Machine Learning and Image Processing Methods for the Segmentation and Quantification of the Corneal Endothelium

Vigueras Guillén, J.P.

DOI
10.4233/uuid:8ebb7d9d-f090-49f7-8c6f-88e0dd31c372

Publication date
2022

Document Version
Final published version

Citation (APA)
Vigueras Guillén, J. P. (2022). *Machine Learning and Image Processing Methods for the Segmentation and Quantification of the Corneal Endothelium*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:8ebb7d9d-f090-49f7-8c6f-88e0dd31c372

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Machine Learning and Image Processing Methods for the Segmentation and Quantification of the Corneal Endothelium

Machine Learning and Image Processing Methods for the Segmentation and Quantification of the Corneal Endothelium

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board of Doctorates
to be defended publicly on Thursday 13 January 2022 at 15:00 o'clock

by

Juan Pedro Vigueras Guillén

Master of Science in Electrical Engineering,
Chalmers University of Technology, Sweden,
&
Master of Engineering in Telecommunications Engineering,
Polytechnic University of Valencia, Spain,
born in Archena, Spain.

This dissertation has been approved by

promotor: Prof.dr.ir. L.J. van Vliet
copromotor: Dr.ir. K.A. Vermeer

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof.dr.ir. L.J. van Vliet,	Delft University of Technology, promotor
Dr.ir. K.A. Vermeer,	Rotterdam Ophthalmic Institute, copromotor
Independent members:	
Prof.dr. B. Rieger	Delft University of Technology
Dr. J.C. van Gemert	Delft University of Technology
Dr. T.T.J.M. Berendschot	Maastricht University
Dr. A. Fabijańska	Lodz University of Technology, Poland
Prof.dr. S. Stallinga	Delft University of Technology, reserve member
Other members:	
Drs. J. van Rooij	Rotterdam Eye Hospital

Printed by:	IPSKAMP Printing
Cover image:	Egor Kamelev (pexels.com)
Cover design:	J.P. Vigueras-Guillén
Financial support:	TopZorg, ZonMw (grants: 842005004 & 842005007); Stichting Combined Ophthalmic Research Rotterdam (grant: 2.1.0).

Printing of this thesis was supported by Stichting Wetenschappelijk Onderzoek Oogziekenhuis (SWOO-Flieringa).

ISBN 978-94-6421-564-9

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Do not think that what is hard for you to master is humanly impossible; and if it is humanly possible, consider it to be within your reach.

Marcus Aurelius, Roman emperor & philosopher, 121–180 AD

CONTENTS

1 Introduction **1**
- 1.1 The human eye . . . 2
- 1.2 The cornea and the corneal layers . . . 3
- 1.3 In-vivo imaging of the corneal endothelium: the specular microscope . . . 4
- 1.4 Assessing the corneal endothelium: the corneal parameters . . . 6
- 1.5 Motivation of the thesis . . . 7
  - 1.5.1 Study 1: Correlation of pre- and post-surgical corneal thickness and visual function . . . 7
  - 1.5.2 Study 2: The postoperative effects of a Baerveldt implant on ocular motility and corneal endothelium . . . 9
- 1.6 Thesis outline . . . 9

References . . . 12

2 Corneal endothelial cell segmentation by classifier-driven merging of over-segmented images **15**
- 2.1 Introduction . . . 17
  - 2.1.1 Related work . . . 18
  - 2.1.2 Our contribution . . . 18
- 2.2 Methodology . . . 19
  - 2.2.1 Method summary . . . 19
  - 2.2.2 Initial oversegmentation . . . 20
  - 2.2.3 Definition of a merger . . . 21
  - 2.2.4 Features . . . 22
  - 2.2.5 Class definition & training data preparation . . . 23
  - 2.2.6 SVM, kernel, & probabilistic model . . . 24
- 2.3 Experiments & evaluation . . . 24
  - 2.3.1 Evaluation of the method . . . 25
  - 2.3.2 Evaluation of the application . . . 28
  - 2.3.3 Computational cost & execution time . . . 33
- 2.4 Discussion . . . 33

References . . . 36

3 Improved accuracy & robustness of a corneal endothelial cell segmentation method based on merging superpixels **39**
- 3.1 Introduction . . . 41
  - 3.1.1 Our aim . . . 41

3.2 Method . . . 42
3.2.1 Materials . . . 42
3.2.2 Description of the oversegmentation method . . . 42
3.2.3 Frequency analysis . . . 43
3.2.4 Generating oversegmentation . . . 44
3.2.5 Evaluating the segmentation . . . 45
3.3 Results & discussion . . . 45
3.3.1 Watershed optimized result . . . 45
3.3.2 Solving the oversegmentation . . . 46
References . . . 48

4 Fully convolutional architecture versus sliding-window CNN for corneal endothelium cell segmentation 49
4.1 Introduction . . . 51
4.2 Methods . . . 53
4.2.1 Materials . . . 53
4.2.2 U-net architecture . . . 54
4.2.3 SW-net architecture . . . 56
4.2.4 Prediction . . . 56
4.2.5 Postprocessing . . . 56
4.2.6 Labels . . . 57
4.2.7 Preprocessing of the intensity images . . . 58
4.2.8 Data augmentation . . . 58
4.2.9 Implementation details, and computational cost . . . 59
4.2.10 Metrics and statistical analysis . . . 59
4.3 Results . . . 61
4.3.1 Evaluation on the CNN performance . . . 61
4.3.2 Evaluation after applying postprocessing . . . 65
4.3.3 Evaluation on the clinical parameters . . . 67
4.4 Discussion . . . 70
4.5 Conclusions . . . 71
References . . . 73

5 Automatic detection of the region of interest in corneal endothelium images using dense convolutional neural networks 77
5.1 Introduction . . . 79
5.2 Previous work and materials . . . 79
5.3 Methods . . . 82
5.4 Results . . . 82
5.4.1 Experimental results for CNN-Edge . . . 82
5.4.2 Experimental results for CNN-ROI . . . 83
5.5 Discussion & conclusions . . . 90
References . . . 93

6 Convolutional neural network-based regression for biomarker estimation in corneal endothelium microscopy images **95**
6.1 Introduction . . . 97
6.2 Materials and previous work . . . 98
6.3 Methods . . . 98
6.3.1 Biomarker estimation from edge images . . . 98
6.3.2 Improvement of the postprocessing method . . . 101
6.3.3 Improvement of the postprocessing by Fourier analysis . . . 103
6.4 Results . . . 103
6.4.1 Experiment for biomarker estimation . . . 103
6.4.2 Experiment for postprocessing . . . 104
6.4.3 Experiment for postprocessing by Fourier analysis . . . 104
6.5 Discussion . . . 105
6.6 Conclusions . . . 106
References . . . 108

7 Deep learning for assessing the corneal endothelium from specular microscopy images up to 1 year after ultrathin-DSAEK surgery **109**
7.1 Introduction . . . 111
7.2 Methods . . . 112
7.2.1 Datasets . . . 112
7.2.2 Development of the network . . . 113
7.2.3 Statistical analysis . . . 115
7.3 Results . . . 116
7.3.1 Quantitative analysis . . . 116
7.3.2 Error analysis over time . . . 119
7.3.3 Clinical analysis . . . 122
7.4 Discussion . . . 123
References . . . 127

8 Notes on the Baerveldt study **131**
8.1 Introduction . . . 132
8.2 Dataset . . . 132
8.3 Accuracy of the fully-automatic method . . . 133
8.4 Clinical findings . . . 134
References . . . 136

9 Conclusions **137**
9.1 Technical developments . . . 138
9.2 Implications of clinical research . . . 140
9.3 General discussion & future directions . . . 140
References . . . 143

Summary **145**

Samenvatting **147**

List of Publications **149**

Acknowledgements **151**

About the author **153**

1

INTRODUCTION

1

1.1. THE HUMAN EYE

The human eye is a complex organ that plays a pivotal sensorial role in human vision. To this end, the eye collects the light reflected from the many surfaces in our surroundings and focuses it onto the retina, a light-sensitive tissue in the back of the eye (Figure 1.1). Specifically, the light enters the eye through the cornea and it travels through the aqueous humor, the pupil, the lens, and the vitreous humor until it reaches the retina. There, million of photo-sensitive nerve cells convert the light into electric impulses, which are sent to the brain through the optic nerve.

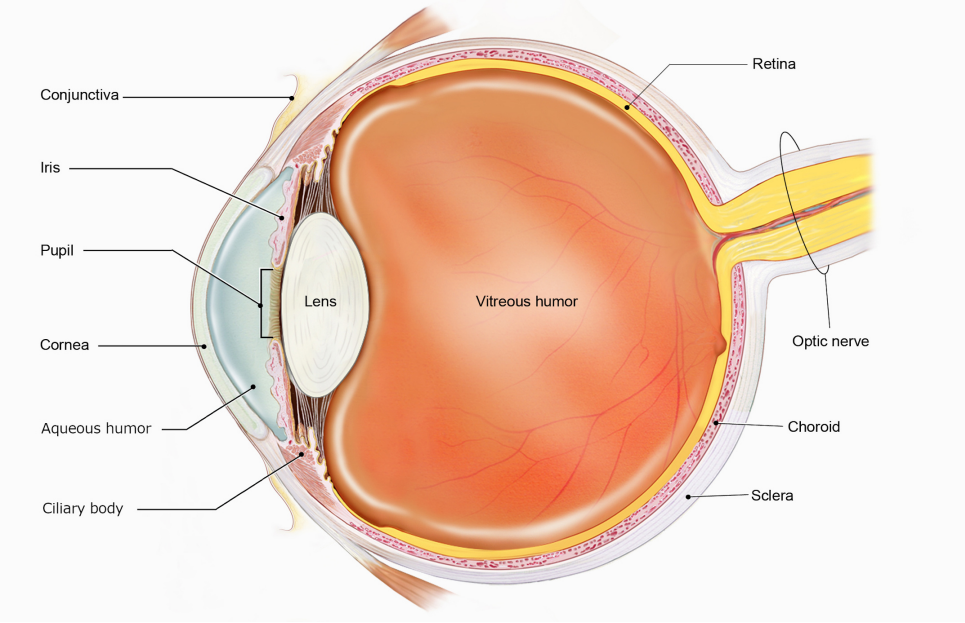

Figure 1.1: Cross-sectional view of the human eye. (Courtesy of: NIH, National Eye Institute [1]).

In order to properly focus the light onto the retina, the cornea and the lens work together as a compound focusing system. The cornea holds a higher but fixed refractive power (approximately 40 diopters), whereas the lens provides a smaller but adaptive refractive power (approximately 20 diopters in the relaxed eye, and up to 33 diopters in the fully accommodated state) [2], which enables focusing light originating from short distances (tens of centimeters) to infinity. This is performed by the ciliary muscles, which control the curvature of the lens in a process called accommodation. The ciliary body also secretes the aqueous humor, a transparent watery fluid (99.9% water) that fills the anterior chamber (between cornea and iris) and the posterior chamber (between iris and lens). The aqueous humor plays an essential role in the eye, as it nourishes the cornea and the lens with nutrients and maintains the intraocular pressure. Indeed, both the cornea and the lens are transparent tissue lacking blood vessels –so that the light is

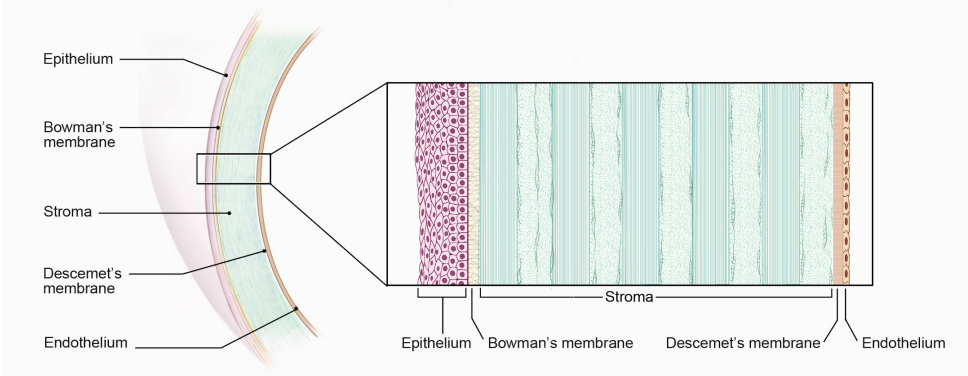

Figure 1.2: Layers of the cornea (Courtesy of: NIH, National Eye Institute [1]). Dua's layer is not depicted in the figure.

not scattered– and therefore they rely mainly on the aqueous humor for their nutrients. In contrast, the vitreous humor is a transparent, gelatinous fluid that fills the vitreous chamber and helps to hold the spherical shape of the eye. In this optic system, the iris is a contractile structure that regulates the size of the pupil, which in turn controls the amount of light entering the posterior eye.

1.2. THE CORNEA AND THE CORNEAL LAYERS

The cornea and the sclera constitute the outer covering of the eye, whose common purpose is to protect the structures inside the eye. However, the tissue arrangement of the cornea differentiates from the sclera in the sense that the latter is opaque due to the irregularity of the collagen fibers whereas the former is transparent due to a near-uniform thickness and parallel arrangement of the collagen [3]. The cornea is convex and aspheric, with a size of 11–12 mm horizontally and 9–11 mm vertically. Its thickness is between 551 and 565 µm in the central cornea and it increases gradually towards the periphery, reaching between 612 and 640 µm [4].

The cornea comprises six layers: epithelium, Bowman's layer, stroma, Dua's layer, Descemet's layer, and endothelium (Figure 1.2). It has cellular components (epithelial cells, keratocytes in the stroma, and endothelial cells) and acellular components (e.g. collagen), and it is densely populated with sensory nerve fibers [4].

The **epithelium** is comprised of 5–7 cell layers of nonkeratinized, stratified squamous epithelial cells and a single layer of basal cells (50 µm thickness in total), which act together as a barrier for microbes and chemicals and provides a smooth optical surface. The epithelial cells have a life expectancy of 7 to 10 days [4].

The **Bowman's layer** is an acellular layer (condensation of collagen and proteoglycans) of 8–12 µm thickness that attaches the epithelium with the stroma and helps maintain the corneal shape. This layer has no regenerative ability; thus, scar tissue results when injured.

The **stroma** is the largest layer, comprising approximately 80–90% of the corneal thickness (approx. 500 µm), and it is made up of keratocytes, collagen, and glycosaminoglycans. The stroma is characteristically transparent due to the precise organization of the stromal fibers. The stroma can heal due to the work of the keratocytes, which can transform into fibroblasts and participate in wound closure and even play a role in nerve regeneration [5].

The **Dua's layer** (15 µm thickness) is located between the stroma and Descemet's layer. Discovered in 2013 By Dua et al. [6], it is characterized by its strength, thinness (15 µm), and impermeability to air.

The **Descemet's layer** is an acellular layer (collagen and laminin) of 7 µm thickness, which are secreted by the endothelial cells, and it attaches the endothelium with the stroma.

The **endothelium** is a single layer of 5 µm thickness of mostly hexagonal cells, which form a honeycomb-like mosaic. This layer controls the transport of fluids and nutrients across the posterior surface of the cornea, allowing them to enter the inner layers of the cornea to nourish the cells, and it maintains a slightly dehydrated state of the cornea (deturgescence) that is necessary for optical transparency. Specifically, the endothelial cells passively allow the leakage of solutes and nutrients from the aqueous humor towards the inner cornea while they actively pumps out the excess of water in the opposite direction. Corneal endothelial cells rarely divide in the post-natal human cornea and, therefore, they lack regeneration. In the presence of injury and/or cell apoptosis, the endothelium heals by sliding and enlarging the adjacent healthy cells rather than performing mitosis [4]. Naturally, cell density decreases with age, from around 4000 cells/mm^2 in 3–5 year-old children, to 2700–2900 cells/mm^2 in healthy middle-aged adults (30–40 years old), and to 2400–2600 cells/mm^2 in healthy elderly people (70–80 years old) [7]. If cell loss is exacerbated due to trauma or certain diseases, the cell density might fall below the threshold needed to maintain corneal deturgescence (400–700 cells/mm^2), which would lead to an excess hydration (edema) of the stroma and epithelium. The stromal edema disrupts the normally uniform organization of the stromal collagen fibers, creating light scattering, and the epithelial edema perturbs the epithelial surface, degrading the optical path and, thus, compromising the visual acuity.

Due to the critical role of the endothelium to maintain a proper optical system and a healthy cornea, it has been important in the clinic to assess its health status.

1.3. IN-VIVO IMAGING OF THE CORNEAL ENDOTHELIUM: THE SPECULAR MICROSCOPE

The first time that the corneal endothelium was observed *in vivo* in the clinic dates back to 1918 [7, 8]. This was performed by the Swiss ophthalmologist Alfred Vogt with the use of a slit-lamp, which is an instrument with a high-intensity light that uses the specular reflex (i.e. regular reflection) to observe different structures of the eye (iris, lens, cornea, etc.). The light is simply applied under a certain angle and the reflected light is then magnified by a lens (40x in Vogt's design) so that the observer can see the reflected tissue.

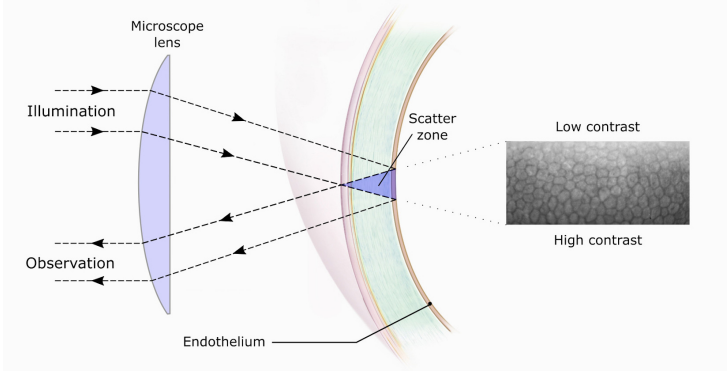

Figure 1.3: Schematic representation of the specular reflex beam on the corneal endothelium. (Background image of the cornea is courtesy of NIH, National Eye Institute [1]).

In order to obtain a specular reflection, a smooth surfaced interface is required. In the case of the endothelium, the cells can be imaged because the refractive index of the endothelium (1.373) is higher than the refractive index of the aqueous humor (1.336), thus reflecting 0.022% of the projected light [7].

In 1968, the concept of imaging the endothelium with a specular reflex was further developed by Maurice, who designed (and named) the first specular microscope [9]. In Maurice's microscope, the incident light and reflected light passed through the same objective lens, whose aperture was divided in two halves (one for each beam). The light beam was very narrow to avoid glare of the overlying stroma, and the magnification was substantially higher than the slit-lamp (400x instead of 40x) [8]. This device was not practical for clinical use (it used a vertical setup), and it was mainly employed in donor eyes to assess the endothelium prior to keratoplasty.

In 1975, Laing et al. introduced a series of modifications to Maurice's microscope, which made it useful for examination in the clinic, becoming the first clinical specular microscope [10]. Among the modifications: (1) a Zeiss slit-lamp base was used to allow horizontal use; (2) the lens was immersed in water to get higher magnification (200x) with simpler lenses (20x); (3) the corneal surface was flattened to add stability; (4) the exposure time was reduced to 1/15 seconds [8].

Due to the invasive nature of Laing's microscope, a different setup that does not require applanation of the cornea was developed in the late 1970s, named noncontact specular microscope [11]. This device was designed to avoid that the reflection from the air-epithelium interface, with a higher reflective power, would obscure the endothelial reflection. In practice, the viewable endothelial area is delimited based on the thickness of the cornea, the curvature of the cornea, and the width of the light beam (Figure 1.3). Furthermore, specular images tend to show different intensity contrasts along both dimensions.

In the late 1980s, a different technology, named confocal microscopy, allowed to image not only the endothelium but any layer in the cornea [12]. The principles of confocal

microscopy were first developed by Minsky in 1957 [13]. Briefly, the basic principle is that a single point of tissue can be illuminated by a light source and simultaneously imaged by a camera when they share the same focal point (hence the name confocal). This would provide very high-resolution images but it has virtually no field of view since it is a single point of illumination-detection. To solve this, the microscope scans a small region with thousands of small beams, reconstructing the final image. Confocal microscopes use a spatial pinhole to block out-of-focus light, thus providing high contrast and resolution (lateral resolution of 1–2 μm) for any spatial plane [12]. For comparative purposes, current specular microscopes also provide a lateral resolution of 1–2 μm, but contrast tend to be lower. The first clinical confocal microscopes also required to have contact with the cornea, but newer instruments allowed for noncontact measurements. Nevertheless, specular microscopes are fast, simple, and cheap, which makes them the preferred option for many clinical tasks.

Nowadays, commercial microscopes are sold along with computers as part of an integrated system, which include built-in image processing and image analysis software that automatically segments the endothelial cells, displaying the image along with several corneal parameters that are considered of interest in clinical practice.

1.4. ASSESSING THE CORNEAL ENDOTHELIUM: THE CORNEAL PARAMETERS

As previously mentioned, the endothelial cell density (ECD) is the most important parameter of the endothelium to assess its health status. Secondary parameters have been defined in order to evaluate the stressed state of the layer, such as the coefficient of variation in cell size (CV, also named polymegethism, defined as the ratio of the standard deviation [SD] of the cell size to its mean size, expressed in percentage) and the percentage of hexagonal cells (HEX, also referred to as pleomorphism). In a healthy cornea, pleomorphism is usually above 60% [14] and polymegethism is below 30% [15]. Biologically, an endothelium affected by acute cell loss would show an increase in CV and a decrease in HEX, whereas an endothelium experiencing cell stabilization after cell loss would show the opposite pattern (decrease in CV and increase in HEX).

Monitoring the corneal parameters is not only important to diagnose and study certain diseases, but also to evaluate and minimize the damage of certain surgical interventions that affect the endothelium. For example:

- Endothelial cell loss after cataract surgery –which needs to go through the cornea to change the lens– varies from 4% to 25% [16]. The subsequent rate of cell loss is 2.5% per year in the central corneal during the ten years after surgery, which is several times the rate in healthy unoperated eyes (0.6 ± 0.5 % per year) [7, 17].
- Patients with primary open-angle glaucoma (POAG) tend to show significantly lower ECD but no significant differences in CV and HEX [18].
- Diabetic patients also tend to display a significantly lower ECD but no changes in CV and HEX [19].
- The use of contact lenses, either hard or soft, have shown long-term changes in polymegethism and pleomorphism [20]. These changes seem to be linked to the

decrease in oxygen (hypoxia) in the cornea due to the presence of the contact lens.
- The success of corneal transplantation, where donor tissue suffers a large cell loss due to surgical trauma and stress, is mainly dependent on the cell density. Penetrating keratoplasty, the main type of surgery until the mid-2000s, had a graft survival of 66% at the 5 years follow-up [21].
- Patients with keratoconus, a corneal disease where the stroma grows abnormally and forms a conical shape, tend to show a significant decrease in ECD and increase in CV [22].

It is indubitable that the assessment of the corneal parameters is very helpful in the clinic. To this end, the segmentation of the endothelial images is necessary to estimate the corneal parameters. However, manual annotations are a very tedious task (up to 30–60 minutes per image) and therefore not practical for large sets of images. For this reason, many of the aforementioned studies with a large number of subjects relied on automatic segmentations. Current microscopes provide automatic segmentation tools, but their accuracy tend to be unsatisfactory not only for pathological corneas but sometimes even for healthy corneas. Several studies compared automatic and manual assessments of ECD (with different devices), and in most cases the automatic values were overestimated [23, 24]. For CV and HEX, no studies have been performed since the current commercial segmentation algorithms do not provide enough accuracy. In fact, only semi-automatic analyses have been reported for CV estimation [25].

1.5. MOTIVATION OF THE THESIS

The research project described in this thesis was a collaboration between Delft University of Technology (Delft, the Netherlands) and Rotterdam Ophthalmic Institute (Rotterdam, the Netherlands). The main goal was to develop and evaluate an automatic cell segmentation framework for corneal endothelium images that estimates the corneal parameters ECD, CV, and HEX with sufficient accuracy for use in clinical practice. Particularly, such framework would be employed in two clinical studies from The Rotterdam Eye Hospital (Rotterdam, the Netherlands): one regarding the transplantation of the cornea and the other regarding the implantation of a drainage device in glaucoma patients. In both cases, images were acquired with a Topcon SP-1P (Topcon Co., Tokyo, Japan) (Figure 1.4) allocated in the Rotterdam Ophthalmic Institute. This microscope had an automatic segmentation program (IMAGEnet i-base, version 1.32); however, its accuracy was not only deficient sometimes (particularly for CV and HEX) but it also could not detect any cell in many images (specifically, 28.5% of the images in the transplantation study and 7.4% in the glaucoma study); thus, the need for a better automatic method. These studies were supported by the program TopZorg from ZonMw, the Dutch organization for health research and healthcare innovation.

1.5.1. STUDY 1: CORRELATION OF PRE- AND POST-SURGICAL CORNEAL THICKNESS AND VISUAL FUNCTION

The aim of this study was to monitor the thickness of the graft after keratoplasty in order to analyze the correlation between pre- and post-surgical thickness and its impact

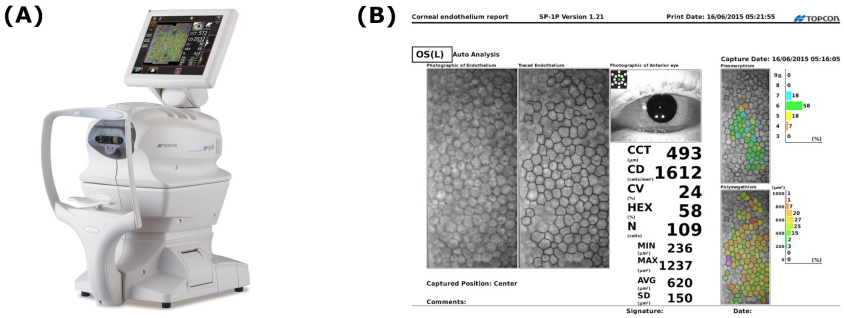

Figure 1.4: (A) The specular microscope Topcon SP-1P (Topcon Co., Tokyo, Japan – Courtesy of: Topcon Co.). This microscope model includes a tactile display where the acquired image and automatic segmentation appears. (B) The analysis made by Topcon's software, which includes several corneal parameters: central corneal thickness (CCT), cell density (CD), coefficient of variation (CV), hexagonality (HEX), number of automatically segmented cells (N), minimum cell size (MIN), maximum cell size (MAX), average cell size (AVG), and standard deviation of the cell size (SD).

on visual acuity. The process of corneal dehydration after keratoplasty can last several months until reaching an equilibrium, and it is still unknown whether such swelling has a role in the long-term visual outcome. Since the corneal swelling is controlled by the endothelium, specular images of the layer were acquired at 1, 3, 6, and 12 months after surgery to estimate ECD. Originally, CV and HEX were not considered in the study design because (1) Topcon's estimates were initially considered to be used for the study and their CV/HEX estimation reliability was expected to be unsatisfactory, and (2) it was unclear whether CV and HEX could provide any knowledge that would help to answer the research question.

Forty-one (41) subjects with Fuchs' endothelial dystrophy (FED) indicated for keratoplasty participated in this study (a total of 383 images). FED is a corneal disease where the endothelium undergoes a degenerative process, triggering an abnormal growth of the Descemet's layer and a rapid loss of endothelial cells. The performed surgical procedure was ultrathin DSAEK (UT-DSAEK), where only Descemet's layer and the endothelium are removed from the patient and the donor tissue contains (besides the two aforementioned layers) a very thin slice of donor stroma, hence the prefix ultrathin. Trial registration: (NL4805), registered 15-12-2014.

Post-keratoplasty specular images are very noisy, with many areas out of focus due to the optical distortions introduced by the corneal swelling, the irregular surface of the graft endothelium, or the graft-recipient interface. In this respect, the work in this thesis aims to develop a framework that can solve very complex cases in the clinic (e.g. post-keratoplasty images), which is currently not feasible with commercial software. Additionally, this work also aims to discuss the patterns and relevance of the three corneal parameters after keratoplasty, since only ECD has been thoroughly discussed in the literature.

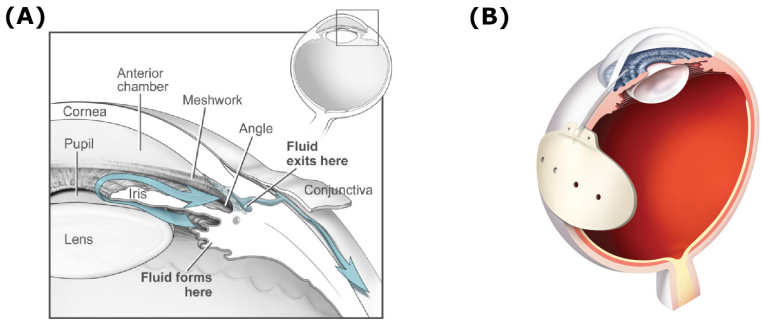

Figure 1.5: (A) An illustration of the flow of the aqueous humor, which is secreted in the ciliary body and absorbed in the trabecular meshwork. If the absorption is compromised, a higher intraocular pressure will occur. (Courtesy of: NIH, National Eye Institute [1]). (B) An illustration of the Baerveldt implant, whose tube reaches the anterior chamber to absorb the excess of fluid. (Courtesy of: Baerveldt, Abbott Laboratories, Johnson & Johnson).

1.5.2. STUDY 2: THE POSTOPERATIVE EFFECTS OF A BAERVELDT IMPLANT ON OCULAR MOTILITY AND CORNEAL ENDOTHELIUM

Glaucoma patients for whom conventional treatment options (medication and/or laser treatment) are inadequate to sufficiently reduce their elevated intraocular pressure usually need surgical intervention, which includes trabeculectomy or the implantation of a permanent drainage device. These devices, which have a tube that reaches the anterior chamber where the excess of aqueous humor is drained (Figure 1.5), are usually preferred over trabeculectomy in secondary glaucoma [26], where drainage devices show a higher success rate after five years [27]. However, two common problems of such devices are their impact on the ocular motility and the degenerative effect on the corneal endothelium triggered by the device's tube. To study the latter, 192 patients that received a Baerveldt implant in The Rotterdam Eye Hospital were imaged before surgery and 3, 6, 12, and 24 months after the implantation, in both the central and the temporal-superior cornea (closer to the device's tube), acquiring a total of 7975 images. Trial registration: NL4823, registered 06-01-2015.

Since glaucoma does not perturb the optical path of the specular microscope, central specular images tend to be of good quality (unless the subject had another corneal pathology). Due to the higher corneal curvature in the periphery, good temporal-superior specular images can be more difficult to acquire, with less contrast and less visible cells. In this respect, the work in this thesis aims to provide accurate estimates of the corneal parameters for these glaucomatous subjects, particularly for the temporal-superior images, which in turn will help to study the effect of the Baerveldt implant on the cornea.

1.6. THESIS OUTLINE

The organization of this thesis is as follows:

Chapter 2 presents a first methodology for segmentation of endothelial cells based

on merging superpixels in oversegmented images. The oversegmentation method is based on stochastic watershed, previously presented by Selig et al. [28], which was originally designed for segmenting confocal microscopy images. Among the several modifications to adapt the method, a scaling factor was added to trigger oversegmentation. The main contribution in the chapter comprises the merging method, based on Support Vector Machines (SVM), which evaluates pairs and triplets of superpixels to determine whether they form whole cells. The method was evaluated on 30 specular images from the Baerveldt study and on three public datasets (confocal microscopy of *in vivo* endothelium, phase-contrast microscopy of *ex vivo* endothelium, and fluorescence confocal microscopy of the retinal pigment epithelium) to illustrate the versatility of the method.

Chapter 3 extends the work presented in Chapter 2 in several ways: (1) it evaluates and discusses the importance of using the correct scaling factor in the oversegmentation method; (2) it proves that the merging method would even improve a segmentation that was designed for a direct, optimal performance by removing any remaining false edge (for this, we tested Selig et al.'s method [28]); and (3) it presents an improved methodology to estimate the most common cell size in an endothelial image from the Fourier domain, which is employed in the oversegmentation method (to set the seeds of the stochastic watershed and to smooth the images). The same 30 Baerveldt images as in Chapter 2 were employed for the evaluation.

Chapter 4 presents a different segmentation methodology, based on convolutional neural networks (CNNs). It evaluates two different approaches: either by pixel classification (CNN classifier) or by direct segmentation (U-net, CNN segmentation). It also introduces the so-called 'postprocessing method' to convert the output of the CNNs into a binary image, necessary to estimate the clinical parameters. This postprocessing is based on the watershed algorithm and Fourier analysis, already introduced in Chapters 2 and 3. A dataset of 50 images from the Baerveldt study was used for the evaluation.

Chapter 5 introduces a CNN-based model to automatically select the region of interest (ROI) in the segmented images where cells are correctly delineated. The fully-automatic framework is finally presented, which includes:

(1) A CNN to generate the edge images (based on Chapter 4).
(2) A CNN to generate the ROI images (this chapter).
(3) A postprocessing method to combine both images, edges and ROI, and generate the binary segmented image (based on Chapters 3–4).
(4) A refinement method to remove any remaining false edge in the segmentation (based on Chapter 2).

The fully-automatic framework was evaluated in an extended dataset of 140 images from the Baerveldt study.

Chapter 6 explores the idea of using a single CNN model to directly estimate the corneal parameters, without the need of performing segmentation. Here, the dataset contained 400 images from the Baerveldt study and 338 images from the DSAEK study, which is assumed to be large enough to explore this hypothesis. This approach is compared to the fully-automatic framework presented in Chapter 5. Alternatively, this chap-

ter also discusses whether Fourier analysis could provide some insight regarding the cell size variation in an image and whether this could be used to improve the postprocessing method.

Chapter 7 presents the final version of the fully-automatic framework, where minor changes were introduced, and it evaluates the 1-year follow-up in the DSAEK study. This chapter highlights the significant improvement to assess the corneal parameters achieved with the proposed method in comparison with Topcon's software. Indeed, the proposed methodology estimates the corneal parameters accurately from practically all images, regardless the challenges observed in the images and, therefore, making the method very useful in a clinical setup. Furthermore, it discusses the existence of clinical patterns in the corneal parameters during the one year post-keratoplasty.

Chapter 8 summarizes the findings related to the corneal parameters on the Baerveldt study. The proposed fully-automatic method was employed to solve the images in that study, which analyses the relation between the progression of corneal parameters with respect to the length and angle of the device's tube.

Chapter 9 discusses the technical developments and limitations presented in this thesis and future research directions.

1

REFERENCES

[1] National Eye Institute, *Licence CC BY 2.0,* https://www.flickr.com/photos/nationaleyeinstitute.

[2] D. Palanker, *Clinical education: optical properties of the eye*, American Academy of Ophthalmology (2013), [Online; accessed on 17 June 2020].

[3] K. M. Meek, *Collagen: structure and mechanics,* (Springer US, 2018) Chap. The cornea and Sclera, pp. 359–396.

[4] M. S. Sridhar, *Anatomy of the cornea and ocular surface,* Indian Journal of Ophthalmology **66**, 190 (2018).

[5] T. N. Lee, *The ins and outs of corneal wound healing,* Review of Optometry (2016), [Online; accessed on 19 June 2020].

[6] H. S. Dua, L. A. Faraj, D. G. Said, T. Gray, and J. Lowe, *Human corneal anatomy redefined: a novel pre-Descemet's layer (Dua's layer)*, Ophthalmology **120**, 1778 (2013).

[7] B. E. McCarey, H. F. Edelhauser, and M. J. Lynn, *Review of corneal endothelial specular microscopy for FDA clinical trials of refractive procedures, surgical devices and new intraocular drugs and solutions*, Cornea **27**, 1 (2008).

[8] C. D. Binkhorst, L. H. Loones, and P. Nygaard, *The clinical specular microscope,* Documenta Ophthalmologica **44**, 55 (1977).

[9] M. D. Maurice, *Cellular membrane activity in the corneal endothelium of the intact eye,* Experientia **24**, 1094 (1968).

[10] R. A. Laing, M. M. Sandstrom, and H. M. Leibowitz, *In vivo photomicrography of the corneal endothelium,* Archives of Ophthalmology **93**, 143 (1975).

[11] B. E. McCarey, *Noncontact specular microscopy: a macrophotography technique and some endothelial cell findings*, Ophthalmology **86**, 1848 (1979).

[12] I. Jalbert, F. Stapleton, E. Papas, D. F. Sweeney, and M. Coroneo, *In vivo confocal microscopy of the human cornea,* British Journal of Ophthalmology **87**, 225 (2003).

[13] M. Minsky, *Memoir on inventing the confocal scanning microscope,* Scanning **10**, 128 (1988).

[14] M. J. Doughty and D. Fonn, *Pleomorphism and endothelial cell size in normal and polymegethous human corneal endothelium,* International Contact Lens Clinic **20**, 116 (1993).

[15] M. J. Doughty, *A prospective analysis of corneal endothelial polymegethism and cell density in young adult Asians,* Clinical and Experimental Optometry **97**, 256 (2014).

[16] M. Cimberle, *Endothelial damage preventable in cataract surgery,* Ocular Surgery News (2008), [Online; accessed on 18 June 2020].

[17] W. M. Bourne, L. R. Nelson, and D. O. Hodge, *Central corneal endothelial cell changes over a ten-year period,* Investigative Ophthalmology & Visual Science **38**, 779 (1997).

[18] Z. Y. Yu, L. Wu, and B. Qu, *Changes in corneal endothelial cell density in patients with primary open-angle glaucoma,* World Journal of Clinical Cases **7**, 1978 (2019).

[19] Y. Chen, S. Huang, G. Jonna, and P. Channa, *Corneal endothelial cell changes in diabetes mellitus,* Investigative Ophthalmology & Visual Science **55**, 2054 (2014).

[20] H. Esgin and N. Erda, *Corneal endothelial polymegethism and pleomorphism induced by daily-wear rigid gas-permeable contact lenses,* Contact Lens Association of Ophthalmologists Journal **28**, 40 (2002).

[21] P. Beckingsale, I. Mavrikakis, N. Al-Yousuf, E. Mavrikakis, and S. Daya, *Penetrating keratoplasty: outcomes from a corneal unit compared to national data,* British Journal of Ophthalmology **90**, 728 (2006).

[22] S. Goebels, T. Eppig, B. Seitz, N. Szentmary, A. Cayless, and A. Langenbucher, *Endothelial alterations in 712 keratoconus patients,* Acta Ophthalmologica **96**, e134 (2018).

[23] M. O. Price, K. M. Fairchild, and F. W. Price, *Comparison of manual and automated endothelial cell density analysis in normal eyes and DSEK eyes,* Cornea **32**, 567 (2013).

[24] J. Huang, J. Maram, T. C. Tepelus, C. Modak, K. Marion, S. R. Sadda, V. Chopra, and L. O. Lee, *Comparison of manual & automated analysis methods for corneal endothelial cell density measurements by specular microscopy,* Journal of Optometry **11**, 182 (2018).

[25] M. J. Doughty and B. M. Aakre, *Further analysis of assessments of the coefficient of variation of corneal endothelial cell areas from specular microscopic images,* Clinical and Experimental Optometry **91**, 438 (2008).

[26] A. B. Joshi, R. K. Parrish 2nd, and W. F. Feuer, *2002 survey of the American Glaucoma Society: practice preferences for glaucoma surgery and antifibrotic use,* Journal of Glaucoma **14**, 172 (2005).

[27] S. J. Gedde, J. C. Schiffman, W. J. Feuer, L. W. Herndon, J. D. Brandt, and D. L. Budenz, *Tube versus trabeculectomy study group. Treatment outcomes in the Tube Versus Trabeculectomy (TVT) study after five years of follow-up,* American Journal of Ophthalmology **153**, 789 (2012).

[28] B. Selig, K. A. Vermeer, B. Rieger, T. Hillenaar, and C. L. Luengo Hendriks, *Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy,* BMC Medical Imaging **15:13** (2015).

2

CORNEAL ENDOTHELIAL CELL SEGMENTATION BY CLASSIFIER-DRIVEN MERGING OF OVERSEGMENTED IMAGES

This chapter is based on the manuscript:

J.P. Vigueras-Guillén, E.R. Andrinopoulou, A. Engel, H.G. Lemij, J. van Rooij, K.A. Vermeer, and L.J. van Vliet, *Corneal endothelial cell segmentation by classifier-driven merging of oversegmented images*, IEEE Transactions on Medical Imaging **37**, 10, pp. 2278–2289 (2018).

2

Abstract

Corneal endothelium images obtained by in vivo specular microscopy provide important information to assess the health status of the cornea. Estimation of clinical parameters, such as cell density, polymegethism, and pleomorphism, requires accurate cell segmentation. State-of-the-art techniques to automatically segment the endothelium are error-prone when applied to images with low contrast and/or large variation in cell size. Here, we propose an automatic method to segment the endothelium. Starting with an oversegmented image comprised of superpixels obtained from a stochastic watershed segmentation, the proposed method uses intensity and shape information of the superpixels to identify and merge those that constitute a cell, using Support Vector Machines. We evaluated the automatic segmentation on a dataset of in vivo specular microscopy images (Topcon SP-1P), obtaining 95.8% correctly merged cells and 2.0% undersegmented cells. We also evaluated the parameter estimation against the results of the vendor's built-in software, obtaining a statistically significant better precision in all parameters and a similar or better accuracy. The parameter estimation was also evaluated on three other datasets from different imaging modalities (confocal microscopy, phase-contrast microscopy, and fluorescence confocal microscopy) and tissue types (ex vivo corneal endothelium and retinal pigment epithelium). In comparison with the estimates of the datasets' authors, we achieved statistically significant better accuracy and precision in all parameters except pleomorphism, where a similar accuracy and precision were obtained.

2.1. INTRODUCTION

The corneal endothelium (CE) is a single layer of closely packed and predominantly hexagonally-shaped cells forming the inner surface of the cornea. It performs an important role maintaining an optimal state of hydration of the cornea [1]. Endothelial cell loss occurs normally due to aging, from a density of approximately 4000–5000 cells/mm^2 at birth down to 2000–3000 cells/mm^2 in a normal adult eye [2]. However, this loss can be exacerbated by trauma, diseases, or intraocular surgery, which might in turn lead to irreversible corneal swelling when the cell density drops below 500–700 cells/mm^2 [3, 4]. Due to its limited cell division capacity, the repair function is usually restricted to the swelling and sliding of the existing cells in order to maintain the barrier properties. Quantitative analysis of the corneal endothelial cell morphology from image data provides clinical information to assist ophthalmologists in diagnosis and treatment of corneal diseases. Currently, endothelial cell density (ECD) is the most important measure to assess the corneal health state, reported as the number of cells per mm^2. Other parameters such as polymegethism –expressed by the coefficient of variation (CV) in cell size– or pleomorphism –quantified by the hexagonality coefficient (HEX) as the percentage of hexagonal cells– are not used in standard clinical practice due to unreliable estimations.

Different instruments are available to assess the cell density and morphometry of the endothelium. Noncontact specular microscopy, developed in the late 1970s, is a fast, noninvasive, *in vivo* imaging method, which permits to record large field-of-view images of the endothelium. In contrast, contact confocal microscopy, developed in the late 1980s, provides an *in vivo*, high-resolution assessment of all corneal layers, but requires physical contact with the cornea and relatively long acquisition time [5]. Due to its noninvasive nature and fast acquisition time, specular microscopy is the current clinical standard technique. While contact confocal microscopy obtains considerably clearer images from edematous corneas [6], both microscopes provide images of sufficient quality from the central and peripheral cornea. Due to the optical principles of the technology, good quality specular microscopy images are limited to corneas that have a smooth endothelium surface [7, 8]. Noncontact confocal microscopy is the most recent modality (early 2000s), and it provides similar image quality as the noncontact specular microscopy for normal corneas but yields a larger field of view [9].

Manual quantification of endothelial images is very labor-intensive, so computer aided techniques to generate the cell segmentation are necessary. Microscope manufacturers currently provide built-in software to automatically segment the recorded images and estimate the parameters. Several studies using different microscope modalities have been performed in order to evaluate the reliability of these fully automated analyses in comparison with semi-automated and manual analyses, especially in the evaluation of cell density. In general, the fully automatic ECD estimations were not consistent with the manual estimations. Some studies indicated an overestimation of ECD in both specular and confocal microscopy images for healthy and glaucomatous eyes exhibiting a high cell density [9], and for healthy and transplanted corneas with normal to low cell density [10, 11], mainly due to the oversegmentation of large cells. In contrast, another study (employing a different microscope vendor's software) reported an underestima-

tion of ECD in both, confocal and specular microscopy, for normal corneas with high cell density due to the identification of false cell borders [12]. When images with large differences in cell density were analyzed with the same software, this was prone to underestimating the cell density in images with high ECD and overestimating it in images with low ECD [13], or *vice versa* [7]. Previous studies on semi-automated analyses reported clinically acceptable results as it produces ECD values comparable to the manual analysis. However, this process is still time consuming since an expert needs to correct the segmentation.

Overall, the automatic detection of cell borders is a complicated task due to the presence of noise, variation in illumination, and optical artifacts. Furthermore, the estimation of polymegethism and pleomorphism can be significantly affected by just a few errors in the segmentation. The aforementioned studies suggest that the built-in software of the currently available commercial microscopes is prone to mistakes in the segmentation, particularly in the presence of a low or high cell density or a high degree of polymegethism. Therefore, there is a need to develop an algorithm that can reliably estimate these parameters from images with such characteristics.

2.1.1. RELATED WORK

Several cell segmentation techniques for corneal endothelium images have been proposed. Selig et al. suggested a stochastic watershed approach to segment *in vivo* confocal microscopy images [14]. Ruggeri and Scarpa segmented *ex vivo* porcine endothelium images obtained with inverse contrast phase microscopy by means of an artificial neural network algorithm [15] and *in vivo* confocal microscopy images by means of a genetic algorithm [16]. Gavet and Pinoli proposed a binarization algorithm to segment *in vivo* specular microscopy images [17]. Finally, Sharif et al. developed a hybrid model based on snake and particle swarm optimization for *in vivo* confocal microscopy images [18].

2.1.2. OUR CONTRIBUTION

Due to the aforementioned challenges, i.e. illumination distortions, optical artifacts, and presence of noise, current segmentation methods fail to identify all cell edges without finding invalid ones. These methods seem to be designed to segment a specific type of endothelium images within a limited range of cell density and variation of cell size. Here, we propose a machine learning framework based on Support Vector Machines (SVM) for segmenting a broad range of endothelium images from different types of microscopes. We hypothesize that, by starting with an oversegmented image where all cell edges have been traced, a trained classifier could identify the fragments (superpixels) that constitute a complete cell. Indeed, the union of fragments belonging to the same cell shows distinctive features (related to shape, intensity, size, etc.) compared to an arbitrary combination of fragments from different cells. Thus, by merging those fragments, the oversegmented image converges towards the correct segmentation. The creation of the initially oversegmented image is not a trivial task, and here we adapted Selig et al.'s approach [14] to generate an initial segmentation without undersegmentation and where hardly any cell is divided in more than three fragments.

Although superpixel merging methods have been proposed in recent years, they were mainly applied to high quality color images and using color histogram features [19, 20], a combination of color and spectral information [21], or color with superpixel size information [22]. Needless to say, all these methods fail to provide acceptable results in specular microscopy images due to the challenges posed by the low image quality. In contrast, the novelty of our method lies in exploiting the morphology of the endothelial cell layer in addition to intensity information of the constituting superpixels and the candidate merger, as well as the edges between them. Moreover, we have exploited the idea that better segmentation results can be achieved if, instead of aiming directly for an optimal segmentation, an oversegmentation followed by a merging process is employed.

To evaluate our method, we used a dataset of endothelial images obtained with *in vivo* specular microscopy (Topcon SP-1P) and we compared our results against the estimates of the microscope's built-in software. To illustrate the versatility of the method, it was also applied to three other datasets: 52 confocal microscopy endothelial images used in Selig et al.'s paper [14], 30 *ex vivo* endothelial images from phase-contrast microscopy published in Ruggeri et al.'s paper [15], and a dataset of 23 confocal microscopy fluorescence images of the central retinal pigment epithelium (RPE) analyzed in Chiu et al.'s paper [23]. RPE cells might show a considerably high variation in cell size, unlikely in corneal endothelial cells, which makes it a good example to test the robustness of the method against polymegethism.

This paper is organized as follows. Section 2.2 provides a description of the proposed method, including how the initial oversegmentation was generated. Section 2.3 first evaluates the merging process and the subsequent cell segmentation of the main dataset. Second, it evaluates the estimates of the clinical parameters for all datasets. Third, a brief explanation about computational cost is provided. Finally, the results are discussed in Section 2.4.

2.2. METHODOLOGY

2.2.1. METHOD SUMMARY

An SVM algorithm is presented, which aims to create a correct segmentation of an endothelial cell image from an initially oversegmented image composed of superpixels. This is accomplished by evaluating whether combinations of adjacent superpixels constitute a complete cell. The SVM classifier uses information about shape, structure, and intensity from the superpixels (both, separately and combined) and their boundaries. All possible combinations of two and three adjacent superpixels in the image are considered simultaneously by means of a dedicated classifier for each type of combination (2 or 3 fragments). In order to compare the combinations from the different classifiers, the signed distance from the classification hyperplane is transformed into a probabilistic output by means of Platt's algorithm [24]. Only the combinations classified as positives and with probability higher than 0.5 are considered. The combination with the highest probability is merged, iterating this process until no more combinations are accepted.

To optimize this iterative process, we create a list of "accepted combinations" such that every time a merger is performed and removed from the list, we also remove all

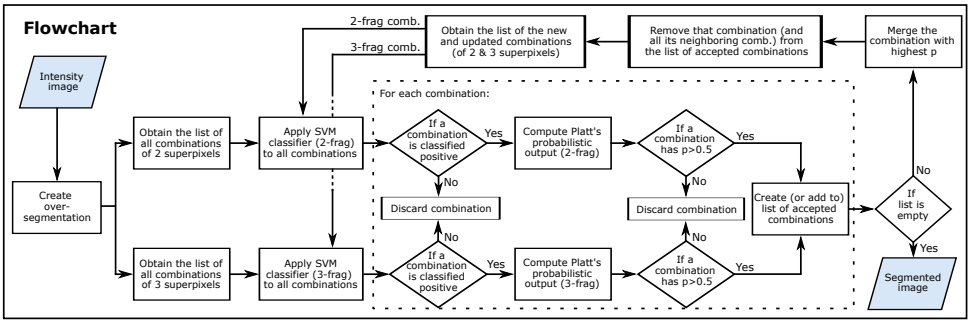

Figure 2.1: Flowchart of the method.

combinations that are no longer possible and all combinations that need to be updated (i.e., neighboring combinations to the merger whose features have changed). Then, all new combinations involving the new superpixel are evaluated as described above, as well as the neighboring combinations that needed to be updated. These combinations are added to the list if their mergers are accepted. Hence, only the strictly necessary (or new) combinations are re-evaluated after each merger, making the process computationally efficient. Once the list of accepted combinations is empty, we have reached the final segmentation. A data-flow diagram is depicted in Figure 2.1.

We chose an SVM classifier with a probabilistic output instead of any other type of classifier or a different method (clustering, regression, etc.) because of the following reasons: (1) SVMs belong to the class of supervised techniques, which usually provide better results than unsupervised methods such as clustering techniques; (2) SVMs can use the features of the constituting superpixels (independently) in combination with the features of the candidate merger in order to determine if a merger is correct, which makes the merging decision more robust; (3) SVM's regularization parameter (C) can be used to avoid over-fitting; (4) SVMs are well known to provide great performance in tasks with highly-unbalanced classes, which happens in our problem; (5) SVMs are robust against the high-dimensional space of the problem; (6) SVM's kernel trick allows building a more complex decision boundary, and in our case it also helps to deal with the arbitrary order of the superpixels in the feature vector; (7) SVMs are computationally very efficient (fast in evaluating), which is necessary in clinical applications with thousands of combinations to be evaluated in one single image; (8) Platt's probabilistic output permits comparing competing types of combinations (2- and 3-fragments), which cannot be done directly.

2.2.2. INITIAL OVERSEGMENTATION

In an ideal oversegmented image, all cell edges are traced with high accuracy and each cell is comprised by a minimum number of superpixels. Due to the low signal-to-noise ratio of specular microscopy images, these two goals are difficult to achieve simultaneously. Since only the removal of edge segments is considered by our method, any undetected cell edge (undersegmentation) produces an error that cannot be corrected *a*

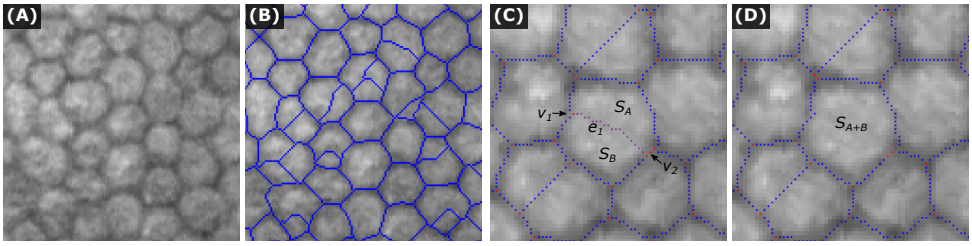

Figure 2.2: (A) Portion of a specular microscopy image. (B) Initial oversegmentation (in blue). Cells are comprised of 1, 2, or 3 superpixels. In the specular microscopy dataset, rarely any cell is divided in more than 3 fragments. (C) Visual representation of a 2-fragment merger. The segmentation (edges in blue, vertices in red) is superimposed on the intensity image. Here, S_A and S_B are merged by removing e_1. (D) After the merger, vertex v_1 is transformed into an edge-pixel, whereas v_2 stays as a vertex.

posteriori. On the other hand, creating a highly oversegmented image might trace all cell edges accurately, but makes the subsequent model too complex and prone to mistakes.

Selig et al. [14] proposed a method to segment endothelial images by applying a seeded watershed algorithm in a stochastic manner. The seed grid has a hexagonal pattern with a spacing equal to the most common cell diameter in the image (l), which is computed as the inverse of the characteristic frequency, $l = 1/f^*$, estimated by Fourier analysis [14]. By creating a denser grid of seeds, we can control the degree of oversegmentation. Here, we divided the estimated cell diameter by $\sqrt{3}$, which corresponds to a grid with three times more seeds, $n_{seeds} = A_I(f^*\sqrt{3})^2 = 3A_I f^{*2}$, being A_I the total image area [14]. This enabled a good detection of all cells edges without creating an excess of oversegmentation (Figure 2.2-B).

For the remaining parameters in the stochastic watershed method (iterations, added noise, blur size, and local minima to ignore), we used the values reported in the original paper [14]. One of these parameters, $k_\sigma = \sigma_{PDF} f^* = 0.17$, is related to a Gaussian smoothing filter (σ_{PDF}) that is applied to the stochastic watershed output. Selig et al. observed that, for large σ_{PDF}, the ridges in the segmentation were shifted with respect to those in the input image. To solve this, Selig et al. added a final, optional step where the classic seeded watershed method was applied to a smoothed version of the input image and the segmented regions of the stochastic watershed were used as seeds. Only two datasets evaluated in this paper showed such a shift in the ridges, namely Selig et al.'s confocal microscope data ($\sigma_{PDF} \approx 6$) and the fluorescence confocal RPE data ($\sigma_{PDF} \approx 4.5$). For the other two datasets ($\sigma_{PDF} \approx 3$), this final step did not substantially change the segmentation (phase-contrast microscope data) or it even degraded the segmentation in areas with blurry cell edges (specular microscope data). Therefore, the edge correction method was only applied to the confocal microscopy datasets.

2.2.3. DEFINITION OF A MERGER

Given a segmented (binary) image, we define the vertices as the branch points of the segmentation, and the edges as the set of 8-connected positive pixels whose endpoints are constrained to vertices. Biologically, a vertex is placed in the point where the cell

edges from three or more cells meet. In order to avoid small edge segments comprised by less than two pixels, these are fused with their respective vertices to form a single vertex (Figure 2.2-C, vertex v_2). The superpixels are defined as the sets of 4-connected negative (non-edge) pixels. We define a 2-fragment merger as the union of two adjacent superpixels by the removal of the edge segment between them. A 3-fragment merger is defined as the union of three adjacent superpixels by the removal of the two or three edge segments between them. This is accomplished by two consecutive 2-fragment mergers.

2.2.4. FEATURES

For every candidate merger, let us consider S_A and S_B as the two fragments to be merged, S_{A+B} as the resulting superpixel, e_1 as the edge segment that needs to be removed to create S_{A+B}, and v_1 and v_2 as the vertices at the ends of e_1 (Figure 2.2-C). A merger is computed as $S_{A+B} = S_A + S_B + e_1$, and v_1 with v_2 are evaluated to identify whether they stay as vertices or they become part of the edge of S_{A+B} (Figure 2.2-D). The following features are computed for each superpixel in the merger (S_A, S_B, and S_{A+B}). Let S be any of those superpixels:

1. Normalized cell size, computed as the number of pixels in S (area) divided by the most common cell size, in pixels (estimated by Fourier Analysis [14]).
2. Normalized cell intensity, computed as the ratio of the average pixel intensity of S to the average pixel intensity of the neighboring superpixels.
3. Convexity, defined as the ratio of the perimeter of the convex hull of S to the perimeter of S.
4. Solidity, defined as the proportion of the pixels in the convex hull that are also in S ($Area/ConvexArea$).
5. Circularity, defined as $4\pi Area/Perimeter^2$.
6. Eccentricity, defined as the eccentricity of the ellipse that has the same second-order moments as S.
7. Number of neighboring superpixels.
8. Standard deviation of the lengths of the edges associated with S.
9. Standard deviation of the inner angles. We define the inner angles as the angles that are formed in the center of mass of S between the straight lines that are traced from that point to the vertices of S.
10. Standard deviation of the outer angles. We define the outer angles as the angles formed in the vertices of S between the straight lines that are traced between the vertices.
11. Ratio of the area of S to the polygon area, where the polygon is formed by connecting the vertices of S with straight lines.
12. Binary value to indicate whether S lies in the border region of the image (this is not included for S_{A+B}).

For S_{A+B}, we compute extra features. Here, we define l_1 as the straight line that is formed by connecting v_1 with v_2. Features related to those vertices provide two values, one per vertex.

13. Ratio of the average intensity of e_1 to the average intensity of S_A and S_B.
14. Ratio of the average intensity of e_1 to the average intensity of the edges of S_{A+B}.

15. Ratio of the area formed between e_1 and l_1 to the area of S_{A+B}.
16. Ratio of the length of l_1 to the length of e_1.
17. For each vertex (v_1, v_2), ratio of the average intensity of half e_1 (the part closest to the vertex) to the average intensity of the edges of S_{A+B} ending in v (only the half of the edges closest to v). In contrast to feature 14, here only the intensity in the vicinity of v is evaluated (2 values).
18. Ratio of the length of e_1 to the average edge length of S_A and S_B (2 values).
19. Standard deviation of the angles formed in the vertices (v_1 and v_2) before the merger occurs (2 values).
20. The angles formed in the vertices (v_1 and v_2) after the merger occurs (2 values).

The concatenation of the features of the involved superpixels creates the feature vector of a merger. For an n-fragment merger, $n!$ possible feature vectors can describe the same merger, depending on the order of the fragments. The SVM kernel was designed to map all those vectors to the same point in feature space (see Section 2.2.6).

For a 2-fragment merger, the feature vector contains 47 elements. For a 3-fragment merger, the extra features are computed twice (one per removed edge segment), and they are processed as follows: For features 13–16, the averaged value is computed; for features 17–20, the maximum and minimum values are kept. In total, a 3-fragment merger is represented with a feature vector of 59 elements.

2.2.5. CLASS DEFINITION & TRAINING DATA PREPARATION

In the 2-fragment SVM classifier, the *good-merger* class is defined as the set of combinations of two superpixels that form a cell. The *bad-merger* class is defined as the set of all other combinations of two adjacent superpixels, with the exception of combinations coming from within cells divided in more than two fragments. The classes in the 3-fragment SVM classifier are defined in an analogous way, but now for sets of three superpixels.

To prepare the training/test data, we retrieved the class elements from the oversegmented images as described above. Since some features are affected by the state of fragmentation of the neighboring cells, the training elements in the good-merger class were computed twice, once when none of the fragmented neighboring cells were merged yet, and again when all were merged. Bad-merger examples were also retrieved from the resulting superpixel of a good-merger with its neighboring superpixels. Hence, the training/test examples not only included all the mergers that could be formed in the initially oversegmented image, but also (some of) the mergers that could appear during the merging process.

To help solve the segmentation of cells divided in more than three fragments without having a classifier specifically built for that purpose, the following training elements in the good-merger class were included: for the n-fragment classifier, a cell divided in m fragments, being $m > n$, will generate combinations of n superpixels where one superpixel is the result of previously merging $m - n + 1$ fragments of the cell.

To deal with cells touching the border of the image, some good and bad mergers were included for those cases. Feature number 12 was added for this purpose. Eventually, the

estimation of the endothelial parameters is done excluding the superpixels touching the border region of the image.

The ratio of bad mergers (negative class) to good mergers (positive class) in the training/test dataset was around 25:1 for the 2-fragment classifier, and 200:1 for the 3-fragment classifier.

2.2.6. SVM, KERNEL, & PROBABILISTIC MODEL

By design, no oversampling or undersampling is applied to the classes, and each element in any of the two classes has the same weight. For computational purposes, the negative class in the 3-fragment classifier was undersampled to the same ratio as in the 2-fragment classifier without losing discriminative power. This was done by selecting the closest negative elements to the positive ones, in Euclidean distance, which removes elements that exert no influence on the SVM hyperplane.

A Gaussian-based kernel was designed, which maps all permutations of the feature vectors of a single merger onto the same point in the feature space. Let us suppose $\mathbf{G}(\mathbf{x}_j, \mathbf{x}_k)$ denotes element (j,k) in the Gram matrix, where $\mathbf{x}_j$ and $\mathbf{x}_k$ are p-dimensional vectors representing elements j and k. For the 2-fragment SVM classifier, let j indicate the merger between superpixels A and B, whose vector can be depicted as $\mathbf{x}_j = [f_A, f_B, f_{A+B}]$ or $\mathbf{x}'_j = [f_B, f_A, f_{A+B}]$, with f_n the features of the superpixel n. Then, the kernel is defined as

$$\mathbf{G}(\mathbf{x}_j, \mathbf{x}_k) = \exp(-\gamma||\mathbf{x}_j - \mathbf{x}_k||^2) + \exp(-\gamma||\mathbf{x}'_j - \mathbf{x}_k||^2), \tag{2.1}$$

with γ the scale parameter of the Gaussian radial basis function. The kernel for the 3-fragment classifier was defined in a similar fashion, covering the six possible permutations that describe the same 3-fragment merger.

In order to compare mergers from different SVM classifiers, the optimal score-to-posterior-probability transformation was computed for each classifier by using Platt's approach [24].

2.3. EXPERIMENTS & EVALUATION

We evaluate the proposed method in Section 2.3.1 for the main dataset. The evaluation is done for the classifiers separately and jointly, and for the full cell segmentation task. In Section 2.3.2 we evaluate the estimates of the clinical parameters from the segmented images for all datasets. The estimates are compared against the gold standard and either the estimates of the microscope's built-in software or the estimates provided by the original authors.

For all datasets, a double 5-fold cross-validation approach was used to separate training, validation, and testing sets, providing a ratio in the size of the sets of approximately 64% training, 14% validation, and 20% testing. Images were completely assigned to a single fold to prevent combinations of fragments from the same image in different sets. For the evaluations in Section 2.3.1, the experiments were repeated five times, reassigning the images to the folds differently each time, and the averaged results were reported.

2.3.1. EVALUATION OF THE METHOD

DATASET

The main dataset consists of 30 corneal endothelium images from the central cornea of 30 glaucomatous eyes, with approximately 250 cells per image. They were obtained with a noncontact specular microscope (SP-1P, Topcon Co, Tokyo, Japan) for an ongoing study regarding the implantation of a Baerveldt glaucoma drainage device in The Rotterdam Eye Hospital. The images were acquired prior to the implantation of the device. They cover an area of 0.25 mm × 0.55 mm and were saved as 241×529 pixels 8-bits grayscale images. The acquisition occurred with informed consent and followed the tenets of the Declaration of Helsinki. One expert created the gold standard for each image by performing manual segmentation of the cell contours using an open-source image manipulation program (GIMP v.2.8).

EVALUATING THE ACCURACY OF THE CLASSIFIERS

The performances of the two classifiers (2- and 3-fragments) were evaluated independently. Fifteen SVMs were computed per classifier by varying the misclassification cost between False Positives (FP) and False Negatives (FN). The cost values FP:FN followed the ratio $5^{n/2}$:1, with $n = 12, 11, \ldots, 1, 0, -1, -2$. Based on the class definition, a FP is defined as a bad-merger classified as positive, whereas a FN is a good-merger classified as negative. A grid search was performed to find the best SVM parameters, γ and C, in the validation set, using the cost function: $\text{argmin}_{\gamma,C}\,(cost_{FP}\text{FP} + cost_{FN}\text{FN})$.

The evaluation was done on the test set, built as described in Section 2.2.5. FROC curves [25] were generated by calculating the true positive rate, TPR = TP/(TP+FN), and the fractional number of false positives per cell, FPs/cell, from the output of each classifier. Two setups were considered: one where the fifteen classifiers performed the task independently, and another for a cascade of classifiers. In the latter, the classifiers were ordered from highest to lowest ratio of cost FP:FN, and any accepted merger at one classifier was automatically accepted by the following classifiers. The classifier with equal costs (FP:FN of 1:1) was marked in all curves (Figure 2.3-A, B).

An inflection point was observed in the curves of the independent classifiers, at those with costs 5:1 (2-frag) and 25:1 (3-frag). Beyond that point, decreasing the misclassification cost of the FP created more FPs, but did not increase the number of TPs. In contrast, the cascade of classifiers benefited from a higher TPR because of the accumulative effect of TPs in a cascade design, but at the expense of a slightly higher number of FPs. It was clear from the graphs that no further improvement occurs beyond the classifier with equal costs (FP:FN of 1:1), as the remaining good-mergers would come at a high price in terms of FPs. At that point (1:1), the 2-fragment classifier in the cascade provided a TPR of 96.5% with 0.037 FPs/cell, and the 3-fragment classifier provided a TPR of 94.9% with 0.046 FPs/cell. Moreover, half of the FPs in both cases came from cells touching the image borders and such cells are not considered when calculating ECD, CV, or HEX.

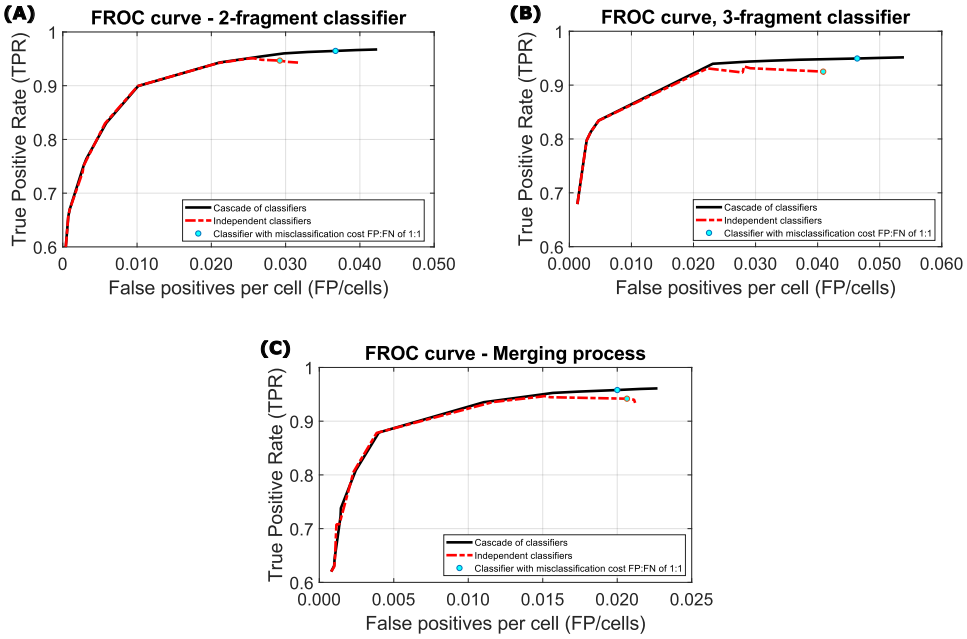

Figure 2.3: FROC curves of the 2-fragment classifier (A), 3-fragment classifier (B), and the outcome of the merging process (C), using 15 classifiers with different costs of FP:FN. The 15 classifiers are evaluated in two ways: independently and connected in a cascade. In all cases, the classifier with equal cost (FP:FN of 1:1) is marked.

EVALUATING THE ACCURACY OF THE MERGING PROCESS

This evaluation differed from the previous one in three aspects: (1) the setup follows the proposed method (Figure 2.1), which means that both classifiers work simultaneously; (2) the evaluation now uses the test images, which means that only mergers that occur during the merging process are involved; (3) the evaluation metric is now defined for the final result at the cell level: a TP is defined as a fragmented cell that is correctly merged, a FN is defined as a fragmented cell that is not completely merged (oversegmentation), and a FP is defined as any cell –fragmented or not– that suffers from an incorrect merger (undersegmentation). A cell can be labeled with only one condition, and the condition FN prevails over FP. For instance, if a fragment from an oversegmented cell (FN) merges with another cell, the former cell is labeled as FN and the latter as FP. If a fragmented cell, after merging correctly, merges with another cell, then both are labeled as FP. Hence, TPR indicates the percentage of oversegmented cells that were correctly merged, whereas the number of FPs/cell indicates the percentage of cells that were undersegmented. This metric will be also used for the remaining datasets.

The same fifteen SVMs were used here, and both setups, independent classifiers and the cascade of classifiers, were considered. In the latter, once all the accepted combinations have been merged in one set of classifiers, the output segmentation is provided as input to the next set of classifiers. The FROC curves (Figure 2.3-C) showed a similar pattern as before. The cascade, whose main goal is to merge the most certain true

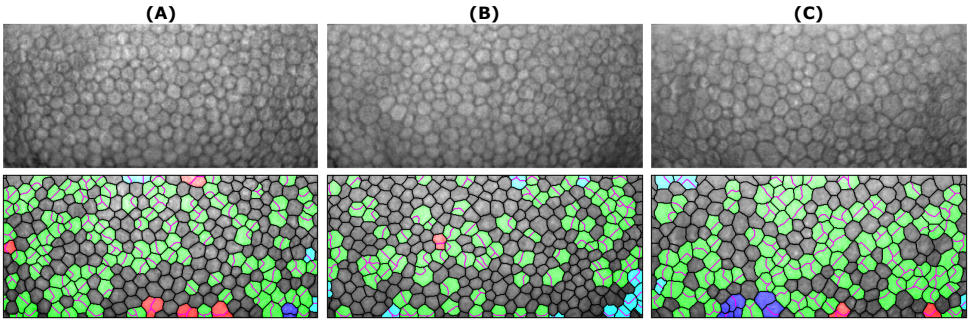

Figure 2.4: Three representative specular microscopy images (top) and their respective visual outcome of the merging (bottom): final segmentation in black lines, and removed edges by merging process in magenta. All mergers are colored: successful mergers in green, undersegmented cells (FP) in red, oversegmented cell (FN) in dark blue, and mergers outside the gold standard area in light blue.

combinations in the first stages, could potentially reduce FPs in the last stages as less false combinations are being evaluated, showing a small but clear effect. At the classifier with equal costs (FP:FN of 1:1) in the cascade, a TPR of 95.8% with 0.020 FPs/cell was obtained. To illustrate the outcome, at that point there are, on average, less than six undersegmented cells and less than two oversegmented cells per image (Figure 2.4), considering that initially 25% of the approximately 250 cells per image were oversegmented. Furthermore, the majority of undersegmented cells were located along the image borders. This supports the idea of having reached the optimal point, since the number of over- and undersegmented cells in the inner part of the image was similar, while undersegmentation was starting to prevail along the image borders.

EVALUATING THE ACCURACY OF THE SEGMENTATION

The segmentation accuracy was measured by using the distance transform. In a first measure, the distance transform was computed on the resulting segmentation. The distance values were averaged over those pixels that were labeled as edge pixels in the gold standard, providing a single value per image. This measure is sensitive to undersegmentation, as missing true edges in the output segmentation increases the value, but is barely affected by oversegmentation. In a second measure, the distance transform was computed on the gold standard, retrieving only the values of those pixels that were labeled as edge pixel in the resulting segmentation. In contrast to the former measure, this is sensitive to oversegmentation, as false edges increase the value, but insensitive to undersegmentation. In almost all images, the first measure was slightly higher (the difference was 0.05 ± 0.05, mean $\pm$ SD), suggesting a very small dominance of undersegmentation. The average of both measures was computed for each image, providing a value of 0.93 ± 0.09 (mean $\pm$ SD). This suggested that the segmentation was on average within one pixel from the gold standard.

2.3.2. EVALUATION OF THE APPLICATION

Based on the previous results, a cascade of four classifiers was established as the final setup, with cost values (FP:FN) of 125:1, 25:1, 5:1 and 1:1.

DATASETS & STATISTICAL ANALYSIS METHOD

Besides the main dataset, three publicly available datasets were evaluated.

The confocal microscopy dataset includes 52 *in vivo* images (768×576 pixels, Nidek Confoscan 4, Italy) of transplanted corneal endothelium (DSAEK), imaged one year after surgery. Images were cropped by the author to remove dark areas. Each image contains approximately 70 cells [14].

The phase-contrast microscopy dataset includes 30 *ex vivo* images (768×576 pixels, Olympus CK 40, USA) of porcine corneal endothelium. These images were also cropped by the author. Each image contains approximately 350 cells [15].

The fluorescence confocal microscopy dataset includes 23 images (1024×1024 pixels, Nikon Eclipse C1, Japan) of central RPE from 17 mice, with approximately 1000 cells per image [23].

The parameters to be estimated were endothelial cell density (ECD), polymegethism (CV), and pleomorphism (HEX) for the endothelium datasets, and the number of cells and the average cell size for the RPE images (as it was done in the original papers). The gold standard and the segmentation images produced by the original authors were publicly available or provided at our request, which allowed us to apply the same algorithm for parameter estimation in all cases. For all images in all datasets, only the cells covered by the area of the gold standard were included for the parameter estimation. The only exception was Topcon's segmentation (specular microscopy dataset), since the microscope's software did not provide any cell segmentation in areas where cells were not correctly detected by the software. In that dataset, the gold standard covered a surface 35% larger. The estimation error was defined as the difference between the estimated value and the gold standard value for the CE datasets (Figure 2.5, Figure 2.6-A–C), and as the percentage of that difference for the RPE dataset (Figure 2.6-D–E). Note that, for polymegethism and pleomorphism, the units are a percentage, whereas the error is the difference of the percentages. The mean value and SD of those estimation errors are indicated in Table 2.1.

The RPE images and their gold standard were originally presented by Ding et al. [26], but Chiu et al. [23] discovered (but did not correct) in the gold standard a few unsegmented cells due to their small size; here those cells were corrected (on average, three cells per image were added). We also noticed that cell edges along the perimeter of the gold standard area were not annotated, and they were corrected. For the phase-contrast microscopy dataset, the cell edges in the gold standard had a thickness of 2-3 pixels, so skeletonization was applied to reduce it to 1 pixel thickness.

A statistical analysis based on linear mixed-effects models [27] was performed to determine, for each parameter, whether there was a statistically significant difference in precision (smaller variance) and in accuracy (smaller absolute mean) between the two estimation errors. To determine whether the variances were different, we used a likeli-

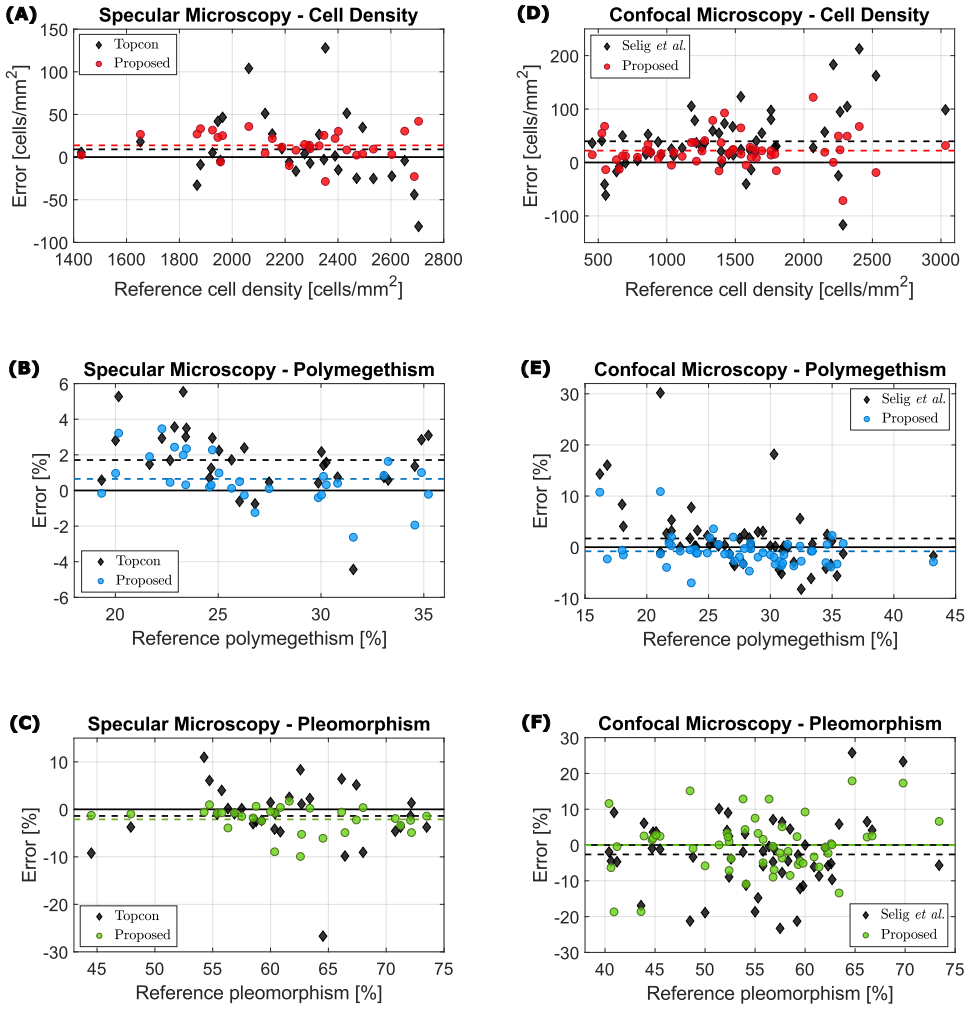

Figure 2.5: Estimates of the clinical parameters in the specular microscopy dataset (A-C), and confocal microscopy dataset (D-F). The *x*-axis indicates the value for the gold standard, and the *y*-axis indicates the error computed as the difference between the estimates and the gold standard. Each point corresponds to one image in the dataset (proposed in colored circles, third-party results in black diamonds). The mean value of the error for each set is drawn with a dashed line.

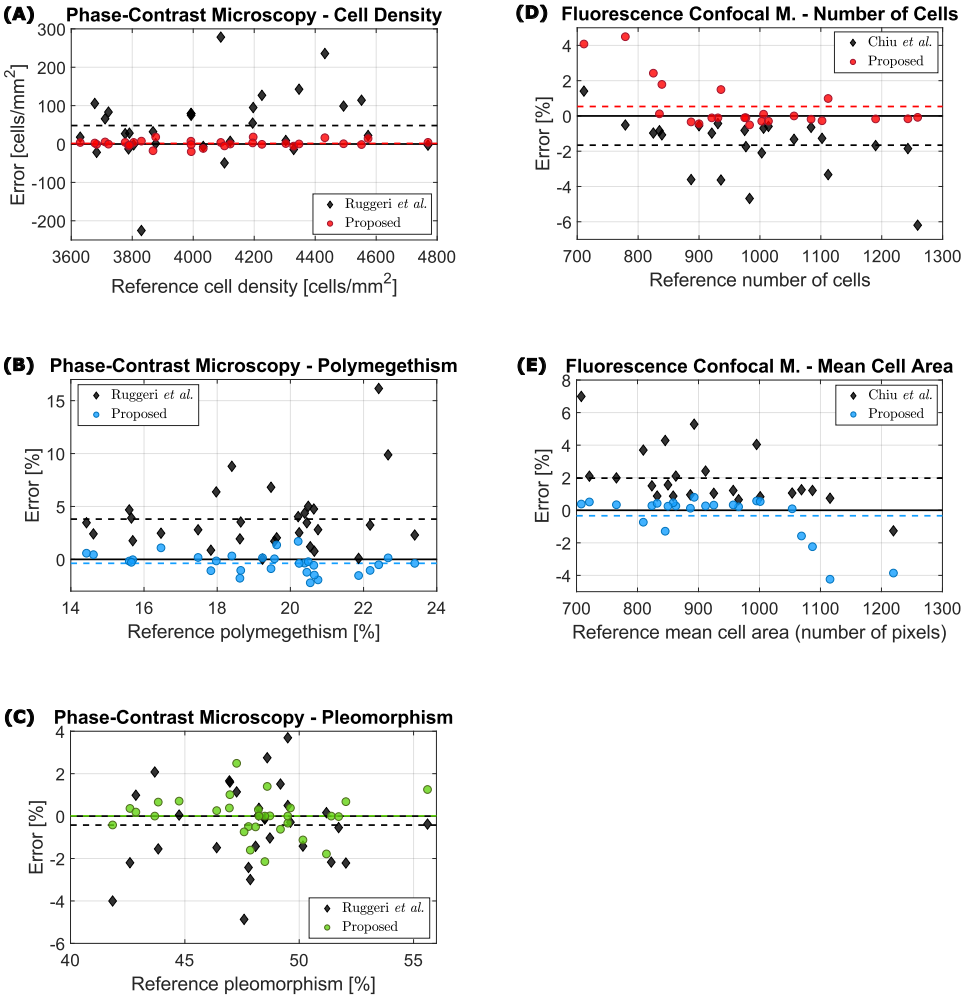

Figure 2.6: Estimates of the clinical parameters in the phase-contrast microscopy (A-C), and fluorescence confocal microscopy (D-E). The *x*-axis indicates the value for the gold standard, and the *y*-axis indicates the error computed as the difference between the estimates and the gold standard for the phase-contrast microscopy, and as the percentage of error for the fluorescence confocal microscopy. Each point corresponds to one image in the dataset (proposed in colored circles, third-party results in black diamonds). The mean value of the error for each set is drawn with a dashed line.

Table 2.1: Mean error (ME) and its standard deviation of the estimated clinical parameters for all datasets.

Dataset	Method	ME of ECD [cells/mm^2]	ME of CV [%]	ME of HEX [%]
Specular Microscopy	Proposed	14 ± 16	0.6 ± 1.3	-2.1 ± 2.7
	Topcon	9 ± 41	1.7 ± 1.8	-1.4 ± 6.8
Confocal Microscopy	Proposed	22 ± 30	-0.8 ± 3.0	0.0 ± 7.8
	Selig et al.	40 ± 57	1.7 ± 6.4	-2.6 ± 9.9
Phase-Contrast Microscopy	Proposed	1 ± 9	-0.4 ± 0.9	0.0 ± 0.9
	Ruggeri et al.	48 ± 88	3.8 ± 3.2	-0.4 ± 1.9
Dataset	**Method**	**ME of no. cells** [%]	**ME of mean cell area [%]**	
Fluorescence-Confocal Microscopy	Proposed at 1:1	0.5 ± 1.4	-0.3 ± 1.4	
	Proposed at 1:25	0.3 ± 1.0	-0.1 ± 1.0	
	Chiu et al.	-1.7 ± 1.6	2.0 ± 1.8	

hood test to compare a model that assumes equal variances between both estimation errors with a model that assumes different variances. From the fixed effects test of the models we evaluated whether the mean values in both estimations were different. No correction for multiple testing was applied, and a P-value of less than 0.05 was considered statistically significant.

SPECULAR MICROSCOPY DATASET

The statistical analysis indicated a significantly better precision in all parameters for the proposed method ($P < 0.001$, $P = 0.020$, and $P < 0.001$, for ECD, CV, and HEX respectively). The analysis also indicated a significantly better accuracy for CV ($P < 0.001$), but comparable accuracy for ECD and HEX ($P = 0.57$, both) (Table 2.1).

CONFOCAL MICROSCOPY DATASET

This dataset was characterized for having significantly smaller images with a rather low cell density (Figure 2.5-D). The larger cell sizes (Figure 2.7-A–D) resulted in an initial oversegmentation in approximately 45% of the cells. In the output, there were on average only one over- and one undersegmented cell per image, the majority of them in the border region of the image. This provided a TPR of 97.5% and 0.016 FPs/cell. The statistical analysis indicated that, for ECD and CV, there was a significantly better precision ($P < 0.001$) and better accuracy ($P = 0.028$, $P = 0.002$, respectively) for the proposed method, but comparable precision and accuracy for HEX ($P = 0.074$, $P = 0.088$, respectively) (Table 2.1). The small image sizes and large cells made the parameter estimation more sensitive to merging errors and small variation in edge location. Overall, the method could successfully cope with images exhibiting large differences in cell density (Figure 2.5-D).

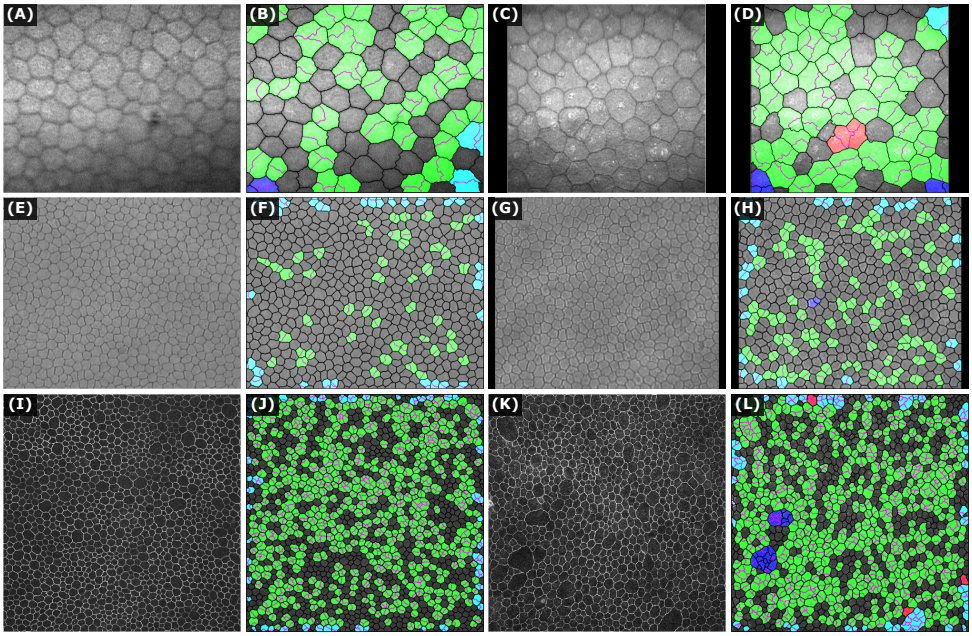

Figure 2.7: Two representative results for each dataset: confocal microscopy (A-D), phase-contrast microscopy (E-H), and fluorescence confocal microscopy (I-L). Segmentation in black lines; removed edges by merging process in magenta. All mergers are colored: successful mergers in green, undersegmented cells (FP) in red, oversegmented cell (FN) in dark blue, and mergers outside the gold standard area in light blue.

PHASE-CONTRAST MICROSCOPY DATASET

This dataset showed a high cell density and a low degree of polymegethism (Figure 2.6-A, B), and the image quality was significantly better (Figure 2.7-E–H) compared to the other datasets. Therefore, only 7.5% of the cells were initially oversegmented. After applying our method, 20 of the images had a perfect merging, and the remaining 10 images showed on average one over- and one under-segmented cell. In total, the method provided a TPR of 98.70% and 0.0009 FPs/cell. The statistical analysis indicated a significantly better precision in all parameters for the proposed method ($P < 0.001$), a significantly better accuracy for ECD and CV ($P = 0.008$, $P < 0.001$, respectively), and similar accuracy for HEX ($P = 0.214$) (Table 2.1). Overall, the high cell density and good quality image made that our method provided highly satisfactory estimates.

FLUORESCENCE CONFOCAL MICROSCOPY DATASET

The last dataset was characterized by a large variation in cell size (Figure 2.7-I–L). Approximately 50% of the cells in each image were oversegmented, and around 5% of the cells were comprised of more than three fragments. In the output, there were on average three over- and three under-segmented cells (from the approx. 1000 cells per image). This provided a TPR of 99.37% and 0.0036 FPs/cell. However, the oversegmented cells were mainly large cells divided in 5-10 fragments (Figure 2.7-K–L), which affected

the parameter estimation considerably. The statistical analysis indicated a significantly better accuracy in both parameters for our method ($p < 0.001$), but comparable precision ($P = 0.428$ and $P = 0.201$, for number of cells and mean cell area, respectively) (Table 2.1). Due to the high intensity contrast between edges and cells, another SVM model with smaller cost (FP:FN of 1:25) could be added to the cascade, which barely introduced any FP but solved the majority of those oversegmented cells (figure not provided; values in Table 2.1). Consequently, this also provided a statistically significant better precision in both parameters ($P = 0.024$ and $P = 0.007$, for number of cells and mean cell area, respectively).

2.3.3. COMPUTATIONAL COST & EXECUTION TIME

Besides the details explained in Section 2.2.1, several considerations were taken into account in order to reduce the computational cost. On average, each superpixel had 7 neighbors in the oversegmented image. Hence, the features of the superpixels were computed only once, stored in memory, and retrieved when building the feature vector of a candidate merger. When the features of a superpixel changed due to neighboring mergers, the features were recomputed and overwritten. Initially, the oversegmented image was translated into a structure where each superpixel, edge, vertex, and pixel was labeled, and where the connection between these elements were listed. This simplified the iterative merging process, as we could directly access to the elements that needed to be updated.

The computational cost of the 2-fragment classifier was linear with respect to the number (N) of initial superpixels, $\mathcal{O}(N)$. For the 3-fragment classifier, the relation was quadratic, $\mathcal{O}(N^2)$. For the main dataset, the entire segmentation process for a single pair of classifiers took on average 2.5 minutes (four times more if we used the cascade of four classifiers). The majority of time was used in evaluating all combinations of 3-fragments. Restricting the segmentation to mergers of two fragments reduced the computation time to less than 30 seconds, but as a consequence not all cells divided in 3 fragments were correctly segmented. We believe that these computation times can be decreased substantially if this method were implemented in a low-level programming language.

2.4. DISCUSSION

We presented a new method to segment corneal endothelium images that requires no user intervention. The method provided very good results in endothelium images obtained from different devices and with large differences in cell density and variation in cell size. Furthermore, our method can be applied to other images of similar closely-packed cells with little extracellular matrix, such as RPE images. To the best of our knowledge, no other technique with such versatility has been reported up to now.

The qualitative results shown in Figures 2.4 and 2.7 indicate accurate cell segmentation despite the presence of image artifacts and blurriness, and the variability in cell size was not a performance-limiting factor. The quantitative results indicate an average error of less than 2.1% (Table 2.1), although this can be considerably smaller depending on the dataset and the estimated parameter. The proposed method outperforms the other

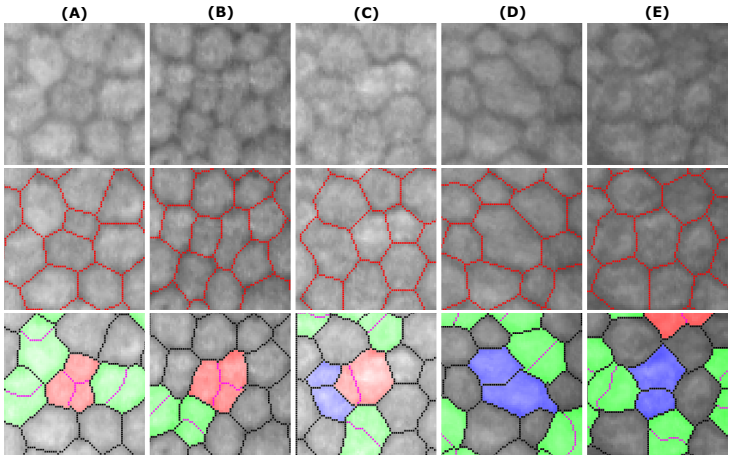

Figure 2.8: Several examples of undersegmented cells (A–C), and oversegmented cells (C–E) in the specular microscopy dataset. (Top) Intensity images. (Middle) Gold standard in red. (Bottom) Segmentation results, with the final segmentation in black lines and edges removed by the merging process in magenta. All mergers are colored: successful mergers in green, undersegmented cells (FP) in red, oversegmented cell (FN) in dark blue.

automatic techniques from the literature, offering a significantly better accuracy and/or precision for the majority of the clinical parameters.

While we have used only classifiers dealing with cells divided in two and three fragments, it is straightforward to extend our method to cells composed of more fragments. Endothelial cells rarely show such a large difference in cell size that inclusion of explicit classifiers for cells divided in more than three fragments is required. Indeed, those cases only represent 0.35% of the cells in the specular microscopy dataset and 2.5% of the cells in the confocal microscopy dataset. In practice, the proposed method achieves the correct segmentation for the majority of those cells as the intensity-related features help to identify the false edges. This was illustrated convincingly in the fluorescence confocal microscopy dataset (RPE), where cells divided in up to 15 fragments were correctly segmented (Figure 2.7-I–L).

Among the datasets presented here, the specular microscopy images generate the highest number of errors due to their lower signal-to-noise ratio and lower image quality in terms of contrast and blurriness. Indeed, lack of contrast between edges and cells produces a rather inaccurate delineation of the edges in the initial oversegmentation, which is the major factor for errors in the merging process (Figure 2.8-A, C, E). Even in the absence of cell fragmentation, the inaccurate edge delineation significantly affects the estimation of CV and, especially, HEX. Other factors, such as oddly shaped cells (Figure 2.8-B, D) or cells touching in the image borders, may produce wrong mergers. We expect that the latter could be avoided with more training examples.

Overall, we observe that fragmented cells with correctly segmented edges are satisfactorily merged. Hence, to further increase the accuracy of the segmentation, improvements should be primarily made in the method to generate the initially oversegmented

images.

In general, the method provides clinically usable results, especially for ECD and CV. As observed in Figures 2.5–2.6, the error of those two biomarkers are relatively low, although the error for HEX is still rather large. Whether this accuracy is clinically acceptable depends more on the actual application and/or disease under study. Nonetheless, a manual correction could be suggested to correct mistakes. This could be done in a user-friendly manner, permitting the user to remove a false edge (or to undo an erroneously removed true edge) with a simple click, without the need of tracing it manually.

REFERENCES

[1] G. O. Waring, W. M. Bourne, H. F. Edelhauser, and K. R. Kenyon, *The corneal endothelium. Normal and pathologic structure and function,* Ophthalmology **89**, 531 (1982).

[2] P. A. Mohammad-Salih, *Corneal endothelial cell density and morphology in normal Malay eyes,* The Medical Journal of Malaysia **66**, 300 (2011).

[3] W. J. Armitage, A. D. Dick, and W. M. Bourne, *Predicting endothelial cell loss and long-term corneal graft survival,* Investigative Ophthalmology & Visual Science **44**, 3326 (2003).

[4] W. M. Bourne, *Biology of the corneal endothelium in health and disease,* Eye **17**, 912 (2003).

[5] I. Jalbert, F. Stapleton, E. Papas, D. F. Sweeney, and M. Coroneo, *In vivo confocal microscopy of the human cornea,* British Journal of Ophthalmology **87**, 225 (2003).

[6] A. S. Kitzmann, E. J. Winter, C. B. Nau, J. W. McLaren, D. O. Hodge, and W. M. Bourne, *Comparison of corneal endothelial cell images from a noncontact specular microscope and a scanning confocal microscope,* Cornea **24**, 980 (2005).

[7] M. L. Salvetat, M. Zeppieri, F. Miani, L. Parisi, M. Felletti, and P. Brusini, *Comparison between laser scanning in vivo confocal microscopy and noncontact specular microscopy in assessing corneal endothelial cell density and central corneal thickness,* Cornea **30**, 754 (2011).

[8] M. Hara, N. Morishige, T. Chikama, and T. Nishida, *Comparison of confocal biomicroscopy and noncontact specular microscopy for evaluation of the corneal endothelium,* Cornea **22**, 512 (2003).

[9] J. Huang, J. Maram, T. C. Tepelus, S. R. Sadda, V. Chopra, and O. L. Lee, *Comparison of noncontact specular and confocal microscopy for evaluation of corneal endothelium,* Eye & Contact Lens **44**, S144 (2018).

[10] S. Jonuscheit, M. J. Doughty, and K. Ramaesh, *In vivo confocal microscopy of the corneal endothelium: comparison of three morphometry methods after corneal transplantation,* Eye **25**, 1130 (2011).

[11] M. O. Price, K. M. Fairchild, and F. W. Price, *Comparison of manual and automated endothelial cell density analysis in normal eyes and DSEK eyes,* Cornea **32**, 567 (2013).

[12] C. M. Klais, J. Buhren, and T. Kohnen, *Comparison of endothelial cell count using confocal and contact specular microscopy,* Ophthalmologica **217**, 99 (2003).

[13] C. Hirneiss, R. G. Schumann, M. Gruterich, U. C. Welge-Luessen, A. Kampik, and A. S. Neubauer, *Endothelial cell density in donor corneas: a comparison of automatic software programs with manual counting,* Cornea **26**, 80 (2007).

[14] B. Selig, K. A. Vermeer, B. Rieger, T. Hillenaar, and C. L. Luengo Hendriks, *Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy,* BMC Medical Imaging **15:13** (2015).

[15] A. Ruggeri, F. Scarpa, M. De Luca, C. Meltendorf, and J. Schroeter, *A system for the automatic estimation of morphometric parameters of corneal endothelium in alizarine red-stained images,* British Journal of Ophthalmology **94**, 643 (2010).

[16] F. Scarpa and A. Ruggeri, *Development of a reliable automated algorithm for the morphometric analysis of human corneal endothelium,* Cornea **35**, 1222 (2016).

[17] Y. Gavet and J. C. Pinoli, *Comparison and supervised learning of segmentation methods dedicated to specular microscope images of corneal endothelium,* International Journal of Biomedical Imaging **2014**, 1 (2014).

[18] M. S. Sharif, R. Qahwaji, E. Shahamatnia, R. Alzubaidi, S. Ipson, and A. Brahma, *An efficient intelligent analysis system for confocal corneal endothelium images,* Computer Methods and Programs in Biomedicine **122**, 421 (2015).

[19] X. Song, L. Zhou, Z. Li, J. Chen, B. Yan, and L. Zeng, *Interactive image segmentation based on hierarchical superpixels initialization and region merging,* in *International Congress on Image & Signal Processing* (Cherbourg, France, 2014) pp. 410–414.

[20] L. Li, J. Yao, J. Tu, X. Lu, K. Li, and Y. Liu, *Edge-based split-and-merge superpixel segmentation,* in *2015 IEEE International Conference on Information and Automation* (Lijiang, China, 2015) pp. 970–975.

[21] C.-Y. Hsu and J.-J. Ding, *Efficient image segmentation algorithm using SLIC superpixels and boundary-focused region merging,* in *9th International Conference on Information, Communications & Signal Processing* (Tainan, Taiwan, 2013) pp. 1–15.

[22] S. Zhu, D. Cao, Y. Wu, and S. Jiang, *Improved accuracy of superpixel segmentation by region merging method,* Frontiers of Optoelectronics **9**, 633 (2016).

[23] S. J. Chiu, C. A. Toth, C. Bowes Rickman, J. A. Izatt, and S. Farsiu, *Automatic segmentation of closed-contour features in ophthalmic images using graph theory and dynamic programming,* Biomedical Optics Express **3**, 1127 (2012).

[24] J. C. Platt, *Probabilistic outputs for Support Vector Machines and comparisons to regularized likelihood methods,* in *Advances in large margin classifiers* (1999) pp. 61–74.

[25] D. P. Chakraborty and K. S. Berbaum, *Observer studies involving detection and localization: modeling, analysis, and validation,* Medical Physics **31**, 2313 (2004).

[26] J. D. Ding, L. V. Johnson, R. Herrmann, S. Farsiu, S. G. Smith, M. Groelle, B. E. Mace, P. Sullivan, J. A. Jamison, U. Kelly, O. Harrabi, S. Subbarao Bollini, J. Dilley, D. Kobayashi, B. Kuang, W. Li, J. Pons, J. C. Lin, and C. Bowes Rickman, *Anti-amyloid therapy protects against retinal pigmented epithelium damage and vision loss in a model of age-related macular degeneration,* Proceedings of the Natural Academy of Science of the USA **108**, E279 (2011).

[27] G. Verbeke and G. Molenberghs, *Linear mixed models for longitudinal data*, Springer series in statistics (Springer, 2000) pp. xxii, 568 p.

3

IMPROVED ACCURACY & ROBUSTNESS OF A CORNEAL ENDOTHELIAL CELL SEGMENTATION METHOD BASED ON MERGING SUPERPIXELS

This chapter is based on the manuscript:

J.P. Vigueras-Guillén, A. Engel, H.G. Lemij, J. van Rooij, K.A. Vermeer, and L.J. van Vliet, *Improved accuracy & robustness of a corneal endothelial cell segmentation method based on merging superpixels*, 15th International Conference on Image Analysis and Recognition (ICIAR), Póvoa de Varzim, Portugal. Lecture Notes in Computer Science **10882**, pp. 631–638 (2018).

Abstract

Clinical parameters related to the corneal endothelium can be estimated by segmenting endothelial cell images. Specular microscopy is the current standard technique to image the endothelium, but its low signal-to-noise ratio make the segmentation a complicated task. Recently, we proposed a method to segment such images by starting with an oversegmented image and merging the superpixels that constitute a cell. Here, we show how our merging method provides better results than optimizing the segmentation itself. Furthermore, our method can provide accurate results despite the degree of the initial oversegmentation, resulting into a precision and recall of 0.91 for the optimal oversegmentation.

3.1. INTRODUCTION

The endothelium of the cornea, a monolayer of hexagonal shaped cells on the posterior corneal surface, plays a key role in keeping an optimal state of corneal hydration [1]. Human endothelial cell density decreases naturally with increasing age, and regeneration of endothelial cells has not been observed under normal circumstances. Instead, endothelial cells grow and migrate to occupy the space that is freed by the dying cells. Intraocular surgery, trauma, and certain diseases may accelerate cell loss. If cell density drops below 700 cells/mm^2, the resulting corneal edema may disrupt vision [2]. Accurate quantification of the endothelial cell morphology is crucial for the assessment of the health status of the cornea.

Noncontact specular microscopy and contact confocal microscopy are commonly used in clinical practice to image the endothelium *in vivo*. Characterized by its noninvasive nature and fast acquisition time, noncontact specular microscopy provides reliable and reproducible measurements [3]. In comparison, contact confocal microscopy requires corneal contact and has longer acquisition time, but it provides higher quality endothelial images in diseased and edematous corneas [4]. Furthermore, ocular microsaccades, respiration, and pulse are important, limiting factors in the acquisition of good quality images *in vivo*.

Endothelial cell density (ECD), coefficient of variation in cell size (CV), and hexagonality (HEX) are the three main parameters to assess the endothelium. Microscope manufacturers provide built-in software to automatically segment endothelium images and estimate these parameters. However, many studies have indicated that current microscope software may provide unreliable estimations, specifically in the cases with low or high cell density or a high degree of CV [3, 5, 6].

In recent years, several cell segmentation techniques for *in vivo* corneal endothelium images were proposed. More studies were focused on segmenting confocal images [7, 8] rather than specular [9], probably because of its better image quality and signal-to-noise ratio (SNR). Hence, there is a need to develop algorithms that can accurately segment specular microscopy endothelial images.

3.1.1. OUR AIM

We have recently developed an automatic algorithm to segment *in vivo* specular microscopy images [10]. Briefly summarized, our method aims to generate an accurate segmentation by starting with an oversegmented endothelial cell image and merging those superpixels that together comprise a whole cell (Figure 3.1). We showed how a machine learning approach (based on Support Vector Machines) can use features related to shape, intensity, size, etc., to identify the superpixels that constitute a complete cell. During the merging process, all combinations of two and three neighboring superpixels are evaluated simultaneously, using a dedicated classifier for the merger of two or three superpixels. The binary classification is then transformed into a probabilistic output, which allows to sort the combinations. In a iterative process, the combination with the highest probability is merged until no more acceptable combinations ($P > 0.5$) remain.

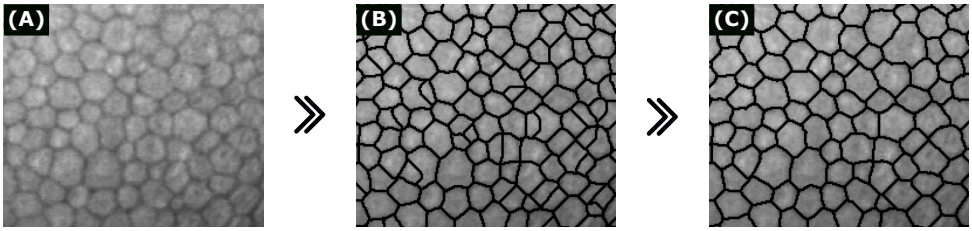

Figure 3.1: (A) Input image. (B) Initial oversegmentation. (C) Output of merging process.

Although any method that can generate oversegmented cell images could be used as a starting point, we chose to adapt Selig et al.'s approach [7] because of its simplicity to generate and adjust oversegmentation. Selig et al. designed a seeded watershed algorithm in a stochastic manner to segment *in vivo* confocal microscopy images, which requires fine-tuning of several parameters in order to achieve a satisfactory result. In our approach [10], parameter tuning was not necessary as we aimed for oversegmentation.

In this paper, we show how our method can improve the accuracy of Selig et al.'s [7] optimal segmentation when applied to *in vivo* specular microscopy images. By using precision and recall to evaluate the segmentation, we obtained the best parameter values for Selig et al.'s method and computed the corresponding segmentation (named "*watershed optimized result*"). After applying our merging method [10] to the watershed optimized result, we evaluated the segmentation results. Furthermore, we generated several oversegmented images with different degrees of oversegmentation, applied our merging method to all of them, and evaluated the resulting segmentation, thereby showing the robustness of our method against various degrees of oversegmentation.

3.2. METHOD

3.2.1. MATERIALS

The dataset contains 30 corneal endothelium images from the central cornea of 30 glaucomatous eyes, acquired in The Rotterdam Eye Hospital with a noncontact specular microscope (SP-1P, Topcon, Tokyo, Japan) for an ongoing study regarding the implantation of a Baerveldt glaucoma drainage device. Each image covers an area of 0.25 mm × 0.55 mm and they were saved as 8-bits grayscale images of 241×529 pixels. The dataset shows a large variability in cell morphology, with a range of 1400–2700 cells/mm^2 in ECD, 19–35% in CV, and 44–73% in HEX. The acquisition occurred with informed consent and followed the tenets of the Declaration of Helsinki. One expert created the gold standard by performing manual segmentation of the edges using an image manipulation program, GIMP.

3.2.2. DESCRIPTION OF THE OVERSEGMENTATION METHOD

Selig et al.'s method [7] employs a seeded stochastic watershed algorithm to segment endothelial cell images. The seeds are arranged in a hexagonal grid to mimic the en-

dothelial cells pattern. The grid size is derived by estimating the most common cell size, which can be computed from the Fourier spectrum of the image (see Section 3.2.3). The seeded watershed algorithm is repeatedly applied, $m = 100$ times (value derived in [11]), to a randomly rotated and translated grid. Uniformly distributed noise in the range of $[0, u]$ is added to the image at each iteration, which avoids the occurrence of spurious segmentation lines [11]. The result of all seeded watershed segmentations are summed together, providing an image named *likelihood map* (or PDF), which indicates how often each pixel was selected as edge. To smooth the result, a Gaussian smoothing filter is applied to the PDF, whose optimal σ_{PDF} is related to the cell size. To adapt σ_{PDF} to the cell size of each image, Selig et al. defined a parameter $k_\sigma = \sigma_{PDF} f^*$, where f^* is the characteristic frequency estimated from the Fourier spectrum of the image. An H-minima transform is then applied to the PDF, $h = k_h m/\sigma_{PDF}$, to discard small minima. Finally, the classical watershed is applied to the PDF, resulting in the final segmentation. Selig et al. observed that the smoothing was very sensitive to errors in the estimation of f^*, and so they suggested to re-estimate f^* from the PDF. In summary, three parameters are to be tuned in the algorithm: noise amplitude u, smoothing size k_σ, and local minima depth k_h.

3.2.3. FREQUENCY ANALYSIS

The 2D Fourier transform (2D FT) of an endothelial image shows a distinctive concentric ring due to the fairly regular size and pattern of the cells [12]. The ring's radius, called *characteristic frequency* (f^*), is related to the most common cell size in the image, $l = 1/f^*$ [7]. To determine the radius, we compute the 1D radial magnitude by angular averaging of the magnitude of the 2D FT,

$$\mathscr{F}_{RM}(f) = \frac{1}{2\pi}\int_0^{2\pi} |\mathscr{F}(f,\theta)| d\theta, \tag{3.1}$$

where $\mathscr{F}(f,\theta)$ is the Fourier transform of the intensity image in polar coordinates. Thus, the ring radius in the 2D FT appears as a peak in the 1D radial magnitude. Since specular microscopy images have lower SNR, lower contrast, and lower resolution than confocal images, the ring appears almost imperceptible in the radial magnitude (Figure 3.2-A, arrow). Selig et al. [7] proposed a method, named reconstruction by dilation, to enhance the peak (Figure 3.2-B) and find its value by fitting a parabola. While the enhancement method works well in specular microscopy images, the parabola fitting must be applied with caution because it is prone to generate wrong estimations. Instead, we propose to fit a function that is comprised of a decaying exponential and a Gaussian,

$$g(f; a, b, c, \mu, \sigma) = a\exp(-bf) + \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(f-\mu)^2}{2\sigma^2}\right) + c, \tag{3.2}$$

where g is the model, f is the spatial frequency, a and b are the scale parameters of the exponential, μ and σ are the mean and standard deviation of the Gaussian respectively, and c is the offset. A non-linear least-squares solver was used to find the parameters. The characteristic frequency is then estimated as the position of the Gaussian (μ). As suggested by Selig et al. [7], the estimation of f^* can be improved if the model is fitted to the Fourier transform of the PDF (Figure 3.2-C).

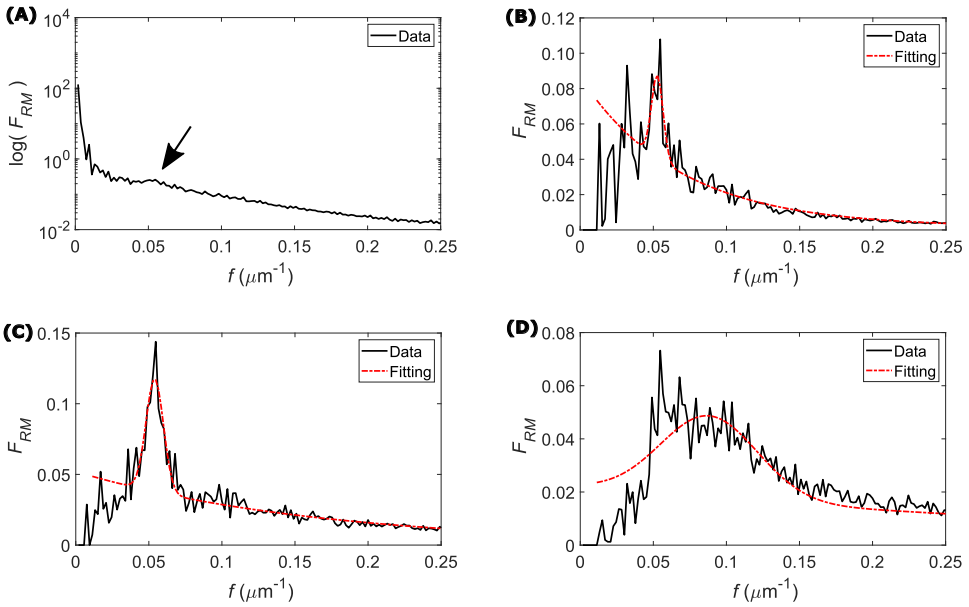

Figure 3.2: The radial magnitude of the FT after angular averaging at different stages of preprocessing in order to estimate f^*. (A) No preprocessing; f^* is almost indiscernible (arrow). (B) After reconstruction by dilation in the FT of the image (fitting gives a $f^* = 0.0527$). (C) After reconstruction by dilation in the FT of the PDF (fitting indicates a $f^* = 0.0541$). (D) After reconstruction by dilation in the FT of the PDF where an oversegmentation with $N = 8$ was applied (fitting provides a $f^* = 0.0880$).

3.2.4. GENERATING OVERSEGMENTATION

Oversegmentation can easily be induced by creating a denser grid of seeds in Selig et al.'s method [7]. Given the total image area A_I, the expected number of cells in the image is $n_{seeds} = NA_I f^{*2}$, with $N = 1$ for the watershed optimized segmentation. In our merging method, we used a value $N = 3$, which is equivalent to a cell density three times higher. In this paper, we generate nine oversegmented images by taking the integer values $N = 1, 2, ..., 9$.

Two scenarios to create oversegmentation were considered. If f^* is re-estimated in the PDF after oversegmentation was applied, the frequency spectrum will show the components of the smaller superpixels in the PDF (Figure 3.2-D). Then, the fitting function will provide an f^* balanced between the average cell size and the smaller superpixels in the PDF. This simply creates a higher degree of oversegmentation with more irregular sizes of superpixels. Alternatively, we can avoid re-estimation (or to re-estimate it using $N = 1$). Nonetheless, both cases were evaluated here, named *oversg. 1* for the former (re-estimation of f^* in the PDF) and *oversg. 2* for the latter. The degree of oversegmentation in both cases is indicated in Table 3.1.

Table 3.1: Estimated precision and recall for different degrees of oversegmentation (N). The percentage indicates the number of superpixels in the initial oversegmented images relative to N=1. Watershed optimized result (WOR) is added for comparison purposes.

N	**WOR**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Oversg. 1 (%)	-	1	1.12	1.24	1.42	1.70	2.04	2.21	2.45	2.68
Precision	0.83	0.89	0.91	0.91	0.91	0.90	0.89	0.88	0.88	0.86
Recall	0.85	0.86	0.90	0.91	0.91	0.91	0.92	0.92	0.92	0.90
Oversg. 2 (%)	-	1	1.10	1.20	1.28	1.35	1.39	1.42	1.44	1.47
Precision	0.83	0.89	0.91	0.91	0.89	0.89	0.88	0.86	0.85	0.85
Recall	0.85	0.86	0.90	0.90	0.89	0.89	0.88	0.86	0.85	0.85

3.2.5. EVALUATING THE SEGMENTATION

In order to evaluate the segmentation, we computed two values from the segmented images, the total number of superpixels (n_{total}) and the number of cells correctly segmented (n_{corr}), as well as one value from the gold standard, the number of real cells (n_{real}). We considered a cell is correctly segmented based on the following rule: given the area of a cell in the segmentation A_S, the area of a cell in the gold standard A_G, and the intersection of those two areas $A_I = A_S \cap A_G$, the cell in the segmentation is correctly segmented if $A_I > 0.75 \times max(A_S, A_G)$. That margin was added to allow small deviations in the cell boundary locations and was selected after visual analysis.

The precision $p = n_{corr}/n_{total}$ and the recall $r = n_{corr}/n_{real}$ were computed and combined into the F-measure, $F = 2pr/(p+r)$. Although both, over- and undersegmented cells affect both metrics, precision decreases more acutely with oversegmentation and recall decreases more strongly with undersegmentation.

3.3. RESULTS & DISCUSSION

3.3.1. WATERSHED OPTIMIZED RESULT

Selig et al.'s algorithm was applied to all images in the dataset for all values of u between 0 and 50 in steps of 10, values of k_σ between 0.10 and 0.25 in steps of 0.01, and values of k_h between 0.000 and 0.010 in steps of 0.001. The F-measure was computed for each image. In a leave-one-out approach, the parameters for image i were estimated as the ones that yielded the largest average F when computed for all images in the dataset excluding i. For all images, the optimal parameters were the same: $u_i = 20$, $k_{\sigma,i} = 0.20$, and $k_{h,i} = 0$.

When evaluated the segmentation without merging, the mean and standard deviation of the precision and recall with the optimal parameters were $p = 0.83 \pm 0.07$ and $r = 0.85 \pm 0.06$, yielding $F = 0.84 \pm 0.07$. This suggested that over- as well as undersegmentation were present.

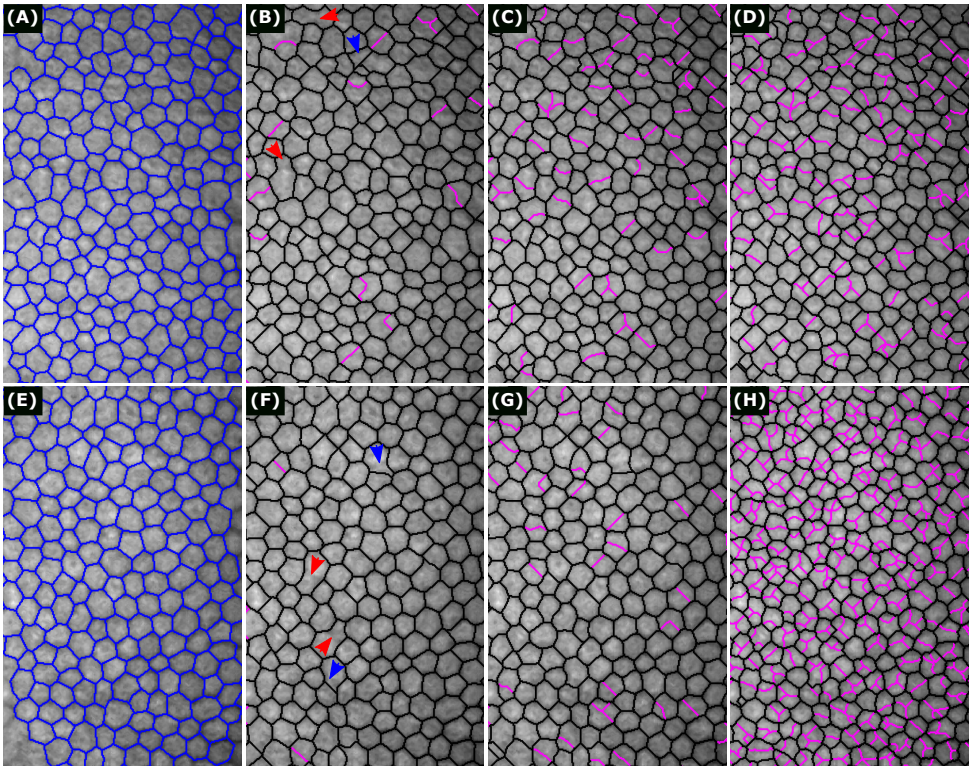

Figure 3.3: Two representative examples: *oversg. 2* in (A–D), and *oversg. 1* in (E–H). (A,E) Gold standard segmentation in blue, superimposed over the intensity image. In the remaining images, resulting segmentation in black and edges removed during merging process in magenta, for an oversegmentation of $N = 1$ in (B,F), $N = 3$ in (C,G), and $N = 9$ in (D,H). Note that watershed optimized result is (B,F) with all edges (black and magenta). Red arrows indicate undersegmentation. Blue arrows indicate inaccurate segmentation.

3.3.2. SOLVING THE OVERSEGMENTATION

Based on the optimal parameters obtained in Section 3.3.1, nine oversegmented images (per image in the dataset) were generated for each type of oversegmentation (as described in Section 3.2.4) after which the merging method [10] was applied to all of them without retraining. The evaluation of the resulting segmentations shows that the watershed optimized result can be directly improved by just applying our merging method (WOR vs. $N = 1$, in Table 3.1). In a visual evaluation (Figure 3.3-B, F), it is clear that the watershed optimized result generates under- (red arrows) and over-segmentation (magenta edges). Whereas the latter is mostly solved by the merging method (large increase in precision from WOR to $N = 1$ in Table 3.1), the former cannot be easily fixed as the number of possible splits grows exponentially with the number of edge pixels per undersegmented superpixel. It is then crucial for our merging method to start with an oversegmented image where no undersegmentation occurs.

The optimal degree of oversegmentation occurs at $N = 3$ for both types of overseg-

mentation ($F_{oversg.1}$ = 0.910, and $F_{oversg.2}$ = 0.906). All true edges seem to be detected at that point, and the degree of oversegmentation is not excessive. For higher values of N, it would be expected that more (presumably unnecessary) initial oversegmentation would only increase the chances of more errors. Interestingly, this hardly happens for the cases in *oversg. 1*. The reason lies in the inaccurate segmentation of some edges due to a limitation of the stochastic watershed: when a strong false edge is detected close to a real edge (blue arrows in Figure 3.3-F), the latter cannot be detected unless we force a large amount of oversegmentation (Figure 3.3-H). Whereas such a degree of oversegmentation would suggest a higher error probability, the merging method can satisfactorily overcome this problem. Nonetheless, this evaluation suggests that the best setup is using *oversg. 1* with $N = 3$, as it is more robust.

In summary, we have shown how a segmentation method based on merging superpixels can improve the accuracy of another segmentation method specifically designed to solve confocal images. Furthermore, we have proven how such a merging method is strongly robust against the degree of initial oversegmentation without requiring any retraining.

REFERENCES

[1] W. M. Bourne, *Biology of the corneal endothelium in health and disease,* Eye **17**, 912 (2003).

[2] W. M. Bourne and J. W. McLaren, *Clinical responses of the corneal endothelium,* Experimental Eye Research **78**, 561 (2004).

[3] M. L. Salvetat, M. Zeppieri, F. Miani, L. Parisi, M. Felletti, and P. Brusini, *Comparison between laser scanning in vivo confocal microscopy and noncontact specular microscopy in assessing corneal endothelial cell density and central corneal thickness,* Cornea **30**, 754 (2011).

[4] M. Hara, N. Morishige, T. Chikama, and T. Nishida, *Comparison of confocal biomicroscopy and noncontact specular microscopy for evaluation of the corneal endothelium,* Cornea **22**, 512 (2003).

[5] C. Hirneiss, R. G. Schumann, M. Gruterich, U. C. Welge-Luessen, A. Kampik, and A. S. Neubauer, *Endothelial cell density in donor corneas: a comparison of automatic software programs with manual counting,* Cornea **26**, 80 (2007).

[6] J. Huang, J. Maram, T. C. Tepelus, S. R. Sadda, V. Chopra, and O. L. Lee, *Comparison of noncontact specular and confocal microscopy for evaluation of corneal endothelium,* Eye & Contact Lens **44**, S144 (2018).

[7] B. Selig, K. A. Vermeer, B. Rieger, T. Hillenaar, and C. L. Luengo Hendriks, *Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy,* BMC Medical Imaging **15:13** (2015).

[8] M. S. Sharif, R. Qahwaji, E. Shahamatnia, R. Alzubaidi, S. Ipson, and A. Brahma, *An efficient intelligent analysis system for confocal corneal endothelium images,* Computer Methods and Programs in Biomedicine **122**, 421 (2015).

[9] F. Scarpa and A. Ruggeri, *Development of a reliable automated algorithm for the morphometric analysis of human corneal endothelium,* Cornea **35**, 1222 (2016).

[10] J. P. Vigueras-Guillén, E. R. Andrinopoulou, A. Engel, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Corneal endothelial cell segmentation by classier-based merging of oversegmented images,* IEEE Transactions on Medical Imaging **37**, 2278 (2018).

[11] K. B. Bernander, K. Gustavsson, B. Selig, I. M. Sintorn, and C. L. Luengo Hendriks, *Improving the stochastic watershed,* Pattern Recognition Letters **34**, 993 (2013).

[12] M. Foracchia and A. Ruggeri, *Automatic estimation of endothelium cell density in donor corneas by means of Fourier analysis,* Medical and Biological Engineering and Computing **42**, 725 (2004).

4

Fully convolutional architecture versus sliding-window CNN for corneal endothelium cell segmentation

This chapter is based on the manuscript:

J.P. Vigueras-Guillén, B. Sari, S.F. Goes, H.G. Lemij, J. van Rooij, K.A. Vermeer, and L.J. van Vliet, *Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation*, BMC Biomedical Engineering **1:4** (2019).

4

Abstract

Background: *Corneal endothelium (CE) images provide valuable clinical information regarding the health state of the cornea. Computation of the clinical morphometric parameters requires the segmentation of endothelial cell images. Current techniques to image the endothelium in vivo deliver low quality images, which makes automatic segmentation a complicated task. Here, we present two convolutional neural networks (CNN) to segment CE images: a global fully convolutional approach based on U-net, and a local sliding-window network (SW-net). We propose to use probabilistic labels instead of binary, we evaluate a preprocessing method to enhance the contrast of images, and we introduce a postprocessing method based on Fourier analysis and watershed to convert the CNN output images into the final cell segmentation. Both methods are applied to 50 images acquired with an SP-1P Topcon specular microscope. Estimates are compared against a manual delineation made by a trained observer.*

Results: *U-net (AUC = 0.9938) yields slightly sharper, clearer images than SW-net (AUC = 0.9921). After postprocessing, U-net obtains a DICE = 0.981 and a MHD = 0.22 (modified Hausdorff distance), whereas SW-net yields a DICE = 0.978 and a MHD = 0.30. U-net generates a wrong cell segmentation in only 0.48% of the cells, versus 0.92% for the SW-net. U-net achieves statistically significant better precision and accuracy than both, Topcon and SW-net, for the estimates of three clinical parameters: cell density (ECD), polymegethism (CV), and pleomorphism (HEX). The mean absolute percentage error in U-net for the parameters is 0.4% in ECD, 2.8% in CV, and 1.3% in HEX. The computation time to segment an image and estimate the parameters is barely a few seconds.*

Conclusions: *Both methods presented here provide a statistically significant improvement over the state of the art. U-net has reached the smallest error rate. We suggest a segmentation refinement based on our previous work to further improve the performance.*

4.1. INTRODUCTION

Convolutional Neural Networks (CNNs) have considerably advanced the state of the art in computer vision in the last years. Although they were introduced 30 years ago [1], it was not until recently that improvements in computer hardware allowed large-scale training of more complex, deep networks [2]. Whilst the typical use of CNNs was aimed at learning classification tasks, segmentation is also a desired outcome in medical imaging. In 2012, Cireşan et al. [3] employed a typical classification architecture to perform tissue segmentation. They segmented neural membranes images from electron microscopy by using a CNN in a sliding-window setup such that in order to predict the class label of a target pixel, a local region (patch) around that pixel was provided as input. Although this strategy yielded great results (it won the ISBI 2012 challenge), it was computationally expensive and did not exploit the redundancy between overlapping patches. In 2015, Ronneberger et al. [4] proposed the U-net, which turned out to be a major contribution to the field of biomedical image segmentation. This network, an extension of a "fully convolutional network" presented in a previous paper [5], had the benefits of faster training by introducing skip-layer connections between layers of the same resolution and by not using fully connected layers. U-nets accept the whole image as input and obtain good results with just a very few annotated images to train on, which made it win the ISBI 2015 challenge. In this paper we aim to adapt, improve, and evaluate a local sliding-window CNN (named SW-net) and a global fully convolutional U-net to segment corneal endothelium (CE) images obtained with specular microscopy.

The CE is a monolayer of closely packed and predominantly hexagonally-shaped cells on the posterior surface of the cornea. Endothelial cells are 4–6 µm in height and 20 µm in width [6], and they play a key role in maintaining an optimal state of corneal hydration [7], but they do not undergo mitosis *in vivo*. Instead, when cells are lost through age-related apoptosis or trauma, the remaining healthy cells grow and migrate to occupy the space of the lost cells. As a result, the CE cell architecture loses its hexagonal appearance. In young adults, the endothelial cell density is within 3000–3500 cells/mm^2, but generally around 2000 cells/mm^2 in elderly people [8]. If the cell density reaches a critical point due to trauma or eye diseases (around 500–700 cells/mm^2), corneal edema occurs. Since edema leads to poor vision, corneal transplantation is usually the treatment in those situations.

Currently, three parameters are used to evaluate the health status of the endothelium: endothelial cell density (ECD), polymegethism (or coefficient of variation in cell size, CV), and pleomorphism (or hexagonality, HEX). To correctly estimate the clinical parameters, an accurate segmentation of the cells is necessary. The current clinical standard technique to image the endothelium *in vivo* is noncontact specular microscopy, which is fast and noninvasive. However, images might appear blurred since this technology requires corneas to have a smooth endothelium surface [9]. In addition, noise, illumination distortions, and optical artifacts are commonly present in specular images.

Manual delineation of the cells is a very labor-intensive task. Existing commercial software for cell segmentation, usually provided by the microscope manufacturers, has limited performance. Several studies using specular microscopy have shown the inaccuracy of the automated analyses [10–13]. For instance, Luft et al. [14] compared four

different noncontact specular microscopes in combination with their built-in segmentation software (models: EM-3000, Tomey; CEM-530, Nidek; CellChek XL, Konan; and Perseus, Bon Optic) in healthy eyes and eyes with corneal grafts, and concluded that all models (except Konan) significantly underestimated ECD in the subgroup of healthy eyes, whereas ECD was significantly overestimated in the corneal graft group for all models.

Several algorithms for *in vivo* corneal endothelium cell segmentation have been proposed in the last three decades. The early approaches (90s and early 00s) used simple methods, such as a combination of thresholding, skeletonization, Gaussian filtering, and morphological operations [6, 15, 16], shape dependent filters [17], and the seeded watershed algorithm [18–20] (each one using different morphological operations to place the seeds). These methods only provided relatively good results for high-quality images and their clinical application was never evaluated. Moreover, many of them suggested the necessity of user interaction to correct errors. In contrast, new clinically applicable methods have been proposed in recent years: Foracchia and Ruggeri [21] developed an algorithm based on Bayesian shape models, which later evolved into a genetic algorithm by Scarpa and Ruggeri [22]; Sharif et al. [23] developed a hybrid model based on a combination of an active contour model (snakes) and a particle swarm optimization approach; Habrat et al. [24] proposed an algorithm based on directional filters, which was clinically evaluated along with other methods [25]; Al-Fahdawi et al. [26] suggested a method based on the watershed algorithm and Voronoi tessellations; Selig et al. [27] employed Fourier analysis and the seeded watershed algorithm in a stochastic manner to segment confocal images; and Vigueras-Guillén et al. [28] proposed a classifier-driven method to generate an accurate segmentation from an oversegmented image, using Selig et al.'s approach [27] to generate the oversegmentation. Among these methods, the ones including a comparison with their respective microscope's estimates were significantly more accurate, yet some mistakes were still present.

Regarding the use of neural networks or CNNs to segment CE images, four algorithms were published in the last year. Fabijańska [29] proposed a feed-forward neural network with one hidden layer to segment 30 *ex vivo* endothelial images from phase-contrast microscopy (dataset published in [30]), achieving an error in cell number detection of 5% and a DICE [31] value of 0.85. Nurzynska [32] further improved the results on the same dataset by employing a CNN in a sliding-window setup, using a similar network as Cireşan et al. [3], and obtaining a precision of 93% and a DICE of 0.94. Phase-contrast microscopy yields *ex vivo* CE images of high quality, which cannot be compared with *in vivo* specular microscopy. In fact, we already solved that dataset, achieving a segmentation error in only 0.28% of the cells and an average error in the clinical parameter estimates of less than 0.4% [28]. Katafuchi et al. [33] also used a CNN in a sliding-window setup to segment human endothelium *in vivo*, although they did not specify the imaging technology. They also employed a similar network as Cireşan et al. [3], and they achieved an error rate of 12%. Since neither of these two papers did a clinical evaluation, no further comparison can be described here. Finally, Fabijańska [34] was the first to apply the U-net to specular images, although using patches as input instead of whole images. She achieved a DICE of 0.85, an AUC (area under the ROC curve) of 0.92, and the error in the clinical parameters were 5.2% in ECD, 11.93% in CV, and 6.2% in HEX. In other image

modalities, different neural networks architectures have been used for image segmentation, such as the use of fuzzy deep neural networks for brain MRI images [35] in order to extract information from both fuzzy and neural representations. Whereas the use of these sophisticated architectures in CE images has not been studied yet, it does not seem to be necessary given the rather low complexity of the cell patterns in CE images.

In summary, two main approaches have been exploited when using CNNs to segment endothelial cell images: via pixel classification (sliding-window setup, SW-net), or via direct segmentation (U-net). Here, we aim to clarify which approach is more optimal, proposing and evaluating two end-to-end solutions to segment *in vivo* CE images acquired with specular microscopy. Specifically, (1) we use a preprocessing technique, a contrast limited adaptive histogram equalization (CLAHE) [36], to enhance the contrast of the images, and evaluate whether any image normalization is beneficial; (2) we propose a modification of the image labels to make them probabilistic instead of binary, which improves the performance; (3) we evaluate several implementation choices of the CNNs; and (4) we suggest a postprocessing method to the CNN output in order to create the final segmented images.

This paper is organized as follows. In the Methods section, we describe the dataset, we illustrate the two networks highlighting the changes we introduce, and we describe the pre- and postprocessing techniques in detail, as well as all the metrics and statistical analysis employed. In the Results section we evaluate the two networks in three ways: the performance of the CNNs and the importance of certain implementation details; the segmentation after applying the postprocessing method, reporting the distance to and similarity with the gold standard, as well as the percentage of correctly detected cells; and the accuracy of the estimated clinical parameters. In the Discussion section, we highlight the main findings and compare the results with some of the aforementioned methods. In the Conclusions section, we summarize the relevance of this study.

4.2. METHODS

4.2.1. MATERIALS

The dataset contains 50 corneal endothelium images from the central cornea of 50 glaucomatous eyes, imaged with a noncontact specular microscope (SP-1P, Topcon Co., Tokyo, Japan). They are part of an ongoing study in The Rotterdam Eye Hospital regarding the implantation of a Baerveldt glaucoma drainage device. Data was collected in accordance with the tenets of the Declaration of Helsinki. Signed informed consent was obtained from all subjects. Approval was obtained from the Medical Ethical Committee of the Erasmus Medical Center, Rotterdam, the Netherlands (MEC-2014-573). Trial registration: NL4823 (https://www.trialregister.nl/trial/4823), registered 06/01/2015.

Glaucoma is a condition related to the buildup of pressure inside the eye, which can eventually damage the optic nerve. In primary open-angle glaucoma (POAG), the eye cannot properly drain the aqueous humor through its drainage system, whereas in primary angle-closure glaucoma (PACG) the iris blocks the entrance of the drainage system. In PACG, surgical intervention is usually required to remove the blockage. In POAG, eye drops are the first treatment option in mild cases, either to reduce the formation of fluid

in the eye or increase the outflow, but surgical intervention is usually considered when these treatment modalities have proven ineffective. Trabeculectomy is a common procedure, which consists of a small hole in the sclera, covered by a thin trap-door, which makes it possible to drain the aqueous humor out of the eye. However, scarring may lead to failure of the trabeculectomy. Therefore, but also because of other possible complications with trabeculectomies, glaucoma drainage devices are often preferred over trabeculectomy. Indeed, in refractory cases, the success rates five years postoperatively of Baerveldt implants are higher than those of trabeculectomies [37].

A common postoperative complication after implantation of a Baerveldt (or similar glaucoma drainage) device is a change in the CE, in both cell count and cell shape [38, 39], due to the proximity of the device's tube. In the study currently ongoing in The Rotterdam Eye Hospital, eyes were imaged before and after the implantation of the device. Here, we focused on solving the cases prior to the implantation, which let us assume that the CE was only affected by the natural aging process. Indeed, it has not been observed that glaucoma has any direct effect in the morphology of the CE cells. In our dataset, the average age is 64.8±9.2 (mean ± SD). Our dataset showed a large variability in cell size and morphology, with a range of 1100–2800 cells/mm^2 in ECD, and 18–36% in CV, and 44–74% in HEX.

Each image covers an area of 0.25 mm × 0.55 mm and was saved as 8-bits grayscale images of 240×528 pixels. According to the manufacturer, pixels have a lateral size of 1.038 µm. On average, there are 240 cells per image. One expert created the gold standard by performing manual segmentation of the cell edges using an open-source image manipulation program (GIMP, v.2.8).

4.2.2. U-NET ARCHITECTURE

The U-net follows a standard fully convolutional architecture, with a contraction and an expansion path, each composed of four resolution steps (Figure 4.1). In the contraction path, each step consists of two 4×4 padded convolutions with a rectified linear unit (ReLU), a dropout layer with a drop rate of 50% between the two convolutions, and a 2×2 max pooling with stride 2 at the end for downsampling. In the expansion path, each step contains a 4×4 transposed convolution with stride 2 for upsampling, a concatenation with the corresponding feature map from the contraction path, two 4×4 padded convolutions with ReLU, and a dropout layer with a drop rate of 50% between the convolutions. The convolutional layers in the first resolution step have 32 feature channels, doubling it at each downsampling step, and halving it at each upsampling step. In the last layer, a 1×1 convolution reduces the channels to the number of classes, which is set to two (cell body and cell edges). A cross-entropy loss function with a pixel-wise soft-max activation is used over the final feature map. No class weighting is employed. The optimizer of our choice is Adam [40] with an initial learning rate ($lr_{i=0}$) of 0.001 and a decay of 0.001, such that $lr_i = lr_{i-1} \cdot (1/(1+\text{decay}\cdot\text{iteration}))$, where i denotes iteration. The network accepts the whole image as input. A batch size of 4 images is used.

Compared to the original U-net architecture, several modifications were introduced. First, we used a kernel size of 4×4 instead of 3×3, and the network was downscaled in width and depth, halving the number of feature channels and removing one resolution

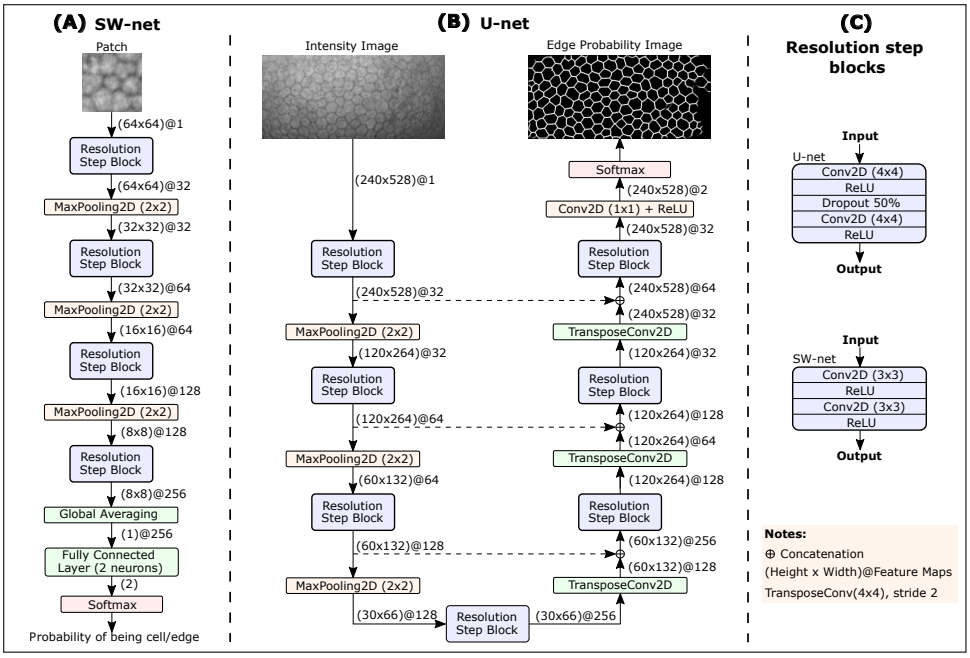

Figure 4.1: Schematic overview of U-net (B) and SW-net (C). The blocks summarizing the layers at the different resolution steps are indicated in (A). For U-net, the network consists of a contraction path and an expansion path. In contrast, SW-net is, in essence, the same contraction path of U-net with a global averaging layer and a fully connected layer of 2 neurons in the end.

step.

Second, dropout layers were added in between the two consecutive convolutions per resolution step. Dropout is a regularization method used to avoid overfitting, originally described for neural networks [41, 42]. It stochastically sets to zero a certain number of activations of hidden units at each training iteration. This prevents the co-adaptation of feature detectors by forcing neurons to rely on population behavior. In CNNs, it simply sets input values (of the feature maps) to zero.

Third, we used transposed convolutions for upsampling in the expansion path. The transposed convolution is described as the operation that forms the same connectivity as the normal convolution but in the opposite direction [43]. Since the weights in the transposed convolution are learned, this avoids to predefine an interpolation method for upsampling. Unfortunately, transposed convolutions can also produce a checkerboard effect due to the uneven overlapping of the filter range in the output pixels [44]. Specifically, the uneven overlapping occurs when the kernel size is not divisible by the stride. While the CNN could, in principle, learn weights to avoid this problem, in practice this effect is often observed, especially in images with strong colors. One practical solution is to use a 4×4 kernel size with a stride of 2 [45]. Nonetheless, we did not observe in our work the checkerboard effect when using filters of 3×3 with a stride of 2.

4.2.3. SW-NET ARCHITECTURE

The SW-net architecture follows the same contraction path as the aforementioned U-net (Figure 4.1). However, instead of the entire image, a patch of size 64×64 is provided as input, and the filter size of the convolutional layers is 3×3. At the end of the contraction path it adds a global averaging pooling layer, where each channel is reduced to its average value, and a fully connected layer of two neurons, which provides the outcome for the two classes regarding the central pixel of the patch. A batch size of 128 patches is used here, which holds a similar amount of data as the batch in our U-net. Moreover, the same loss function and optimizer is employed.

The original Cireşan et al.'s architecture [3] consisted of four stages of one convolutional layer followed by max-pooling. All convolutional layers had 48 feature maps and filters of size 4×4 (one of 5×5). The network ended with two fully connected layers: one of 200 neurons followed by another with 2 neurons to obtain the class labels. In comparison, our network has doubled the number of convolutional layers, albeit with a smaller kernel size, increasing the receptive field (61 pixels instead of 48 pixels). Moreover, we substituted the large fully connected layer with a global averaging pooling. This idea was originally suggested by Lin et al. [46], where they argued that fully connected layers at the end of a CNN are prone to overfitting, whereas global averaging layers are more native to the convolution structure, overfitting is avoided, and the feature maps can be interpreted as categories confidence maps.

4.2.4. PREDICTION

For U-net, the segmentation was retrieved directly from the network output. In the SW-net, one patch per each pixel was retrieved, building up the segmentation image with the classification value of each patch. Images were mirrored in order to extract the patches that reached beyond the image borders.

4.2.5. POSTPROCESSING

To obtain the final segmentation, we smoothed the CNN output and applied the classic watershed algorithm [47]. Specifically, we first estimated the average cell size in the image by Fourier analysis in order to built a Gaussian smoothing filter whose standard deviation was related to that size. It is well known how the 2D Fourier Transform (FT) of a CE image shows a distinctive concentric ring due to the fairly regular pattern of the cells [27], and for the output of the CNN that ring is clearly noticeable (Figure 4.2-A). Selig et al. described in [27] how the radius of the ring, called *characteristic frequency* (f^*), is related to the most common cell size in the image, $l = 1/f^*$. We estimated the radius by first applying a method called "reconstruction by dilation" to remove the low frequencies (defined by Selig et al. in [27]) and later computing the 1D radial magnitude, defined as the angular averaging of the magnitude of the 2D FT of the images,

$$\mathscr{F}_{RM}(f) = \frac{1}{2\pi}\int_0^{2\pi} |\mathscr{F}(f,\theta)|d\theta, \tag{4.1}$$

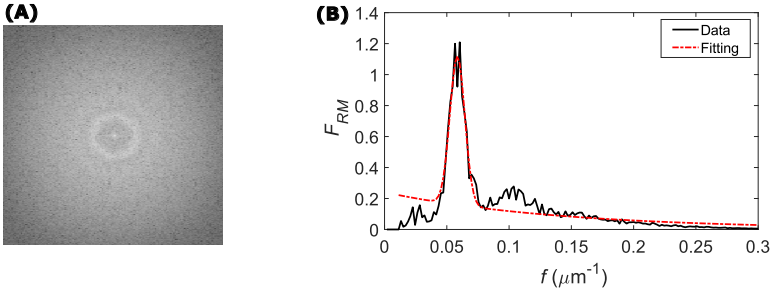

Figure 4.2: (A) 2D FT of the U-net output of a CE image (up to $f = 0.3$). (B) The magnitude of the FT after reconstruction by dilation and angular averaging (black), and the fitted model (red) in order to estimate the peak.

where $\mathscr{F}(f,\theta)$ is the FT of the image in polar coordinates (Figure 4.2-B). In our previous work [48], we described a fitting function to estimate the peak position (f^*) and also derived a parameter, $k_\sigma = 0.20$, used to adapt the filter σ to each image, $\sigma = k_\sigma / f^*$. Once images were smoothed, the watershed algorithm was applied, and the clinical parameters were estimated from the resulting images. The classic watershed does not require any parameter tuning, but it is expected that each object (cell) to detect has a single local minimum, otherwise cells will be oversegmented.

4.2.6. LABELS

The gold standard, a binary image where value 1 indicates a cell edge and value 0 represents a cell body, was defined such that cell edges are 8-connected-pixel lines of 1 pixel-width (Figure 4.3-B). In the intensity image, the cell edges might appear thicker, with a steep but clear transition in intensity from the peak of the edge towards the inner cell. However, this thickness might vary considerably even in the same image (Figure 4.3-A). Hence, instead of using the gold standard images as labels, we proposed to use probabilistic labels where edges appear thicker and in which the aforementioned intensity transition between edges and cells is preserved. There are three reasons for doing so: (1) it is counterproductive to teach the network that the pixels adjacent to the annotated 1-pixel-width edge are cell pixels as they usually have the same characteristics as the annotated edge; (2) mimicking the intensity transition in the labels is a more natural approach and helps the network in its classification task; (3) as the network will learn to replicate this pattern (gradual intensity transition between edges and cell bodies), this will be beneficial when applying the watershed algorithm in the postprocessing step.

To create the probabilistic labels, we convolved the gold standard images with a 7×7 isotropic unnormalized Gaussian filter of standard deviation 1 pixel. This allowed all pixels with label 1 (edges) in the gold standard to keep a value equal to 1 in the probabilistic label image, with increasingly smaller probabilities for pixels further away from the annotated cell edge (Figure 4.3-C). Hence, the pixels in the label image can be regarded as the probability of being part of an edge. This is used as the target output of the networks to be trained. During evaluation, the edge class was considered any pixel with $p > 0.5$.

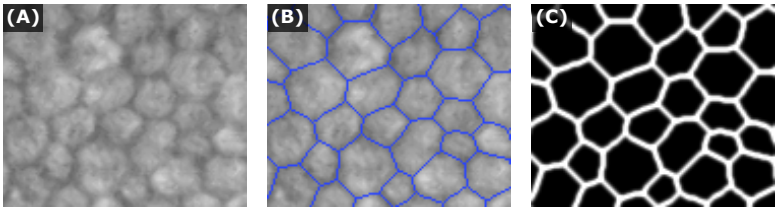

Figure 4.3: (A) Raw intensity image, size 100×120 pixels. (B) Gold standard superimposed on the image. (C) Label image.

In practice this means that we accept a 1 pixel error in the location of the edge. For comparative purposes, we also evaluated the outcome segmentation when the "hard", binary gold standard labels are used as target output.

4.2.7. PREPROCESSING OF THE INTENSITY IMAGES

Specular microscopy images usually have a non-uniform luminosity across the image and low contrast (Figure 4.4-A). Here, we want to evaluate whether the CNN can benefit from some kind of image enhancement. Furthermore, it is common practice in neural networks to standardize the input images,

$$image_{stand} = \frac{\text{image} - mean(\text{image})}{std(\text{image})}, \tag{4.2}$$

or normalize them,

$$image_{norm} = \frac{\text{image} - min(\text{image})}{max(\text{image}) - min(\text{image})}. \tag{4.3}$$

To enhance local image contrast, we proposed to use contrast limited adaptive histogram equalization (CLAHE) [36] with a kernel of 24×24 (Figure 4.4-B). This kernel size matches approximately the area of the average cell. A kernel with a size less than half of a cell would overamplify noise, whereas a kernel too large would reduce the benefits of local contrast enhancement. In earlier work on aneurysm detection in fundus images, we achieved a much better performance with intensity normalization than without it [49].

In summary, we tested the influence of preprocessing by analyzing five possible scenarios: feeding the raw images, normalizing them, standardizing them, and enhancing them by CLAHE (with and without standardization, since the output of CLAHE is already normalized).

4.2.8. DATA AUGMENTATION

Given the nature of the images, flipping them horizontally and/or vertically was a natural way of augmenting the training data by a factor of four. We avoided other transformations, such as rotation or elastic deformations [50], for two reasons: (1) the images show a small degree of distortions only in horizontal and vertical lines, hence rotating or deforming the images would create new noise patterns that do not exist in the original

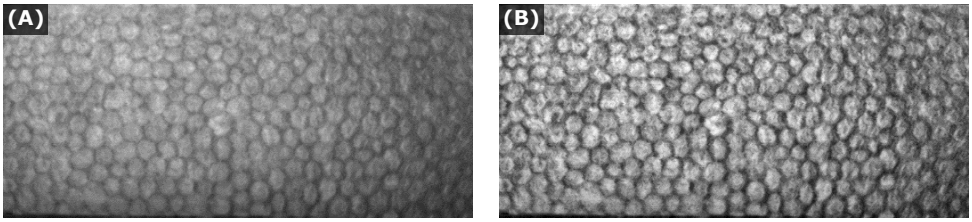

Figure 4.4: (A) A specular microscopy image, size 240×528 pixels. (B) The intensity image after CLAHE.

images; (2) when rotating, the image corners need to be filled, either by mirroring the image or setting that area in black; either way, we are introducing new patterns to be solved by the network.

4.2.9. IMPLEMENTATION DETAILS, AND COMPUTATIONAL COST

The data set was divided in 5 folds of 10 images each. To obtain the optimal network parameters, we used 4 folds for training and 1 for validation/test. For the evaluation of the CNN segmentation and the clinical parameters, a 5-fold cross-validation approach was employed in order to test all the remaining folds, using the parameters determined in the first test set.

Regarding class weighting in U-net, we evaluated whether adding weights in the loss function was advantageous. Here, the edge class has 4 times less pixels than the cell class. For the SW-net, we sampled the same amount of patches per class in each batch.

Other loss functions were tested, specifically mean-squared and mean-absolute loss, but with very similar performance as using cross-entropy. Batch normalization layers [51] were also tested by including them after every ReLU, but this created slightly more overfitting and degraded the performance. Similarly to what Springenberg et al. reported in [52], no differences were observed in SW-net if max-pooling layers were substituted with a stride of 2 in the previous convolutional layer.

CNN filter weights were initialized from an uniform distribution of mean = 0 and width ≈ 1 (glorot uniform initializer in Keras). Networks were coded in Python 3.6 programming language, using the Keras library and Tensorflow as backend. Experiments were run in the free research tool Google Colaboratory, which includes GPU support (Tesla K80), taking roughly 0.8 seconds per training iteration in U-net and 0.5 seconds for the SW-net. The testing took less than 1 second per image for U-net. However, for the SW-net, evaluating all patches in an image took around 1 minute. The postprocessing and parameter estimation took barely 1–2 seconds per image.

4.2.10. METRICS AND STATISTICAL ANALYSIS

In the evaluation of the CNNs performance, accuracy and AUC were provided. However, due to the probabilistic nature of the labels, pixels with label values p close to 0.5 are not relevant for our ultimate goal. Indeed, the most important pixels are either at the crest of the cell edge ($p = 1$) or at the inner cell body ($p = 0$). Furthermore, the class imbalance

makes it important to evaluate each class performance independently. Hence, we also reported the precision (PRE), sensitivity (SEN), and specificity (SPE) for the final designs, but only considering the pixels with values 0 and 1 in the label images. For clarification purposes, we placed an asterisk (*) in the metrics that followed this rule.

In the evaluation of the post-processed segmentation, only the cells within the area of the gold standard were kept, discarding all cells in contact with the image borders. We used the modified Hausdorff distance (MHD) [53] to measure the distance between the gold standard and the proposed segmentation. MHD is defined as

$$\mathrm{MHD}(\mathcal{U},\mathcal{V}) = max(hd(\mathcal{U},\mathcal{V}), hd(\mathcal{V},\mathcal{U})), \tag{4.4}$$

where

$$hd(\mathcal{U},\mathcal{V}) = \frac{1}{|\mathcal{U}|}\sum_{a\in\mathcal{U}}\min_{b\in\mathcal{V}}||a-b||_2, \tag{4.5}$$

$\mathcal{U}$ is the gold standard segmentation, and $\mathcal{V}$ the proposed segmentation.

DICE [31] was used to assess the segmentation at the cell level. We computed the DICE for each cell independently, reporting the average DICE. Specifically, for each cell (C_i) in the gold standard images, we select the superpixel (S_j) in the proposed segmentation with the largest overlap to C_i, such that TP = $C_i \cap S_j$ (True Positive), FN = $C_i \setminus S_j$ (False Negative), FP = $S_j \setminus C_i$ (False Positive),

$$\mathrm{DICE}_{ith\ cell} = \frac{2\cdot TP}{2\cdot TP + FP + FN}, \tag{4.6}$$

$$\mathrm{DICE}_{image} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{DICE}_{ith\ cell}, \tag{4.7}$$

where n is the number of cells in the image.

We also evaluated the number of cells correctly segmented, reporting the number of cells that were oversegmented (divided in more than one superpixel) and undersegmented (within a superpixel that covers more than one cell). We considered a cell was correctly segmented if its $TP_i > 0.80\cdot max(C_i, S_j)$. That margin was added to allow small deviations in the cell boundary locations and was selected after visual analysis.

For the three previous metrics, either the parametric paired t-test or the nonparametric Wilcoxon signed-rank test was performed to determine which method, U-net or SW-net, was more accurate. We used the non-parametric test when the distributions did not fulfill the normality assumption (Shapiro–Wilk normality test). A P-value of less than 0.05 was considered statistically significant.

In the evaluation of the clinical parameters, a statistical analysis based on linear mixed-effects models [54] was performed to determine, for each parameter, whether there was a statistically significant difference in accuracy (smaller absolute mean) and in precision (smaller variance) between the two estimation errors. To determine whether the variances were different, we used a likelihood test to compare a model that assumes equal variances between both estimation errors with a model that assumes different variances. From the fixed effects test of the models we evaluated whether the absolute mean values in both estimations were different. No correction for multiple testing was applied, and a P-value of less than 0.05 was considered statistically significant.

Table 4.1: Accuracy and AUC from the test fold for different types of preprocessing methods in both networks

	Accuracy		AUC	
	SW-net	**U-net**	**SW-net**	**U-net**
Raw	95.45	97.65	0.9932	0.9958
Normalize	95.49	97.67	0.9933	0.9954
Standardize	95.54	97.64	0.9937	0.9956
CLAHE	95.82	97.63	0.9938	0.9957
CLAHE + Standardize	95.88	97.65	0.9935	0.9953

4.3. RESULTS

4.3.1. EVALUATION ON THE CNN PERFORMANCE

PREPROCESSING METHOD

Our experiments showed two main conclusions: (1) networks fed with raw images took slightly more time to converge, especially for SW-net; (2) either enhancing or standardizing/normalizing the images did not lead to prominent improvements in the performance (Table 4.1).

SW-net provided higher accuracy when using CLAHE but similar AUC, which suggested that enhancing the images helps in the classification of those pixels whose p is closer to 0.5, but no significant changes occur in the proper edge ($p = 1$) and body ($p = 0$) pixels. For U-net, the differences were even smaller. In fact, the case with raw images provided the largest AUC. This suggested that U-net does not need any type of preprocessing to perform at its best.

In conclusion, we selected the type of preprocessing with the largest AUC: raw images for U-net, and CLAHE for SW-net.

OVERFITTING, ELASTIC DEFORMATIONS, AND DROPOUT LAYERS

We observed that overfitting was an important problem in training U-net. We could either tackle the issue by adding dropout layers (our approach), by using more data augmentation (elastic deformations), or both.

While elastic deformations could create an artificially large training set, dropout layers were already optimal, removing any effect of overfitting in U-net and increasing the accuracy (Figure 4.5-B). If elastic deformations were added on top of that, the accuracy decreased from 97.65 to 97.22, which made us discard that approach.

In contrast, SW-net was not affected by overfitting (Figure 4.5-A). In fact, the network diverged and classified all pixels as cell body when dropout layers with a drop rate of 50% were added. Furthermore, we investigated whether substituting the global averaging layer for a fully connected layer had any effect in performance. We observed that overfitting was also not present when using a fully connected layer of 200 neurons (as Cireşan et al.'s network [3]), but performance degraded (Table 4.2).

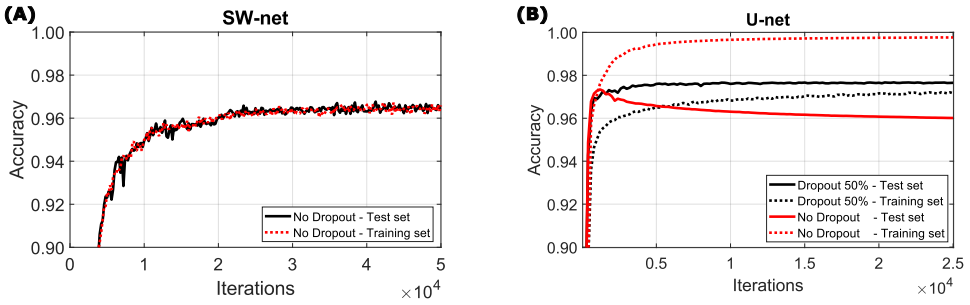

Figure 4.5: Effect of using dropout layers for SW-net (A) and U-net (B), where accuracy is plotted as a function of iterations. Note that SW-net required more iterations in order to stabilize its accuracy.

Figure 4.5 also shows the difference between both networks in terms of stability and convergence. Training U-nets yields much faster convergence and is more stable than training SW-nets. The latter shows a relatively large accuracy fluctuation, probably due to the large variation between patches. However, it is worth noting that, for the SW-net, we only sampled randomly 200 batches (25600 patches) from the test set every 200 training iterations, whereas the whole test set (10 images) was evaluated for the U-net at the same iterations. Using the whole test set for SW-net would entail to evaluate 12 million patches, which was extremely expensive computationally if evaluated so frequently. This was only done once the training was finished. Regarding the results for the training set in Figure 4.5, they indicate the average accuracy in the 200 training batches previous to each test evaluation. Since batches in both networks had similar amount of data, it is possible to conclude that U-net is more stable. Nonetheless, both networks did not show any type of performance degradation as the number of iterations increases.

RECEPTIVE FIELD AND FILTER SIZE

A key discrepancy between the two networks was the difference in receptive field size (Table 4.2). It is believed that a cell only has a direct effect in the shape of its adjacent cells. Indeed, it was observed a long time ago how the endothelial cells elongate and pull their neighboring cells when they need to cover a large space of dying cells [55]. Hence, it was expected that, in order to classify one pixel, only the shape and intensity information of the neighboring cells was required. Given that the average cell diameter is 25-30 pixels, a receptive field of 75–90 pixels would be optimal. Indeed, our experiments suggested that for U-net: the performance degraded when decreasing the receptive field, either by using filters of 3×3 or removing one resolution step, but also when increasing the receptive field, either by using larger filters of 5×5 or adding another resolution step (Table 4.2). Based on the cell size, more than 5 resolutions steps would be counterproductive, as cells would be unrecognizable at the last resolution ($2^5 = 32 >$ average cell size).

It could be argued that a different network composition with different filter sizes, but reaching the desired receptive field, would also be optimal. To evaluate this, we built networks reaching comparable receptive fields: for the 3×3 filters, we added another convolutional layer at each resolution step of the contraction path (receptive field of 93

Table 4.2: Receptive field (in pixels), accuracy, and AUC from the test fold for different types of filter sizes, number of filters, depth of the network (resolution steps), using class weighting or binary labels (for U-net), and patch size (for SW-net). Best performing (default) networks are indicated in bold.

		Receptive field	Accuracy	AUC
SW-net	**Patch 64pix, 32 filters of 3×3**	**61**	**95.82**	**0.9938**
SW-net	Default, Fully Connected Layer	61	94.97	0.9916
SW-net	Patch 96pix, 32 filters of 3×3	61	94.03	0.9888
SW-net	Patch 96pix, 32 filters of 4×4	91	95.39	0.9931
U-net	32 filters of 3×3, 4 steps	61	97.55	0.9949
U-net	32 filters of 3×3, 5 steps	125	97.62	0.9955
U-net	**32 filters of 4×4, 4 steps**	**91**	**97.65**	**0.9958**
U-net	32 filters of 5×5, 4 steps	121	97.46	0.9954
U-net	32 filters of 4×4, 3 steps	43	97.48	0.9951
U-net	32 filters of 4×4, 5 steps	187	96.92	0.9939
U-net	16 filters of 4×4, 4 steps	91	97.32	0.9951
U-net	64 filters of 4×4, 4 steps	91	97.61	0.9956
U-net	Default, weighted class	91	96.65	0.9958
U-net	Default, binary labels	91	93.92	0.9919

pixels); for the 5×5 filters, we removed the last convolutional layer of the contraction path (receptive field of 89 pixels). Still, accuracy and AUC for the network using filters of 4×4 were always slightly higher (data not included). Moreover, visual evaluation indicated that filters of 4×4 (Figure 4.6-C) were somehow better than 3×3 (Figure 4.6-D) or 5×5 (Figure 4.6-E) in segmenting complex areas where the contrast was low. We believe this is due to the transposed convolutional layers and their problems in handling filter sizes not divisible by the stride, as discussed in the Methods section. This hypothesis was reinforced when the same experiment was done using SW-net, where no transposed convolutions were present, obtaining similar noisy results for both filter sizes, 3×3 ((Figure 4.6-G) and 4×4 (Figure 4.6-H).

In comparison with U-net, SW-net generated a “grainy” effect in those complex areas. Moreover, it was observed that increasing the patch size to 96 pixels did not improve the performance (Table 4.2). Thus, the receptive field of SW-net was significantly smaller than that of U-net. This might be linked to the inherent nature of the patch-based approach, where increasing the patch size also increases the variation between patches, which in turn would take higher efforts for the CNN to distinguish patches of different classes.

Finally, we also tested the number of filters in U-net, halving or doubling them, obtaining slightly less accuracy in both cases (Table 4.2). In general, we observed that modifying the depth and width of our U-net did not drastically degraded the performance. Considering that the postprocessing corrects some mistakes and enhances the final segmentation, most probably all these networks would give similar clinical estimates.

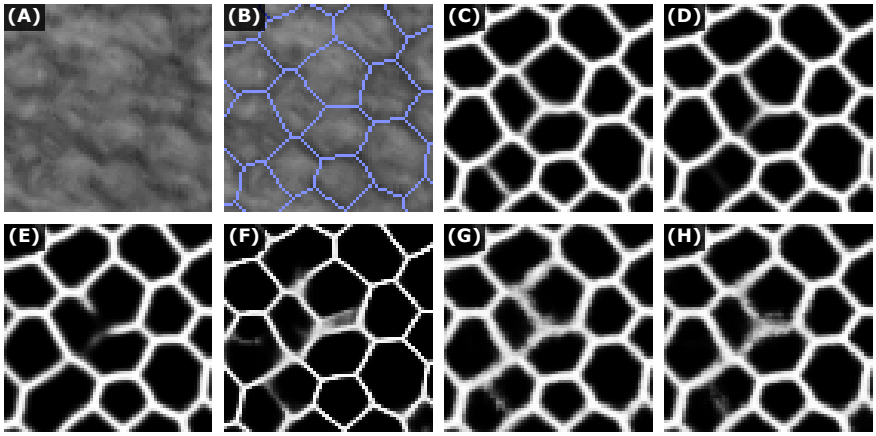

Figure 4.6: (A) Small, blurred area of a specular image (size 68×68 pixels) where the identification of small cells is difficult. (B) The gold standard (in blue) superimposed on the intensity image. (C) U-net output for a filter size of 4×4. (D) U-net output for a filter size of 3×3 with similar receptive field. (E) U-net output for a filter size of 5×5 with similar receptive field. (F) Default U-net output for a filter size of 4×4, but using the original binary labels. (G) SW-net output for a filter size of 4×4. (H) SW-net output for a filter size of 3×3.

WEIGHTED CLASSES AND BINARY LABELS

Two distinctive decisions were taken when designing the network: not weighting the classes for U-net, and using probabilistic labels instead of the binary gold standard images.

Weighting the classes did not change the AUC in U-net, but the accuracy decreased (Table 4.2). The visible effect was slightly thicker edges, which in turn provided higher sensitivity* (0.9954 instead of 0.9940), but lower precision* (0.9907 instead of 0.9938).

On the contrary, the use of binary, weighted labels was clearly a mistake in terms of performance (Table 4.2). Furthermore, it created a "halo" effect in complex areas (Figure 4.6-F), with no clear intensity pattern, which would create many artifacts in the post-processing step.

THE EFFECT OF THE AMOUNT OF TRAINING DATA

Large training sets are important to achieve good results in CNNs. To evaluate this, we defined an experiment where the training set was comprised of the following number of images, $n_{training} = [1, 2, 3, 5, 10, 15, 20, 25, 30, 35, 40, 45]$, while the remaining images were assigned to the test set. AUC was retrieved for each case (Figure 4.7-A). The experiment showed the following: (1) although overfitting was present when less than 25 training images were used, no degradation in the performance of the test set was observed; (2) both networks could perform reasonably well with just one training image; (3) the performance of U-net improved more acutely than that of SW-net as more training images were included. In summary, this experiment suggested than building a larger training dataset might be the best choice to improve the overall performance.

Table 4.3: Accuracy, AUC, precision*, sensitivity*, and specificity* from all images (i.e. using a 5-fold cross-validation) in both networks, SW-net and U-net.

	Accuracy	**AUC**	**Precision***	**Sensitivity***	**Specificity***
SW-net	95.48	0.9921	0.9585	0.9906	0.9914
U-net	97.33	0.9938	0.9855	0.9892	0.9971

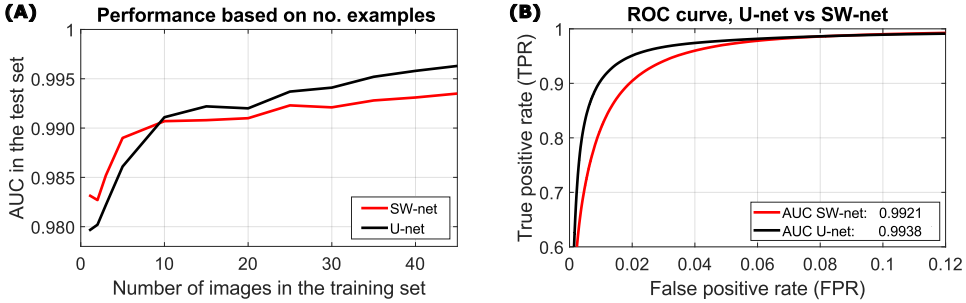

Figure 4.7: (A) Network performance (AUC) based on the number of training examples (without cross-validation). (B) The ROC curves (zoomed) and the corresponding AUC values for both networks for the whole dataset (5-fold cross-validation).

COMPARISON BETWEEN U-NET AND SW-NET

Finally, we tested all images in both networks by employing a 5-fold cross-validation, using their respective best design parameters indicated above. The computed metrics clearly showed a higher performance for U-net (Table 4.3), with a considerably larger accuracy and precision. The ROC (Receiver Operating Characteristic) curves are displayed in Figure 4.7-B.

4.3.2. EVALUATION AFTER APPLYING POSTPROCESSING

The postprocessing method was particularly effective when the cell size in the image was rather regular, as it detected weak edges in the CNN output and "filled" discontinuities in the visual appearance of some edges (Figure 4.9, green arrows). On the contrary, it sometimes reinforced weak, false edges in large cells (Figure 4.9, red arrow) or smoothed away small cells in images with a large variation in cell size (Figure 4.9, blue arrows). Furthermore, it was exceptionally beneficial for SW-net, as it corrected the "grainy" edges. In Figure 4.9, we reported the CNN output and final segmentation for three representative examples, along with the segmentation of the microscope's built-in software. The gold standard images were not included, but instead the errors were indicated with red or blue arrows.

The modified Hausdorff distance (MHD) [53] indicated very low values for both networks (Table 4.4), which is in favor of concluding we achieved a very precise segmentation. To compare both networks, we applied the Wilcoxon signed-rank test since neither of both passed the Shapiro-Wild normality test ($P < 0.0001$), achieving a statistically sig-

Table 4.4: Average MHD (±SD), average DICE (±SD), and percentage of over- and under-segmented cells, in both networks (SW-net and U-net), for $\alpha = 1$.

	MHD	DICE	Overseg. Cells (%)	Underseg. Cells (%)
SW-net	0.30 ± 0.09	0.978 ± 0.006	0.537	0.382
U-net	0.22 ± 0.04	0.981 ± 0.003	0.220	0.260

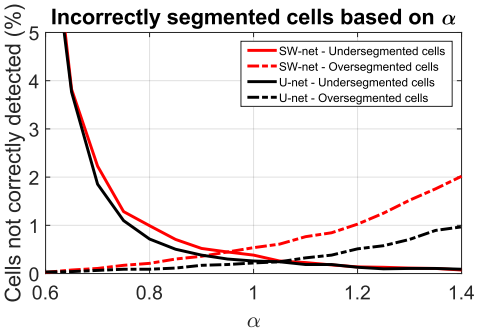

Figure 4.8: Percentage of wrongly detected cells (both, under- and over-segmented cells) for different scaling values (α) applied to f^*, for both networks.

nificant difference in favor of U-net ($P < 0.0001$).

The DICE metric [31] showed higher values for U-net (Table 4.4). Wilcoxon signed-rank test was also applied since the SW-net distribution did not pass the Shapiro-Wild normality test ($P < 0.0001$), achieving a statistically significant better performance for U-net ($P < 0.0001$).

Regarding the number of over- and undersegmented cells, U-net correctly segmented 99.52% of the cells. In contrast, SW-net achieved 99.08% success rate (Table 4.4). The distributions of "percentage of correctly segmented cells" from both assessments failed the Shapiro–Wilk normality test ($P < 0.0001$). The Wilcoxon signed-rank test indicated a statistically significant difference in favor of U-net ($P = 0.0006$).

Furthermore, we evaluated the robustness of the postprocessing method by adding a scaling factor (α) to the estimated characteristic frequency, $\sigma = k_\sigma/(\alpha f^*)$ (see Methods section). Specifically, we evaluated the method for both networks and values of α between 0.60 and 1.40 in steps of 0.05 (Figure 4.8). Overall, both approaches yielded optimal results for values of $\alpha \approx 1$, but the error for SW-net rose much faster as α increased. In comparison with the Topcon output segmentation (Figure 4.9-F), both our methods did significantly better, detecting all the cells in the image (roughly 70% more cells than Topcon).

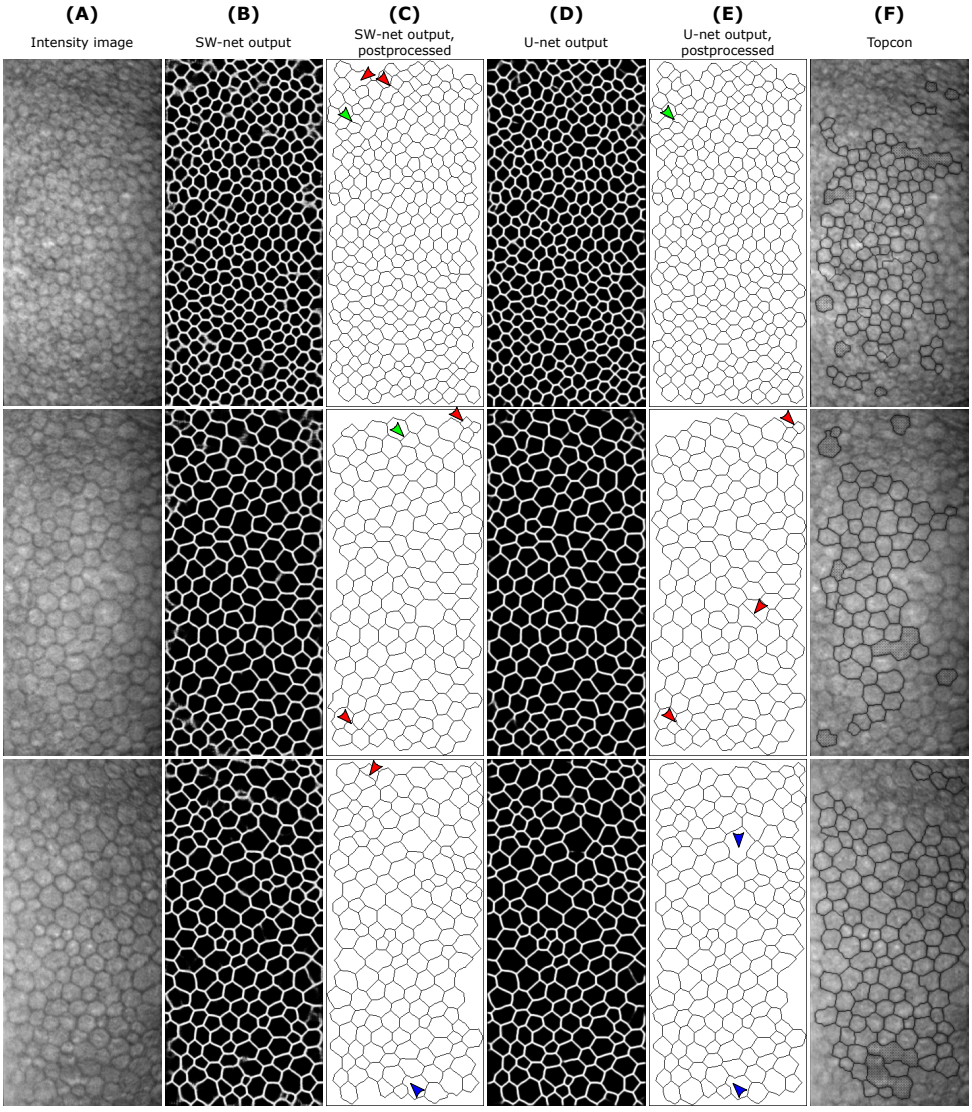

Figure 4.9: Three representative examples (high ECD in top row, low ECD in middle row, high CV in bottom row) for both networks. (A) Intensity images. (BN) Outcome of the SW-net. (C) Segmentation after postprocessing of the SW-net outcome. (D) Outcome of the U-net. (E) Segmentation after postprocessing of the U-net outcome. Green arrows indicate true edges that were weak in the CNN output but detected by the postprocessing. Blue arrows denote true edges that were missed by the postprocessing, either because they were weak edges or because a small cell surrounded by large cells was smoothed away. Red arrows indicate false edges and mistakes in general. (F) Segmentation provided by the Topcon microscope's built-in software.

4.3.3. EVALUATION ON THE CLINICAL PARAMETERS

The clinical parameters for both methods were determined from the final segmentation results and compared to the corresponding values calculated based upon the gold stan-

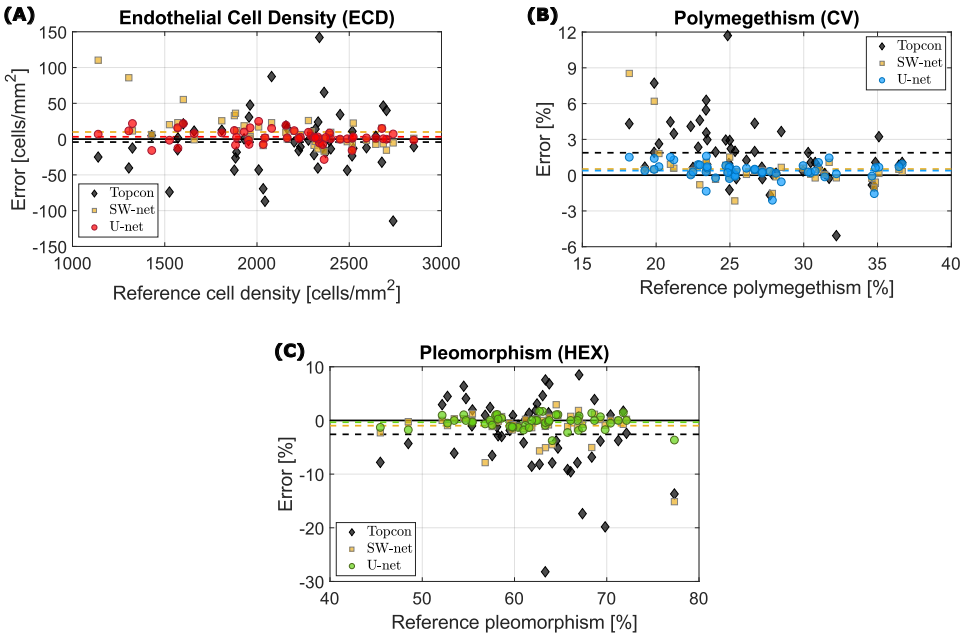

Figure 4.10: Estimates of the clinical parameters in both networks –ECD (A), CV (B), and HEX (C)–. The x-axis indicates the value for the gold standard, and the y-axis indicates the error computed as the difference between the network estimates and the gold standard estimates. Each point corresponds to one image in the dataset (U-net in colored circles, SW-net in brown squares, and Topcon in black diamonds). The mean value of the error for each set is drawn with a dashed line.

dard. The same algorithm for parameter estimation was used in all sets, including Topcon's segmentation images. For all images, only the cells covered by the area of the gold standard were included for the parameter estimation. The only exception was Topcon's segmentation, since the microscope's software did not provide any cell segmentation beyond the segmented area (Figure 4.9). In that set, the gold standard covered twice its segmented area.

The clinical parameters were defined as follows. For cell density,

$$\mathrm{ECD} = \frac{n}{\sum_{i=1}^{n} S_i}, \tag{4.8}$$

where n denotes the number of cells, and S_i the area (in pixels) of the ith cell, defined as $S_i = B_i + E_i/2$, where B is the cell body and E the cell edge. Polymegethism was defined as

$$\mathrm{CV} = 100\% \frac{1}{\bar{S}} \sqrt{\frac{\sum_{i=1}^{n} (S_i - \bar{S})^2}{n}}, \tag{4.9}$$

where $\bar{S}$ stands for the average cell size. Finally, pleomorphism was defined as

$$\mathrm{HEX} = 100\% \frac{n_{hex}}{n}, \tag{4.10}$$

Table 4.5: Mean and standard deviation of the estimation error (mean error [ME], mean absolute error [MAE], and mean absolute percentage error [MAPE]) of the clinical parameters for both networks and Topcon's built-in software. Error and absolute error are computed as the difference (and absolute difference) between estimates and gold standard values. Percentage error is computed as the percentage of the absolute error with respect to the gold standard values.

Type of error	Method	ECD [cells/mm^2]	CV [%]	HEX [%]
ME	**Topcon**	-4.1 ± 41.7	1.9 ± 2.6	-2.2 ± 7.2
	SW-net	9.9 ± 23.1	0.5 ± 1.6	-0.7 ± 2.0
	U-net	3.2 ± 10.1	0.4 ± 0.7	-0.2 ± 1.0
MAE	**Topcon**	29.8 ± 29.6	2.3 ± 2.2	5.3 ± 5.3
	SW-net	14.9 ± 20.2	0.8 ± 1.4	1.6 ± 2.5
	U-net	7.8 ± 7.2	0.6 ± 0.5	0.9 ± 0.8
	Method	**ECD (%)**	**CV (%)**	**HEX (%)**
MAPE	**Topcon**	1.3 ± 1.4	10.1 ± 9.0	8.0 ± 8.9
	SW-net	0.8 ± 1.3	3.6 ± 7.3	2.1 ± 2.9
	U-net	0.4 ± 0.4	2.8 ± 2.6	1.3 ± 1.0

where n_{hex} denotes the number of six-sided cells.

The estimation error was defined as the difference between the estimated value and the gold standard value. The absolute error was defined as the absolute difference. Note that, for polymegethism (Figure 4.10-B) and pleomorphism (Figure 4.10-C), the parameter values were provided as a percentage, and the error was the difference of the percentages. The mean value and standard deviation (SD) of those estimation errors are indicated in Table 4.5. To statistically evaluate the precision, we used the SD of the error, whereas the absolute error was employed to evaluate the accuracy.

The statistical analysis between U-net and Topcon indicated a significantly better precision and accuracy in all parameters for U-net ($P < 0.0001$). For SW-net, the statistical analysis also indicated a significantly better precision ($P < 0.0001$, $P = 0.0002$, and $P < 0.0001$ for ECD, CV and HEX, respectively) and a significantly better accuracy ($P = 0.0054$, $P < 0.0001$, and $P < 0.0001$ for ECD, CV and HEX, respectively) than Topcon for all parameters.

Finally, we compared U-net against SW-net. The statistical analysis denoted a significantly better precision for U-net in all parameters ($P < 0.0001$). The analysis also showed a significantly better accuracy in ECD for U-net ($P = 0.013$) and HEX ($P = 0.048$), but comparable for CV ($P = 0.30$).

One of the main differences between SW-net and U-net was the robustness of U-net against images of different cell density. Indeed, SW-net tends to overestimate ECD as ECD decreases, whereas the ECD error for U-net is rather constant regardless of the cell density (Figure 4.10-A). This problem of SW-net might be explained by the large percentage of images of high ECD in the dataset, which in turn might lead the network to infer that cells are normally of a small size. Interestingly, U-net can overcome this

drawback, probably because U-net can exploit the overlapping features between nearby pixels. Nonetheless, a more inhomogeneous and larger dataset would certainly improve this.

Clinically, it is more important to achieve better precision than accuracy, as the latter could be mitigated by adding a bias to all measures. Moreover, it is desired to obtain more precise, accurate estimates in the images with low ECD, as those are the cases where clinical decisions are more critical. In this sense, U-net is preferred over SW-net.

4.4. DISCUSSION

All the experiments regarding the CNNs architectures clearly indicated a quantitatively better performance in U-net. In contrast, the qualitative results were quite similar for the two networks, with only subtle differences, such as the "grainy" effect on the SW-net output (Figure 4.6). Overall, SW-net did not detect more false edges than U-net (Figure 4.9), but the presence of blurred, faded edges in SW-net was manifest. Interestingly, those subtle differences had a significant effect in the parameters estimation. This highlights the importance of the postprocessing method, which in our case was designed to minimize those problems. A simpler postprocessing approach, such as thresholding and skeletonization, could potentially create many small false cells, sometimes of just a few pixels. This would require to define morphological operations *ad hoc* that would remove them. Given the large variation in cell size between images –or even in the same image (Figure 4.9, bottom image)–, such operations would be prone to mistakes. In this respect, our postprocessing method does not require to define or tune any variable. Indeed, the 1D radial magnitude of the 2D Fourier Transform (FT) of the CNN output shows a clearly distinctive peak (Figure 4.2-B), which makes it easy to estimate the most common cell size in the image and adapt the Gaussian smoothing filter of the postprocessing to that size. The only drawback of this approach occurs when an image shows a large variation in cell size (as in Figure 4.9, bottom image), where very small cells can be smoothed away. As we showed in Figure 4.8, adding a scaling factor to create a thinner smoothing filter does not reduce the overall error in cell detection since oversegmented cells would rapidly increase if α is increased. However, we could tackle this problem by employing a refinement method. In our previous work [28], we performed the segmentation of CE images by employing a merging method that is applied to oversegmented CE images. There, we defined several features based on cell size, shape, and intensity, which were used to identify and remove false edges. Moreover, we showed how the errors mainly originated from wrong edge delineations in the oversegmented images and that the method was robust against a high degree of oversegmentation [48]. For those reasons, both methods could be combined in order to provide an even better performance. In this sense, the aforementioned problem could be simply solved by reducing the filter σ in order to generate a small degree of oversegmentation, and afterwards applying the merging method from our former study [28] (this refinement method was not tested in this paper).

Regardless of this suggestion for refinement, the proposed method achieves a mean absolute percentage error in U-net of 0.4% in ECD, 2.8% in CV, and 1.3% in HEX (Ta-

ble 4.5). When comparing the percentage error of CV and HEX in both networks with the Topcon estimates (Table 4.5), the improvement is outstanding, reducing the error in less than one third. In comparison with Fabijańska's U-net paper [34], our U-net error is more than 4 times smaller. We believe that this large difference is not only due to the result of changes in the U-net architecture, but also due to the use of probabilistic labels in combination with a more sophisticated postprocessing method.

In comparison with other methods from the literature described in the Introduction section, we either achieved the smallest error rate in parameter estimation and/or the smallest error in segmentation accuracy (only a few papers performed a full clinical evaluation). For instance, Scarpa and Ruggeri [22], who developed an algorithm that mimics biological evolution in order to detect the endothelial cells in specular microscopy images, achieved a mean absolute percentage error of 0.6% in ECD, 5.33% in CV, and 3.11% in HEX; Selig et al. [27], who employed stochastic watershed to segment endothelial cells in confocal microscopy images, obtained a mean absolute percentage error of 4.2% in ECD, 22.3% in CV, and 14.4% in HEX; or in our previous work regarding the merging method [28] we achieved an error of 0.8% in ECD, 4.5% in CV, and 3.9% in HEX. While the current work clearly indicates that we have achieved state-of-the-art results, the same dataset should be evaluated in all the previous methods in order to validate that conclusion.

Finally, it is important to highlight that we evaluated a dataset of relatively healthy endothelial cell layers, whose main common factor (besides all being from glaucomatous eyes) was the old age of the subjects. Whereas these cases are the most commonly observed in the clinic, several cornea diseases, such as Fuchs' dystrophy syndrome, bullous keratopathy, or keratoconus, provide heavily blurred, noisy specular images, sometimes with large portions of the image out of focus. Further work would be required to assess the performance of the proposed method in such cases. Moreover, it would be beneficial to develop a method that could automatically select the region of interest in the images from where to estimate the clinical parameters, discarding the excessively blurred or unfocused areas. Currently, this is manually performed by the user.

4.5. CONCLUSIONS

We have presented and evaluated two end-to-end methods for segmenting CE images, a global approach based on U-net and a local approach based on a sliding-window CNN (named SW-net). We have demonstrated excellent results with both approaches, outperforming the current segmentation that the microscope's built-in software provides. Overall, U-net is the preferred approach, as it provides higher accuracy/precision and faster convergence in network training.

Up to now, the inability to provide an accurate segmentation made it difficult to use morphological parameters (CV or HEX) in clinical studies with large amount of data, even though it was observed decades ago that there is a direct link between these biomarkers and certain diseases [56, 57]. Indeed, cell density is currently the only endothelial parameter used in the majority of clinical studies due to the limited accuracy of the current segmentation techniques. Deep learning now opens new opportunities to further

analyze a large number of endothelial images.

REFERENCES

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, *Backpropagation applied to handwritten zip code recognition,* Neural Computation **1**, 541 (1989).

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet classification with deep convolutional neural networks,* in *Advances in Neural Information Processing Systems 25* (2012) pp. 1097–1105.

[3] D. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, *Deep neural networks segment neuronal membranes in electron microscopy images,* in *Advances in Neural Information Processing Systems 25* (2012) pp. 2843–2851.

[4] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: convolutional networks for biomedical image segmentation,* in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. 9351 (2015) pp. 234–241.

[5] E. Shelhamer, J. Long, and T. Darrell, *Fully convolutional networks for semantic segmentation,* IEEE Transactions on Pattern Analysis and Machine Intelligence **39**, 640 (2016).

[6] G. Ayala, M. E. Díaz, and L. Martínez-Costa, *Granulometric moments and corneal endothelium status,* Pattern Recognition **34**, 1219 (2001).

[7] W. M. Bourne, *Biology of the corneal endothelium in health and disease,* Eye **17**, 912 (2003).

[8] P. A. Mohammad-Salih, *Corneal endothelial cell density and morphology in normal Malay eyes,* The Medical Journal of Malaysia **66**, 300 (2011).

[9] M. Hara, N. Morishige, T. Chikama, and T. Nishida, *Comparison of confocal biomicroscopy and noncontact specular microscopy for evaluation of the corneal endothelium,* Cornea **22**, 512 (2003).

[10] J. Huang, J. Maram, T. C. Tepelus, S. R. Sadda, V. Chopra, and O. L. Lee, *Comparison of noncontact specular and confocal microscopy for evaluation of corneal endothelium,* Eye & Contact Lens **44**, S144 (2018).

[11] W. van Schaick, B. T. H. van Dooren, P. G. H. Mulder, and H. J. M. Völker-Dieben, *Validity of endothelial cell analysis methods and recommendations for calibration in Topcon SP-2000P specular microscopy,* Cornea **24**, 538 (2005).

[12] C. Hirneiss, R. G. Schumann, M. Gruterich, U. C. Welge-Luessen, A. Kampik, and A. S. Neubauer, *Endothelial cell density in donor corneas: a comparison of automatic software programs with manual counting,* Cornea **26**, 80 (2007).

[13] M. O. Price, K. M. Fairchild, and F. W. Price, *Comparison of manual and automated endothelial cell density analysis in normal eyes and DSEK eyes,* Cornea **32**, 567 (2013).

[14] N. Luft, N. Hirnschall, S. Schuschitz, P. Draschl, and O. Findl, *Comparison of 4 specular microscopes in healthy eyes and eyes with cornea guttata or corneal grafts,* Cornea **34**, 381 (2015).

[15] R. Nadachi and K. Nunokawa, *Automated corneal endothelial cell analysis,* in *5th Annual IEEE Symposium on Computer-Based Medical Systems* (Durham, NC, USA, 1992) pp. 450–457.

[16] F. J. Sanchez-Marin, *Automatic segmentation of contours of corneal cells,* Computers in Biology and Medicine **29**, 243 (1999).

[17] M. R. Mahzoun, K. Okazaki, H. Mitsumoto, H. Kawai, Y. Sato, S. Tamura, and K. Kani, *Detection and complement of hexagonal borders in corneal endothelial cell image,* Medical Imaging Technology **14**, 56 (1996).

[18] L. Vincent and B. Masters, *Morphological image processing and network analysis of cornea endothelial cell images,* in *Proceedings of SPIE, Image Algebra and Morphological Image Processing III*, Vol. 1769 (San Diego, CA, USA, 1992) pp. 212–226.

[19] Y. Gavet and J. C. Pinoli, *Visual perception based automatic recognition of cell mosaics in human corneal endothelium microscopy images,* Image Analysis & Stereology **27**, 53 (2008).

[20] J. Angulo and S. Matou, *Automatic quantification of in vitro endothelial cell networks using mathematical morphology,* in *5th IASTED International Conference on Visualization, Imaging, and Image Processing* (Benidorm, Spain, 2005) pp. 51–56.

[21] M. Foracchia and A. Ruggeri, *Corneal endothelium cell field analysis by means of interacting bayesian shape models,* in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)* (Lyon, France, 2007) pp. 6035–6038.

[22] F. Scarpa and A. Ruggeri, *Development of a reliable automated algorithm for the morphometric analysis of human corneal endothelium,* Cornea **35**, 1222 (2016).

[23] M. S. Sharif, R. Qahwaji, E. Shahamatnia, R. Alzubaidi, S. Ipson, and A. Brahma, *An efficient intelligent analysis system for confocal corneal endothelium images,* Computer Methods and Programs in Biomedicine **122**, 421 (2015).

[24] K. Habrat, M. Habrat, J. Gronkowska-Serafin, and A. Piórkowski, *Cell detection in corneal endothelial images using directional filters,* Advances in Intelligent Systems and Computing **389**, 113 (2016).

[25] A. Piórkowski, K. Nurzynska, J. Gronkowska-Serafin, B. Selig, C. Boldak, and D. Reska, *Influence of applied corneal endothelium image segmentation techniques on the clinical parameters,* Computerized Medical Imaging and Graphics **55**, 13 (2017).

[26] S. Al-Fahdawi, R. Qahwaji, A. S. Al-Waisy, S. Ipson, M. Ferdousi, R. A. Malik, and A. Brahma, *A fully automated cell segmentation and morphometric parameter system for quantifying corneal endothelial cell morphology,* Computer Methods and Programs in Biomedicine **160**, 11 (2018).

[27] B. Selig, K. A. Vermeer, B. Rieger, T. Hillenaar, and C. L. Luengo Hendriks, *Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy,* BMC Medical Imaging **15:13** (2015).

[28] J. P. Vigueras-Guillén, E. R. Andrinopoulou, A. Engel, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Corneal endothelial cell segmentation by classier-based merging of oversegmented images,* IEEE Transactions on Medical Imaging **37**, 2278 (2018).

[29] A. Fabijańska, *Corneal endothelium image segmentation using feedforward neural network,* in *Proceedings of the 2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Vol. 11 (Prague, Czech Republic, 2017) pp. 629–637.

[30] A. Ruggeri, F. Scarpa, M. De Luca, C. Meltendorf, and J. Schroeter, *A system for the automatic estimation of morphometric parameters of corneal endothelium in alizarine red-stained images,* British Journal of Ophthalmology **94**, 643 (2010).

[31] L. R. Dice, *Measures of the amount of ecologic association between species,* Ecology **26**, 297 (1945).

[32] K. Nurzynska, *Deep learning as a tool for automatic segmentation of corneal endothelium images,* Symmetry **10**, 60 (2018).

[33] S. Katafuchi and M. Yoshimura, *Convolution neural network for contour extraction of corneal endothelial cells,* in *Proceedings of SPIE, Thirteenth International Conference on Quality Control by Artificial Vision*, Vol. 10338 (Tokyo, Japan, 2017).

[34] A. Fabijańska, *Segmentation of corneal endothelium images using a U-net-based convolutional neural network,* Artificial Intelligence in Medicine **88**, 1 (2018).

[35] Y. Deng, Z. Ren, Y. Kong, F. Bao, and Q. Dai, *A hierarchical fused fuzzy deep neural network for data classification,* IEEE Transactions on Fuzzy Systems **25**, 1006 (2017).

[36] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. H. Romeny, J. B. Zimmerman, and K. Zuiderveld, *Adaptive histogram equalization and its variations,* Computer Vision, Graphics, and Image Processing **39**, 355 (1987).

[37] S. J. Gedde, L. W. Herndon, J. D. Brandt, D. L. Budenz, W. J. Feuer, and J. C. Schiffman, *Postoperative complications in the Tube Versus Trabeculectomy (TVT) study during five years of follow-up,* American Journal of Ophthalmology **153**, 804 (2012).

[38] N. Nassiri, N. Nassiri, M. Majdi-N, M. Salehi, N. Panahi, A. R. Djalilian, and G. A. Peyman, *Corneal endothelial cell changes after Ahmed valve and Molteno glaucoma implants,* Ophthalmic surgery, lasers & imaging **92**, 394 (2011).

[39] E. K. Lee, Y. J. Yun, J. E. Lee, J. H. Yim, and C. S. Kim, *Changes in corneal endothelial cells after Ahmed glaucoma valve implantation: 2-year follow-up,* American Journal of Ophthalmology **148**, 361 (2009).

[40] D. P. Kingma and J. Ba, *Adam: a method for stochastic optimization,* in *3rd International Conference for Learning Representations* (San Diego, CA, USA, 2015).

[41] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *Improving neural networks by preventing co-adaptation of feature detectors,* CoRR (2012).

[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *Dropout: a simple way to prevent neural networks from overfitting,* Journal of Machine Learning Research **15**, 1929 (2014).

[43] V. Dumoulin and F. Visin, *A guide to convolution arithmetic for deep learning,* CoRR (2016).

[44] A. Odena, V. Dumoulin, and C. Olah, *Deconvolution and checkerboard artifacts,* Distill (2016), 10.23915/distill.00003.

[45] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, *Adversarially learned inference,* in *International Conference on Learning Representations (ICLR)* (Toulon, France, 2017).

[46] M. Lin, Q. Chen, and S. Yan, *Network in network,* CoRR (2013).

[47] S. Beucher and F. Meyer, *Mathematical morphology in image processing,* (Taylor & Francis Group, 1993) Chap. The morphological approach to segmentation: the watershed transformation, pp. 433–481.

[48] J. P. Vigueras-Guillén, A. Engel, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Improved accuracy and robustness of a corneal endothelial cell segmentation method based on merging superpixels,* in *15th International Conference Image Analysis and Recognition (ICIAR)*, Lecture Notes in Computer Science, Vol. 10882 (Póvoa de Varzim, Portugal, 2018) pp. 631–638.

[49] K. M. Adal, P. G. van Etten, J. P. Martinez, K. Rouwen, K. A. Vermeer, and L. J. van Vliet, *Detection of retinal changes from illumination normalized fundus images using convolutional neural networks,* in *Proceedings of SPIE, Medical Imaging 2017: Computer-Aided Diagnosis*, Vol. 10134 (Orlando, FL, United States, 2017).

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

[51] S. Ioffe and C. Szegedy, *Batch normalization: accelerating deep network training by reducing internal covariate shift,* in *Proceedings of the 32nd International Conference on International Conference on Machine Learning (ICML)*, Vol. 37 (Lille, France, 2015) pp. 448–456.

[52] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller, *Striving for simplicity: the all convolutional net,* in *International Conference on Learning Representations (ICLR) Workshop* (San Diego, CA, USA, 2015).

[53] M. P. Dubuisson and A. K. Jain, *A modified Hausdorff distance for object matching,* in *Proceedings of 12th International Conference on Pattern Recognition*, Vol. 1 (Jerusalem, Israel, 1994) pp. 566–568.

[54] G. Verbeke and G. Molenberghs, *Linear mixed models for longitudinal data*, Springer series in statistics (Springer, 2000) pp. xxii, 568 p.

[55] H. Honda, Y. Ogita, S. Higuchi, and K. Kani, *Cell movements in a living mammalian tissue: long-term observation of individual cells in wounded corneal endothelia of cats,* Journal of morphology **174**, 25 (1982).

[56] W. M. Bourne, L. R. Nelson, and D. O. Hodge, *Central corneal endothelial cell changes over a ten-year period,* Investigative Ophthalmology & Visual Science **38**, 779 (1997).

[57] H. S. Leem, K. J. Lee, and K. C. Shin, *Central corneal thickness and corneal endothelial cell changes caused by contact lens use in diabetic patients,* Yonsei Medical Journal **52**, 322 (2011).

5

AUTOMATIC DETECTION OF THE REGION OF INTEREST IN CORNEAL ENDOTHELIUM IMAGES USING DENSE CONVOLUTIONAL NEURAL NETWORKS

This chapter is based on the manuscript:

J.P. Vigueras-Guillén, H.G. Lemij, J. van Rooij, K.A. Vermeer, and L.J. van Vliet, *Automatic detection of the region of interest in corneal endothelium images using dense convolutional neural networks*, Proceedings of SPIE, Medical Imaging 2019: Image Processing **10949**, 1094931, San Diego, CA, USA (2019).

5

Abstract

In images of the corneal endothelium (CE) acquired by specular microscopy, endothelial cells are commonly only visible in a part of the image due to varying contrast, mainly caused by challenging imaging conditions as a result of a strongly curved endothelium. In order to estimate the morphometric parameters of the corneal endothelium, the analyses need to be restricted to trustworthy regions –the region of interest (ROI)– where individual cells are discernible. We developed an automatic method to find the ROI by Dense U-nets, a densely connected network of convolutional layers. We tested the method on a heterogeneous dataset of 140 images, which contains a large number of blurred, noisy, and/or out of focus images, where the selection of the ROI for automatic biomarker extraction is vital. By using edge images as input, which can be estimated after retraining the same network, Dense U-net detected the trustworthy areas with an accuracy of 98.94% and an area under the ROC curve (AUC) of 0.998, without being affected by the class imbalance (9:1 in our dataset). After applying the estimated ROI to the edge images, the mean absolute percentage error in the estimated endothelial parameters was 0.80% for ECD (cell density), 3.60% for CV (polymegethism), and 2.55% for HEX (pleomorphism).

5.1. INTRODUCTION

The endothelium is a monolayer of quasi-hexagonal cells on the posterior surface of the cornea, which plays a key role in maintaining an optimal state of corneal hydration. The endothelium can be imaged *in vivo* by using specular microscopy, a fast and non-invasive method that unfortunately provides low-contrast, noisy images. By performing cell segmentation in these images, several clinical parameters can be estimated – endothelial cell density (ECD), coefficient of variation in cell size (CV), and hexagonality (HEX)–, which are used in the clinic to assess the health status of the cornea. To provide high-quality images, specular microscopy requires that the corneal endothelium surface is rather smooth and flat over the entire field of view, which is uncommon in diseased corneas. In those cases, large fractions of the recorded images are out of focus or blurred, rendering cells in these areas indistinguishable (Figure 5.1). In order to reliably estimate the endothelial biomarkers, the selection of the region of interest (ROI) in which cells are visible is essential. Currently, the user manually performs this task. Here, we aim to automatically determine the ROI in endothelium images, developing and evaluating a dense, fully convolutional neural network (CNN), Dense U-net. In order to select the ROI, Dense U-net uses as input the edge probability images from our previous work [1], which were obtained with a U-net whose input is the endothelium intensity images. Moreover, we explore whether the same Dense U-net can be used to estimate the edge probability images. Finally, we evaluate the whole end-to-end framework to estimate the biomarkers, which includes the CNNs to infer the edges (CNN-Edge) and the ROI (CNN-ROI) and a postprocessing method to combine both. A refinement method that merges oversegmented cells (also from previous work [2]) is added to further improve the biomarker accuracy.

The idea of making a densely connected network was originally presented by Huang et al. [3]. Two years before, Ronnenberg et al. [4] proposed the concept of U-nets. A U-net is a fully convolutional neural network that is particularly suitable to learn image-to-image mapping such as image segmentation. Incorporating a densely connected network into a U-net was a natural progression, which turned out to be very effective in solving a variety of image processing and image segmentation tasks [5–8]. One main issue in these dense networks was the bottlenecks produced by piling up layers, which Huang solved by adding a 1×1 convolutional layer before each 3×3 convolution to reduce the number of input layers. In our case, we modified the network, removing those 1×1 convolutional layers, which did not incur in a high computational cost. Overall, many details were modified with respect to all previous proposals in order to adapt it to our problem.

5.2. PREVIOUS WORK AND MATERIALS

Previously, we designed a CNN based on U-net to segment endothelial cells in specular microscopy images [1] using manually annotated edge images as target and the intensity images as input (Figure 5.2, CNN-Edge block). This provided the edge probability images as depicted in Figure 5.1-(middle row). That work [1] also described a postprocessing method based on Fourier analysis and watershed to transform the edge probability

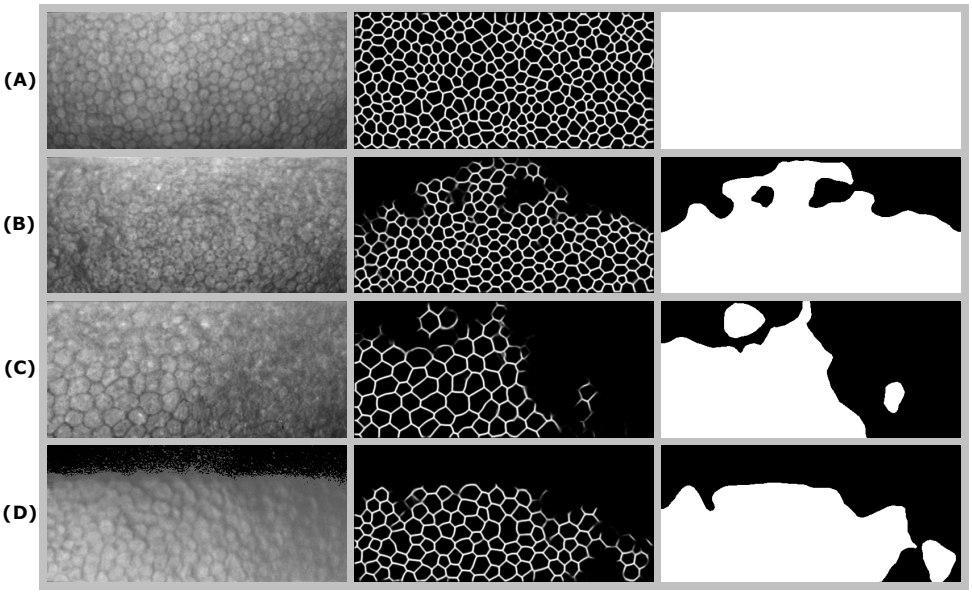

Figure 5.1: Four representative images obtained with specular microscopy: a good quality image (A), images from a curved (non-flat) endothelium (B–C), and an image with poor focus and affected by artifacts (D). (Left column) Intensity images. (Middle column) Edge probability images as obtained by the CNN-Edge [1]. (Right column) The annotated ROI (in white) where cells are recognizable.

images into binary edge images (Figure 5.2, postprocessing block), which were used to estimate the endothelial parameters. We achieved state-of-the-art results: AUC = 0.9938 (area under the ROC curve), DICE = 0.981, and a mean absolute percentage error (MAPE) of 0.4% in ECD, 2.8% in CV, and 1.3% in HEX. The clinical parameters were defined as follows. For cell density,

$$\mathrm{ECD} = \frac{n}{\sum_{i=1}^{n} S_i}, \tag{5.1}$$

where n denotes the number of cells, and S_i the area of the ith cell, defined as $S_i = B_i + E_i/2$, where B is the cell body and E the cell edge. Cell size variation was defined as

$$\mathrm{CV} = 100\% \frac{1}{\bar{S}} \sqrt{\frac{\sum_{i=1}^{n} (S_i - \bar{S})^2}{n}}, \tag{5.2}$$

where $\bar{S}$ stands for the average cell size. Finally, hexagonality was defined as

$$\mathrm{HEX} = 100\% \frac{n_{hex}}{n}, \tag{5.3}$$

where n_{hex} denotes the number of six-sided cells. MAPE was defined as

$$\mathrm{MAPE} = \frac{100\%}{m} \sum_{i=1}^{m} \left| \frac{A_i - E_i}{A_i} \right|, \tag{5.4}$$

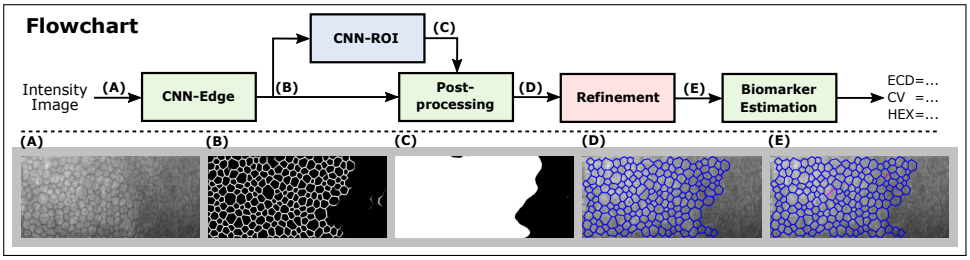

Figure 5.2: Flowchart of the entire biomarker estimation process. The different images at each step (A–E) are displayed below the dashed line. For (D) and (E), the results are binary images, depicted in blue (superimposed to the intensity images for display purposes). The green boxes (CNN-Edge, postprocessing, and biomarker estimation) were already presented in a previous paper [1]. In this work, we add the selection of the ROI (CNN-ROI) to make the process fully automatic. Before estimating the biomarkers, we add a refinement process [2] to detect and remove potential false edges (magenta lines in E).

where A_i is the gold standard value of the parameter, E_i the estimated value, and m is the number of images. That dataset [1] contained 50 relatively healthy endothelial images (as in Figure 5.1-A), with a high variation in cell size, but where cells were recognizable across the entire image. Hence, no ROI selection was necessary. In the current work, we extended the dataset to 140 images by adding challenging images encountered in standard clinical practice: heavily blurred, extensively affected by noise, and with large fractions out of focus (Figure 5.1-B–D). This required a new method to select the ROI before estimating the morphometric parameters (Figure 5.2, CNN-ROI block). By using our previous method [1], we retrieved the edge probability images of all the intensity images (Figure 5.1, middle row). To obtain the ROI, we updated our U-net [1] into a Dense U-net because our experiments showed that training the U-net on our heterogeneous dataset caused convergence problems (see Results section). An expert annotated the ROI from the obtained edge probability images (Figure 5.1, bottom row). This was used as the target output of the network to be trained. The same expert initially created the gold standard of the edge images (cell segmentation) based on the intensity images.

In earlier work [2], we designed a machine learning approach to remove false edges in oversegmented corneal endothelium images, called *merging method*. In order to create the oversegmented images, we employed a method based on stochastic watersheds [9]. While the use of CNNs to solve this problem has shown superior performance [1], the merging method can still be used to remove potential false edges from the binary edge map derived from the CNN output, thereby further improving the biomarker estimation (Figure 5.2, refinement block). Here, we also evaluate the benefits of using such refinement, whose robustness was previously evaluated [10].

All images used in this work are part of an ongoing study in the Rotterdam Eye Hospital regarding the implantation of a Baerveldt glaucoma drainage device. Images were retrieved before the surgical implantation and 3, 6, 12, and 24 months after surgery, in both the central and the temporal superior (TS) cornea. Due to the peripheral location and the curvature of the cornea, TS images are prone to have large fractions of the field-of-view out of focus. Images were obtained with a noncontact specular microscopy, Top-

con SP-1P (Topcon Co., Tokyo, Japan). Each image covers an area of 0.25 mm × 0.55 mm and was saved as 8-bits grayscale images of 240×528 pixels.

5.3. METHODS

The proposed Dense U-net (Figure 5.3) has a similar architecture as a typical U-net. It contains one contraction (downsampling) path and one expansion (upsampling) path, each composed of four resolution steps, and with short-connections between layers in both paths of the network that have the same image resolution. However, each resolution step is comprised of dense blocks, where a layer receives the feature maps of all preceding layers. These connections improve the direct information flow to all preceding layers, helping back-propagating gradients and thereby facilitating better network learning during training. The output of each layer in a dense block has k feature maps, where k is the growth rate parameter. Initially, we set $k = 24$, doubling it at each downsampling step, and halving it at each upsampling step. After each dense block, the feature maps are reduced to k by *Conv2D*(1×1) or *TransposeConv2D*. Furthermore, we introduce the edge probability image, properly scaled, at the beginning of each resolution step. Since the short-connections between paths are made before the *Conv2D*(1×1), the edge probability image is also present in the upsampling path. For data augmentation, only horizontal and vertical flipping were used. A cross-entropy loss function with a pixel-wise softmax activation was employed over the final feature maps. Specifically, the loss function was defined as follows:

$$C(l,s) = -w_{l_i} \sum_{i=0}^{B} l_i \log(s_i) + (1-l_i)\log(1-s_i), \tag{5.5}$$

where s is the assigned pixel score, l the reference pixel label, B the total number of samples in one batch, and w_{l_i} the pixel weight. The latter was computed based on the ratio of elements per class. In our case, the positive class (ROI) is 9 times larger than the negative class (non-ROI). Thus, we set $w_{l_i=0} = 9$ and $w_{l_i=1} = 1$ to balance the classes. The optimizer of our choice was Adam, with an initial learning rate (lr) of 0.001 and a decay of 0.001, such that $lr_{i+1} = lr_i \cdot (1\,/\,(1 + decay \cdot iteration))$. The batch size was four images. For more details, check Figure 5.3.

For comparison purposes, our original U-net [1] had a similar overall architecture. It lacked the dense connections and the *Conv2D*(1×1)+*ReLU* in the downsampling path. It only employed a single dropout layer with a 50% drop rate sandwiched between convolutional layers and only introduced the input image in the first resolution step, which started with 32 feature channels.

5.4. RESULTS

5.4.1. EXPERIMENTAL RESULTS FOR CNN-EDGE

Before addressing the CNN-ROI, we evaluated whether using a Dense U-net to obtain the edge probability images was any different from using our previous U-net [1]. First, we observed that weighting the classes was detrimental, preventing Dense U-net to con-

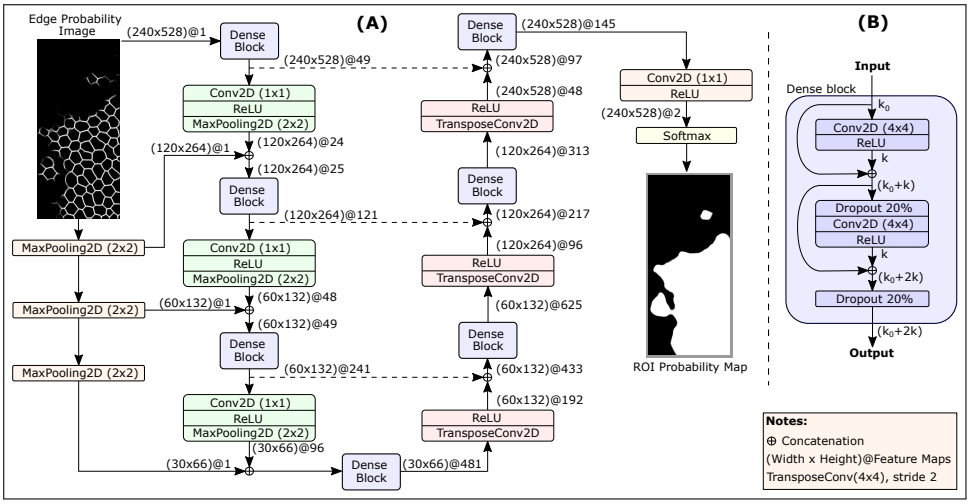

Figure 5.3: (A) A schematic overview of the proposed Dense U-net. The dashed arrow-lines indicate the short-connections; and (B) a schematic overview of the design of one dense block.

5

verge during training. For U-net, it provided a solution where the edges were excessively enhanced in blurred areas, delivering inaccurate results (Figure 5.4-B). On the contrary, the nonweighted solution was satisfactory, providing similar quantitative results for both networks: an accuracy of 96.98% and 96.64% for U-net and Dense U-net, respectively. Qualitatively, the difference was manifest: the Dense U-net provided a more conservative solution, avoiding any inference in heavily blurred areas (Figure 5.4-D), whereas U-net seemed to infer correctly many edges in those areas with only minor errors (Figure 5.4-C). Since the estimation of endothelial parameters can be largely affected by just a few segmentation errors, it could not be concluded at this point which solution would be preferred. Thus, the edge images provided by both networks were further tested in the following experiments.

We also observed that increasing the dropout rate in the Dense U-net made the network behave even more conservative without degrading the segmentation in areas with high contrast between cell edges and cell bodies (Figure 5.4-E). While not useful for us, this behavior could be interesting when applied to other problems. On the contrary, not using any dropout would provide a predominant binary solution, with no gradual inference in heavily blurred areas (Figure 5.4-F), which can be a source of problems for our postprocessing method. Quantitatively, the Dense U-net provided the largest accuracy with a drop rate of 20%, being this our final choice.

5.4.2. EXPERIMENTAL RESULTS FOR CNN-ROI

The dataset was divided in five folds, using four folds for training and one for testing. Images were randomly placed in the folds, having on average the same number of "easy" images (ROI covers the entire or almost the entire image) and "complicated" images (ROI covers less than ~80% of the image) in each fold. In a 5-fold cross-validation, all folds

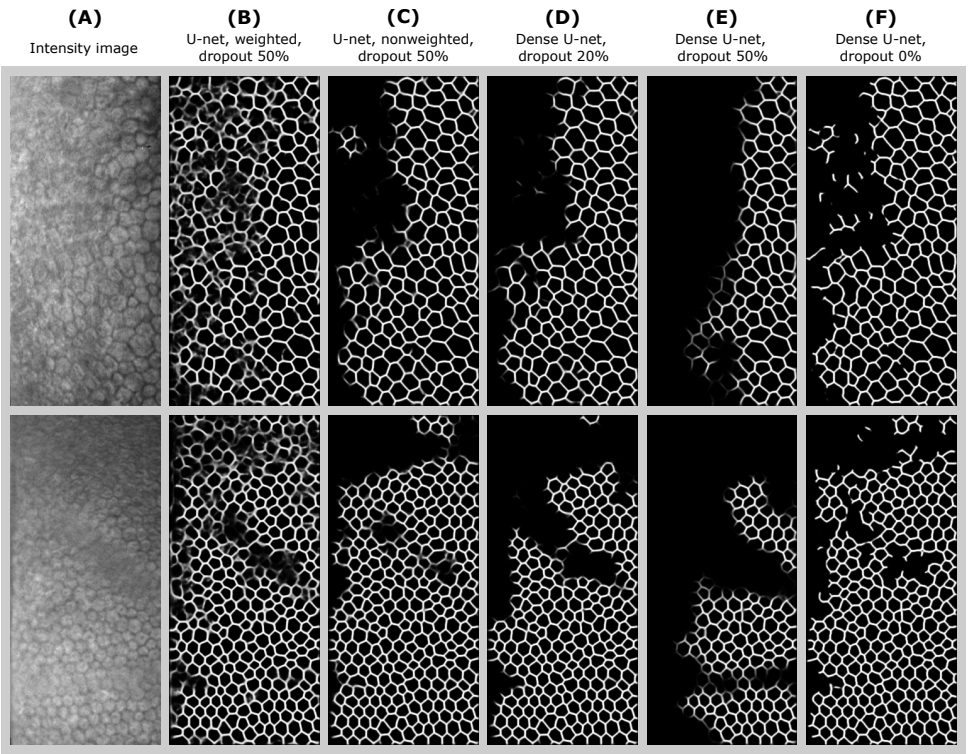

Figure 5.4: Representative output images for five CNN-Edge configurations based on either U-net (B–C) or Dense U-net (D–F). For all networks, the intensity image (A) is the input of the network.

were tested for all network configurations.

In a first experiment, we compared the performance of Dense U-nets versus U-nets [1], and we studied different configurations for both architectures: (1) the effect of weighting the loss function; and (2) the choice of input: the edge probability image, the intensity image, or both (two channels). In a second experiment, we evaluated the benefit of using the refinement method [2] once the ROI was applied to the resulting edge images. In a third experiment, the endothelial parameters were retrieved from these three datasets: (1) the gold standard images; (2) the edge images as obtained with a U-net (CNN-Edge) after applying the estimated ROI; and (3) the edge images as obtained with a Dense U-net (CNN-Edge) after applying the estimated ROI. Topcon estimates were also collected. We considered that a cell was inside the ROI if more than 75% of its area overlapped with the ROI.

EXPERIMENT 1

By default, the input to the CNN-ROI was the edge probability image (Figure 5.5-B) and the target was the annotated ROI, which also served as gold standard (Figure 5.5-C). Initially, no differences were perceived between the weighted and nonweighted cases for

both, U-net and Dense U-net. However, for some folds, we observed that configurations with a weighted loss function had problems to converge to an optimal solution (Figure 5.5-D, H). Clearly, the networks with a weighted loss function inferred many false non-ROI areas (black) within the ROI, sometimes resulting into an utterly wrong solution (Figure 5.5-H, top), and the transition between areas was blurry (Figure 5.5-D, bottom) instead of the expected sharpness of the gold standard. In contrast, this problem was not observed for the nonweighted cases in any of the folds. This made us to discard the use of a weighted loss function.

Quantitatively, the nonweighted U-net provided similar accuracy as the nonweighted Dense U-net (Figure 5.6-A). Qualitatively, U-net showed a peculiar behavior in a few folds (Figure 5.5-E, bottom) without showing a clear hint of convergence problems in the metrics (Figure 5.6). Specifically, the non-ROI class (black, value 0 in a scale 0–1) was classified to a value of 0.49 (gray). To understand this, the following needs to be clarified:

- To create the ROI gold standard, we used the edge probability image to determine which areas should be included. Selecting the ROI was a rather subjective decision, and we did not follow any specific rule. Instead, we roughly delineating the areas.
- We avoided very small blurred areas in the edge images to appear as 'black holes' inside the ROI. Only when those inner blurred areas were large enough (larger than an average cell), they were annotated as non-ROI (thereby punching a hole in the ROI). The purpose was to allow the postprocessing method to correctly identify the cell edges in the small blurred areas. Otherwise, the biomarker estimation, especially HEX, would have been highly affected if many inner cells would have been discarded unnecessarily. In contrast, if blurred areas occurred along the border of the ROI, they were set as non-ROI.

In summary, our ROI gold standard was flawed by design. Nonetheless, we expected the network to provide a solution where the inferred areas would roughly agree with the gold standard and, at the same time, no false black spots (false negatives) would be detected within the ROI. We observed that U-net replicated with higher precision the shape of the gold standard but at the expense of not correctly classifying the minority class in some images (Figure 5.5-E, bottom). Having a heavily unbalanced dataset (9:1) had a major impact on this behavior. In contrast, Dense U-net provided a more generalized solution. Since ROI images were eventually binarized by thresholding at 0.5, U-net could still be used with no inconvenience. Nonetheless, this experiment showed that, for this specific problem, a Dense U-net was more robust and reliable than a U-net.

Finally, we evaluated whether introducing the intensity image to the network could benefit the ROI segmentation. Clearly, the intensity image on its own produced unreliable segmentations (Figure 5.5-F). In contrast, adding the intensity image to the edge probability image as a second channel neither improved nor degraded the performance (Figure 5.6), although we noted a few more holes inside the ROI (Figure 5.5-G).

In conclusion, the nonweighted Dense U-net with only the probability edge image as input was our choice. The average metrics among the five folds were: accuracy = 98.94,

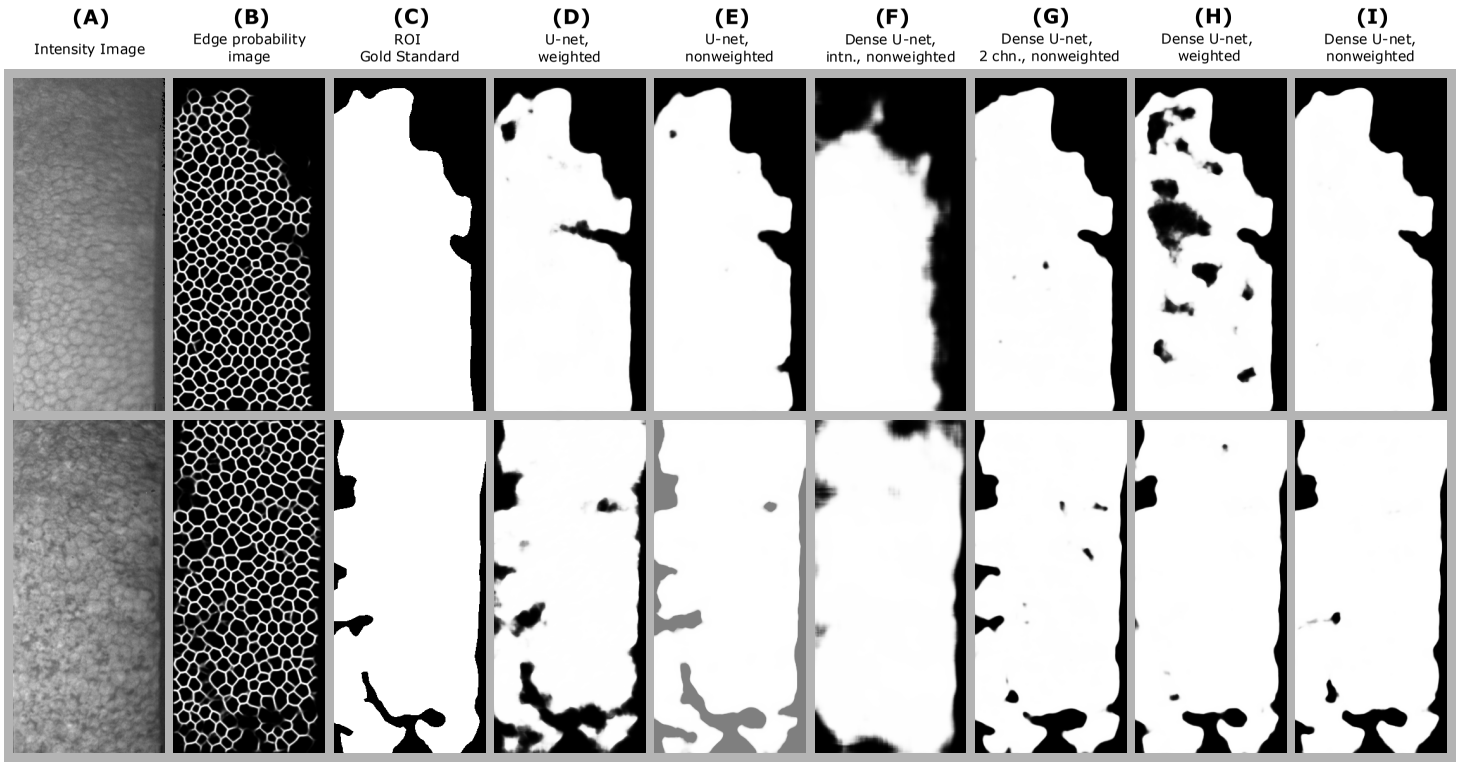

Figure 5.5: Representative output images for six CNN-ROI configurations based on either U-net (D–E) or Dense U-net (F–I). For all networks, the edge probability image (B) is the input of the network, except for (F) –input is the intensity image (A)– and (G) –input with 2 channels, intensity and edge images–. The ROI gold standard (C) is the target of the network. Top image comes from fold 3 and bottom image from fold 4.

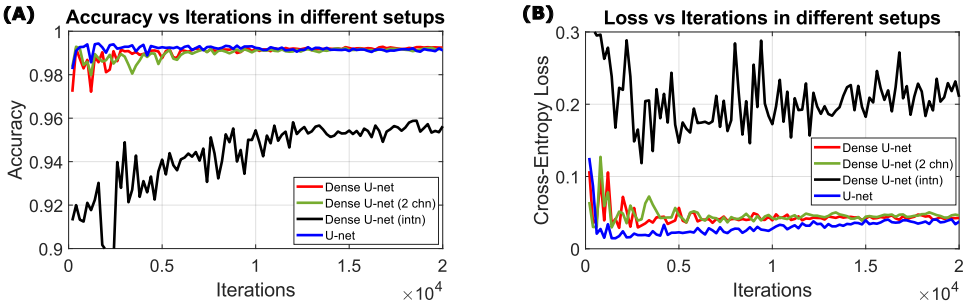

Figure 5.6: (A) Accuracy and (B) cross-entropy loss in the test set (fold 4) for different setups, all employing a nonweighted loss function.

AUC = 0.998, precision = 0.994, sensitivity = 0.989, specificity = 0.980.

5

EXPERIMENT 2

Once we applied the estimated ROI to the edge images, the postprocessing method [1] was performed to produce the final binary edge images and the biomarkers were estimated. The postprocessing involved estimating the average cell size ($l = 1/f^*$) in the edge probability images by using Fourier Analysis (f^* is called the characteristic frequency), applying a Gaussian smoothing filter to the edge probability images whose standard deviation $\sigma = k_\sigma/(\alpha f^*)$, and finally applying the classic watershed [11], which does not require any parameter. The scaling factor of the sigma, $k_\sigma = 0.20$, was estimated in previous work [10], which assumed $\alpha = 1$.

At this point, we considered that including our refinement method [2] based on merging superpixels could further improve the segmentation. This merging method is based on Support Vector Machines (SVM), and it evaluates every possible combination of two superpixels to infer whether they form a complete cell. Several features, based on shape, intensity, and size, were used. Thus, this refinement method could only remove edges but not create new ones.

In order to evaluate this addition, we computed two values from the output binary edge images, the total number of cells detected (n_{total}) and the number of cells correctly segmented (n_{corr}), and one value from the gold standard, the number of real cells (n_{real}). To determine whether a cell was correctly segmented, we used the following rule: given the area of a cell in the segmentation A_S, the area of a cell in the gold standard A_G, and the intersection of those two areas $A_I = A_S \cap A_G$, the cell in the segmentation is correctly segmented if $A_I > 0.75 \times max(A_S, A_G)$. The 75% margin was added to allow small deviations in the cell boundaries. Precision, $p = n_{corr}/n_{total}$, and recall, $r = n_{corr}/n_{real}$, were computed and combined into the F-measure, $F = 2pr/(p+r)$. Roughly, precision decreases more acutely with oversegmentation and recall is more affected by undersegmentation. Finally, several SVMs were trained with different ratios of cost for False Positives and False Negatives (FP:FN). For this evaluation, we used the edge images provided by the U-net.

This experiment proved that the merging method barely removes true edges, i.e. the recall for the different refined cases is hardly lower than that of the non-refined case for $\alpha < 1$ (Figure 5.7-C). In contrast, it removes a few false edges, i.e. precision always increases (Figure 5.7-B). It also showed that it is not necessary to build over-conservative SVMs with higher cost for FPs in order to avoid false mergers (FP:FN of 1:1 is the optimal). Furthermore, it clearly showed that, as we increase α (smaller smoothing filter), oversegmentation appears (precision decreases acutely), but the merging method can amend it to a certain extent. One key point is that the highest precision occurs at $\alpha = 0.8$, whereas in our previous work [1] there was no need to scale the f^*, i.e. $\alpha = 1$ was the optimal. This is related to the larger presence of blurry, false edges (made by U-net), which can be smoothed away if the Gaussian filter is larger (smaller α). For comparison purposes, the largest F-measure occurred at $\alpha = 1$ if a Dense U-net was used for CNN-Edge. The implications of selecting α will be further addressed in the Discussion section.

We performed a similar experiment to evaluate the effect of the refinement in the estimation of biomarkers. Edge images created by either U-net or Dense U-net were tested. Both provided comparable performance but for different values of α (Figure 5.7, right column). Therefore, the fact that U-net (as CNN-Edge) inferred cells in larger areas of the image did not directly cause an improvement in the estimation of biomarkers. Since no significant differences were further observed, the remaining experiments were with a Dense U-net for both the CNN-Edge and CNN-ROI networks (reasons addressed in the Discussion section).

EXPERIMENT 3

The MAPE of the biomarkers estimates (Dense U-net, postprocessing with $\alpha = 1$, refinement with FP:FN of 1:1) was 0.80% for ECD, 3.60% for CV, and 2.55% for HEX (Figure 5.8). Without refinement, it was 0.96% for ECD, 3.75% for CV, and 2.76% for HEX. While being a modest improvement, the benefit of adding the refinement process was unequivocal. Since there are on average 200 cells per image, the effect of removing 2–3 false edges only had a small impact.

In contrast, Topcon provided a MAPE of 6.43% for ECD, 14.23% for CV, and 23.78% for HEX (Figure 5.8). Topcon failed to estimate HEX in 13% of the images since it was required to detect 6 inner cells to provide an estimation (we gave a relative error of 100% in those cases). Inner cells were defined as the segmented cells that are completely surrounded by other segmented cells.

None of the distributions of ECD, CV, and HEX for the different methods passed the Shapiro-Wilk normality test ($P < 0.01$, all cases), and thus the non-parametric Mann-Whitney U-test was performed. It revealed no statistically significant difference in ECD ($P = 0.890$), CV ($P = 0.585$), and HEX ($P = 0.832$) between our estimates and the gold standard, whereas it only showed no significant difference for ECD ($P = 0.202$) between Topcon and gold standard ($P = 0.014$ for CV, $P < 0.001$ for HEX). A P-value less than 0.05 was considered statistically significant. Therefore, only Topcon's ECD estimates could be trusted to certain extent, whereas our three estimates were highly reliable. When directly comparing the error produced by Topcon and our method, the Wilcoxon signed rank test revealed a statistically significant difference ($P < 0.0001$, $P = 0.0013$, $P < 0.0001$, for

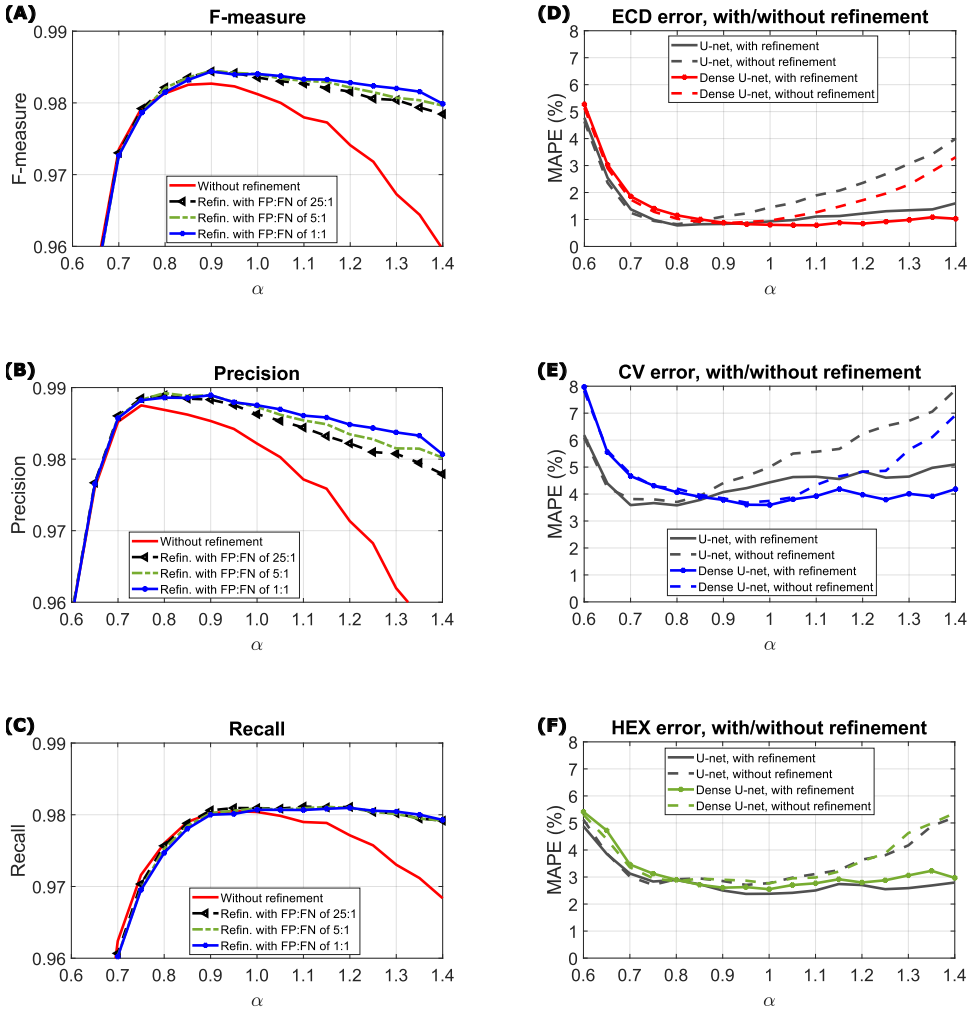

Figure 5.7: (Left column) F-measure (A), precision (B), and Recall (C) computed for the edge images (created with the U-net in the CNN-Edge) once the ROI has been applied, for different values of α in the postprocessing and cost ratios FP:FN of the SVMs in the refinement method. (Right column) MAPE of the biomarkers estimates –ECD (D), CV (E), and HEX (F)– for different values of α in four setups: using U-net (colored lines) or Dense U-net (black lines) to obtain the edge images with or without refinement. For the latter, the SVM with a cost ratio FP:FN of 1:1 was employed.

ECD, CV, and HEX, respectively), thereby suggesting that our average error is significantly smaller.

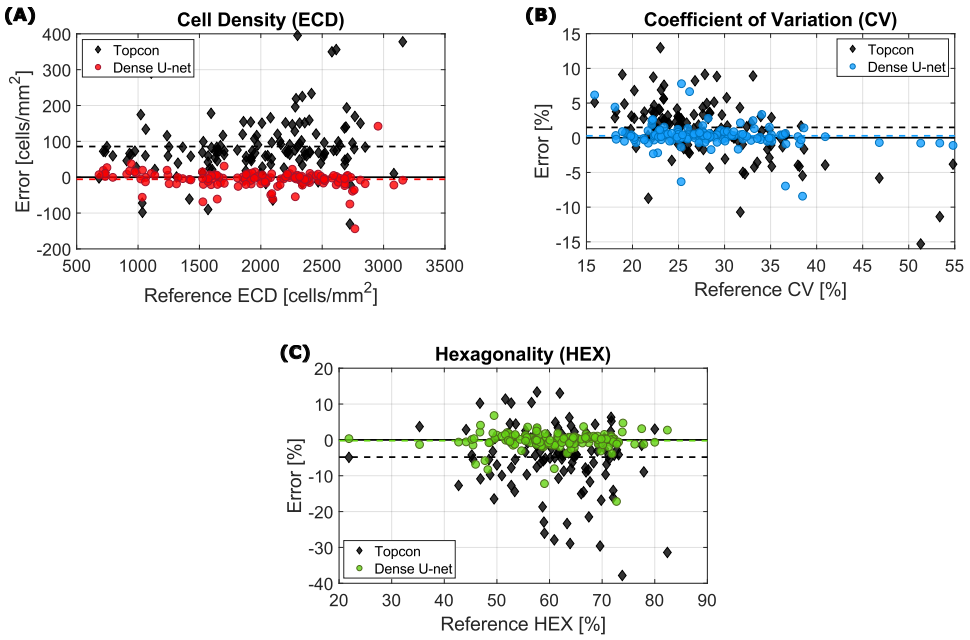

Figure 5.8: Error of the biomarkers estimates –ECD (A), CV (B), and HEX (C)–, using Dense U-net in both CNN-Edge and CNN-ROI, a postprocessing with $\alpha = 1$, and a refinement method with FP:FN of 1:1. Topcon errors are displayed in black diamonds. The x-axis indicates the value for the gold standard, and the y-axis indicates the error computed as the difference between the proposed estimates and the gold standard. Each point corresponds to one image in the dataset. Dashed lines indicate the average error.

5.5. DISCUSSION & CONCLUSIONS

The proposed Dense U-net has proven to be a reliable, robust network to infer the confident zone in corneal endothelium images. While taking twice the time than U-net in each training iteration and being slower in convergence (Figure 5.6), our experiments showed that a dense architecture allows more stable optimization results and it can deal with unbalanced datasets more proficiently than U-net.

Although the present work was meant to introduce a CNN architecture to automatically infer the ROI, this work could only have a clinical relevance when evaluated along with the other methods that, jointly, allow for the automatic estimation of clinical parameters (Figure 5.2). For that, we incorporated methods from our previous work and evaluated the whole end-to-end framework. To the best of our knowledge, this is the first time that a fully automatic method to estimate the corneal endothelium parameters has been proposed, but also the first time that the segmentation of diseased, complicated endothelial images have been evaluated.

Our framework has a number of potential benefits. First, it is a system with a structure that closely matches the clinical decision-making process, separating judgements about the inference of the edges, the selection of the ROI, and the biomarker estimation. This allows a clinician to inspect both segmentations, and manual corrections of the ROI

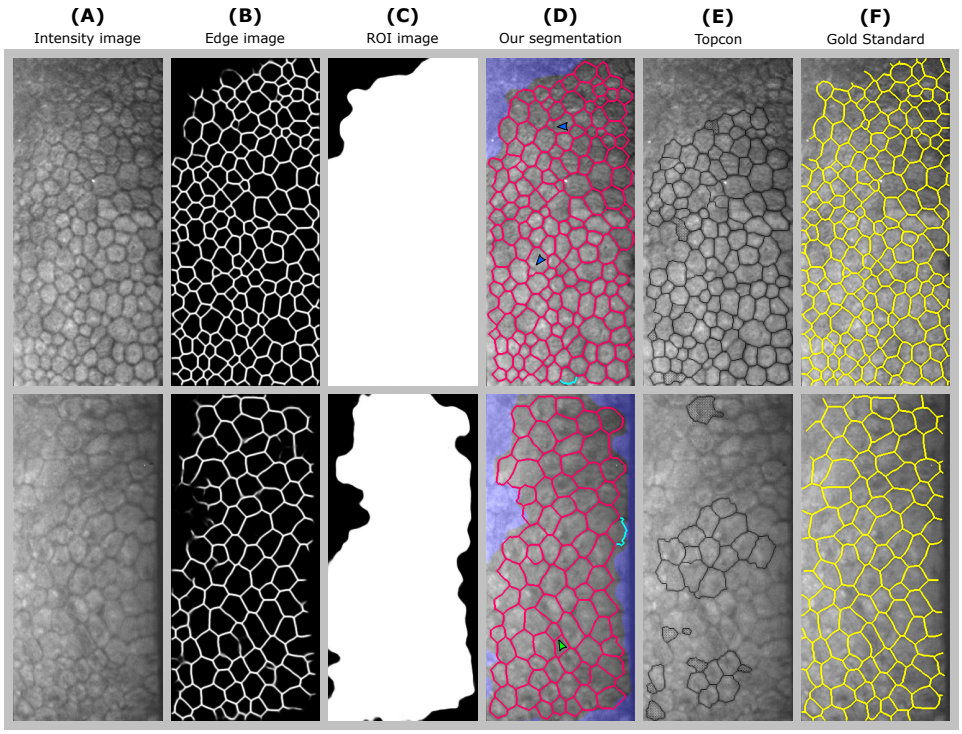

Figure 5.9: Two representative images facing different requirements in the postprocessing method: a large CV (top) would require a smaller smoothing filter, whereas a larger presence of faded false edges (bottom) would benefit from a larger smoothing filter. (A) Intensity images. (B) Edge probability images. (C) Estimated ROI. (D) Final segmentation (*red*) superimposed on the intensity image, with the edges removed by the refinement method (*cyan*) and the non-ROI area (*blue*). The blue arrows denote true edges that were smoothed away by the postprocessing. The green arrow denotes a false edge, nonexistent in the probability edge image, but created by the postprocessing. (E) Topcon's segmentation. (F) Gold standard images.

could be easily implemented. Second, our framework decouples the two problems and solves them independently. One key benefit of this is the independence of the second stage, which is not necessary to retrain depending on the first stage. As we proved in Section 5.4.1, U-net and Dense U-net can provide very different edge images, but this is not an obstacle as the purpose of the CNN-ROI is to interpret them.

A limitation of our framework is the assumption that the cell size is rather regular within the image and thus a postprocessing method whose filter is based on the average cell size should suffice. In reality, a small percentage of images have a very large variation in cell size (Figure 5.9, top). In those cases, the filter should be reduced in order to avoid smoothing away the small cells (blue arrows in Figure 5.9, top; here, the error was minimal for $\alpha = 1.10$). Analogously, some images might have a regular cell size but some faded, weak false edges (Figure 5.9, bottom), in which case a slightly larger filter would be beneficial. There, a non-existent edge in the probability edge image was created by the watershed of the postprocessing (green arrow in Figure 5.9, bottom; here, the error was

minimal for $\alpha = 0.80$ because it removed the false edge). Given all those factors, we chose a compromise solution, i.e. to use $\alpha = 1$. For that, we selected a Dense U-net for both, the cell segmentation (CNN-Edge) and the ROI segmentation (CNN-ROI), as it showed a well-balanced performance among images of different properties. This is shown in Figure 5.8 where a similar error is obtained for all images regardless the reference ECD or CV (images with very low ECD or high CV are more prone to errors).

Overall, it was important to design a ROI segmentation network (CNN-ROI) that would exclude from the ROI large blurred areas but would include in the ROI small inner blurred edges (either true or false), thus allowing later the refinement method to remove the false ones. That goal was achieved satisfactorily (Figure 5.9-C).

In comparison with Topcon's automatic segmentation, our method not only provides better quantitative results (Figure 5.8) but also better qualitative results (Figure 5.9-F). Indeed, Topcon's performance degrades significantly with complex images, detecting less cells and creating more segmentation mistakes (Figure 5.9-F).

In conclusion, we have presented a novel framework that automatically estimates several morphometric parameters from corneal endothelium images, the novel core of this work being the automatic selection of the ROI. The experimental results demonstrate that our approach achieves a state-of-the-art performance. Computationally, it takes only a fraction of a second to obtain all different segmented images (the refinement method is the bottleneck, which requires 10–20 seconds). Future work can address an adaptive postprocessing method in order to solve the aforementioned limitations. Clinically, this work demonstrate that high parameter accuracy can be reached in endothelium images from diseased corneas, bringing new opportunities in studies with very large number of images.

REFERENCES

[1] J. P. Vigueras-Guillén, B. Sari, S. F. Goes, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation,* BMC Biomedical Engineering **1:4** (2019).

[2] J. P. Vigueras-Guillén, E. R. Andrinopoulou, A. Engel, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Corneal endothelial cell segmentation by classier-based merging of oversegmented images,* IEEE Transactions on Medical Imaging **37**, 2278 (2018).

[3] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, *Densely connected convolutional networks,* in *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI, USA, 2017) pp. 2261–2269.

[4] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: convolutional networks for biomedical image segmentation,* in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. 9351 (2015) pp. 234–241.

[5] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, *H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes,* IEEE Transactions on Medical Imaging **37**, 2663 (2018).

[6] S. Guan, A. Khan, S. Sikdar, and P. V. Chitnis, *Fully Dense UNet for 2D sparse photoacoustic tomography artifact removal,* IEEE Journal of Biomedical and Health Informatics **24**, 568 (2020).

[7] G. Zeng and G. Zheng, *Multi-scale fully convolutional DenseNets for automated skin lesion segmentation in dermoscopy images,* in *15th International Conference Image Analysis and Recognition (ICIAR)*, Lecture Notes in Computer Science, Vol. 10882 (Póvoa de Varzim, Portugal, 2018) pp. 513–521.

[8] S. Jégou, M. Drozdzal, D. Vázquez, A. Romero, and Y. Bengio, *The one hundred layers tiramisu: fully convolutional DenseNets for semantic segmentation,* in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Honolulu, HI, USA, 2017) pp. 1175–1183.

[9] B. Selig, K. A. Vermeer, B. Rieger, T. Hillenaar, and C. L. Luengo Hendriks, *Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy,* BMC Medical Imaging **15:13** (2015).

[10] J. P. Vigueras-Guillén, A. Engel, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Improved accuracy and robustness of a corneal endothelial cell segmentation method based on merging superpixels,* in *15th International Conference Image Analysis and Recognition (ICIAR)*, Lecture Notes in Computer Science, Vol. 10882 (Póvoa de Varzim, Portugal, 2018) pp. 631–638.

[11] S. Beucher and F. Meyer, *Mathematical morphology in image processing,* (Taylor & Francis Group, 1993) Chap. The morphological approach to segmentation: the watershed transformation, pp. 433–481.

6

CONVOLUTIONAL NEURAL NETWORK-BASED REGRESSION FOR BIOMARKER ESTIMATION IN CORNEAL ENDOTHELIUM MICROSCOPY IMAGES

This chapter is based on the manuscript:

J.P. Vigueras-Guillén, J. van Rooij, H.G. Lemij, K.A. Vermeer, and L.J. van Vliet, *Convolutional neural network-based regression for biomarker estimation in corneal endothelium microscopy images*, 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 876–881, Berlin, Germany (2019).

6

Abstract

The morphometric parameters of the corneal endothelium –cell density (ECD), coefficient of variation in cell size (CV), and hexagonality (HEX)– provide clinically relevant information about the cornea. To estimate these parameters, the endothelium is commonly imaged with a noncontact specular microscope and cell segmentation is performed to these images. In previous work, we have developed several methods that, combined, can perform an automatic estimation of the parameters: the inference of the cell edges, the detection of the region of interest (ROI), a postprocessing method that combines both images (edges and ROI), and a refinement method that removes false edges. In this work, we first explore the possibility of using a CNN-based regressor to directly infer the parameters from the edge images, simplifying the framework. We use a dataset of 738 images coming from a study related to the implantation of a Baerveldt glaucoma device and a standard clinical care regarding DSAEK corneal transplantation, both from the Rotterdam Eye Hospital and both containing images of unhealthy endothelium. This large dataset allows us to build a large training set that makes this approach feasible. We achieved a mean absolute percentage error of 4.32% for ECD, 7.07% for CV, and 11.74% for HEX. These results, while promising, do not outperform our previous work. In a second experiment, we explore the use of the CNN-based regressor to improve the postprocessing method of our previous approach in order to adapt it to the specifics of each image. Our results showed no clear benefit and proved that our previous postprocessing is already highly reliable and robust.

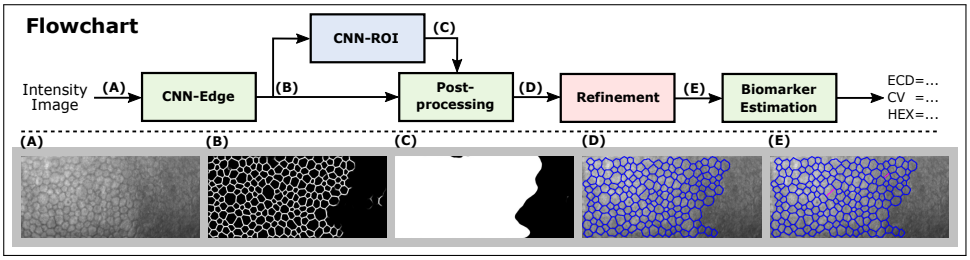

Figure 6.1: Flowchart of our previous framework for biomarker estimation. CNN-Edge infers the edge probability images, CNN-ROI estimates the region of interest, the postprocessing combines both images to provide a binary edge image, and the refinement method removes potential false edges (magenta lines in E) before estimating the biomarkers. The different images at each step (A-E) are displayed below the dashed line. For (D) and (E), the results are binary images, depicted in blue (superimposed to the intensity images for display purposes).

6.1. INTRODUCTION

In the last decade, Convolutional Neural Networks (CNNs) have become a very popular approach for image classification [1–3] and image segmentation, either via pixel classification or semantic segmentation [4]. Regression problems in images have also been solved with CNNs by adding fully connected layers with linear activations functions at the end of the network [5]. Overall, CNNs have led to a series of breakthroughs in the recent years, being employed with high success in many disciplines.

In corneal endothelium images, it is clinically relevant to determine the morphometric parameters of the cells in order to assess the quality and health status of the cornea. A quantitative analysis can provide information on endothelial cell density (ECD, reported as the number of cells per square millimeter), coefficient of variation in cell size (CV, expressed in percentage), and hexagonality (HEX, defined as the percentage of 6-sided cells). These image-based biomarkers can be easily estimated if the cell boundaries are identified in the image, therefore image segmentation has been the approach most commonly employed for solving this task [6–8]. In previous work, we achieved state-of-the-art results by employing a CNN U-net to segment the cell boundaries [9], but in order to make the process completely automatic we developed a dense U-net to infer the area in the image (region of interest, ROI) in which the biomarkers could be reliably estimated [10]. Furthermore, it is not a trivial task to combine both CNN output images (the edge probability image and the ROI image) and transform them into the final binary edge image to perform the biomarker estimation. Indeed, any small error or discontinuity in an edge could cause a significant error in the resulting estimates. This step was called *postprocessing* in our previous work [9].

In this work, we explore the possibility of using a CNN-based regressor, named *DenseNet-R35*, to directly estimate the biomarkers from the edge probability images, substituting the inference of the ROI, the postprocessing, and refinement. In addition, we explore the use of DenseNet-R35 to improve the postprocessing itself, and we compare it against a simpler approach based on Fourier analysis.

6.2. MATERIALS AND PREVIOUS WORK

The images used in this work are part of two ongoing studies in the Rotterdam Eye Hospital (Rotterdam, the Netherlands): one regarding the implantation of a Baerveldt glaucoma drainage device, and another regarding the visual function after corneal transplantation (Descemet Stripping Automated Endothelial Keratoplasty, DSAEK). For the Baerveldt study, images were retrieved before the surgical implantation and 3, 6, 12, and 24 months after surgery, in both the central and the temporal superior (TS) cornea. Here, we used 400 images from 100 patients. For the DSAEK study, 338 images from 41 patients were acquired 1, 3, 6, and 12 months after surgery, only from the central cornea. In both cases, images were obtained with a noncontact specular microscope Topcon SP-1P (Topcon Co., Tokyo, Japan). Data was collected in accordance with the tenets of the Declaration of Helsinki. Signed informed consent was obtained from all participants, and approval was obtained from the Medical Ethical Committee of the Erasmus Medical Center (Rotterdam, The Netherlands). Trial registrations:

- Baerveldt study (NL4823 (https://www.trialregister.nl/trial/4823), registered 06-01-2015)
- DSAEK study (NL4805 (https://www.trialregister.nl/trial/4805), registered 15-12-2014).

The images were saved as grayscale images of 240×528 pixels, covering an area of 0.25 mm × 0.55 mm. Images from the Baerveldt study generally show a good contrast between cell edges and cell bodies. The main challenge is that large fractions of the TS images are out of focus due to the curvature of the cornea at that position (as in Figure 6.1-A). In contrast, images from the DSAEK study are more prone to appear blurred, with low contrast, and slightly out of focus due to the optical distortions produced by the graft-recipient interface and the rough surface of the graft. An expert created the gold standard for each image by manually delineating the cell boundaries.

Previously, we presented a fully automatic framework to estimate the endothelial biomarkers from specular microscopy images (Figure 6.1). This framework is subdivided into four different methods: a CNN based on a Dense U-net to segment endothelial cells using an intensity image as input and providing an edge probability image as output (CNN-Edge) [9, 10]; a CNN based on a Dense U-net to segment the region of interest where cells are correctly detected using the edge probability image as input (CNN-ROI) [10]; a postprocessing method based on Fourier analysis and watershed that combines both output images and yields the binary edge image [9, 11]; and a machine learning approach based on Support Vector Machines (SVM) that removes potential false edges from the binary edge image [8]. The biomarkers could then be estimated from the final binary image.

6.3. METHODS

6.3.1. BIOMARKER ESTIMATION FROM EDGE IMAGES

In the first approach, we aim to use the DenseNet-R35 (Figure 6.2) after the CNN-Edge (Figure 6.1) in order to directly infer the biomarkers from the edge probability image.

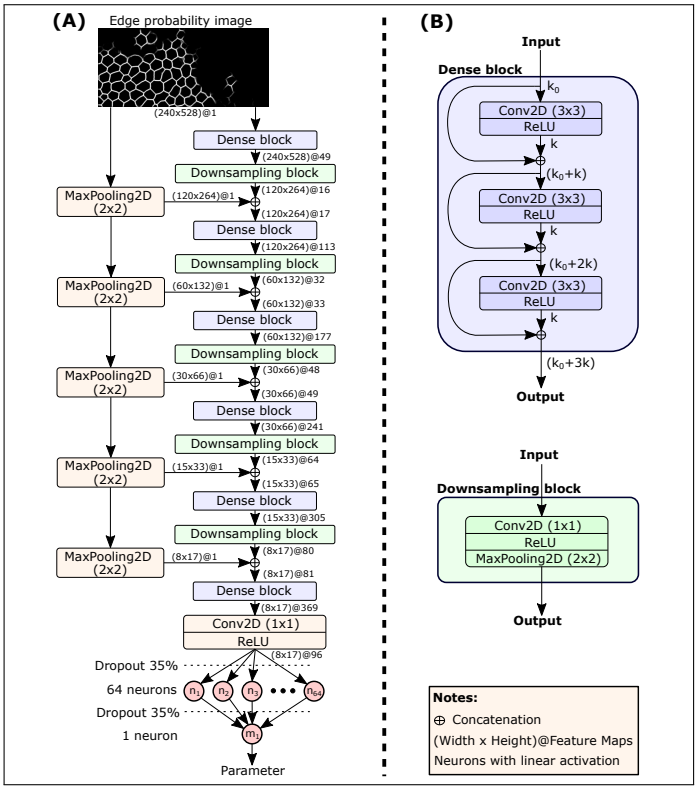

Figure 6.2: (A) A schematic overview of the proposed DenseNet-R35 ; and (B) a schematic overview of one dense block and one downsampling block.

Ideally, it would be preferred to use DenseNet-R35 as a substitution of the whole previous framework, but that would require a much larger annotated dataset. This possibility was tested, but failed to obtain reliable results. As described below, we can perform more data augmentation on the edge probability images, thereby making this approach feasible.

The proposed DenseNet-R35 (Figure 6.2) borrows the concept of densely connected convolutional layers from Huang et al. [3]. In dense blocks, a layer receives the feature maps of all preceding layers, which improves the direct information flow, helping back-propagating gradients and thereby facilitating better network learning during training. Huang et al. [3] used dozens of convolutional layers in each dense block, which creates bottlenecks due to the excessive number of piled up layers. To solve this, they added one 1×1 convolutional layer before each 3×3 convolutional layer in order to reduce the number feature maps. Here, we use only three 3×3 convolutional layers in each dense block, hence the 1×1 convolutional layers are not needed. Each layer in a dense block provides k feature maps, where k is the growth rate parameter. We set $k = 16$ in the first dense block, adding 16 more features at each new resolution step. After each dense block, the

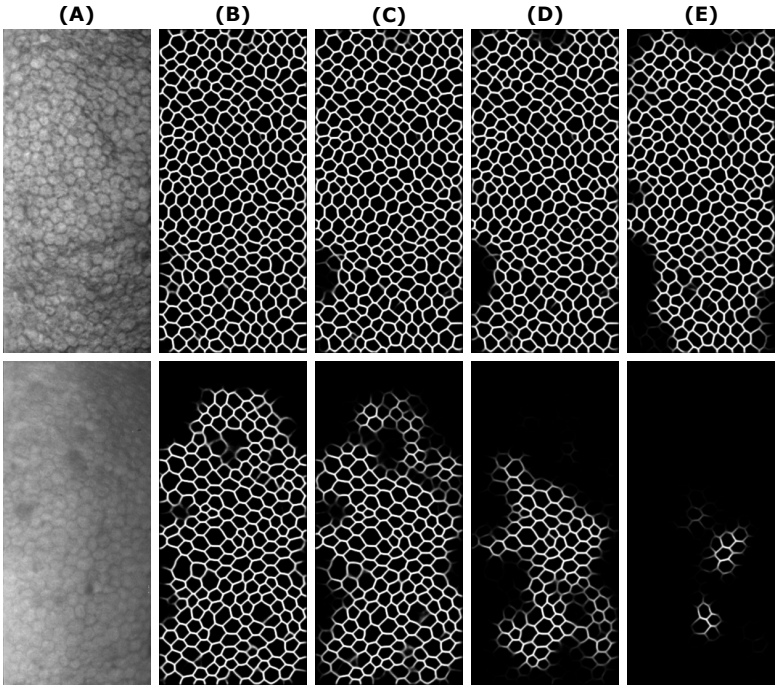

Figure 6.3: Representative output images of our previous Dense U-net (CNN-Edge) for four configurations based on the dropout rate of the dropout layers: 5% (B), 20% (C), 35% (D), and 50% (E), given (A) as input. Top image belongs to the Baerveldt study, whereas bottom comes from the DSAEK study.

feature maps are reduced to k by a 1×1 convolutional layer in the downsampling block. Furthermore, we introduce the edge probability image, properly scaled, at the beginning of each resolution step. Finally, the dense layer of neurons and the output neuron employ a linear activation function. The chosen loss function is the mean squared error (MSE), and the mean absolute percentage error (MAPE) is computed for analysis.

Another key aspect is the absence of batch normalization or dropout layers in the convolutional part, which either increased the MSE loss or prevented the network to converge, respectively. In contrast, the use of dropout in the layer of densely connected neurons was highly beneficial, obtaining the best performance for a dropout rate of 35%. This architecture was computationally efficient, avoiding bottlenecks during network training. It has a receptive field of 379×379 (approx. (height + width)/2), which showed to be optimal as either adding or removing a convolutional layer in each block was detrimental.

Data augmentation was a critical aspect in our training. Horizontal and vertical flipping multiplied the available data by four. Rotation had the inconvenience of losing the corners of the image once rotated, and thus it was evaluated. We obtained that uniform sampling of the rotation angle between $\pm35°$ yielded the best performance. Elastic deformations and translation were detrimental once rotation was employed. Finally, we

further increased the training data by employing, for the same intensity image, slightly different versions of the edge probability images, obtained as discussed in our previous work [10]. Briefly explained, the Dense U-net (CNN-Edge, Figure 6.1) could be tuned to provide more conservative edge images by simply increasing the dropout rate, thereby not inferring potentially spurious cell boundaries in highly blurred and/or noisy areas (Figure 6.3). For images with high contrast (Figure 6.3-A, top), the resulting edge images at different dropout rates did not show major differences unless the dropout was largely increased. However, for images with low contrast (Figure 6.3-A, bottom), larger areas of the images were rapidly not inferred after just small increments of the dropout rate. We argued that the biomarkers estimated from those images should, in principle, be very similar if enough cells were detected. The DenseNet-R35 performance improved drastically as we were including more images built with larger dropout rates, reaching the minimum test loss when images up to a dropout rate of 35% were used. Therefore, they were added to the training with the same labels and were named *twin sets*. Considering all augmentation, the original 738 images could be converted to 1.6 million images.

We employed a 10-fold cross-validation, and the same fold order was applied to all twin sets. The validation set contained images from all twin sets, but the test set only contained the original dataset so that it could be compared to our previous framework. The original dataset (here and in our previous work) was obtained with a dropout rate of 5%.

6.3.2. IMPROVEMENT OF THE POSTPROCESSING METHOD

In a second approach, the DenseNet-R35 was used to improve the postprocessing method from our previous framework. The postprocessing involved estimating the most common cell size in the edge probability images by using Fourier analysis ($l = 1/f^*$, being f^* the characteristic frequency), then smoothing the edge probability images with a Gaussian filter whose standard deviation (SD) is $\sigma = k_\sigma/(\alpha f^*)$, and finally applying the classic watershed [12], which does not require any parameter. The scaling factor of the sigma, $k_\sigma = 0.20$, was estimated in previous work [11], which assumed $\alpha = 1$.

The idea of estimating the cell size by Fourier analysis was proposed after observing that the 2D Fourier Transform (FT) of an endothelium image shows a distinctive ring related to the regular hexagonal patterns of the cells [13]. If the 2D FT was applied to the edge probability image instead, we got rid of the noisy patterns of the intensity image that could obscure the ring in Fourier domain. By simply computing the radial mean of the magnitude of the 2D FT [11], we could estimate the peak of the ring, f^* (Figure 6.4). In previous work [11], we proposed a model based on an exponential and a Gaussian to fit the radial mean, thereby estimating the ring size with the fitted Gaussian mean. This method is highly reliable to estimate the ECD of the image (Figure **??**-A), obtaining a high linear correlation between the estimated f^* and the true ECD (Pearson correlation coefficient $R = 0.965$, $P < 0.001$).

The main benefit of our postprocessing step is the detection of faded, true edges in the edge probability images. Since watershed requires a single local minimum per cell in order to yield a good segmentation, the former smoothing of the edge image is crucial. Given the robustness of the estimation of f^*, our postprocessing is simple yet

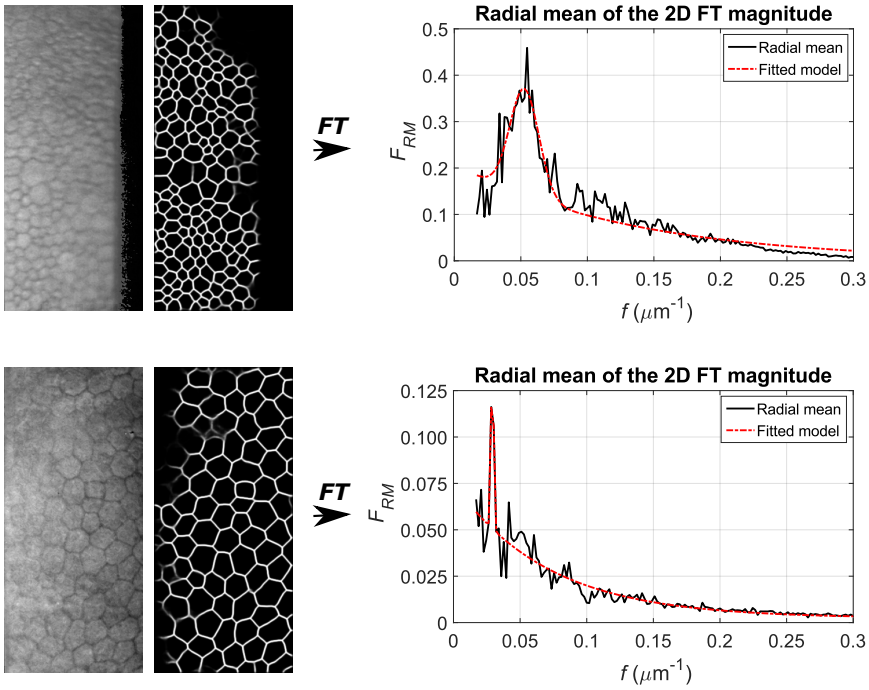

Figure 6.4: The radial mean of the magnitude of the 2D FT (in black) of the probability edge image (intensity images are only depicted for illustrative purposes). A fitted model based on a Gaussian and an exponential (in red) is used to estimate the peak (Gaussian mean, referred as characteristic frequency f^*) and Gaussian standard deviation. (Top row) Image with high CV. (Bottom row) Image with low CV.

very reliable. However, two flaws are observed in this approach: it does not take into account the (1) variation in cell size and (2) the blurriness of the edge probability image. Indeed, an image with high variation in cell size requires to reduce the filter size in order to avoid merging nearby edges and thereby missing small cells (Figure 6.4, top), whereas edge images with some blurred areas and regular cell size would benefit from a larger smoothing filter (Figure 6.4, bottom). To correct this, we define a scaling factor, α, to f^*.

In summary, our goal is to employ DenseNet-R35 to determine the best α for each image. We hypothesize that our network would be able to learn the characteristics of each image and therefore able to adapt the postprocessing. For this, we estimated the best α for each image (twin sets included). Specifically, we computed two values from the binary edge output images: the total number of cells detected (n_{total}) and the number of cells correctly segmented (n_{corr}); as well as one value from the gold standard: the number of real cells (n_{real}). Precision, $p = n_{corr}/n_{total}$, and recall, $r = n_{corr}/n_{real}$, were computed and combined into the F-measure, $F = 2pr/(p+r)$. We performed this evaluation for values of α between 0.7 and 1.3 in steps of 0.05, being the best α the one that provides the largest F-measure. Therefore, DenseNet-R35 would have an edge probability image as input and its best α as target. The same 10-fold cross-validation was used.

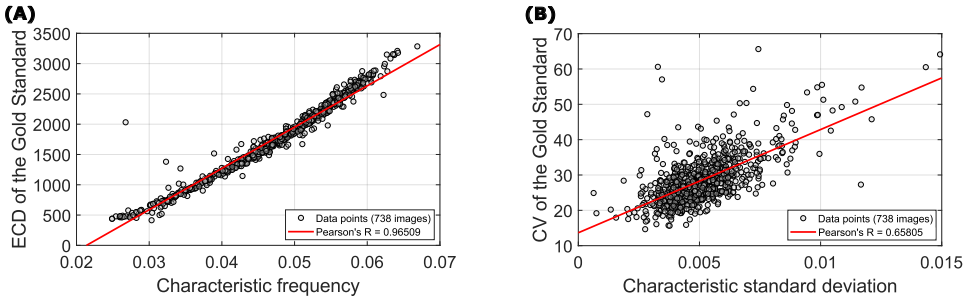

Figure 6.5: (A) Relation between the estimated characteristic frequency in Fourier domain and the true ECD. (B) The relation between the estimated characteristic SD and the true CV. Each circle is an image and the red line is the first degree polynomial that best fits the data (in a least-squares sense).

6.3.3. IMPROVEMENT OF THE POSTPROCESSING BY FOURIER ANALYSIS

The Fourier Analysis also showed a weak correlation between the fitted Gaussian SD (named "characteristic SD") and the biomarker CV (Figure 6.5-B, Pearson correlation coefficient $R = 0.658$, $P < 0.0001$). Indeed, if the endothelium exhibits a large variation in cell size, many frequency components appear in the FT (Figure 6.4-top), whereas if the cells are rather uniform in size, only a very narrow range of frequencies are enhanced (Figure 6.4-bottom). Therefore, we consider the use of the characteristic SD, which seems to roughly encode information about the cell size variation, to estimate the best scaling factor α in our postprocessing. For this, we employ a sigmoid model, defined as

$$\alpha = \frac{a}{1+\exp(-b\cdot \mathrm{SD}+c)} + d, \tag{6.1}$$

where a, b, c, and d are the parameters of the model. We set $d = 1 - a/2$ to center the sigmoid at $\alpha = 1$. The remaining parameters are estimated by employing a grid search and the F-measure defined above.

6.4. RESULTS

6.4.1. EXPERIMENT FOR BIOMARKER ESTIMATION

The training did not show over-fitting and converged rapidly. The error produced in each biomarker was significantly different, being HEX the most complicated parameter to predict (Figure 6.6). In the test set, the MAPE for the DenseNet-R35 was 4.32% for ECD, 7.07% for CV, and 11.74% for HEX, whereas our previous framework, with postprocessing and solving for oversegmentation, achieved a MAPE of 2.34% for ECD, 5.56% for CV, and 6.08% for HEX (Figure 6.7). The non-parametric Wilcoxon signed rank test on the MAPE distributions revealed a statistically significant difference ($P < 0.0001$, all biomarkers), thereby suggesting that the error in our previous framework was significantly smaller.

The endothelium of a healthy adult usually shows an ECD around 2500 cells/mm^2 (or higher), a CV around 30% (or lower), and a HEX around 60% (or higher) [14]. It is

clinically more relevant to assess an accurate estimation in images with low ECD or HEX, or high CV. In this respect, Figure 6.7-A shows an ECD error with similar magnitude for images with low and high ECD in both methods (within ±100 cells/mm^2, slightly higher for the DenseNet-R35), with the exception of some outliers (around 5% of images). While that error can be considered high for cases with very low ECD, both methods present a robust, stable inference of the ECD along different types of images, which means they are not affected by either a high or a low number of detectable cells. In contrast, the CV error is large with respect to the reference CV (Figure 6.7-B), and there is a small CV overestimation in images with low CV and an underestimation in images with high CV (same with HEX). Regression analysis determined that the errors were significantly correlated with the reference parameter ($P < 0.001$ for all biomarkers in both methods), yielding a slope of −0.03, 0.04, and 0.29 for ECD, CV, and HEX (respectively) in DenseNet-R35, and −0.01, 0.07, and 0.15 for ECD, CV, and HEX in our old framework. Thus, the correlation was very small for ECD and CV, but not for HEX. In previous work, these problems were the results of just a few minor errors in the segmentation since CV and HEX are biomarkers highly sensitive to segmentation errors. In this sense, DenseNet-R35 cannot overcome this problem.

6.4.2. EXPERIMENT FOR POSTPROCESSING

The histogram distribution of the true α showed a Gaussian shape (Figure 6.8-B), which suggested that, for the majority of images, an α close to 1 yields the best segmentation. However, DenseNet-R35 was unable to detect the pattern in the image that would determine the best α and instead estimated all values around $\alpha \approx 1$ (Figure 6.8).

6.4.3. EXPERIMENT FOR POSTPROCESSING BY FOURIER ANALYSIS

The grid search provided the largest F-measure = 0.947 for the parameters $a = 0.4$, $b = 300$, and $c = 1.8$. The sigmoid resembled a linear response for an SD of 0–0.01 with an α range 0.85–1.10. In contrast, an F-measure = 0.949 was obtained if $\alpha = 1$ was employed in the postprocessing method. Therefore, it was counterproductive to employ the char-

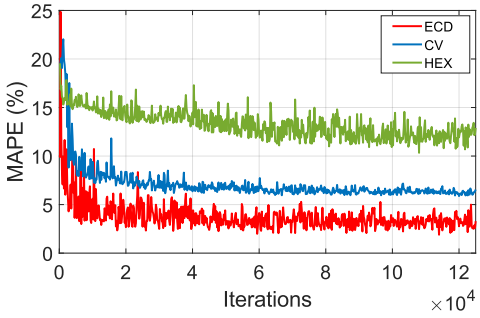

Figure 6.6: Mean absolute percentage error (MAPE) in the validation set (fold 1) during training (DenseNet-R35) for the three corneal parameters.

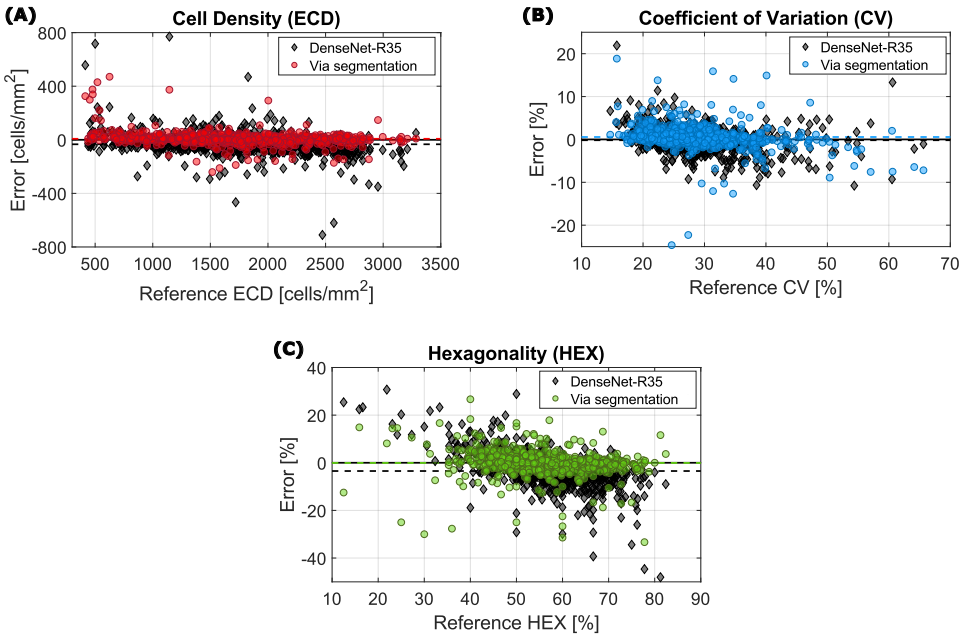

Figure 6.7: Error of the biomarkers estimates –ECD (A), CV (B), and HEX (C)–, for the DenseNet-R35 (black diamonds) and our old framework based on segmentation (colored circles). The x-axis indicates the value for the gold standard, and the y-axis indicates the error computed as the difference between the proposed estimates and the gold standard. Each point corresponds to one image in the dataset (738 images in total). Dashed lines indicate the average error.

acteristic SD for the inference of α.

6.5. DISCUSSION

The use of a CNN regressor to directly infer the biomarkers from edge probability images is a simple and fast method, yielding good results considering the complexity of the dataset under study, but it does not outperform our previous method [10]. In the latter, the selection of the ROI, the postprocessing, and the subsequent refinement (Figure 6.1) were designed to detect, discard, and/or minimize any mistake in the edge probability images yielded by the CNN-Edge. The substitution of those three methods by a single CNN-based regressor would simplify this task, and the main challenge was to build a large annotated dataset (in the order of millions of images). In this respect, the creation of twin sets was a simple way to increase the data. Different from other types of data augmentation, twin sets did not provide unrealistic cases, and they had enough differences to make a profitable contribution to the training. However, due to the time cost of obtaining the true biomarkers from each image in the twin sets, we assumed they had the same parameter estimates as their respective original edge images. We believed this might be the major cause of our modest results, which also can explain why HEX had a considerably larger error. Indeed, since the twin sets show less detectable cells (Fig-

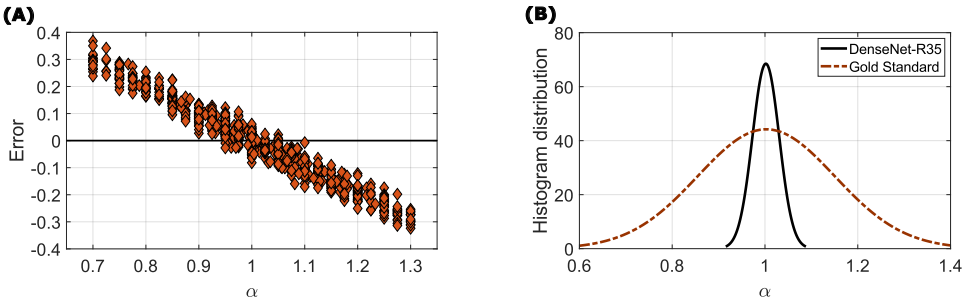

Figure 6.8: (A) Error of the estimation of alpha (each point correspond to one image). (B) Distributions of the histograms of two sets, the estimated alphas by DenseNet-R35 and the true alphas (Gold Standard).

ure6.3-C, D), the real HEX in those images could be significantly different, whereas ECD would be barely affected.

Once we concluded that DenseNet-R35 could not outperform our previous method, we considered its use to improve it instead. As discussed in Section 6.3.2, our postprocessing method does not consider the blurriness of the image or its variation in cell size. Although the latter could be roughly inferred by Fourier analysis (Figure 6.5-B), it had a significant estimation error. The experiment in Section 6.4.3 indicated that the inference of α based on the characteristic SD from Fourier analysis degraded the performance, therefore suggesting that blurriness might be the major factor to consider. Subsequently, we tested the use of DenseNet-R35 to infer the scaling factor α, thereby allowing the network to find the relevant features directly from the edge images. Different from the first experiment, we computed the target α for each independent image in the twin sets, thus avoiding any bias. However, DenseNet-R35 could not converge towards a reasonable solution. It is indubitable that this problem implies a larger complexity. For instance, blurriness might appear only in a fraction of the image while having high intensity contrast in the rest, and a CNN should be able to find the appropriate α for that small blurred portion while overlooking the characteristics of the rest of the image. Probably, a dataset of dozens of millions of images would be necessary to solve this problem.

The lack of convergence and similar behavior as in Figure 6.8-A was also observed if we attempted to estimate the biomarkers from the intensity images directly. Even with the use of more data augmentation (adding blurriness or noise), the network could not perform regression from the intensity images.

6.6. CONCLUSIONS

This study has shown that our CNN-based regressor, which directly infers corneal endothelium biomarkers from the edge probability images (created by our previous work), does not provide a better solution than aiming for the binary segmentation instead. Based on our experiments, the amount of annotated data that would be required to provide similar accuracy seems unreasonable. Indeed, the augmentation of the data up to 1.6 million of annotated images were not enough to outperform our previous work.

While it is possible that the limitation of the proposed data augmentation might have jeopardized the success, it is not clear whether a flawless augmentation would have made a large impact. Further work needs to be done to this respect. Nonetheless, considering that these biomarkers are defined based on the shape of the cells, it is reasonable to conclude that cell segmentation is still the preferable approach.

This study has also shown that the postprocessing method of our previous work, which uses the average cell size estimated by Fourier analysis to smooth the edge images such that a subsequent watershed provides the proper segmentation, is simple and yet very robust. Our attempts to improve the postprocessing by using more information from the Fourier analysis (the characteristic SD, which encodes the variation in cell size) have been unfruitful. We believe this is because the correlation between the characteristic SD and the cell size variation is not strong enough and the blurriness of the edge images should also be considered. This conclusion made us hypothesize that our CNN-based regressor could be satisfactorily employed for this task, as it could encode both factors (blurriness and cell size variation) to adapt the smoothing filter of the postprocessing. However, our experiments were unsuccessful, suggesting that these features might be too subtle to identify. A possible alternative could be to adapt α to a local window instead of inferring a global α for each image.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet classification with deep convolutional neural networks,* in *Advances in Neural Information Processing Systems 25* (2012) pp. 1097–1105.

[2] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition,* in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV, USA, 2016) pp. 770–778.

[3] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, *Densely connected convolutional networks,* in *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI, USA, 2017) pp. 2261–2269.

[4] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: convolutional networks for biomedical image segmentation,* in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. 9351 (2015) pp. 234–241.

[5] T. Shimobaba, T. Kakue, and T. Ito, *Convolutional neural network-based regression for depth prediction in digital holography,* in *27th IEEE International Symposium on Industrial Electronics (ISIE)* (2018) pp. 1323–1326.

[6] F. Scarpa and A. Ruggeri, *Development of a reliable automated algorithm for the morphometric analysis of human corneal endothelium,* Cornea **35**, 1222 (2016).

[7] B. Selig, K. A. Vermeer, B. Rieger, T. Hillenaar, and C. L. Luengo Hendriks, *Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy,* BMC Medical Imaging **15:13** (2015).

[8] J. P. Vigueras-Guillén, E. R. Andrinopoulou, A. Engel, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Corneal endothelial cell segmentation by classier-based merging of oversegmented images,* IEEE Transactions on Medical Imaging **37**, 2278 (2018).

[9] J. P. Vigueras-Guillén, B. Sari, S. F. Goes, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation,* BMC Biomedical Engineering **1:4** (2019).

[10] J. P. Vigueras-Guillén, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Automatic detection of the region of interest in corneal endothelium images using dense convolutional neural networks,* in *Proceedings of SPIE, Medical Imaging 2019: Image Processing*, Vol. 10949 (San Diego, CA, USA, 2019).

[11] J. P. Vigueras-Guillén, A. Engel, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Improved accuracy and robustness of a corneal endothelial cell segmentation method based on merging superpixels,* in *15th International Conference Image Analysis and Recognition (ICIAR)*, Lecture Notes in Computer Science, Vol. 10882 (Póvoa de Varzim, Portugal, 2018) pp. 631–638.

[12] S. Beucher and F. Meyer, *Mathematical morphology in image processing,* (Taylor & Francis Group, 1993) Chap. The morphological approach to segmentation: the watershed transformation, pp. 433–481.

[13] M. Foracchia and A. Ruggeri, *Automatic estimation of endothelium cell density in donor corneas by means of Fourier analysis,* Medical and Biological Engineering and Computing **42**, 725 (2004).

[14] B. E. McCarey, H. F. Edelhauser, and M. J. Lynn, *Review of corneal endothelial specular microscopy for FDA clinical trials of refractive procedures, surgical devices and new intraocular drugs and solutions,* Cornea **27**, 1 (2008).

7

DEEP LEARNING FOR ASSESSING THE CORNEAL ENDOTHELIUM FROM SPECULAR MICROSCOPY IMAGES UP TO 1 YEAR AFTER ULTRATHIN-DSAEK SURGERY

This chapter is based on the manuscript:

J.P. Vigueras-Guillén, J. van Rooij, A. Engel, H.G. Lemij, L.J. van Vliet, and K.A. Vermeer, *Deep learning for assessing the corneal endothelium from specular microscopy images up to 1 year after ultrathin-DSAEK surgery*, Translational Vision Science & Technology **9**(2):49 (2020).

7

Abstract

Purpose: *To present a fully automatic method to estimate the corneal endothelium parameters from specular microscopy images and to use it to study a one-year follow-up after ultrathin Descemet stripping automated endothelial keratoplasty.*

Methods: *We analyzed 383 post ultrathin Descemet stripping automated endothelial keratoplasty images from 41 eyes acquired with a Topcon SP-1P specular microscope at 1, 3, 6, and 12 months after surgery. The estimated parameters were endothelial cell density (ECD), coefficient of variation (CV), and hexagonality (HEX). Manual segmentation was performed in all images.*

Results: *Our method provided an estimate for ECD, CV, and HEX in 98.4% of the images, whereas Topcon's software had a success rate of 71.5% for ECD/CV and 30.5% for HEX. For the images with estimates, the percentage error in our method was 2.5% for ECD, 5.7% for CV, and 5.7% for HEX, whereas Topcon's software provided an error of 7.5% for ECD, 17.5% for CV, and 18.3% for HEX. Our method was significantly better than Topcon's ($P < 0.0001$) and was not statistically significantly different from the manual assessments ($P > 0.05$). At month 12, the subjects presented an average ECD = 1377 ± 483 [cells/mm2], CV = 26.1 ± 5.7 [%], and HEX = 58.1 ± 7.1 [%].*

Conclusions: *The proposed method obtains reliable and accurate estimations even in challenging specular images of pathologic corneas.*

Translational Relevance: *CV and HEX, not currently used in the clinic owing to a lack of reliability in automatic methods, are useful biomarkers to analyze the postoperative healing process. Our accurate estimations allow now for their clinical use.*

7.1. INTRODUCTION

Descemet stripping automated endothelial keratoplasty (DSAEK) is a cornea transplant procedure that involves the removal of the cornea's Descemet membrane and endothelium, followed by the transplantation of donor tissue containing the endothelium, Descemet membrane, and a part of corneal stroma. This procedure was introduced in 2006 [1, 2] and has replaced conventional penetrating keratoplasty for most cases of endothelial dysfunction [3, 4]. The most common indication for DSAEK is corneal edema in cases of Fuchs endothelial dystrophy (FED) or pseudophakic bullous keratopathy [5]. FED is characterized by degenerative changes in the endothelial cells, which trigger an abnormal, irregular growth of the Descemet membrane and a subsequent rapid loss of endothelial cells. Graft survival after DSAEK is reported to be good, especially in eyes with FED [6], although the outcome with respect to visual function is variable [7]. Newer techniques have aimed to reduce the thickness of the donor stroma (ultrathin [UT]-DSAEK), reporting better visual outcomes than conventional DSAEK in terms of speed of visual recovery and visual acuity [8]. A different procedure, Descemet membrane endothelial keratoplasty, aims to transplant donor tissue with corneal endothelium and Descemet membrane but without additional stromal tissue.

Regardless of the type of surgical procedure, long-term survival of corneal grafts is mainly dependent on the endothelial cell density (ECD) [9], which is defined as the number of cells per square millimeter. ECD decreases naturally with age, from an average density of 2800 to 3000 cells/mm^2 in healthy young people to a density of 2200 to 2600 cells/mm^2 for healthy elderly people [9–11]. This loss can be exacerbated by surgical trauma and certain diseases or stress factors, and it is assumed that 400 to 700 cells/mm^2 is the density at which corneal decompensation occurs and, thus, it defines the limit for graft survival [9, 12]. In addition to the loss of endothelial cells, the remaining healthy cells lose their regular hexagonal form and become irregular in shape and size. The two parameters used to describe this irregularity are the coefficient of variation in cell size (CV or polymegethism, expressed as the ratio of the standard deviation [SD] of the cell size to its mean size, in percentage), and the hexagonality (HEX or pleomorphism, defined as the percentage of six-sided cells).

To measure these parameters, it is necessary to image the endothelium. A clinically used method is specular microscopy, a noncontact, noninvasive technique that sends light toward the cornea at a small angle of incidence and captures the light reflected from the interface between the endothelium and the aqueous humor [10]. The specular reflex requires a regular, smooth endothelial surface to provide high-quality images, which is the norm in healthy corneas. Furthermore, the curvature of the cornea limits the area that can be imaged, which translates into a rather small field of view. Postkeratoplasty specular images are prone to show low contrast, high image noise, and areas that are heavily blurred or out of focus because of the many optical distortions that may arise owing to a variable thickness of the graft tissue, an irregular surface of the graft endothelium, or a graft-recipient interface haze [11]. This factor can substantially decrease the number of visible cells in the images, especially during the early postoperative phase.

Current specular microscopy manufacturers usually provide built-in software that automatically segments the images and estimates the endothelial parameters. However,

many recent studies have shown a lack of agreement between the estimates from different microscope's methods and sometimes an overestimation when compared with a manual segmentation, suggesting that the automatic results should be used with caution [13–17]. These studies mainly focused in ECD, in both healthy and diseased corneas. Accurate estimation of CV and HEX is even more challenging because it requires good quality segmentations, which is not reachable in most clinical cases. Several improved approaches have been proposed in the last years [18–27]. However, most of these methods show two main limitations that prevent them to be used in the clinic: (1) they require user intervention to manually discard the areas where cells are not distinguishable after segmentation and (2) they are mainly tested in rather high-quality images and/or healthy corneas.

In this study, we present a deep learning (DL) approach that automatically segments post–UT-DSAEK endothelial images, selects the trustworthy area where cells can reliably be detected, and computes the endothelial parameters with high accuracy, all in a few seconds, without the need for any user intervention. This method was compared with the estimates provided by the microscope's software and a manual segmentation (error analysis), and it was used to study the 1-year evolution of the endothelial parameters in post–UT-DSAEK corneas (clinical analysis).

7.2. METHODS

7.2.1. DATASETS

The main dataset contained 383 images of the central corneal endothelium from 41 eyes (41 patients) that underwent UT-DSAEK surgery in the Rotterdam Eye Hospital (Rotterdam, the Netherlands), acquired at 1, 3, 6, and 12 months after surgery. The included population for the study were patients over 18 years old with FED indicated for keratoplasty for visual rehabilitation and a visual acuity of less than 0.6 (Snellen chart). At the time of surgery, patients had an average age of 73 ± 7 years. The cornea grafts were provided by the Euro Tissue Bank (Beverwijk, the Netherlands). Written informed consent was obtained from all participants, and the study was conducted in accordance with the principles of the Declaration of Helsinki (October, 2013). Trial registration NL4805 (https://www.trialregister.nl/trial/4805, registered on 15-12-2014). Endothelium images were obtained with the noncontact specular microscope Topcon SP-1P (Topcon Co., Tokyo, Japan), which included the software IMAGEnet i-base (version 1.32) to estimate the endothelial parameters. The images covered an area of approximately 0.25 mm × 0.55 mm and were saved as 8-bit grayscale images of 240×528 pixels. Our protocol for image acquisition stated to take one image of the cornea and to repeat the acquisition up to a maximum of five images if the quality was unsatisfactory. Following this protocol, images were reacquired in 76% of the cases, with on average 2.3 images per session. For the clinical analysis, we used the parameters obtained from the image with the highest number of cells, and two patients were excluded because they missed one visit.

A secondary dataset was used to assist the training phase of the DL models. This dataset came from a clinical study concerning the implantation of a Baerveldt glaucoma device in the Rotterdam Eye Hospital (trial registration NL4823), referred as the Baer-

veldt dataset hereafter. This dataset contained 400 specular images of the central and temporal superior endothelium, obtained from 100 patients who were imaged before surgical implantation and 3, 6, 12, and 24 months after surgery, with the aim of observing whether the implant's tube was affecting the endothelium (the original Baerveldt dataset contained almost 8000 images from 200 patients, but only 400 images were manually annotated). The images were acquired with the same Topcon SP-1P microscope. Patients in this clinical study had no corneal disease, showed an average cell density of 2200 cells/mm^2, and the image quality was significantly better than the UT-DSAEK dataset. Thus, these images were useful for building robust DL models because they provided examples not present in the UT-DSAEK dataset.

All images were manually segmented to create the gold standard, using the open-source image manipulation program GIMP (version 2.10). The DL networks were programmed in Python 3.7, and we used Tensorflow (version 2.10) to train and test the models. The parameter estimation and statistical analyses were done in Matlab 2018a (MathWorks, Natick, MA). Confidence intervals were computed with the bootstrap method "bias corrected and accelerated percentile" in Matlab.

7.2.2. DEVELOPMENT OF THE NETWORK

DL is a class of machine learning algorithms that uses a sequence of mathematical operations (encoded in layers) to progressively transform the input data into more abstract representations to ultimately perform a task, such as classification, segmentation, or regression [28]. In our case, cell segmentation was done by a convolutional neural network (CNN), which are sliding-window filters that execute operations on images. This method basically entails transforming a specular endothelial image into another image where the pixels corresponding with the cell edges are given a high value (Figure 7.1). In contrast with classic machine learning techniques, DL does not require to set specific rules to extract the relevant features from the images. Instead, DL will do both tasks —feature extraction and segmentation— on its own, in what is called the "black box paradigm". To this end, DL requires plenty of labeled data to learn the transformations.

An earlier version of the DL network was presented in a previous work [27], in which we used a U-net [29] and 50 images from the Baerveldt dataset to produce the edge images (CNN-Edge), obtaining a pixel accuracy of 97.33%. Subsequently, we presented a different network [30] based on Dense U-Nets [31] to perform the segmentation of the region of interest (ROI) given the edge images (CNN-ROI); that is, the CNN-ROI identified the areas in an edge images where the cells were correctly detected. For that network, we extended the dataset up to 140 challenging images from the Baerveldt dataset, and we manually created the ROI gold standard images by selecting the trustworthy areas based on the edge images. Hence, CNN-ROI would mimic the evaluation that a human would do when observing the edge images. Our Dense U-Net [30] was a direct update from our previous U-Net [27], achieving a pixel accuracy of 98.94% on the ROI problem.

In this work, we improved our Dense U-Net in the following way (Figure 7.1): (1) we decreased the number of feature maps per convolutional block and added more blocks, therefore allowing for a higher reuse of features; (2) batch renormalization [32] layers were added in between the convolutional layers and the activation layers, which helped

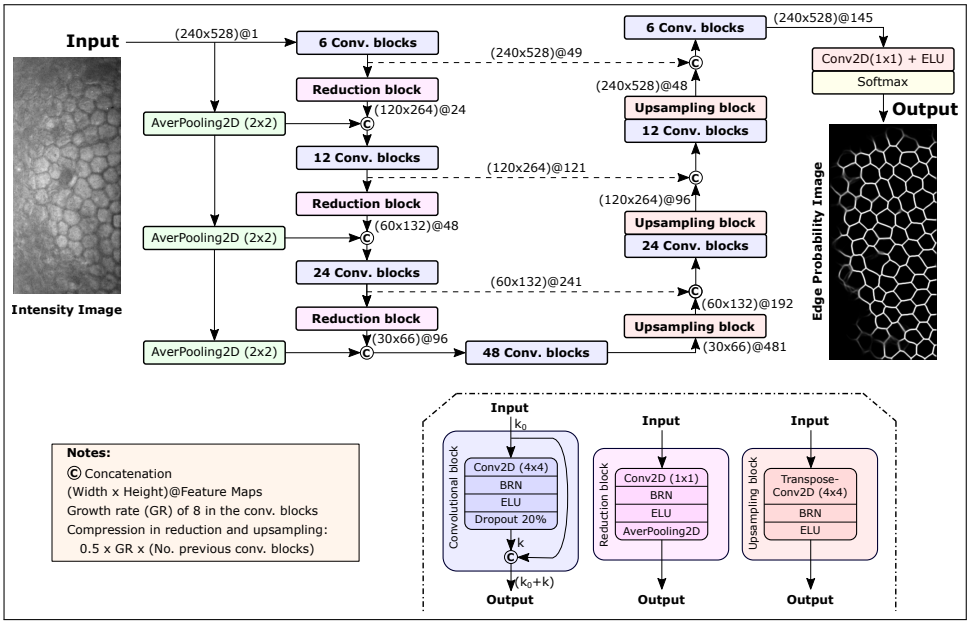

Figure 7.1: A schematic overview of the DL network. The different blocks in the network (convolutional, reduction, and upsampling) are depicted in the bottom. A dense block is defined by the concatenation of several convolutional blocks: 6 in the first resolution block, 12 in the second, etc. Each convolutional layer within the convolutional blocks has a growth rate GR = 8 (feature maps created by the layer), whereas the convolutional layers within the reduction and upsampling blocks create 0.5 × GR × (number of previous convolutional blocks) feature maps.

to stabilize the training; (3) we used exponential linear unit activations [33] and average pooling layers to improve performance; and (4) dropout layers [34] were placed before the concatenation to avoid additive dropout to earlier maps.

This network was used in both DL methods, the CNN-Edge and CNN-ROI (Figure 7.2). That is, CNN-Edge was trained with intensity images as input and the manually segmented images as target, providing the edge probability images as output, whereas the CNN-ROI was trained with edge probability images as input and the manually annotated ROI as target. Both networks were trained independently. Subsequently, a post-processing method [27, 35] combined both images (edge and ROI) to produce the final binary segmentation image. Briefly, the edge image was smoothed based on the average cell size (automatically obtained by Fourier analysis) and subsequently the watershed algorithm [36] was applied to create the binary segmentation. Finally, cells at the image border or with less than 75% of their area within the ROI were discarded. Once the final binary segmentation was created, the endothelial parameters were computed. To allow for a fair comparison with Topcon's software, we used the same Topcon's restrictions: ECD and CV were computed when at least six cells were segmented, whereas HEX was estimated from the inner cells (defined as segmented cells surrounded by other segmented cells) when there were at least six inner cells.

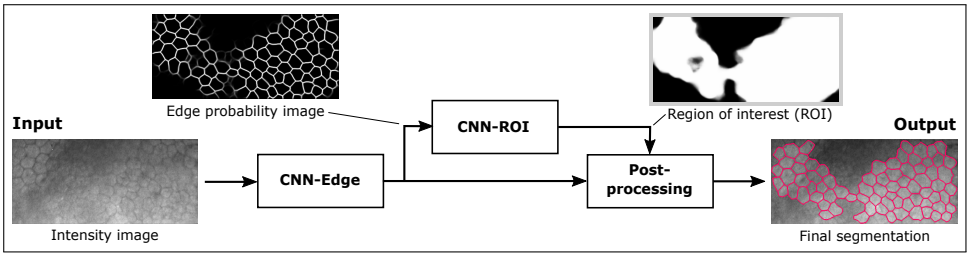

Figure 7.2: Overview diagram of the fully automatic method. The intensity image is the input of the first network (CNN-Edge), which outputs an edge (probability) image. This is introduced as input to the second network (CNN-ROI), which determines the region of interest. Both images are input for the postprocessing method, providing the final binary segmented image (the final segmentation is superimposed to the intensity image for display purposes).

Although having three different stages to first perform the segmentation and later infer the clinical parameters might seem cumbersome, it provided a robust aproach. Earlier experiments on the design of a single DL network that would directly estimate the endothelial parameters from the specular images indicated a substantial decrease in accuracy [37].

To evaluate the DL algorithms, a five-fold cross-validation was performed on the UT-DSAEK dataset: the images were randomly divided into five subsets (with all images from one eye in the same subset), using four subsets (plus the whole Baerveldt dataset) for training and the remaining subset for testing, and repeating the same process for the other subsets. The DL network hyperparameters were categorical cross-entropy as loss function, nadam optimizer [38], 150 epochs, and flipping left-right and up-down as data augmentation. A Jupyter notebook with the code of the DL models, the weights, and a few example images can be found in GitHub:
https://github.com/jpviguerasguillen/corneal_endothelium_TVST

7.2.3. STATISTICAL ANALYSIS

To assess the statistical significance of our method against Topcon's software, we used the paired Wilcoxon test to compare the percentage error after assigning a 100% error if no parameter estimate was produced. To assess the clinical statistical significance of our method against the gold standard, we used the one-way analysis of variance. For this end, we used the Shapiro-Wilk normality test to establish normality and Levene's test for homogeneity. The 95% limits of agreement from the Bland-Altman analysis were also used to compare our estimates against the gold standard. Changes in the estimates between visits were evaluated with paired t-test. A Pearson correlation coefficient was used to evaluate the correlation among the changes over the 1-year follow-up. All tests used a statistical significance established at $\alpha = 0.05$.

7.3. RESULTS

7.3.1. QUANTITATIVE ANALYSIS

Our fully automatic method detected, on average, 115 cells per image whereas Topcon only detected 30 cells (paired Wilcoxon test, $P < 0.0001$). As a result, our method and Topcon's software were able to estimate the ECD/CV in 98.4% and 71.5% of the images, respectively. According to the gold standard, insufficient cells were captured in 25 images (6.5%) to be able to estimate HEX. For the remaining images, our method and Topcon's software provided HEX estimates in 99.7% and 30.5% of the images, respectively. This large discrepancy originated from the necessity of detecting at least 25 to 50 cells in an image to have six or more inner cells, which Topcon's software failed to reach in most cases. These differences are illustrated in Figure 7.3 and Figure 7.4, showing how our method was able to detect as many cells as the gold standard in very challenging images, whereas Topcon's performance was unsatisfactory unless the image had good quality (Figure 7.4-H), which was the exception in this dataset.

An error analysis on the endothelial parameters was performed (Figure 7.5-A–C). Topcon's estimates displayed an ECD overestimation for the cases with high ECD and a large spread of the error in the three parameters. In contrast, our estimates showed an overall smaller error and a similar error spread for different ECD/CV/HEX reference values. Therefore, our method exhibited a robust performance regardless of the characteristics of the cells in the image.

A key factor in the method accuracy was the number of detected cells in the image, as a segmentation mistake (or discrepancy in the cells segmented) would distort the estimation significantly for images with a low number of cells. This finding can be observed in Figure 7.3-(A, E), where no major mistakes were detected but several cells in the border of the ROI were not included in the gold standard but appeared in our segmentation (or vice versa), and as a result there was a significant difference between our and gold standard's estimates, particularly in HEX. To model the behavior of our method in this respect, the error was plotted as a function of the number of cells (Figure 7.5-D–F). Because the images in the UT-DSAEK dataset contained a rather low number of cells, we included the estimates that would be obtained from the images of the Baerveldt dataset so that we had an overview of the method's performance for images containing a large number of cells (200–400 cells). For this, we trained new models, now applying a similar five-fold cross-validation to the Baerveldt dataset and using the UT-DSAEK dataset as training assistance. Subsequently, we fitted two exponentials to the mean and SD of the error using the least-squares method (Figure 7.5-D–F). The error showed a normal distribution along the y-axis for all three parameters and, thus, we could assume that the area within two SDs covered approximately the 95% of the error. This analysis showed that (1) for all three parameters, our error spread decreased as more cells were detected, (2) HEX required more cells to reduce the error spread, and (3) there was a small overestimation in CV for the images with less than 50 cells.

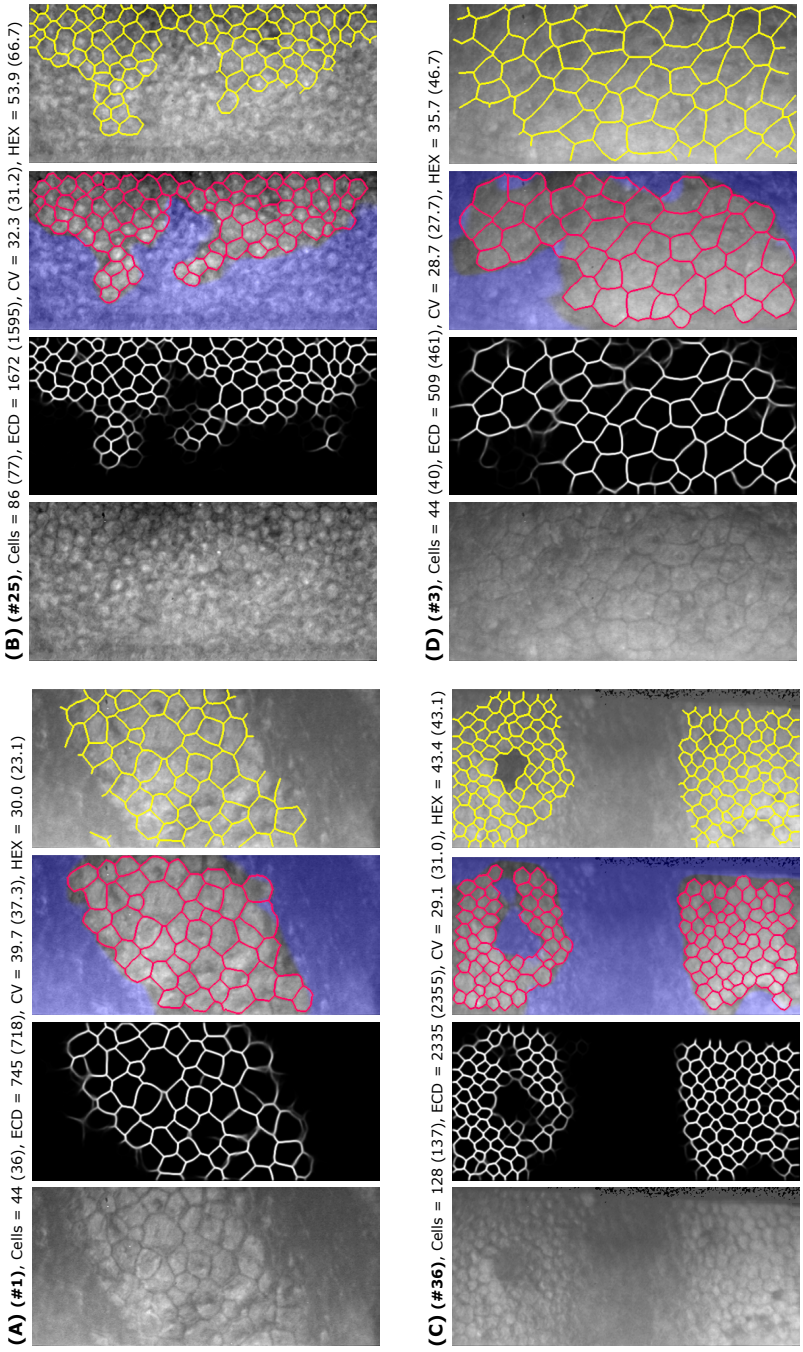

Figure 7.3: (i) Eight endothelial microscopy images from different patients (patient number # is in relation to Figure 7.8), acquired at month 1 (A, B), month 3 (C, D), month 6 (E, F), and month 12 (G, H). Our estimates are indicated above (gold standard in parenthesis). (ii) The output of the CNN-Edge.

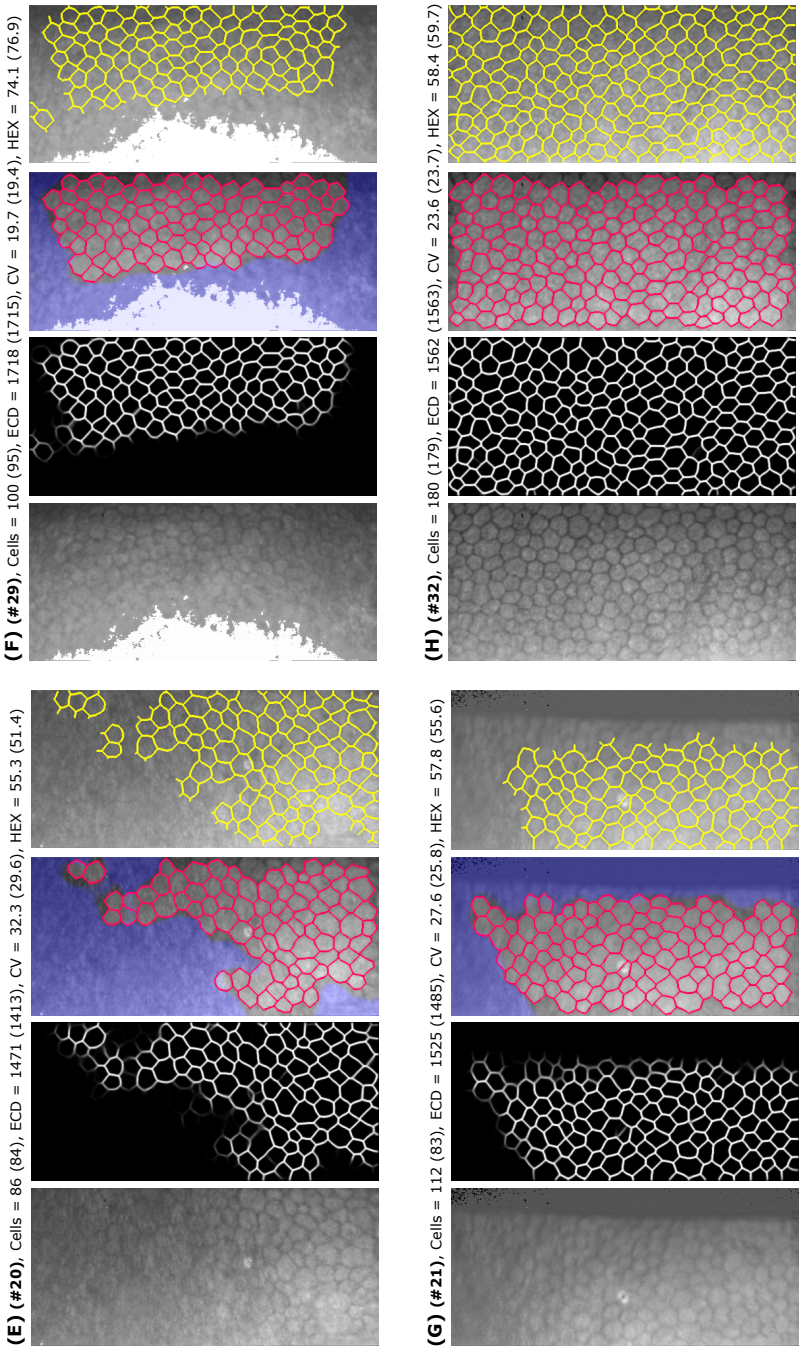

Figure 7.3: ... (iii) Our fully automatic DL segmentation (*red*) superimposed on the specular image along with the areas that CNN-ROI identifies as not trustworthy (*blue*). (iv) The manual gold standard annotations (*yellow*).

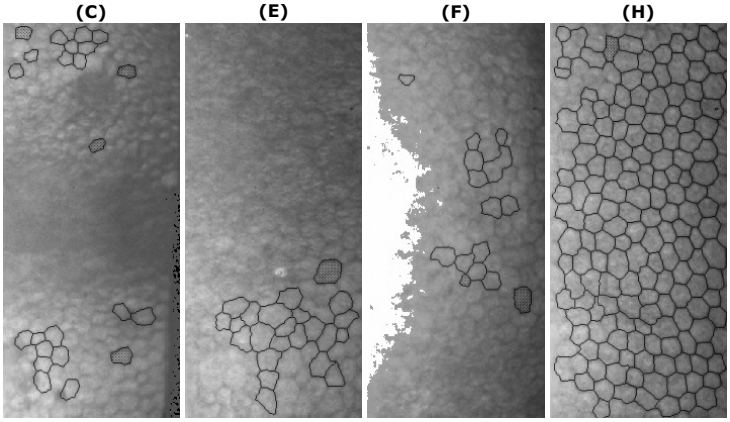

Figure 7.4: The only images from Figure 7.3 where Topcon's software could perform cell segmentation. Topcon's estimates are indicated underneath (gold standard values in parenthesis). N/A, not acquired.

7.3.2. ERROR ANALYSIS OVER TIME

As stated in the Methods, images were reacquired in the clinic when the quality was unsatisfactory. Specifically, we selected the image with more cells in each session, discarding the remaining images. For our method, a satisfactory image was obtained with the first acquisition in 50% of the visits, after the second in 27%, after the third in 20%, after the fourth in 2%, and after the fifth in 1%, which suggested that taking up to three images ensures acquiring a good clinical measure with high certainty. The average number of cells detected by our method was 124 ± 74, 132 ± 61, 134 ± 55, and 134 ± 55 cells for months 1, 3, 6, and 12, respectively, whereas Topcon's software detected 32 ± 42, 37 ± 36, 47 ± 39, and 54 ± 41 cells, respectively (paired Wilcoxon test, $P < 0.0001$). The mean absolute error (Table 7.1) and mean absolute percentage error (Figure 7.6-A–C) were computed for both, our and Topcon's estimates. The gold standard did not display enough cells in 6.2% of images to calculate the HEX, and our method was able to estimate HEX in all other cases.

Our method provided a significantly smaller error than Topcon (Figure 7.6-A–C). The statistical analysis yielded a statistically significant difference in favor of our approach for all parameters at all time points (paired Wilcoxon test, $P < 0.0001$). The percentage error was highest in the first month after surgery for all three parameters and decreased with time after surgery (Figure 7.6-A–C). This behavior was mainly due to the improvement in image quality over the months, which resulted in more images with estimates, more visible cells, less segmentation mistakes, and hence better estimation accuracy (Figure 7.7).

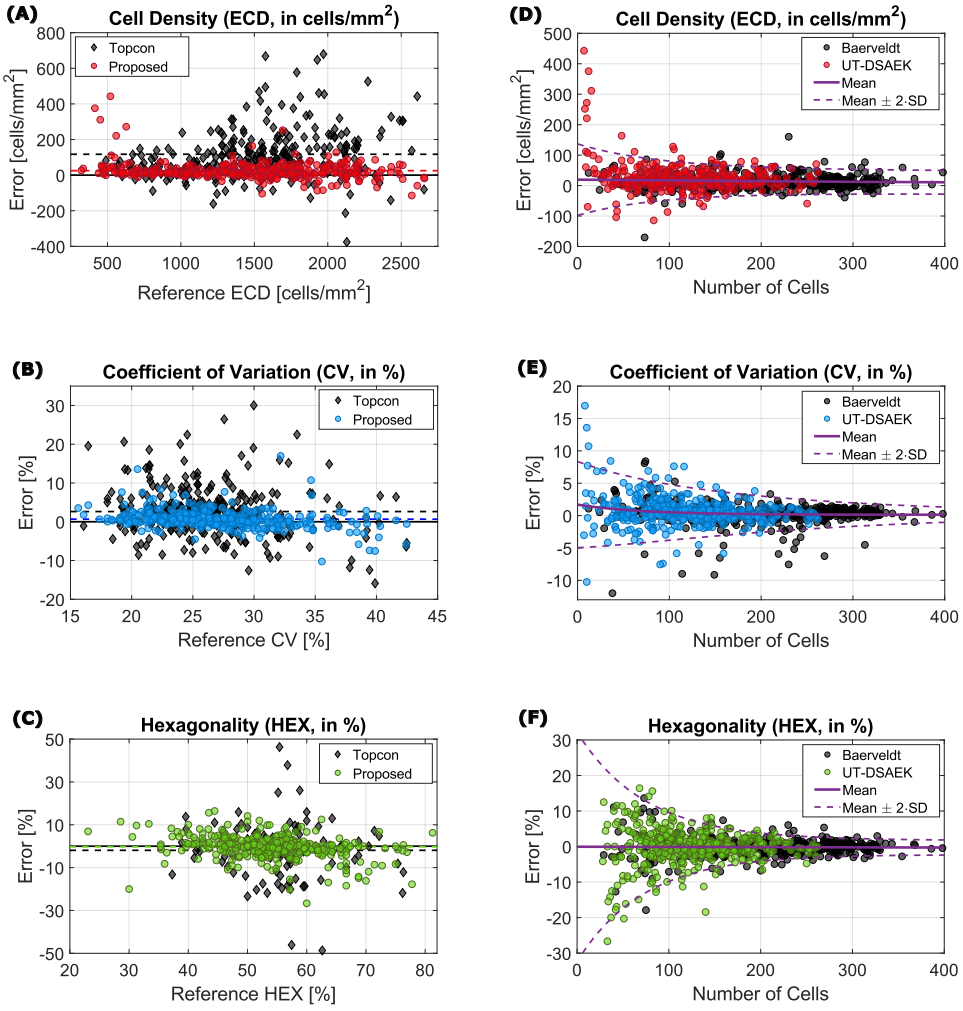

Figure 7.5: (Left column) Error of the estimates of ECD (A), CV (B), and HEX (C) in the UT-DSAEK dataset for our DL approach (colored circles) and Topcon (black diamonds). The x-axis indicates the value for the gold standard images and the y-axis indicates the error computed as the difference between the estimates and the gold standard. The average error for each approach is depicted with a dashed line. (Right column) Error of the estimates of ECD (D), CV (E), and HEX (F) with our DL approach in the UT-DSAEK (colored circles, 383 images) and Baerveldt datasets (black circles, 400 images), displayed as a function of the number of detected cells. The y-axis indicates the error computed as the difference between the estimates and the gold standard. The mean (solid line) and two SDs (dashed lines) of the error function were modeled with exponentials.

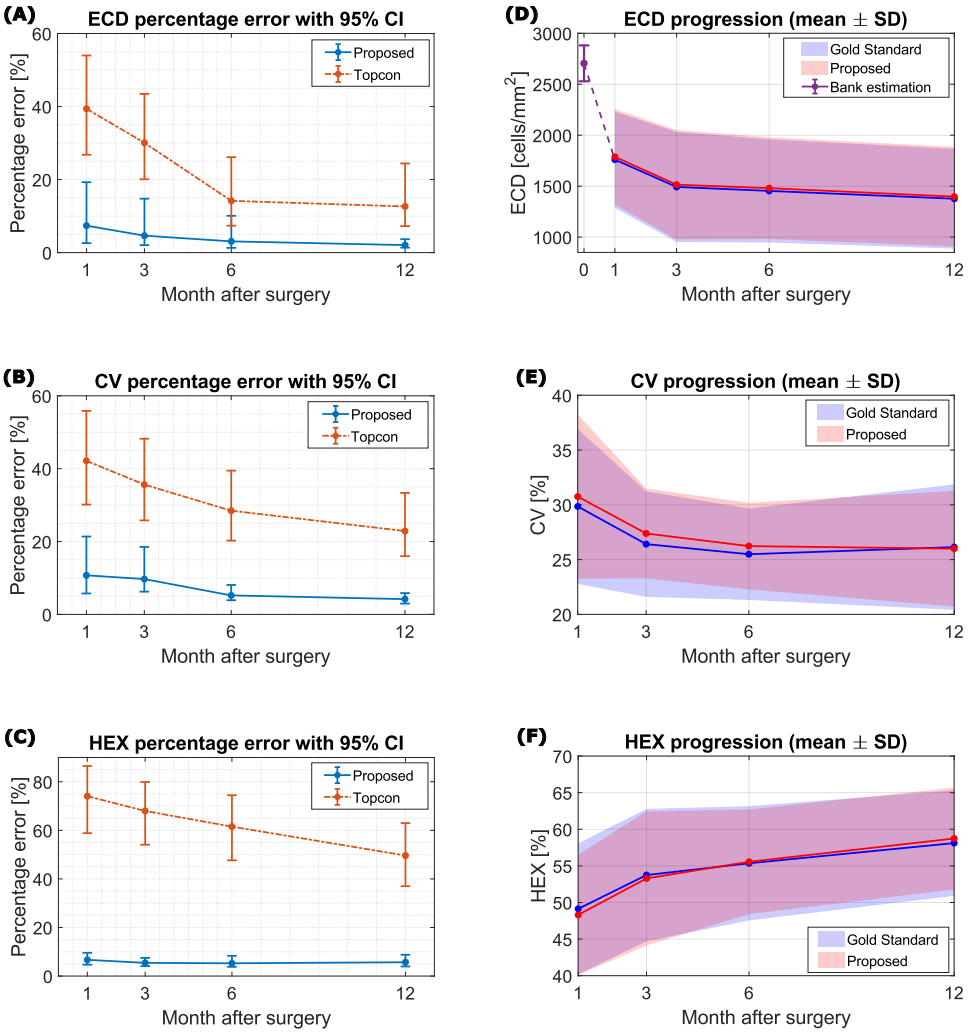

Figure 7.6: (Left column) The mean absolute percentage error (MAPE, computed as $\frac{100\%}{n}\sum_{i=1}^{n}|E_i - A_i|/A_i$, where A_i is the actual value, E_i is the estimated value, and n is the number of images) in the parameters ECD (A), CV (B), and HEX (C) with the 95% confidence intervals, at the different time points, for Topcon (*red*) and our method (*blue*), in the selected images. Images without a parameter estimation were assigned a 100% percentage error (for HEX, the cases where the gold standard did not provide an estimate were discarded). If the images without estimates were not considered, the overall mean absolute percentage error would be 2.5% for ECD, 5.7% for CV, and 5.7% for HEX in our method, and 7.5% for ECD, 17.5% for CV, and 18.3% for HEX in Topcon's software. (Right column) The evolution of the clinical parameters –ECD (D), CV (E), and HEX (F)– in the UT-DSAEK dataset over time, displayed as the average value (solid line) and one SD (colored area). Gold standard is depicted in *blue* and our proposed method in *red*. For ECD, the estimates provided by the cornea bank are included.

Table 7.1: The mean absolute error (MAE) and the SD of the endothelial parameters for Topcon and our method, in the selected images. The mean absolute error is computed as $\frac{1}{n}\sum_{i=1}^{n}|E_i - A_i|$, where A_i is the actual value, E_i is the estimated value, and n is the number of images. Only the cases with estimates were used to obtain the error (% indicated in 5th column).

	ECD MAE [cells/mm^2]	CV MAE [%]	HEX MAE [%]	% images with estimates (ECD/CV, HEX)
Our method	30.0 ± 38.5	1.5 ± 2.0	3.0 ± 3.2	98.1%, 93.8%
Topcon	118.7 ± 166.2	4.6 ± 4.8	10.3 ± 11.1	82.1%, 42.6%

7.3.3. CLINICAL ANALYSIS

Our method depicted the same evolution over time in the parameters as the gold standard except for CV, where a slight overestimation occurred (Figure 7.6-D–F). The distributions of the estimated parameters from our method and the gold standard at the different time points resembled a normal distribution, and they passed the Shapiro-Wilk normality test after excluding one outlier for CV at month 1 and for HEX at month 3. All distributions also passed the Levene's test for homogeneity. A one-way analysis of variance indicated no statistically significant difference between the manual and our automatic assessments for any case ($P > 0.05$). A Bland-Altman analysis showed that more than 95% of the ECD estimates were within the 95% limit of agreement for all months except month 1 (94.8%). For CV, month 1 (92.1%), and month 3 (92.5%) did not reach the 95% limit of agreement. HEX was below the limit for all months (91.2%–94.9%).

The average estimated ECD at the cornea bank was 2705 ± 174 cells/mm^2, which decreased abruptly after the surgical intervention and it stabilized after the third month (Figure 9), although the changes were statistically significant between all visits (paired t-test, $P < 0.03$). The average CV presented a similar evolution to ECD, with a substantial decrease between the first and third months ($P < 0.01$) and a subsequent stabilization ($P > 0.1$). In contrast, the average HEX displayed a continuous increase, with a statistically significant change between months 1 and 3 ($P = 0.01$).

We observed that the group of patients with the largest loss in ECD (Figure 7.8, patients #1–#15, except #2) had a consistent large decrease in CV (-11.1 ± 9.9) and increase in HEX ($+13.2 \pm 12.4$) between months 1 and 12 (values from gold standard estimates), which suggested a possible stabilization in the cellular health. The exception, patient #2, was the only case developing FED after surgery, with a bad progression in CV (+6.2) and no estimates for HEX. Overall, we found a statistically significant correlation between the changes in ECD and CV from month 1 to month 12 among all patients (Pearson's $R = 0.356$; $P = 0.031$) and between CV and HEX ($R = -0.538$; $P = 0.001$), but no correlation between ECD and HEX. This was also observed with our estimates.

Finally, we also observed that the patients that displayed a smaller ROI at month 1 usually had a larger ECD loss between months 1 and 12 (Pearson's $R = -0.374$; $P = 0.019$). This agrees with the idea that a smaller area with visible cells is an indication of an unhealthy tissue, which will in turn have an increased progressive loss of ECD.

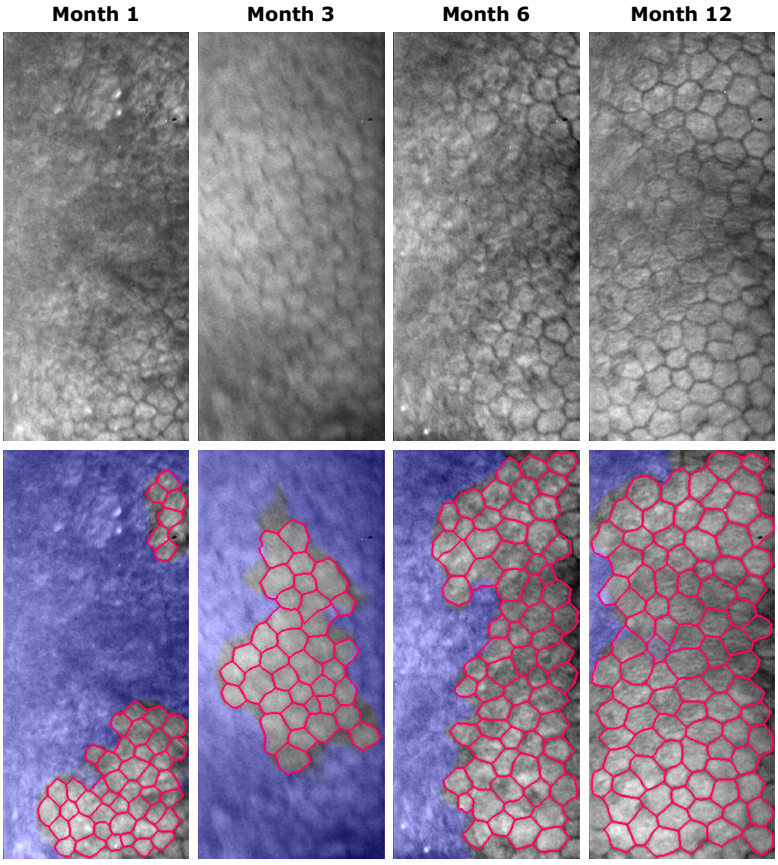

Figure 7.7: A representative example (patient #7) of the progression of the endothelium over the months, with a clear improvement in image quality and detected cells. Top row displays the intensity image and bottom row indicates our DL segmentation in *red* (non-trustworthy areas in *blue*). Further information on the ECD for patient #7 in Figure 7.8.

7.4. DISCUSSION

We have presented a robust, fully-automatic method for the estimation of the corneal endothelium density and morphometric parameters from specular microscopy images in very challenging cases (i.e. post–UT-DSAEK images). The parameters estimates were in very high agreement with the gold standard. In comparison with Topcon's software, the improvement was considerable: our method detected cells in almost all images (Topcon failed in more than one-fourth of the images); the mean absolute error was more than three times smaller in the three parameters (Table 7.1); and the number of detected cells was almost three times higher, hence decreasing the estimation variability. This variability, particularly significant for HEX, is a well-known problem [39], being widely accepted that 75 cells are required to estimate ECD with high reliability [40].

As mentioned, just a few segmentation mistakes can significantly affect the param-

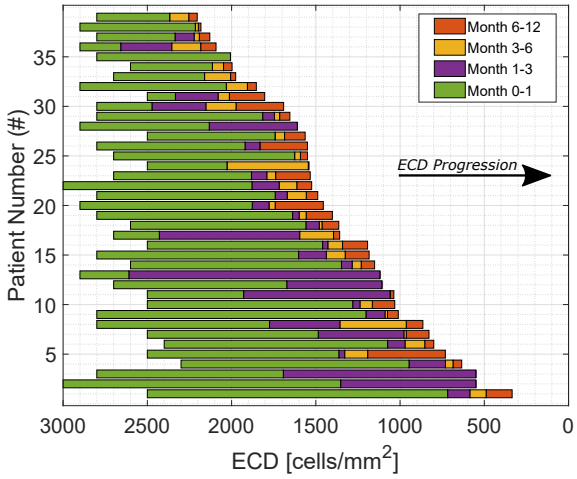

Figure 7.8: The ECD progression during the one-year follow-up for each patient (to read from left to right). The progression (decrease in ECD) for each time interval is depicted with a bar (values from the gold standard estimates). Patients are sorted based on the ECD at month 12. If there is an increase in ECD during a time interval, the corresponding bar is not depicted and instead the bar of the previous period is shortened.

eters estimates. Therefore, the use of three stages (CNN-Edge, CNN-ROI, and postprocessing) is necessary to not only infer the corneal parameters from the trustworthy areas but to also fix potential mistakes in the edge images. For example, the CNN-ROI in Figure 2 detected (in the left–central area) three cells with weak and/or duplicate edges (small dark blob), but the CNN-ROI combined with the postprocessing was able to fix this problem. A single DL network to generate both images simultaneously (edges and ROI) has been used in the literature [41], but this approach provided unreliable ROI images for our image data. Indeed, selecting the trustworthy area based on the intensity images generated blurred ROI images that did not sufficiently match the edge images [37]. Alternatively, a single DL model to directly infer the parameters [30] may seem attractive because of its simplicity, but (1) it would require considerably more annotated data for training, which is very expensive to create, and (2) it would be impossible for the user to know whether the provided estimates are reliable (black box paradigm). In contrast, a segmentation can be visually inspected, allowing the user to assess its reliability and even allowing for manual corrections.

Previous work on automated endothelial cell parameter estimation found in the literature can be classified into two main groups: pre-DL and DL methods. Pre-DL methods [18–22] aimed mainly to segment the whole image with classic machine learning techniques, which would be clinically usable if cells were visible in the whole image or if the user would select the ROI manually. The DL approaches [23–27], presented in the last 2 years, mainly focused on proving that a DL network is capable of segmenting endothelial images, addressing the segmentation accuracy on relatively good quality images, but rarely the clinical parameters. The closest work to ours was presented by Daniel et al. [24], who used the U-net to segment 385 “real-world” specular images, including differ-

ent ophthalmologic diseases and image qualities. Their ground truth was made by manually dotting the center of each cell and manually selecting the ROI, which only allows for ECD evaluation, and hence it was difficult to establish a comparison. Furthermore, their postprocessing was based on simple thresholding (prone to mistakes in edges not completely delineated), and their approach lacked a method to automatically select the ROI (blurred areas surrounded by detected cells would be considered cells).

One strength of our method is its ability to deal with a variety of imaging artifacts (Figure 7.3). Soon after surgery, the irregular, unsmoothed surface of the endothelial graft makes it difficult to have an image in which all parts are in focus (Figure 7.3-A–C). Furthermore, studies in postoperative corneal thickness have indicated a presence of edema in the recipient corneal stroma soon after surgery and a continuous decrease over time [42], which agrees with the improvement in image quality over time. In general, any lesion or abnormal tissue structure in the upper layers of the cornea can distort the optic path, creating all kind of noisy patterns. This includes saturation problems (Figure 7.3-F, G), in which the microscope internal software fails to transform the specular reflex into a suitable image because of those light aberrations. In contrast, ocular microsaccades, the movement of the patient, or simply respiration can produce heavy blurriness (Figure 7.7, month 3), which is unrelated to the state of the endothelium. In all these cases, Topcon failed to detect any cell (or detected barely a few) whereas our method performed almost as good as the gold standard. For the most complex cases, such as in images with FED, our approach needs further improvements, which probably could be tackled with more labeled images containing guttae to teach the DL models the structural and visual changes introduced by FED.

Clinically, we obtained an average ECD loss of 46.3% at month 6 and 49.1% at month 12, which was similar (or slightly higher) than other cases found in the literature. For instance, the ECD loss reported after 12 months from UT-DSAEK surgery was 49% by Feng et al. [43], 38.9% by Graffi et al. [44], and 35.6% by Busin et al. [8]. For DSAEK, van Cleynenbreugel et al. [45] reported a 40.2% loss after 6 months, Guerra et al. [46] a 34.9% loss after 12 months, and Javadi et al. [47] a 42.8% at 2 years of follow-up. We observed that our cohort showed a large variability on ECD loss at the different time intervals, without a clear pattern (Figure 10). Indeed, some patients showed a deceleration on ECD loss over time with an almost non-existent progression within the last 6 months, whereas others depicted a small acceleration on ECD loss at that time. Regarding CV and HEX, barely any post-keratoplasty study includes them in their analysis because of the lack of reliability (up to now) in automatic methods. In our case, we estimated from the manual assessments a CV of 26.1 ± 5.7 and a HEX of 58.1 ± 7.1 at month 12. Existing literature in various ethnic groups indicates that the average CV is 26 ± 4 [48, 49] and HEX is within 58% to 74% [50] in healthy population, and it is widely accepted that a CV of less than 30% and a HEX of greater than 60% is usually a sign of a healthy, stable endothelium. Hence, our post-transplant cohort showed an overall good outcome in terms of CV and HEX. Furthermore, we have shown that (1) our automatic method provides a high accuracy in CV and HEX (Table 7.1), (2) the existence of an evolution pattern after UT-DSAEK surgery (decrease of CV, increase of HEX; Figure 7.6-D–F), particularly significant for the cases with the largest loss in ECD, and (3) a correlation between some of the parameters' evolution (stronger between CV and HEX), which might be an indication of good healing

and cell loss stabilization. Indeed, the transplants that suffered a larger ECD loss in the first month had a better improvement in CV and HEX later. These two parameters can tell us something about the distress of the cells, but this is merely an extrapolation from biological science lacking proper confirmation from clinical trials. In contrast, ECD as a clinical parameter to evaluate the functionality of the cornea in terms of total thickness and clarity has been studied extensively, both in the natural state and after corneal transplantation. As we have developed a better method to analyze CV and HEX, it would be interesting to study these two parameters, for example in future follow-up studies of corneal grafting.

In summary, the results reported here demonstrated the ability of this DL method to estimate the endothelial parameters from images with different qualities and noise patterns. It also indicated a potential usefulness of employing all three endothelial parameters to study the evolution of the tissue after keratoplasty. Hence, our DL method presents itself as a valuable tool to be used in studies of corneal transplantation programs with more patients and larger follow-ups to analyze the relevance of CV and HEX and their potential roles as predictors for graft survival.

REFERENCES

[1] M. O. Price and F. W. J. Price, *Descemet's stripping with endothelial keratoplasty: comparative outcomes with microkeratome-dissected and manually dissected donor tissue,* Ophthalmology **113**, 1936 (2006).

[2] M. S. Gorovoy, *Descemet-stripping automated endothelial keratoplasty,* Cornea **25**, 886 (2006).

[3] C. Y. Park, J. K. Lee, P. K. Gore, C. Y. Lim, and R. S. Chuck, *Keratoplasty in the United States: a 10-year review from 2005 through 2014,* Ophthalmology **122**, 2432 (2015).

[4] A. Anshu, M. O. Price, and D. T. T. F. W. J. Price, *Endothelial keratoplasty: a revolution in evolution,* Survey of Ophthalmology **57**, 236 (2012).

[5] M. Busin, *DSAEK for the treatment of endothelial disease: results in the initial 100 cases,* Klinische Monatsbl'atter f'ur Augenheilkunde **226**, 757 (2009).

[6] M. Ang, Y. Soh, H. M. Htoon, J. S. Mehta, and D. Tan, *Five-year graft survival comparing Descemet stripping automated endothelial keratoplasty and penetrating keratoplasty,* Ophthalmology **123**, 1646 (2016).

[7] A. M. Turnbull, M. Tsatsos, P. N. Hossain, and D. F. Anderson, *Determinants of visual quality after endothelial keratoplasty,* Survey of Ophthalmology **61**, 257 (2016).

[8] M. Busin, S. Madi, P. Santorum, V. Scorcia, and J. Beltz, *Ultrathin Descemet's stripping automated endothelial keratoplasty with the microkeratome double-pass technique (two-year outcomes).* Ophthalmology **120**, 1186 (2013).

[9] W. J. Armitage, A. D. Dick, and W. M. Bourne, *Predicting endothelial cell loss and long-term corneal graft survival,* Investigative Ophthalmology & Visual Science **44**, 3326 (2003).

[10] B. E. McCarey, H. F. Edelhauser, and M. J. Lynn, *Review of corneal endothelial specular microscopy for FDA clinical trials of refractive procedures, surgical devices and new intraocular drugs and solutions,* Cornea **27**, 1 (2008).

[11] H. B. Hindman, K. R. Huxlin, S. M. Pantanelli, C. L. Callan, R. Sabesan, S. S. Ching, B. E. Miller, T. Martin, and G. Yoon, *Post-DSAEK optical changes: a comprehensive prospective analysis on the role of ocular wavefront aberrations, haze, and corneal thickness,* Cornea **32**, 1567 (2013).

[12] C. S. Foster, D. T. Azar, and C. H. Dohlman, *Smolin and Thoft's the cornea: scientific foundations & clinical practice* (Lippincott Williams & Wilkins, Philadelphia, PA, 2004) pp. 46–48.

[13] J. Huang, J. Maram, T. C. Tepelus, C. Modak, K. Marion, S. R. Sadda, V. Chopra, and L. O. Lee, *Comparison of manual & automated analysis methods for corneal endothelial cell density measurements by specular microscopy,* Journal of Optometry **11**, 182 (2018).

[14] M. O. Price, K. M. Fairchild, and F. W. Price, *Comparison of manual and automated endothelial cell density analysis in normal eyes and DSEK eyes,* Cornea **32**, 567 (2013).

[15] N. Luft, N. Hirnschall, S. Schuschitz, P. Draschl, and O. Findl, *Comparison of 4 specular microscopes in healthy eyes and eyes with cornea guttata or corneal grafts,* Cornea **34**, 381 (2015).

[16] L. Gasser, T. Reinhard, and D. B'ohringer, *Comparison of corneal endothelial cell measurements by two non-contact specular microscopes,* BMC Ophthalmology **15:87** (2015).

[17] A. S. Kitzmann, E. J. Winter, C. B. Nau, J. W. McLaren, D. O. Hodge, and W. M. Bourne, *Comparison of corneal endothelial cell images from a noncontact specular microscope and a scanning confocal microscope,* Cornea **24**, 980 (2005).

[18] F. Scarpa and A. Ruggeri, *Development of a reliable automated algorithm for the morphometric analysis of human corneal endothelium,* Cornea **35**, 1222 (2016).

[19] S. Al-Fahdawi, R. Qahwaji, A. S. Al-Waisy, S. Ipson, M. Ferdousi, R. A. Malik, and A. Brahma, *A fully automated cell segmentation and morphometric parameter system for quantifying corneal endothelial cell morphology,* Computer Methods and Programs in Biomedicine **160**, 11 (2018).

[20] A. Piórkowski and J. Gronkowska-Serafin, *Towards precise segmentation of corneal endothelial cells,* in *Conference on Bioinformatics and Biomedical Engineering (IWBBIO)*, Lecture Notes in Computer Science, Vol. 9043 (Granada, Spain, 2015) pp. 240–249.

[21] B. Selig, K. A. Vermeer, B. Rieger, T. Hillenaar, and C. L. Luengo Hendriks, *Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy,* BMC Medical Imaging **15:13** (2015).

[22] J. P. Vigueras-Guillén, E. R. Andrinopoulou, A. Engel, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Corneal endothelial cell segmentation by classier-based merging of oversegmented images,* IEEE Transactions on Medical Imaging **37**, 2278 (2018).

[23] A. Fabijańska, *Automatic segmentation of corneal endothelial cells from microscopy images,* Biomedical Signal Processing and Control **47**, 145 (2019).

[24] M. C. Daniel, L. Atzrodt, F. Bucher, K. Wacker, S. B´ohringer, T. Reinhard, and D. B´ohringer, *Automated segmentation of the corneal endothelium in a large set of 'real-world' specular microscopy images using the U-net architecture,* Nature Scientific Reports **9**, 4752 (2019).

[25] C. Kolluru, B. A. Benetz, N. Joseph, H. J. Menegay, J. H. Lass, and D. Wilson, *Machine learning for segmenting cells in corneal endothelium images,* in *Proceedings of SPIE, Medical Imaging 2019: Computer-Aided Diagnosis*, Vol. 10950 (San Diego, CA, USA, 2019).

[26] K. Nurzynska, *Deep learning as a tool for automatic segmentation of corneal endothelium images,* Symmetry **10**, 60 (2018).

[27] J. P. Vigueras-Guillén, B. Sari, S. F. Goes, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation,* BMC Biomedical Engineering **1:4** (2019).

[28] C. Angermueller, P´arnamaa, L. Parts, and O. Stegle, *Deep learning for computational biology,* Molecular Systems Biology **12** (2016), 10.15252/msb.20156651.

[29] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: convolutional networks for biomedical image segmentation,* in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. 9351 (2015) pp. 234–241.

[30] J. P. Vigueras-Guillén, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Automatic detection of the region of interest in corneal endothelium images using dense convolutional neural networks,* in *Proceedings of SPIE, Medical Imaging 2019: Image Processing*, Vol. 10949 (San Diego, CA, USA, 2019).

[31] S. Jégou, M. Drozdzal, D. Vázquez, A. Romero, and Y. Bengio, *The one hundred layers tiramisu: fully convolutional DenseNets for semantic segmentation,* in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Honolulu, HI, USA, 2017) pp. 1175–1183.

[32] S. Ioffe, *Batch renormalization: towards reducing minibatch dependence in batch-normalized models,* in *31st Conference on Neural Information Processing Systems (NIPS)* (Long Beach, CA, USA, 2017) pp. 1942–1950.

[33] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, *Fast and accurate deep network learning by exponential linear units (ELUs),* in *International Conference on Learning Representations (ICLR)* (San Juan, Puerto Rico, 2016).

[34] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *Dropout: a simple way to prevent neural networks from overfitting,* Journal of Machine Learning Research **15**, 1929 (2014).

[35] J. P. Vigueras-Guillén, A. Engel, H. G. Lemij, J. van Rooij, K. A. Vermeer, and L. J. van Vliet, *Improved accuracy and robustness of a corneal endothelial cell segmentation method based on merging superpixels,* in *15th International Conference Image Analysis and Recognition (ICIAR)*, Lecture Notes in Computer Science, Vol. 10882 (Póvoa de Varzim, Portugal, 2018) pp. 631–638.

[36] S. Beucher and F. Meyer, *Mathematical morphology in image processing,* (Taylor & Francis Group, 1993) Chap. The morphological approach to segmentation: the watershed transformation, pp. 433–481.

[37] J. P. Vigueras-Guillén, J. van Rooij, H. G. Lemij, K. A. Vermeer, and L. J. van Vliet, *Convolutional neural network-based regression for biomarker estimation in corneal endothelium microscopy images,* in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Berlin, Germany, 2019) pp. 876–881.

[38] T. Dozat, *Incorporating Nesterov momentum into Adam,* in *International Conference on Learning Representations (ICLR) Workshop*, Vol. 1 (San Juan, Puerto Rico, 2016) pp. 2013–2016.

[39] L. W. Hirst, F. L. Ferris, W. J. Stark, and J. A. Fleishman, *Clinical specular microscopy,* Investigative Ophthalmology & Visual Science **19**, 2 (1980).

[40] M. J. Doughty, A. M´uller, and M. L. Zaman, *Assessment of the reliability of human corneal endothelial cell-density estimates using a noncontact specular microscope,* Cornea **19**, 148 (2000).

[41] S. Heinzelmann, M. C. Daniel, P. C. Maier, T. Reinhard, and D. B´ohringer, *Automated cell counting using "deep learning" in donor corneas from organ culture achieves high precision and accuracy,* Klinische Monatsblatter fur Augenheilkunde **236**, 1407 (2019).

[42] J. van Rooij, A. Engel, L. Remeijer, H. van Vleijnenbreugel, and R. Wubbels, *Long-term functional and anatomical outcome after Descemet stripping automated endothelial keratoplasty: a prospective single-center study,* Journal of Ophthalmology **7320816**, 1 (2018).

[43] Y. Feng, H. Q. Qu, J. Ren, P. Prahs, and J. Hong, *orneal endothelial cell loss in femtosecond laser-assisted Descemet's stripping automated endothelial keratoplasty: a 12-month follow-up study,* Chinese Medical Journal **130**, 2927 (2017).

[44] S. Graffi, P. Leon, Y. Nahum, S. Gutfreund, R. Spena, L. Mattioli, and M. Busin, *Outcomes of ultrathin Descemet stripping automated endothelial keratoplasty (UT-DSAEK) performed in eyes with failure of primary Descemet membrane endothelial keratoplasty (DMEK),* British Journal of Ophthalmology **103**, 599 (2019).

[45] H. van Cleynenbreugel, L. Remeijer, and T. Hillenaar, *Descemet stripping automated endothelial keratoplasty: effect of intraoperative lenticule thickness on visual outcome and endothelial cell density,* Cornea **30**, 1195 (2011).

[46] F. P. Guerra, A. Anshu, M. O. Price, and F. W. Price, *Endothelial keratoplasty: fellow eyes comparison of Descemet stripping automated endothelial keratoplasty and Descemet membrane endothelial keratoplasty,* Cornea **30**, 1382 (2011).

[47] M. A. Javadi, S. Feizi, R. Jafari, S. B. Hosseini, and S. Safapour, *Factors influencing graft endothelial cell density after Descemet stripping automated endothelial keratoplasty,* Journal of Ophthalmic and Visual Research **13**, 10 (2018).

[48] M. J. Doughty, *A prospective analysis of corneal endothelial polymegethism and cell density in young adult Asians,* Clinical and Experimental Optometry **97**, 256 (2014).

[49] M. J. Doughty and B. M. Aakre, *Further analysis of assessments of the coefficient of variation of corneal endothelial cell areas from specular microscopic images,* Clinical and Experimental Optometry **91**, 438 (2008).

[50] M. J. Doughty and D. Fonn, *Pleomorphism and endothelial cell size in normal and polymegethous human corneal endothelium,* International Contact Lens Clinic **20**, 116 (1993).

8

NOTES ON THE BAERVELDT STUDY

This chapter summarizes the role of the fully-automated cell segmentation framework on the Baerveldt study. Some parts are based on the manuscript:

E. Islamaj, P.W.T. de Waard, T.J.P. Nieuwendijk, J.G. Bollemeijer, **J.P. Vigueras-Guillén**, J. van Rooij, K.A. Vermeer, and H.G. Lemij, *Baerveldt glaucoma drainage device implantation affects the corneal endothelial cells*, Manuscript submitted for publication.

8.1. INTRODUCTION

Glaucoma is an eye condition that damages the optic nerve, often caused by an elevated intraocular pressure (IOP). In a healthy eye, IOP is kept within an acceptable range, which is controlled by the production of the aqueous humor in the posterior chamber and its drainage in the anterior chamber. If the aqueous humor cannot be drained at a proper rate, IOP increases, compromising the optic nerve. Glaucoma does not usually cause any symptoms in the early stages, and it tends to develop slowly over many years [1]. If not treated early, it might result in an irreversible vision loss. Initial treatments involve eye drops to reduce the IOP and laser therapy to open the blocked drainage system, but if those fail, surgery (trabeculectomy) or the implantation of a permanent drainage device are performed. The drainage devices (Baerveldt, Ahmed, Molteno, etc.), which have a tube that reaches the anterior chamber, are designed to divert aqueous humor to an external reservoir. However, the proximity of the tube to the corneal endothelium is linked to endothelial cell loss [2]. While the cause has not been fully determined yet, several hypotheses exist: chronic inflammation, tube-cornea contact, and aqueous flow changes [3].

Between January 2015 and September 2019, a clinical study to analyze the impact of a Baerveldt glaucoma implant (BGI) in the eye was conducted in The Rotterdam Eye Hospital (Rotterdam, The Netherlands) (trial registration NL4805, registered 15-12-2014). The study involved two main research questions: the impact of the BGI on (i) the eye motility and (ii) the corneal endothelium. In regards to the latter, specular microscopy images were acquired and the fully automated cell segmentation algorithm presented in Chapter 7 was used to estimate the corneal parameters. Subsequently, these were employed in the clinical analysis of the study, presented by Islamaj et al. in [4].

This chapter briefly discusses the impact of the proposed algorithm on the study, highlighting the obtained improvement (in accuracy and cell detection) with respect to using the estimates provided by the microscope's software. Furthermore, it discusses the clinical findings regarding the three corneal parameters, commenting on the patterns depicted through time.

8.2. DATASET

Two hundred and four (204) patients (average age 66 ± 10 years) scheduled for a BGI were included in the aforementioned clinical study, which had a 2-year follow-up. Among those, 192 patients completed the study. The eyes were imaged with anterior segment optical coherence tomography, AS-OCT (SS-1000, CASIA, Tomey, Nagoya, Japan), to measure the implant's tube length, angle, and distance to the cornea, and the central (CE) and temporal-superior (TS) corneal endothelium –closer to the implant's tube– were imaged with specular microscopy (Topcon SP-1P, Topcon Co., Tokyo, Japan). Specular images were acquired before implantation and at 3, 6, 12, and 24 months after implantation. The protocol indicated to take five specular images in each area for each visit; however, it was common to have difficulties to reach that number of gradable images even after several attempts, particularly for the TS cornea. On average, 4.7 images per visit were obtained from the CE cornea, whereas only 3.6 images per visit were acquired

from the TS cornea. In total, 7975 images were collected.

8.3. ACCURACY OF THE FULLY-AUTOMATIC METHOD

Four hundred images from the first 100 patients (four images per patient) were manually annotated, with a ratio of 47%/53% between CE/TS images. Images were selected based on properties such as noise, blurriness, cell morphology, etc., so that a diverse annotated dataset could be created. Through this selection process, complex images were overrepresented in the annotated dataset.

The fully-automatic method presented in Chapter 7 was applied to these images as described in Section 7.3.1, and the error as a function of the number of cells was already depicted in Figure 7.5-D–F. The mean absolute error was computed for both Topcon's and our estimates (Table 8.1). The benefits of using our proposed method instead of Topcon's software were the following:

(1) Our method detected almost twice the number of cells compared to Topcon. As discussed in the previous chapter, the estimation error decreases with an increasing number of detected cells.
(2) Topcon failed to detect any cell in 6.5% of the images, and more than 20% of Topcon's segmentations did not have enough cells to assess the hexagonality. In contrast, our method performed properly in all cases: cells were detected in every image and in only three images an insufficient number of cells were detected to estimate the hexagonality, but this was also the case in the manual annotations.
(3) Our estimation error was less than one third of Topcon's error for all parameters, thus allowing for a more accurate analysis of the corneal progression.

The individual discrepancies between the manual and our estimates were very low: mean absolute percentage error (MAPE) was 3.7% for CV, 3.5% for HEX, and only 1.2% for ECD. Part of this error could be explained to a certain extent by the cell variability: as depicted in Figure 7.5, the error for the images with 200 or more cells (the case of this subset of images) have a similar magnitude as the number of cell increases, which suggests that the cell variability dominates the magnitude of the error at that detection rate.

Table 8.1: The mean absolute error (MAE, computed as $\frac{1}{n}\sum_{i=1}^{n}|E_i - A_i|$, where A_i is the gold standard value, E_i is the estimated value, and n is the number of images) and the standard deviation of the endothelial parameters for Topcon and our method in the 400 annotated images. Only the cases with estimates were used to obtain the error (% indicated in 4th column). Average number of cells per image in 5th column (the manual annotations had 200 ± 79 cells per image).

	ECD MAE [cells/mm^2]	CV MAE [%]	HEX MAE [%]	% images with estimates (ECD/CV, HEX)	Number of cells per image
Our method	20.6 ± 45.4	1.1 ± 2.3	1.8 ± 2.2	100%, 99.3%	209 ± 79
Topcon	114.3 ± 108.3	3.7 ± 4.3	7.3 ± 7.7	93.5%, 79.0%	107 ± 71

8.4. CLINICAL FINDINGS

All the images in the dataset were processed with the proposed method. Islamaj et al. [4] performed the clinical analysis using only the specular image with the most detected cells in each session and corneal area and discarding the remaining ones. No significant differences were observed if the average values among all the acquired images were used instead (Table 8.2). Particularly, we would have observed a mean absolute percentage difference between those two options of 1.4%, 0.3%, and 0.5% for in ECD, CV, and HEX, respectively.

The main findings of the clinical analysis performed by Islamaj et al. [4], which focused on ECD, were:

- Before BGI implantation, the cell density was similar in both corneal areas. Two years after implantation, the cell density in the peripheral area was considerably smaller than in the central area (Table 8.2).
- The rate of cell loss per annum (2.6% in CE and 11.0% in TS) was larger than the one observed in healthy corneas (0.6% in central area [5]).
- There was a statistically significant relation between the tube characteristics and the ECD loss. Particularly, ECD loss in the central cornea was statistically significantly related to two tube characteristics (angle and length), whereas ECD loss in the periphery was statistically significantly related to all three tube characteristics (angle, length, and distance). As indicated by Islamaj et al. [4], similar (or higher) loss rates were reported by other studies.

Regarding the other corneal parameters, CV depicted a significant decrease in the periphery but remained stable in the central cornea, and HEX did not depict any clear progression, remaining stable within a limited range in both areas (Table 8.2). However, it is worth noting that, at baseline, both parameters were significantly different between the areas. Indeed, CV and HEX in the central cornea at baseline were within the range of (what is considered) healthy corneas (HEX above 60% [6] and CV below 30% [7]), but the values in the periphery were out of that range. Since all studies in pleomorphism and polymegethism have been done with measurements from the central cornea, it is not

Table 8.2: The average (± SD) corneal parameters from **our proposed** estimates, measured before implantation (baseline) and up to 2-year follow-up, computed by considering (for each patient and session) the average values obtained among all the acquired images.

		Baseline	**3 months**	**6 months**	**1 year**	**2 years**
CE cornea	ECD [cells/mm^2]	2198 ± 410	2160 ± 430	2176 ± 408	2146 ± 427	2085 ± 454
	CV [%]	28.0 ± 4.4	28.1 ± 5.3	27.7 ± 4.5	27.7 ± 5.5	27.9 ± 6.5
	HEX [%]	61.4 ± 7.0	60.8 ± 6.9	61.5 ± 6.7	61.6 ± 6.5	61.6 ± 7.4
TS cornea	ECD [cells/mm^2]	2213 ± 554	2151 ± 569	2122 ± 582	1984 ± 594	1755 ± 648
	CV [%]	34.2 ± 8.2	32.4 ± 6.7	31.7 ± 6.8	31.0 ± 6.4	29.8 ± 6.7
	HEX [%]	57.7 ± 7.7	57.6 ± 7.7	58.3 ± 7.7	58.0 ± 7.6	56.8 ± 9.3

possible to ascertain whether those values are abnormal, but it is most plausible that the peripheral cornea simply has different tissue properties in a stable state.

From the point of view of the assessment tool, these observations highlight that it is key to obtain accurate estimates from peripheral images, as the phenomenon to observe is occurring in that area. In contrast with Topcon's software, our method was able to segment images of both corneal areas equally well. However, the main obstacle was the existence of less discernible cells in TS images. Specifically, there were 20% less visible cells in TS than in CE images, which was observed in both the manual annotations and our assessments, whereas Topcon detected 34% less cells in TS than in CE images.

Other limitations in Topcon's software were the lack of any estimates in some sessions (considering that a session involved to take up to five images per area and one single image with estimates would suffice). Specifically:

- In the CE area, our algorithm obtained estimates in ECD/CV in 1.2% more sessions and estimates in HEX in 5% more sessions than Topcon's software.
- In the TS area, our algorithm obtained estimates in ECD/CV in 8.3% more sessions and estimates in HEX in 20% more sessions than Topcon's software.

This particularly highlighted the limitations of Topcon's software in the TS area. Nevertheless, Topcon's algorithm was able to show a similar pattern in the biomarkers over the months (Table 8.3) than our method. Some interesting observations were that Topcon's software depicted (i) an overestimation in ECD, (ii) an underestimation and a larger fluctuation over the months in HEX, (iii) a larger fluctuation over the months in CV in the CE area, and (iv) a smaller decrease over the months in CV in the TS area (a -2.9 versus our estimated -4.4). These larger fluctuations in CV and HEX, along with the higher error rate in the subset of annotated images (Table 8.1), seemed to suggest a low reliability in Topcon's software for the estimations of CV and HEX.

Overall, our proposed fully-automatic method provided such a high accuracy that the distinctive patterns in peripheral CV and HEX were clearly observable. Whether these observations can have an actual impact in the clinic or in clinical research is still something that needs further research.

Table 8.3: The average (± SD) corneal parameters from **Topcon**'s estimates, measured before implantation (baseline) and up to 2-year follow-up, computed by considering (for each patient and session) the average values obtained among all the acquired images.

		Baseline	**3 months**	**6 months**	**1 year**	**2 years**
CE cornea	ECD [cells/mm^2]	2299 ± 399	2278 ± 190	2283 ± 408	2244 ± 428	2203 ± 455
	CV [%]	28.8 ± 3.9	29.6 ± 4.4	28.6 ± 4.2	28.5 ± 4.5	28.3 ± 4.7
	HEX [%]	58.1 ± 7.5	55.9 ± 9.2	58.0 ± 8.5	59.1 ± 8.4	60.2 ± 9.0
TS cornea	ECD [cells/mm^2]	2363 ± 584	2311 ± 580	2291 ± 587	2139 ± 560	1969 ± 618
	CV [%]	33.6 ± 7.0	32.8 ± 6.3	31.4 ± 5.8	31.3 ± 5.6	30.7 ± 5.6
	HEX [%]	54.0 ± 9.2	53.0 ± 11.3	55.8 ± 10.7	55.5 ± 9.9	56.5 ± 9.6

REFERENCES

[1] United Kingdom National Health Service (NHS), *Glaucoma (Overview),* https://www.nhs.uk/conditions/glaucoma/ (2021), [Online; accessed on 20 January 2021].

[2] M. S. Kim, K. N. Kim, and C.-S. Kim, *Changes in corneal endothelial cell after Ahmed glaucoma valve implantation and trabeculectomy: 1-year follow-up,* Korean Journal of Ophthalmology **30**, 416 (2016).

[3] M. L. McDermott, R. P. Swendris, D. H. Shin, M. S. Juzych, and J. W. Cowden, *Corneal endothelial cell counts after Molteno implantation,* American Journal of Ophthalmology **115**, 93 (1993).

[4] E. Islamaj, P. W. T. de Waard, T. J. P. Nieuwendijk, J. G. Bollemeijer, J. P. Vigueras-Guillén, J. van Rooij, K. A. Vermeer, and H. G. Lemij, *Baerveldt glaucoma drainage device implantation affects the corneal endothelial cells,* Manuscript submitted for publication (2020).

[5] W. M. Bourne, L. R. Nelson, and D. O. Hodge, *Central corneal endothelial cell changes over a ten-year period,* Investigative Ophthalmology & Visual Science **38**, 779 (1997).

[6] M. J. Doughty and D. Fonn, *Pleomorphism and endothelial cell size in normal and polymegethous human corneal endothelium,* International Contact Lens Clinic **20**, 116 (1993).

[7] M. J. Doughty, *A prospective analysis of corneal endothelial polymegethism and cell density in young adult Asians,* Clinical and Experimental Optometry **97**, 256 (2014).

9

Conclusions

The corneal endothelium, a non-regenerative layer of cells controlling the state of corneal hydration, is the most critical tissue of the cornea. Quantifying its health status is not only important to diagnose and treat certain corneal diseases but also relevant to the execution and evaluation of many eye surgeries.

By means of specular or confocal microscopy, the endothelium can be visualized and evaluated in vivo, in a non-invasive manner. The estimation of the corneal endothelium parameters, particularly cell density, provides a valuable input for disease diagnosis, prognosis, and monitoring. Manual estimation, which requires cell segmentation, is time consuming and tedious.

In this thesis, we have presented and evaluated several automatic techniques for the segmentation and quantification of the corneal endothelium.

9.1. TECHNICAL DEVELOPMENTS

In **Chapter 2**, we introduced a machine learning method that aimed for an accurate cell segmentation by first generating oversegmentation and then identifying and merging the superpixels that form whole cells. The oversegmentation method was based on stochastic watershed (an adaptation of Selig et al.'s work [1]) and the merging method was based on support vector machines. This approach generated very accurate cell segmentation in different types of images (confocal, specular, and phase-contrast microscopy), achieving state-of-the-art performance. In the case of our dataset (Baerveldt study, 30 images), the proposed approach was significantly better than the microscope manufacturer's software (Table 9.1).

In **Chapter 3**, we further evaluated the methodology presented in Chapter 2. First, it was proven that the degree of oversegmentation proposed in Chapter 2 was indeed optimal. Second, it was shown how the merging method was beneficial even when applied to an *a priori* optimized segmentation, as it detected and removed any remaining false edge. This would make the merging method a good candidate to refine the solution of any other segmentation algorithm. Finally, we presented a new, more accurate methodology to estimate the average cell size from Fourier analysis, which was necessary in the oversegmentation to set the seeds in the stochastic watershed.

In **Chapter 4**, we proposed a completely new approach to segment the endothelial images. Two deep learning methodologies based on Convolutional Neural Networks (CNNs) were presented: either by pixel classification or by direct segmentation. For the classification network, a patch approach was used, where the CNN would classify the center-pixel of a patch. This would require extracting as many patches as pixels in the image, which was not optimal computationally. For the segmentation network, the U-net was employed, a recent architecture where the encoding and decoding parts of the network were short-connected. This network proved to be better than the patch approach, quantitatively, qualitatively, and computationally. Furthermore, the U-net obtained better accuracy than the "oversegmentation + merging" method from Chapter 2 (Table 9.1) and it was notably faster: segmenting an image took the U-net a fraction of a second, whereas the merging method took 20-40 seconds. Thus, the U-net became the selected methodology from this point forward. Nonetheless, the merging method was

still used in the subsequent Chapters 5 & 6 as a refinement method given the findings in Chapter 3.

This chapter also proposed a method to transform the CNN output (probability images) into binary images, which was necessary to estimate the corneal parameters. This postprocessing was not trivial, as edges in the CNN output could be blurry or not completely connected. The proposed solution made use of all the previous work in watershed and Fourier analysis.

In **Chapter 5**, a key aspect to make the method fully automatic was addressed: the selection of the region of interest (ROI). This would be necessary for images with large areas without visible cells and/or out of focus. To this goal, a Dense U-net architecture was proposed, whose input was the edge probability images. This chapter discussed how the ROI selection (a conceptually simple task) caused significant network stability problems, which were solved by employing dense connections in the CNN. In this chapter, the fully-automatic framework was presented for the first time, which included the estimation of the edge image (CNN-Edge, Chapter 4), the detection of the ROI (CNN-ROI, this chapter), the postprocessing (Chapters 3 & 4, updated to apply the ROI to the edge images), and the refinement method (Chapter 2). In comparison with Topcon's software, the error yielded by the fully-automatic method in the corneal parameters (Baerveldt study, 140 images) was 8, 4, and 9 times smaller for ECD, CV, and HEX, respectively, although the most complex images (from DSAEK study) were still to be tested in Chapter 7.

In **Chapter 6**, we explored a different approach to solve the specular images: to use a single CNN to directly estimate the corneal parameters. Considering the limitations of the fully-automatic method, where each part had their own drawbacks, a single CNN designed for regression would simplify the methodology. The large annotated dataset at this point (738 images, plus augmentation) allowed for this experiment. The results showed that inferring the parameters directly from the intensity images was still unfeasible. However, inferring them from the edge images (U-net output) was possible, although the accuracy was not better than simply applying the postprocessing to the edge images. In a second experiment, the CNN-regressor was considered as a tool to improve the postprocessing by estimating a specific smoothing factor for each image. Alternatively, the chapter also explored estimating that factor by Fourier analysis. How-

Table 9.1: The mean absolute error (MAE, computed as $\frac{1}{n}\sum_{i=1}^{n}|E_i - A_i|$, where A_i is the gold standard value, E_i is the estimated value, and n is the number of images) and the standard deviation of the endothelial parameters for only the 30 images presented in Chapter 2 (Baerveldt dataset), either solved by Topcon's software, by the "oversegmentation + merging" method (Chapter 2), or by the CNN for segmentation (U-net, Chapter 4). Average number of cells per image in 4th column (the manual annotations had 259 ± 38 cells per image, same as estimated by the U-net).

	ECD MAE [cells/mm^2]	**CV MAE [%]**	**HEX MAE [%]**	**Number of cells**
Topcon	26.7 ± 33.3	1.9 ± 1.5	4.6 ± 5.2	186 ± 45
Merging (Chapter 2)	16.0 ± 8.8	1.1 ± 1.0	2.6 ± 2.9	251 ± 36
Unet (Chapter 4)	6.4 ± 5.7	0.6 ± 0.5	0.8 ± 0.8	259 ± 38

ever, both ideas provided unsuccessful results. In summary, this chapter discarded the idea of a single CNN for inference and helped to consolidate the fully-automatic method (from Chapter 5) as a proficient approach.

In **Chapter 7**, the fully-automatic framework was updated with some modifications and it was applied to the DSAEK study. The main change was the removal of the refinement method, as this became not only the computational bottleneck but was also counterproductive in some images. Within the CNN section, minor changes were made to improve the overall performance; these modifications included adding new convolutional layers (or changing them), but the overall architecture remained as presented in Chapter 5. The final framework proved to be very accurate and it significantly outperformed Topcon's software.

In **Chapter 8**, no further technical improvements were introduced, and the fully-automatic method was simply applied to the Baerveldt study.

9.2. IMPLICATIONS OF CLINICAL RESEARCH

In **Chapter 7**, the proposed method estimated the corneal parameters in practically all images of the DSAEK dataset (images with extensive FED were still problematic to solve) and with almost the same accuracy as if they were done manually. Given the complexity of our post-keratoplasty dataset, the performance was remarkable. Chapter 7 also showed the importance of detecting as many cells a possible in an image in order to reduce the method's error. However, it is still unclear what percentage of the error is based on segmentation mistakes and what percentage is due to the cell variability, which is not an error but an inherent uncertainty that should be considered. Further studies should be done in order to answer that question.

In both studies, DSAEK (**Chapter 7**) and Baerveldt (**Chapter 8**), the proposed method achieved accurate estimates for CV and HEX, which are two biomarkers barely addressed in the current literature due to the lack of suitable tools to estimate them. The present work opens new opportunities to revise old studies –and to design new clinical studies– with them into consideration.

Regarding the patterns of the corneal parameters, CV and HEX in the DSAEK study showed a positive progression as the ECD loss was decelerating, which suggested that the tissue was stabilizing and recovering from the surgical trauma. In contrast, the Baerveldt study depicted a similar and (*a priori*) positive progression in peripheral CV and HEX as the ECD loss was accelerating, hence suggesting that the external trauma was still persisting. This highlights that it is difficult to have a whole understanding of the state of the tissue if only ECD is considered.

9.3. GENERAL DISCUSSION & FUTURE DIRECTIONS

This thesis has described several methods to estimate the corneal endothelium parameters, with two major methodologies.

The first one, "oversegmentation + merging" (Chapters 2–3), was already an improve-

ment with respect to the state of the art. However, the success of the merging method was highly dependent upon the accuracy of the oversegmentation, which became the weakest part of the process. Indeed, the merging process was highly accurate as long as the edges were correctly delineated, regardless the number of superpixels needed to be merged (proven in Chapter 3). In contrast, a wrongly delineated edge from an oversegmented cell (even just a small portion of the edge) could hinder the identification of those superpixels as the fragments that form a cell and, thus, the expected merger could not happen. Another disadvantage of this method was the high time cost due to the need of evaluating every possible combination of superpixels in the image. If this line of research would have been continued, more emphasis would have been made in improving the oversegmentation method (i.e., refining the edges before starting the merging process) and in reducing the time cost (i.e., only evaluating superpixels which clearly were not whole cells).

The second methodology, segmentation by CNNs (Chapters 4–7), made a large improvement in the performance. Endothelial images from subjects with rather healthy corneas could be segmented with practically perfect accuracy, images from diseased corneas could be segmented rather satisfactorily (and considerably better than the current state of the art), and CNNs enabled full automation of the process by selecting the relevant area within the images. This work has shown that CNNs are faster (and sometimes better) than humans to detect cell edges in the presence of moderate blurriness, inferring cells in areas where our annotator had difficulties and doubts. Another benefit of CNNs is its quick computational time due to the use of graphics processing units (GPUs) to perform the convolutional operations, otherwise model training would have been unattainable within a reasonable time-frame.

Future work should address the most challenging images (e.g., with guttae, with a large number of artifacts, or extremely blurred). In these cases, there are many areas where the endothelial cells are not visible. In the case of small blurred areas, a person is sometimes able to observe the surroundings and infer the most plausible tessellation in that area. If a machine learning method could mimic this human cognitive process, it could be applied to images with Fuchs' endothelial dystrophy (FED), which is a current unsolved problem in the literature.

Nevertheless, we believe that, with the use of the presented (or similar) methodology, future studies can address the progression of the corneal parameters in many pathological corneas that currently are not been studied because of lack of reliability in automatic segmentation methods or because manual annotations is too time-consuming to consider. In this respect, most studies referenced along this thesis either performed manual annotations (usually designing studies with a low number of patients to make the annotation process feasible in time) or relied on the values provided by their respective microscope's software, in which case the accuracy was unknown, which might raise questions on the reliability of the results.

To make that scenario even more attractive, the presented CNN segmentation method could be easily implemented along with a simple graphic user interface that would allow the user to correct the very few segmentation mistakes that still persist. This would bring a higher certainty of reaching an accurate biomarker estimation.

Hopefully, the tool presented in this thesis can be employed in many large clinical studies, bringing small but meaningful changes to clinical practice in the future.

REFERENCES

[1] B. Selig, K. A. Vermeer, B. Rieger, T. Hillenaar, and C. L. Luengo Hendriks, *Fully automatic evaluation of the corneal endothelium from in vivo confocal microscopy,* BMC Medical Imaging **15:13** (2015).

SUMMARY

Trauma, eye diseases, or eye surgery can compromise the endothelium of the cornea and its function, which may lead to loss of transparency and, ultimately, the necessity for corneal transplantation. Non-contact *in vivo* imaging of the endothelial cell layer allows the assessment of its health status. This thesis presents methods based on advanced image processing, machine learning, and deep learning for the estimation of image-based biomarkers to characterize the cell architecture of the corneal endothelium.

Because of the important role that the corneal endothelium plays in human vision, it is clinically relevant to image it and quantify its health status. Thanks to specular microscopy, the endothelium can be imaged *in vivo* in the clinic, quickly and in a noninvase manner. The specular images could then be segmented in order to estimate several corneal parameters (endothelial cell density [ECD], hexagonality of the cells [HEX], and cell size variation [CV]), which are used to assess the state of the tissue. Because manual segmentation is very tedious and time-consuming, there is a need for automatic tools, particularly in clinical studies with a large number of subjects. However, current segmentation tools are not accurate enough and make significant mistakes, or are unable to detect enough cells. This is due to the poor quality of the specular images, which tend to have low contrast and many noisy artifacts, particularly for pathological corneas.

Initially, a machine learning approach based on Support Vector Machines was proposed (Chapters 2 & 3), which exploited the idea of merging superpixels. Specifically, the method would start by generating an oversegmented image comprised of superpixels followed by a merging process that would evaluate all possible combinations of two and three superpixels, merging those that would form a whole cell. However, a different methodology based on Convolutional Neural Networks (CNN, Chapter 4) provided significantly better performance, thus becoming the preferred option. This methodology used a deep learning network named U-net, which was capable of inferring the cell segmentation in a fraction of a second with very high accuracy. The framework was then extended with another CNN to select the areas in the image where cells were correctly delineated (Chapter 5), which allowed for a fully-automatic approach. Given the great performance of CNNs to solve this problem, the idea of using a single CNN to estimate the corneal parameters directly was also explored (Chapter 6). However, the performance degraded significantly, therefore consolidating the approach of using a segmentation tool to solve the problem.

In this thesis, we have presented a fully-automatic method based on deep learning that segments the specular images and estimates the corneal parameters. The ultimate goal was to apply such methodology in two clinical studies from the Rotterdam Eye Hospital. The first study dealt with the transplantation of the cornea, following the recovery of 41 patients during the first year after surgery. The second study dealt with the implan-

tation of a drainage device in 192 glaucomatous eyes, with the purpose of studying the effect of the device's tube on the endothelium during the 2 years after implantation.

The fully-automatic deep learning-based method was applied to all images in both studies. In the transplantation study (Chapter 7), the proposed method was able to estimate the corneal parameters in practically all images, resulting in an estimation error that was more than three times smaller than the one from the microscope's software, and the number of correctly segmented cells per image was three times higher in the proposed method. This was a remarkable improvement. Similarly, the method was applied to the glaucoma study (Chapter 8), achieving even better performance since these images had significantly better quality. CV and HEX, two biomarkers barely used in clinical studies due to the difficulties to automatically estimate them, were now obtained with high accuracy, and they depicted peculiar patterns during the healing process in both studies. The clinical relevance of such patterns is still something that needs further research. Nonetheless, the work presented here opens new opportunities to study the corneal parameters in clinical studies with a high number of patients and follow-ups, where manual annotations are not a feasible solution.

SAMENVATTING

Trauma, oogaandoeningen of oogchirurgie kunnen het endotheel van het hoornvlies en zijn functie aantasten, wat kan leiden tot verlies van transparantie en uiteindelijk tot de noodzaak van een hoornvliestransplantatie.

De beoordeling van de gezondheidsstatus van de endotheelcellaag wordt vergemakkelijkt door *in vivo*, contactloze beeldvorming. Contactloos in vivo afbeelding van het de endotheel cellaag maakt het mogelijk om de gezondheid hiervan te beoordelen. Dit proefschrift presenteert methoden die zijn gebaseerd op geavanceerde beeldverwerking, lerende systemen en diep leren voor het schatten van biomarkers om de celarchitectuur van het hoornvliesendotheel te karakteriseren.

Vanwege de belangrijke rol die het hoornvliesendotheel speelt in het menselijk gezichtsvermogen is het klinisch relevant om het in beeld te brengen en de gezondheidstoestand ervan te kwantificeren. Dankzij reflectiemicroscopie kan het endotheel in de kliniek *in vivo* snel en op niet-invasieve wijze in beeld worden gebracht. De reflectiebeelden kunnen vervolgens worden gesegmenteerd om verschillende corneale parameters te schatten (endotheliale celdichtheid [ECD], hexagonaliteit van de cellen [HEX] en celgroottevariatie [CV]) die worden gebruikt om de toestand van het weefsel te beoordelen. Omdat handmatige segmentatie erg vervelend en tijdrovend is, er is behoefte aan automatisch gereedschap, vooral bij klinische onderzoeken met een groot aantal proefpersonen. De huidige segmentatietools zijn echter niet nauwkeurig genoeg en maken aanzienlijke fouten, of kunnen niet genoeg cellen detecteren. Dit komt door de slechte kwaliteit van de reflectiebeelden, die meestal een laag contrast en veel ruisartefacten hebben, vooral bij pathologische hoornvliezen.

Aanvankelijk werd een lerend systeem aanpak op basis van steunvectormachines voorgesteld (Hoofdstukken 2 & 3), waarbij gebruik werd gemaakt van het idee om superpixels samen te voegen. Specifiek zou de methode beginnen met het genereren van een overgesegmenteerde afbeelding die bestaat uit superpixels, gevolgd door een samenvoegproces dat alle mogelijke combinaties van twee en drie superpixels zou evalueren en die combinaties zou samenvoegen die een hele cel zouden vormen. Een andere methodologie gebaseerd op convolutionele neurale netwerken (CNN, Hoofdstuk 4) leverde echter significant betere prestaties, waardoor het de voorkeursoptie werd. Deze methodologie maakte gebruik van een dieplerend-netwerk genaamd U-net, dat in staat was om de celsegmentatie in een fractie van een seconde met een zeer hoge nauwkeurigheid af te leiden. Het raamwerk werd vervolgens uitgebreid met een andere CNN om de gebieden in de afbeelding te selecteren waar cellen correct waren afgebakend (Hoofdstuk 5), wat een volledig automatische benadering mogelijk maakte. Gezien de geweldige prestaties van CNN's om dit probleem op te lossen, werd ook het idee onderzocht om een enkel CNN te gebruiken om de hoornvliesparameters direct te schatten (Hoofdstuk

6). De prestaties gingen echter aanzienlijk achteruit, waardoor de benadering van het gebruik van een segmentatietool om het probleem op te lossen werd geconsolideerd.

In dit proefschrift hebben we een volledig automatische methode gepresenteerd op basis van diep leren die de spiegelbeelden segmenteert en de parameters van het hoornvlies schat. Het uiteindelijke doel was om een dergelijke methodiek toe te passen in twee klinische onderzoeken van Het Oogziekenhuis Rotterdam. De eerste studie had betrekking op de transplantatie van het hoornvlies en volgde het herstel van 41 patiënten tijdens het eerste jaar na de operatie. De tweede studie had betrekking op de implantatie van een drainage-implant in 192 ogen met glaucoom, met als doel het effect van het buisje van het hulpmiddel op het hoornvliesendotheel te bestuderen gedurende de 2 jaar na implantatie.

Ten slotte werd de volledig automatische op dieplerende netwerken toegepast op alle afbeeldingen in beide onderzoeken. In de transplantatiestudie (Hoofdstuk 7) was de voorgestelde methode in staat om de hoornvliesparameters in praktisch alle afbeeldingen te schatten, wat resulteerde in een schattingsfout die meer dan drie keer kleiner was dan die van de software van de microscoop, en het aantal correct gesegmenteerde cellen per afbeelding was drie keer hoger in de voorgestelde methode. Dit was een opmerkelijke verbetering. Op dezelfde manier werd de methode toegepast in de glaucoomstudie (Hoofdstuk 8), waarbij nog betere prestaties werden behaald omdat deze beelden een significant betere kwaliteit hadden. CV en HEX, twee biomarkers die nauwelijks werden gebruikt in klinische onderzoeken vanwege de moeilijkheden om ze automatisch te schatten, werden nu met hoge nauwkeurigheid verkregen en ze toonden in beide onderzoeken eigenaardige patronen tijdens het genezingsproces. De klinische relevantie van dergelijke patronen is nog iets dat nader onderzoek behoeft. Desalniettemin biedt het hier gepresenteerde werk nieuwe mogelijkheden om de cornea-parameters te bestuderen in klinisch onderzoek met een groot aantal patiënten en nacontroles, waar handmatige annotaties geen haalbare oplossing bieden.

List of Publications

Publications contributing to this thesis

7. E. Islamaj, P.W.T. de Waard, T.J.P. Nieuwendijk, J.G. Bollemeijer, **J.P. Vigueras-Guillén**, J. van Rooij J, K.A. Vermeer, and H.G. Lemij, *Baerveldt glaucoma drainage device implantation affects the corneal endothelial cells*, In Review.

6. **J.P. Vigueras-Guillén**, J. van Rooij, A. Engel, H.G. Lemij, L.J. van Vliet, and K.A. Vermeer, *Deep learning for assessing the corneal endothelium from specular microscopy images up to one year after ultrathin-DSAEK surgery*, Translational Vision Science & Technology **9**(2):49 (2020).

5. **J.P. Vigueras-Guillén**, J. van Rooij, H.G. Lemij, K.A. Vermeer, and L.J. van Vliet, *Convolutional neural network-based regression for biomarker estimation in corneal endothelium microscopy images*, 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 876–881, Berlin, Germany (2019).

4. **J.P. Vigueras-Guillén**, H.G. Lemij, J. van Rooij, K.A. Vermeer, and L.J. van Vliet, *Automatic detection of the region of interest in corneal endothelium images using dense convolutional neural networks*, Proceedings of SPIE, Medical Imaging 2019: Image Processing **10949**, 1094931, San Diego, CA, USA (2019).

3. **J.P. Vigueras-Guillén**, B. Sari, S.F. Goes, H.G. Lemij, J. van Rooij, K.A. Vermeer, and L.J. van Vliet, *Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation*, BMC Biomedical Engineering **1:4** (2019).

2. **J.P. Vigueras-Guillén**, A. Engel, H.G. Lemij, J. van Rooij, K.A. Vermeer, and L.J. van Vliet, *Improved accuracy and robustness of a corneal endothelial cell segmentation method based on merging superpixels*, 15th International Conference on Image Analysis and Recognition (ICIAR), Póvoa de Varzim, Portugal. Lecture Notes in Computer Science **10882**, pp. 631–638 (2018).

1. **J.P. Vigueras-Guillén**, E.R. Andrinopoulou, A. Engel, H.G. Lemij, J. van Rooij, K.A. Vermeer, and L.J. van Vliet, *Corneal endothelial cell segmentation by classifier-driven merging of oversegmented images*, IEEE Transactions on Medical Imaging **37**, 10, pp. 2278–2289 (2018).

Other publications

4. A. Engel, R. Wubbels, T. van Goor, L. Remeijer, A. Geerards, **J.P. Vigueras-Guillén**, and J. van Rooij, *Predicting the need for combined phacoemulsification and endothelial keratoplasty in eyes with both Fuchs' endothelial dystrophy and cataract*, In Review.

3. A. Engel, J. van Rooij, P. Steijger-Vermaat, A. Molenaar, **J.P. Vigueras-Guillén**, and R. Wubbels, *Donor thickness after Ultrathin Decemet Stripping Automated Endothelial Keratoplasty (UT-DSAEK) and its relation to postoperative visual acuity and preoperative lamella measures*, In Review.

2. J.H. de Jong, **J.P. Vigueras-Guillén**, R.J. Wubbels, R. Timman, K.A. Vermeer, and J.C. van Meurs, *The influence of prolongation of interruptions of preoperative posturing and other clinical factors on the progress of macula-on retinal detachment*, Ophthalmology Retina **3**, 11, pp. 938–946 (2019).

1. J.H. de Jong, **J.P. Vigueras-Guillén**, T. Simon, R. Timman, K.A. Vermeer, and J.C. van Meurs, *Preoperative posturing of patients with macula-on retinal detachment retards progression towards the fovea*, Ophthalmology **124**, 10, pp. 1510–1522 (2017).

ACKNOWLEDGEMENTS

Pursuing this PhD degree was a challenging but gratifying task, made possible with the help of many great people.

First, I would like to thank my promotor Prof. Lucas van Vliet and co-promotor Dr. Koen Vermeer for giving me the opportunity to work on this project, their support, and patient supervision. I was lucky and honored to be under your guidance.

To Dr. Jeroen van Rooij and Dr. Hans Lemij, for their role in developing and carrying the clinical studies that provided the images that form the core of this thesis. To Angela, Esma, Annemiek, and Caroline for their invaluable time in handling the patients and acquiring the images. To my former colleagues from the Rotterdam Ophthalmic Institute, thank you for the great moments and support along the way. To my officemates Henk, Angela, Sonia, Kedir, Elrozy, Mia, Olympia, and Valentina; and to all the other people from the department: Babak, Jelena, Grzegorz, Juleke, Lisette, Susan, Heijan, Verena, Robin, Monica, Gijs, Stijn, Aline, René, Aletta, and Netty.

I am also grateful to have been part of the Quantitative Imaging group in TU Delft. To my officemates Robiel, Willem, Joor, Martijn, thank you for creating a great work environment. To my research colleagues Busra, Sten, Tom, Anna, Jelle, Lena, Nadya, Jos, Gyllion, Leon, Robert M, Robert N, Jeroen H, Jeroen S, Hamidreza, Motjaba, Mohammad, Rasmus, Yan, Boling, Tian, Zhang, Christiaan, Pierre, Marco, Richard, Davide, Bernd, Sjoerd, Jeroen K, Frans, Ronald, Annelies, and all the others who I might miss, thank you for the great time together.

Finally, I thank my friends and family in Spain for their love and support during these long four years. *Un viaje de mil millas comienza con un paso, y paso a paso se llega a la meta.*

ABOUT THE AUTHOR

Juan Pedro Vigueras-Guillén was born in Archena, Murcia, Spain on May 11, 1982. He obtained his BEng. in Telecommunications Engineering from the Polytechnic University of Valencia (Spain) in June 2006. He later pursued the MEng. in Telecommunications Engineering at the same university, obtaining his degree in February 2010.

After a few years, he decided to change his career, moving to Göteborg, Sweden in August 2012 to study a Master of Science in Electrical Engineering (Biomedical programme) at Chalmers University of Technology. He specialized in medical imaging and performed his master's thesis in the research group MedTech West at Sahlgrenska Hospital (Göteborg, Sweden). The project, a collaboration with the company Medetect (Lund, Sweden), dealt with the automated detection of immune cells in immunohistochemical images, defending it on February 2015.

Just a week later, in March 2015, he started his PhD in a joint project between Delft University of Technology and Rotterdam Ophthalmic Institute. His research dealt with the developing of segmentation tools in specular images of the corneal endothelium, which resulted in this thesis.

Since June 2019, he is doing a postdoc in the pharmaceutical company AstraZeneca in Göteborg, Sweden, working in the development of new deep learning methodologies (Capsule Networks) to be applied in histopathology images.