centroïde et leur forme varient également peu. Par conséquent, une application naïve de U-Net prend le risque de passer à côté de ces propriétés de régularité. De ce fait, nous proposons un réseau prenant en entrée 3 images adjacentes de volumes H/V/fusionnées, et nous produisons une seule carte de segmentation pour l’image du milieu.

Table 8.2: Résultats de la segmentation sur les données de test. Nous privilégions la métrique *Recall* car la vérité de terrain est incomplète.

Modèle	Pixel-level		Object-level	
	Dice	Jaccard	Precision	Recall
Chan_Vese	0.084	0.056	0.126	0.677
W-Net	0.101	0.056	0.127	0.694
U-Net	0.265	0.169	0.262	0.656
Attention U-Net	0.269	0.173	0.283	0.697
Notre méthode	0.293	0.178	0.275	0.746

Les résultats expérimentaux, affichés dans le tableau 8.2, montrent que l’approche proposée est robuste pour identifier les vrais pores avec une valeur de rappel (*Recall*) élevée. Les pores manquants dans la segmentation automatique sont principalement situés en périphérie de la LC, là où le SNR est le plus dégradé, et où il est difficile, même pour des experts, d’identifier des pores sans se référer à la continuité. Avec les images fusionnées supplémentaires pour l’entraînement, nous obtenons un score de rappel plus élevé qu’avec les images horizontales ou verticales.

8.5 Suivi en 3D des Trajets Axonaux dans la LC

Avec les cartes de segmentation précises fournies par la section précédente, nous sommes capable de relier des pores dans les images fusionnées adjacentes. Pour cela, nous proposons une méthode de suivi fondée sur un modèle de contour actif paramétrique dérivé du modèle initial de (Kass et al., 1988). Partant du plan de référence où les pores segmentés sont les plus nombreux et les plus contrastés, nous progressons selon l’axe vertical (abscisse x vers $x + 1$ ou $x - 1$) pour relier les pores segmentés et former les chemins axonaux. Pour cela, nous proposons d’initialiser au centre de gravité de chaque pore l du plan x un petit segment vertical $V_x^{(l)}$ et de le déformer par minimisation d’une fonctionnelle d’énergie $E(V_x^{(l)})$ similaire à celle de (Kass et al., 1988):

$$E(V_x^{(l)}) = \int_{x-\delta}^{x+\delta} \alpha |V_x^{(l)'}(s)|^2 + \beta |V_x^{(l)''}(s)|^2 + I_{fus}(V_x^{(l)}(s)) ds \quad (8.1)$$

Où I_{fus} est l'image fusionnée avec la méthode proposée précédemment. Le modèle classique de contour actif a été défini pour des images 2D et n'est pas directement extensible aux images 3D. Cependant, ici nous considérons une courbe discrète, avec un unique point par plan en-face et chaque point évolue horizontalement dans son propre plan. δ détermine la longueur des chemins. Les deux premiers termes de (8.1) imposent à la courbe d'être régulière, et le dernier la pousse vers les minima d'intensité, donc les pores. Notre implémentation permet de traiter simultanément tous les chemins d'un plan en-face. Comme les chemins sont petits ($\delta = 3$) et les segments par construction proches des minima recherchés, seulement quelques itérations sont nécessaires pour faire converger la courbe, de sorte que l'algorithme n'est pas lourd. On progresse verticalement en réinitialisant le modèle dans le plan suivant ($x + 1$ ou $x - 1$) à partir des positions trouvées, pour toutes celles qui tombent dans un pore segmenté par le U-Net.

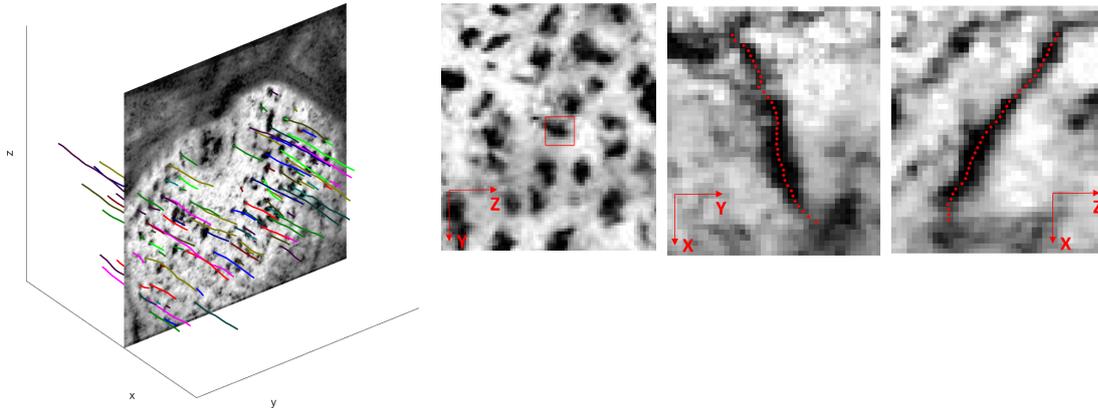


Figure 8.5: Reconstruction 3D des trajets axonaux. Rendu final (gauche) et focus sur un pore (droite). Les deux images de droite sont calculées par interpolation.

Cette méthode est simple mais efficace, puisqu'elle modélise explicitement les caractéristiques principales des trajets axonaux, à savoir leur faible intensité et leur régularité. Elle opère moins localement que d'autres approches comme le filtre de Kalman (Rossant et al., 2017), ce qui la rend plus robuste et plus précise. L'exemple dans la Figure 8.5 (gauche) montre que la méthode proposée est capable de reconstruire les trajets principaux par lesquels passent la plupart des axones. Dans la Figure 8.5 à

droite, le trajet est très contrasté et la trajectoire est déterminée sans ambiguïté.

8.6 Conclusion et perspectives

Dans ce manuscrit, nous présentons nos travaux de recherches visant à proposer une méthode automatique de reconstruction en 3D des pores de la LC en utilisant des images OCT en 3D. Les résultats sont très encourageants étant donnée la difficulté du problème (images peu résolues et très bruitées, annotations partielles en apprentissage). Pour la première fois, deux volumes orthogonaux peuvent être exploités conjointement pour analyser les trajets des pores de la LC, ce qui facilite beaucoup la lecture des images et l'interprétation médicale.

Tout au long du manuscrit, nous fournissons des résultats expérimentaux permettant d'évaluer quantitativement chaque étape de notre méthode et son apport. Nous avons montré que notre algorithme permet de traiter les données OCT 3D de patients atteints de glaucome, et qu'elle aboutit à une reconstruction 3D fiable et précise des principaux trajets axonaux. Considérant l'état de l'art, nous pensons que notre travail a permis des progrès significatifs dans l'analyse automatisées des données OCT 3D de la lame criblée, et que l'application de notre algorithme dans un contexte clinique va pouvoir fournir des informations précieuses aux médecins pour l'étude du glaucome.

Les perspectives de ce travail sont nombreuses. Dans un premier temps, nous pouvons travailler sur des schémas de fusion des données H et V plus avancés, afin d'exploiter localement le meilleur des deux images. Ces schémas de fusion peuvent se faire en amont du réseau de neurones, avec des méthodes plus sophistiquées qu'un simple opérateur ET. On peut aussi tester des schémas de fusion en entrée ou dans les couches intermédiaires d'un réseau de neurones profond. D'autres perspectives concernent l'exploitation des segmentations obtenues, par le calcul de biomarqueurs caractérisant les trajets axonaux, puis l'analyse des déformations de ces trajets pour le suivi longitudinal. Enfin, nous pouvons enrichir notre travail d'une analyse multi-modale, en intégrant une image 2D de haute résolution acquise en optique adaptative. Cette image permettrait de sélectionner les pores les plus significatifs et d'extraire des informations complémentaires sur leur taille et leur forme.

Bibliography

- Adler, D. C., Ko, T. H., and Fujimoto, J. G. (2004). Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter. *Optics letters*, 29(24):2878–2880.
- Aja, S., Alberola, C., and Ruiz, A. (2001). Fuzzy anisotropic diffusion for speckle filtering. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, volume 2, pages 1261–1264. IEEE.
- Anvari, P., Ashrafkhorasani, M., Habibi, A., and Falavarjani, K. G. (2021). Artifacts in optical coherence tomography angiography. *Journal of Ophthalmic & Vision Research*, 16(2):271.
- Avanaki, M. R., Cernat, R., Tadrous, P. J., Tatla, T., Podoleanu, A. G., and Hojjatoleslami, S. A. (2013). Spatial compounding algorithm for speckle reduction of dynamic focus oct images. *IEEE Photonics Technology Letters*, 25(15):1439–1442.
- Bastelica, P. (2022). *Microarchitecture 3D des pores de la lame criblée chez des sujets sains et glaucomateux en tomographie en cohérence optique*. PhD thesis.
- Belghith, A., Bowd, C., Medeiros, F. A., Weinreb, R. N., and Zangwill, L. M. (2015). Automated segmentation of anterior lamina cribrosa surface: How the lamina cribrosa responds to intraocular pressure change in glaucoma eyes? In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 222–225. IEEE.
- Bennett, T. J. (2020). Scanning laser ophthalmoscopy.
- Chan, T. and Vese, L. (2001). Active contour without edges. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 10:266–77.
- Chen, Z., Li, D., Shen, H., Mo, H., Zeng, Z., and Wei, H. (2020). Automated segmen-

- tation of fluid regions in optical coherence tomography b-scan images of age-related macular degeneration. *Optics & Laser Technology*, 122:105830.
- Chen, Z., Peng, P., Shen, H., Wei, H., Ouyang, P., and Duan, X. (2019). Region-segmentation strategy for bruch’s membrane opening detection in spectral domain optical coherence tomography images. *Biomedical optics express*, 10(2):526–538.
- Cheng, Y., Chu, Z., and Wang, R. K. (2021). Robust three-dimensional registration on optical coherence tomography angiography for speckle reduction and visualization. *Quantitative Imaging in Medicine and Surgery*, 11(3):879.
- Chhablani, J., Krishnan, T., Sethi, V., and Kozak, I. (2014). Artifacts in optical coherence tomography. *Saudi Journal of Ophthalmology*, 28(2):81–87.
- Chiu, S. J., Li, X. T., Nicholas, P., Toth, C. A., Izatt, J. A., and Farsiu, S. (2010). Automatic segmentation of seven retinal layers in sdopt images congruent with expert manual segmentation. *Optics express*, 18(18):19413–19428.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer.
- Dandona, L., Quigley, H. A., Brown, A. E., and Enger, C. (1990). Quantitative regional structure of the normal human lamina cribrosa: a racial comparison. *Archives of ophthalmology*, 108(3):393–398.
- Devalla, S. K., Chin, K. S., Mari, J.-M., Tun, T. A., Strouthidis, N. G., Aung, T., Thiéry, A. H., and Girard, M. J. (2018a). A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head. *Investigative ophthalmology & visual science*, 59(1):63–74.
- Devalla, S. K., Pham, T. H., Panda, S. K., Zhang, L., Subramanian, G., Swaminathan, A., Yun, C. Z., Rajan, M., Mohan, S., Krishnadas, R., et al. (2020). Towards label-free 3d segmentation of optical coherence tomography images of the optic nerve head using deep learning. *Biomedical optics express*, 11(11):6356–6378.
- Devalla, S. K., Renukanand, P. K., Sreedhar, B. K., Subramanian, G., Zhang, L., Perera, S., Mari, J.-M., Chin, K. S., Tun, T. A., Strouthidis, N. G., et al. (2018b).

- Drunet: a dilated-residual u-net deep learning network to segment optic nerve head tissues in optical coherence tomography images. *Biomedical optics express*, 9(7):3244–3265.
- Ding, N., Rossant, F., Urien, H., Sublime, J., Bastelica, P., Baudouin, C., and Paques, M. (2024). A complete method for the 3d reconstruction of axonal pathways from 2 orthogonal 3d oct images of the lamina cribrosa. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2360–2364. IEEE.
- Ding, N., Rossant, F., Urien, H., Sublime, J., and Paques, M. (2023). 3d orthogonal sd-oct volumes registration for the enhancement of pores in lamina cribrosa. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE.
- Ding, N., Urien, H., Rossant, F., Sublime, J., Bastelica, P., and Paques, M. (2022a). Attention u-net pour la segmentation des pores de la lame criblée. In *GRETSI*.
- Ding, N., Urien, H., Rossant, F., Sublime, J., and Paques, M. (2022b). Context-aware attention u-net for the segmentation of pores in lamina cribrosa using partial points annotation. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 537–542. IEEE.
- Downs, J. C. and Girkin, C. A. (2017). Lamina cribrosa in glaucoma. *Current opinion in ophthalmology*, 28(2):113.
- Dubra, A., Sulai, Y., Norris, J. L., Cooper, R. F., Dubis, A. M., Williams, D. R., and Carroll, J. (2011). Noninvasive imaging of the human rod photoreceptor mosaic using a confocal adaptive optics scanning ophthalmoscope. *Biomedical optics express*, 2(7):1864–1876.
- Faridi, O. S., Park, S. C., Kabadi, R., Su, D., De Moraes, C. G., Liebmann, J. M., and Ritch, R. (2014). Effect of focal lamina cribrosa defect on glaucomatous visual field progression. *Ophthalmology*, 121(8):1524–1530.
- Ferguson, R. D., Zhong, Z., Hammer, D. X., Mujat, M., Patel, A. H., Deng, C., Zou, W., and Burns, S. A. (2010). Adaptive optics scanning laser ophthalmoscope with integrated wide-field retinal imaging and tracking. *JOSA A*, 27(11):A265–A277.

- Fitzke, F. and Masters, B. (1993). Three-dimensional visualization of confocal sections of in vivo human fundus and optic nerve. *Current eye research*, 12(11):1015–1018.
- Girard, M. J., Tun, T. A., Husain, R., Acharyya, S., Haaland, B. A., Wei, X., Mari, J. M., Perera, S. A., Baskaran, M., Aung, T., et al. (2015). Lamina cribrosa visibility using optical coherence tomography: comparison of devices and effects of image enhancement techniques. *Investigative ophthalmology & visual science*, 56(2):865–874.
- Girkin, C. A., Fazio, M. A., Yang, H., Reynaud, J., Burgoyne, C. F., Smith, B., Wang, L., and Downs, J. C. (2017). Variation in the three-dimensional histomorphometry of the normal human optic nerve head with age and race: lamina cribrosa and peripapillary scleral thickness and position. *Investigative ophthalmology & visual science*, 58(9):3759–3769.
- Golabbakhsh, M., Rabbani, H., and Esmaeili, M. (2012). Detection and registration of vessels of fundus and oct images using curvelet analysis. In *2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE)*, pages 594–597. IEEE.
- Gonzalez, R. C. (2009). *Digital image processing*. Pearson education india.
- Greig, E. C., Duker, J. S., and Waheed, N. K. (2020). A practical guide to optical coherence tomography angiography interpretation. *International journal of retina and vitreous*, 6(1):1–17.
- Grytz, R., Meschke, G., Jonas, J. B., and Downs, J. C. (2016). Glaucoma and structure-based mechanics of the lamina cribrosa at multiple scales. *Structure-Based Mechanics of Tissues and Organs*, pages 93–122.
- Grzybowski, A., Och, M., Kanclerz, P., Leffler, C., and De Moraes, C. G. (2020). Primary open angle glaucoma and vascular risk factors: a review of population based studies from 1990 to 2019. *Journal of clinical medicine*, 9(3):761.
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

- Hongwei, Z., Baowang, L., and Juan, F. (2011). Adaptive wavelet transformation for speckle reduction in optical coherence tomography images. In *2011 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 1–5. IEEE.
- Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., Hee, M. R., Flotte, T., Gregory, K., Puliafito, C. A., et al. (1991). Optical coherence tomography. *science*, 254(5035):1178–1181.
- Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211.
- Ivers, K. M., Sredar, N., Patel, N. B., Rajagopalan, L., Queener, H. M., Twa, M. D., Harwerth, R. S., and Porter, J. (2015). In vivo changes in lamina cribrosa microarchitecture and optic nerve head structure in early experimental glaucoma. *PloS one*, 10(7):e0134223.
- Jin, K., Yan, Y., Chen, M., Wang, J., Pan, X., Liu, X., Liu, M., Lou, L., Wang, Y., and Ye, J. (2022). Multimodal deep learning with feature level fusion for identification of choroidal neovascularization activity in age-related macular degeneration. *Acta Ophthalmologica*, 100(2):e512–e520.
- Jonas, J., Mardin, C. Y., Schlötzer-Schrehardt, U., and Naumann, G. (1991). Morphometry of the human lamina cribrosa surface. *Investigative ophthalmology & visual science*, 32(2):401–405.
- Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331.
- Kenry, Duan, Y., and Liu, B. (2018). Recent advances of optical imaging in the second near-infrared window. *Advanced materials*, 30(47):1802394.
- Khansari, M. M., Zhang, J., Qiao, Y., Gahm, J. K., Sarabi, M. S., Kashani, A. H., and Shi, Y. (2019). Automated deformation-based analysis of 3d optical coherence tomography in diabetic retinopathy. *IEEE transactions on medical imaging*, 39(1):236–245.
- Kim, J. and Fessler, J. A. (2004). Intensity-based image registration using robust correlation coefficients. *IEEE transactions on medical imaging*, 23(11):1430–1444.

- Kim, Y. W., Jeoung, J. W., Kim, D. W., Girard, M. J., Mari, J. M., Park, K. H., and Kim, D. M. (2016). Clinical assessment of lamina cribrosa curvature in eyes with primary open-angle glaucoma. *PloS one*, 11(3):e0150260.
- Kiumehr, S., Park, S. C., Dorairaj, S., Teng, C. C., Tello, C., Liebmann, J. M., and Ritch, R. (2012). In vivo evaluation of focal lamina cribrosa defects in glaucoma. *Archives of ophthalmology*, 130(5):552–559.
- Kobayashi, M., Hanafusa, H., Takada, K., and Noda, J. (1991). Polarization-independent interferometric optical-time-domain reflectometer. *Journal of Lightwave Technology*, 9(5):623–628.
- Kolar, R. and Tasevsky, P. (2010). Registration of 3d retinal optical coherence tomography data and 2d fundus images. In *International Workshop on Biomedical Image Registration*, pages 72–82. Springer.
- Kraus, M. F., Liu, J. J., Schottenhamml, J., Chen, C.-L., Budai, A., Branchini, L., Ko, T., Ishikawa, H., Wollstein, G., Schuman, J., et al. (2014). Quantitative 3d-oct motion correction with tilt and illumination correction, robust similarity measure and regularization. *Biomedical optics express*.
- Leske, M. C., Heijl, A., Hyman, L., Bengtsson, B., Dong, L., Yang, Z., group, E., et al. (2007). Predictors of long-term progression in the early manifest glaucoma trial. *Ophthalmology*, 114(11):1965–1972.
- Liu, Y., Chen, X., Peng, H., and Wang, Z. (2017). Multi-focus image fusion with a deep convolutional neural network. *Information Fusion*, 36:191–207.
- Löffler, K., Scherr, T., and Mikut, R. (2021). A graph-based cell tracking algorithm with few manually tunable parameters and automated segmentation error correction. *PloS one*, 16(9):e0249257.
- Mao, Z., Miki, A., Mei, S., Dong, Y., Maruyama, K., Kawasaki, R., Usui, S., Matsushita, K., Nishida, K., and Chan, K. (2019). Deep learning based noise reduction method for automatic 3d segmentation of the anterior of lamina cribrosa in optical coherence tomography volumetric scans. *Biomedical optics express*, 10(11):5832–5851.
- Maragos, P. and Pessoa, L. F. (1999). Morphological filtering for image enhancement and detection. *The Image and Video Processing Handbook*,, pages 135–156.

- Mari, J. M., Strouthidis, N. G., Park, S. C., and Girard, M. J. (2013). Enhancement of lamina cribrosa visibility in optical coherence tomography images using adaptive compensation. *Investigative ophthalmology & visual science*, 54(3):2238–2247.
- Marques, R., De Jesus, D. A., Barbosa-Breda, J., Van Eijgen, J., Stalmans, I., van Walsum, T., Klein, S., Vaz, P. G., and Brea, L. S. (2022). Automatic segmentation of the optic nerve head region in optical coherence tomography: A methodological review. *Computer Methods and Programs in Biomedicine*, 220:106801.
- Miri, M. S., Abràmoff, M. D., Kwon, Y. H., and Garvin, M. K. (2016). Multimodal registration of sd-oct volumes and fundus photographs using histograms of oriented gradients. *Biomedical Optics Express*, 7(12):5252–5267.
- Mokhtari, M., Rabbani, H., and Mehri-Dehnavi, A. (2017). Alignment of optic nerve head optical coherence tomography b-scans in right and left eyes. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2279–2283. IEEE.
- Nadler, Z., Wang, B., Wollstein, G., Nevins, J. E., Ishikawa, H., Kagemann, L., Sigal, I. A., Ferguson, R. D., Hammer, D. X., Grulkowski, I., et al. (2013). Automated lamina cribrosa microstructural segmentation in optical coherence tomography scans of healthy and glaucomatous eyes. *Biomedical optics express*, 4(11):2596–2608.
- Niblack, W. (1985). *An introduction to digital image processing*. Strandberg Publishing Company.
- Oktay, O., Schlemper, J., Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N., Kainz, B., Glocker, B., and Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas.
- Omodaka, K., Horii, T., Takahashi, S., Kikawa, T., Matsumoto, A., Shiga, Y., Maruyama, K., Yuasa, T., Akiba, M., and Nakazawa, T. (2015). 3d evaluation of the lamina cribrosa with swept-source optical coherence tomography in normal tension glaucoma. *PLOS ONE*, 10:e0122347.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*.
- Pan, L., Guan, L., and Chen, X. (2019). Segmentation guided registration for 3d spectral-domain optical coherence tomography images. *IEEE Access*, 7:138833–138845.

- Pan, L., Shi, F., Xiang, D., Yu, K., Duan, L., Zheng, J., and Chen, X. (2020). Octrexpert: a feature-based 3d registration method for retinal oct images. *IEEE Transactions on Image Processing*, 29:3885–3897.
- Park, H.-Y. L., Jeon, S. H., and Park, C. K. (2012). Enhanced depth imaging detects lamina cribrosa thickness differences in normal tension glaucoma and primary open-angle glaucoma. *Ophthalmology*, 119(1):10–20.
- Pekala, M., Joshi, N., Liu, T. A., Bressler, N. M., DeBuc, D. C., and Burlina, P. (2019). Deep learning based retinal oct segmentation. *Computers in biology and medicine*, 114:103445.
- Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639.
- Pircher, M., Goetzinger, E., Leitgeb, R., Fercher, A. F., and Hitzenberger, C. K. (2003). Speckle reduction in optical coherence tomography by frequency compounding. *Journal of biomedical optics*, 8(3):565–569.
- Ploner, S. B., Kraus, M. F., Moulton, E. M., Husvogt, L., Schottenhamml, J., Alibhai, A. Y., Waheed, N. K., Duker, J. S., Fujimoto, J. G., and Maier, A. K. (2021). Efficient and high accuracy 3-d oct angiography motion correction in pathology. *Biomedical Optics Express*, 12(1):125–146.
- Puvanathan, P. and Bizheva, K. (2009). Interval type-ii fuzzy anisotropic diffusion algorithm for speckle noise reduction in optical coherence tomography images. *Optics express*, 17(2):733–746.
- Quigley, H. and Broman, A. (2006). The number of people with glaucoma worldwide in 2010 and 2020. *The British journal of ophthalmology*, 90:262–7.
- Quigley, H. A., Addicks, E. M., Green, W. R., and Maumenee, A. (1981). Optic nerve damage in human glaucoma: II. the site of injury and susceptibility to damage. *Archives of ophthalmology*, 99(4):635–649.
- Quigley, H. A., Hohman, R. M., Addicks, E. M., Massof, R. W., and Green, W. R. (1983). Morphologic changes in the lamina cribrosa correlated with neural loss in open-angle glaucoma. *American journal of ophthalmology*, 95(5):673–691.

- Rahman, M. H., Jeong, H. W., Kim, N. R., and Kim, D. Y. (2021). Automatic quantification of anterior lamina cribrosa structures in optical coherence tomography using a two-stage cnn framework. *Sensors*, 21(16):5383.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.
- Ren, R., Yang, H., Gardiner, S. K., Fortune, B., Hardin, C., Demirel, S., and Burgoyne, C. F. (2014). Anterior lamina cribrosa surface depth, age, and visual field sensitivity in the portland progression project. *Investigative ophthalmology & visual science*, 55(3):1531–1539.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation.
- Roorda, A., Romero-Borja, F., Donnelly III, W. J., Queener, H., Hebert, T. J., and Campbell, M. C. (2002). Adaptive optics scanning laser ophthalmoscopy. *Optics express*, 10(9):405–412.
- Roorda, A. and Williams, D. R. (1999). The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520–522.
- Rossant, F., Ding, N., Urien, H., Sublime, J., Bastelica, P., Baudouin, C., and Paques, M. (2023). Recalage des volumes oct 3d orthogonaux pour le rehaussement des pores de la lame criblée. In *GRETI*.
- Rossant, F., Grieve, K., Zwillinger, S., and Paques, M. (2017). Detection and tracking of the pores of the lamina cribrosa in three dimensional sd-oct data. pages 651–663.
- Royer, C., Sublime, J., Rossant, F., and Paques, M. (2021). Imaging unsupervised approaches for the segmentation of dry amd lesions in eye fundus cslo images. *Journal of Imaging*, 7:143.
- Schmitt, J. (1997). Array detection for speckle reduction in optical coherence microscopy. *Physics in Medicine & Biology*, 42(7):1427.
- Serra, J. (1982). Image analysis and mathematical morphology. (*No Title*).
- Shehryar, T., Akram, M. U., Khalid, S., Nasreen, S., Tariq, A., Perwaiz, A., and Shaikat, A. (2020). Improved automated detection of glaucoma by correlating fundus

- and sd-oct image analysis. *International Journal of Imaging Systems and Technology*, 30(4):1046–1065.
- Shi, F., Chen, X., Zhao, H., Zhu, W., Xiang, D., Gao, E., Sonka, M., and Chen, H. (2014). Automated 3-d retinal layer segmentation of macular optical coherence tomography images with serous pigment epithelial detachments. *IEEE transactions on medical imaging*, 34(2):441–452.
- Simard, P., Steinkraus, D., and Platt, J. (2003). Best practices for convolutional neural networks applied to visual document analysis. pages 958–962.
- Sredar, N., Ivers, K. M., Queener, H. M., Zouridakis, G., and Porter, J. (2013). 3d modeling to characterize lamina cribrosa surface and pore geometries using in vivo images from normal and glaucomatous eyes. *Biomedical optics express*, 4(7):1153–1165.
- Stuurman, N. (2003). Mtrack2 software. <http://valelab.ucsf.edu/~nico/IJplugins/MTrack2.html/>.
- Sudre, C., Li, W., Vercauteren, T., Ourselin, S., and Cardoso, M. J. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations.
- Suzuki, Y., Iwase, A., Araie, M., Yamamoto, T., Abe, H., Shirato, S., Kuwayama, Y., Mishima, H. K., Shimizu, H., Tomita, G., et al. (2006). Risk factors for open-angle glaucoma in a japanese population: the tajimi study. *Ophthalmology*, 113(9):1613–1617.
- Syga, P., Sielużycki, C., Krzyżanowska-Berkowska, P., and Iskander, D. R. (2018). A fully automated 3d in-vivo delineation and shape parameterization of the human lamina cribrosa in optical coherence tomography. *IEEE Transactions on Biomedical Engineering*, 66(5):1422–1428.
- Tan, M., Ong, S., Thakku, S., Cheng, C.-y., Aung, T., and Girard, M. (2015). Automatic feature extraction of optical coherence tomography for lamina cribrosa detection. *Journal of Image and Graphics*, 3.
- Tan, O., Li, G., Lu, A. T.-H., Varma, R., Huang, D., for Glaucoma Study Group, A. I., et al. (2008). Mapping of macular substructures with optical coherence tomography for glaucoma diagnosis. *Ophthalmology*, 115(6):949–956.

- Tennakoon, R., Gostar, A. K., Hoseinnezhad, R., and Bab-Hadiashar, A. (2018). Retinal fluid segmentation in oct images using adversarial loss based convolutional neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE.
- Tezel, G., Trinkaus, K., and Wax, M. (2004). Alterations in the morphology of lamina cribrosa pores in glaucomatous eyes. *British journal of ophthalmology*, 88(2):251–256.
- Tham, Y.-C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., and Cheng, C.-Y. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11):2081–2090.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need.
- Vilupuru, A., Rangaswamy, N., Frishman, L., Smith, E., Harwerth, R., and Roorda, A. (2007). Adaptive optics scanning laser ophthalmoscopy for in vivo imaging of lamina cribrosa. *Journal of the Optical Society of America. A, Optics, image science, and vision*.
- Wang, B., Lucy, K., Schuman, J., Sigal, I., Bilonick, R., Lu, C., Liu, J., Grulkowski, I., Nadler, Z., Ishikawa, H., Kagemann, L., Fujimoto, J., and Wollstein, G. (2018). Tortuous pore path through the glaucomatous lamina cribrosa. *Scientific Reports*, 8.
- Wang, R. K. (2005). Reduction of speckle noise for optical coherence tomography by the use of nonlinear anisotropic diffusion. In *Coherence Domain Optical Methods and Optical Coherence Tomography in Biomedicine IX*, volume 5690, pages 380–385. SPIE.
- Webb, R. H., Hughes, G. W., and Delori, F. C. (1987). Confocal scanning laser ophthalmoscope. *Applied optics*, 26(8):1492–1499.
- Welch, G., Bishop, G., et al. (1995). An introduction to the kalman filter.
- Weng, X., Wang, Y., Man, Y., and Kitani, K. M. (2020). Gnn3dmot: Graph neural network for 3d multi-object tracking with 2d-3d multi-feature learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6499–6508.

- Wylee, O. (2023). Glaucoma: definition, symptoms, types, causes, diagnosis, and treatments.
- Xiang, D., Tian, H., Yang, X., Shi, F., Zhu, W., Chen, H., and Chen, X. (2018). Automatic segmentation of retinal layer in oct images with choroidal neovascularization. *IEEE Transactions on Image Processing*, 27(12):5880–5891.
- Xie, S., Okuwobi, I. P., Li, M., Zhang, Y., Yuan, S., and Chen, Q. (2020). Fast and automated hyperreflective foci segmentation based on image enhancement and improved 3d u-net in sd-oct volumes with diabetic retinopathy. *Translational Vision Science & Technology*, 9(2):21–21.
- Zagorchev, L. and Goshtasby, A. (2006). A comparative study of transformation functions for nonrigid image registration. *IEEE transactions on image processing*.
- Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., and Zhang, L. (2020). Ifcnn: A general image fusion framework based on convolutional neural network. *Information Fusion*, 54:99–118.
- Zwillinger, S., Paques, M., Safran, B., and Baudouin, C. (2016). In vivo characterization of lamina cribrosa pore morphology in primary open-angle glaucoma. *Journal Français d’Ophtalmologie*, 39(3):265–271.