# FOURIERNET: SHAPE-PRESERVING NETWORK FOR HENLE'S FIBER LAYER SEGMENTATION IN OPTICAL COHERENCE TOMOGRAPHY IMAGES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

By
Selahattin Cansız
September 2022

FOURIERNET: SHAPE-PRESERVING NETWORK FOR HENLE'S FIBER LAYER SEGMENTATION IN OPTICAL COHERENCE TOMOGRAPHY IMAGES

By Selahattin Cansız

September 2022

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Çiğdem Gündüz Demir(Advisor)

Selim Aksoy

Celal Murat Hasanreisoğlu

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

# ABSTRACT

## FOURIERNET: SHAPE-PRESERVING NETWORK FOR HENLE'S FIBER LAYER SEGMENTATION IN OPTICAL COHERENCE TOMOGRAPHY IMAGES

Selahattin Cansız
M.S. in Computer Engineering
Advisor: Çiğdem Gündüz Demir
September 2022

Henle's fiber layer (HFL), a retinal layer located in the outer retina between the outer nuclear and outer plexiform layers (ONL and OPL, respectively), is composed of uniformly linear photoreceptor axons and Müller cell processes. However, in the standard optical coherence tomography (OCT) imaging, this layer is usually included in the ONL since it is difficult to perceive HFL contours on OCT images. Due to its variable reflectivity under an imaging beam, delineating the HFL contours necessitates directional OCT, which requires additional imaging. This thesis addresses this issue by introducing a shape-preserving network, *FourierNet*, which achieves HFL segmentation in standard OCT scans with the target performance obtained when directional OCT is available. *FourierNet* is a new cascaded network design that puts forward the idea of benefiting the shape prior of the HFL in the network training. This design proposes to represent the shape prior by extracting Fourier descriptors on the HFL contours and defining an additional regression task of learning these descriptors. *FourierNet* then formulates HFL segmentation as concurrent learning of regression and classification tasks, in which Fourier descriptors are estimated from an input image to encode the shape prior and used together with the input image to construct the HFL segmentation map. Our experiments on 1470 images of 30 OCT scans reveal that quantifying the HFL shape with Fourier descriptors and concurrently learning them with the main task of segmentation leads to significantly better results. These findings indicate the effectiveness of designing a shape-preserving network to facilitate HFL segmentation by reducing the need to perform directional OCT imaging.

*Keywords:* Cascaded neural networks, Fourier descriptors, fully convolutional

networks, Henle's fiber layer segmentation, optical coherence tomography, shape-preserving network.

# ÖZET

# FOURIERNET: OPTİK KOHERANS TOMOGRAFİ GÖRÜNTÜLERİNDE HENLE LİFİ TABAKASI BÖLÜTLEMESİ İÇİN ŞEKİL KORUYAN AĞ

Selahattin Cansız
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Danışmanı: Çiğdem Gündüz Demir
Eylül 2022

Dış nükleer tabaka (DNT) ve dış pleksiform tabaka (DPT) arasındaki dış retinada bulunan bir tabaka olan Henle lifi tabakası (HLT), düzgün doğrusal fotoreseptör aksonlarından ve Müller hücre çıkıntılarından oluşur. Öte yandan, standart optik koherans tomografi (OKT) görüntülerinde bu tabaka genellikle DNT tabakasına dahil edilir. Bunun nedeni, OKT görüntülerinde HLT konturlarının algılanmasındaki zorluktur. Görüntüleme ışınları altında gösterdiği yansıma farklılığından ötürü HLT konturlarını belirlemek için yöneltmeli OKT'ye ihtiyaç duyulur. Bu ise ek görüntüleme gerektiren bir süreçtir. Bu tez, *FourierNet* adını verdiğimiz bir şekil koruyan ağ modelini sunarak bu soruna çözüm önerir. Bu model, HLT bölütlemesinde yöneltmeli OKT görüntüleri kullanılarak ulaşılacak hedef performansı, standart OKT görüntüleri ile başarır. *FourierNet*, ağ eğitme sürecinde HLT şekil bilgisini kullanan yeni bir kademeli ağ tasarımıdır. Bu tasarım, HLT konturları üzerinde Fourier tanımlayıcıları hesaplayarak ve bu tanımlayıcıları öğrenmek için ek bir regresyon işi tanımlayarak, şekil bilgisinin ifade edilmesini önerir. *FourierNet*, HLT bölütlemesini regresyon ve sınıflandırma işlerinin eşzamanlı öğrenimi olarak formüle eder. Bu işlerde, şekil bilgisini kodlamak için girdi görüntüsünden Fourier tanımlayıcılar tahmin edilir ve bu tanımlayıcılar girdi görüntüsü ile birlikte HLT bölütleme haritasını oluşturmak için kullanılır. Otuz OKT taramasındaki toplam 1470 OKT görüntüsü üzerinde yapılan deneyler, HLT şeklinin Fourier tanımlayıcılar ile nicelleştirilmesinin ve bunların bölütleme ana işi ile birlikte eş zamanlı olarak öğrenilmesinin önemli ölçüde daha iyi sonuçlar verdiğini göstermiştir. Bu bulgular, yöneltmeli OKT görüntülerine olan ihtiyacı azaltarak HLT bölütlemesini kolaylaştıran şekil koruyan bir ağ tasarlamanın verimliliğini göstermektedir.

*Anahtar sözcükler*: Kademeli sinir ağları, Fourier tanımlayıcılar, tam evrişimsel

sinir ağları, Henle lifi tabakası bölütlemesi, optik koherans tomografi, şekil koruyan ağlar.

# Acknowledgement

Foremost, I would like to express my heartfelt thanks to my supervisor Prof. Dr. Çiğdem Gündüz Demir. I could not have undertaken this journey without her continuous support, patience, and encouragement. Her immense knowledge and experience have always helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to extend my sincere thanks to my thesis committee, Prof. Dr. Selim Aksoy and Prof. Dr. Murat Hasanreisoğlu, for reviewing the thesis and improving the quality of the work with their feedback. I am also extremely grateful to Prof. Dr. Murat Hasanreisoğlu once more as well as to Dr. Cem Kesim for providing the medical data and their expertise for this study. This endeavor would not have been possible without their support.

I had the pleasure of studying at Bilkent University. My special thanks should also go to all faculty members and administrative staff for providing a tremendous learning environment. I am grateful to be a part of this university. I am also thankful for the partly support to this thesis work by the Scientific and Technological Research Council of Turkey through the project TÜBİTAK 120E497.

Last but not the least, I would like to express my deepest appreciation to my family and friends for their endless love, support, help, and being source of motivation all the time.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This chapter is based on our manuscript submitted for publication [1].

Optical coherence tomography (OCT) is a retinal imaging modality that allows tomographic visualization of the posterior pole of the eye. In the current clinical practice, OCT imaging has become the essential tool for diagnosis and treatment of patients with retinal diseases. OCT provides axial retinal images, which are created from the coherence patterns of the imaging light beam projected and reflected back from the retinal surface. In this method, each retinal layer is presented on the acquired image with a definite signal intensity that corresponds to its own reflectivity property. However, a layer that is mainly composed of uniformly aligned photoreceptor axons and Müller cell processes, known as the Henle's fiber layer (HFL), shows an ambivalent reflectivity pattern that depends on the angle between the incident OCT light beam and the alignment of fibers. Due to this intrinsic property of variable reflectivity of the fibers in the HFL under a single imaging beam, to date it remains a fundamental challenge to perform a separate segmentation of the HFL on standard OCT scans. On the other hand, directional OCT, which obtains multiple images by altering the entry position of the imaging beam at the pupil, emerges as a useful technique for HFL segmentation [2]. However, this technique is not a routine clinical procedure, mostly because it necessitates a significant amount of additional examination

time. As a result, in the common practice, the HFL is considered rather as a part of the outer nuclear layer (ONL). Therefore, an automatic HFL segmentation lacks both in commercially available OCT software [3] and in scholarly studies [4, 5].

To address this gap, this thesis introduces a new cascaded neural network, which we name *FourierNet*, for the purpose of HFL segmentation in standard OCT scans. The proposed *FourierNet* model uses only the standard scans as its inputs and achieves the performance of a method that uses additional scans when directional OCT is available. Such a comparison method is intended to provide us with the achievable target performance, which the proposed model aims to reach without requiring any non-routinely used imaging.

To enable HFL segmentation, we propose to employ prior knowledge on the shape of the HFL in the network training. The proposed *FourierNet* model quantifies this shape prior with a function defined on the HFL contours. To this end, it expands the function in a Fourier series, uses the harmonic amplitudes of its Fourier coefficients as the Fourier descriptors of the HFL, and represents this shape prior in the network design by defining a regression task of learning these descriptors. *FourierNet* then formulates HFL segmentation as concurrent learning of regression and classification tasks, in which Fourier descriptor maps are estimated from an input image to represent the shape prior of the HFL and used along with the input image to estimate a segmentation label for each image pixel. *FourierNet* achieves this learning by designing a cascaded fully convolutional network (FCN) that consists of an intermediate regression and a final classification task.

## 1.1 Contributions

The contributions of this thesis are three-fold:

- To the best of our knowledge, this is the first study that has automatically segmented the HFL in standard OCT scans. Although layer-wise retina segmentation in OCT scans has been studied widely, none of the previous studies segment the HFL separately but instead include it in the ONL.
- *FourierNet* presents a cascaded FCN design where the intermediate task of Fourier descriptor estimation contributes to the final segmentation task by providing it with the shape-related information about what it needs to estimate. Since the network weights are updated at the same time by minimizing a joint loss function, this concurrent learning more likely imposes the shape on the network. This, in turn, leads to better preserving the HFL shape, and thus, enhances HFL segmentation.
- *FourierNet* quantifies the HFL shape with a set of Fourier descriptors and devises a cascaded FCN design for their estimation. These Fourier descriptors, which were first suggested by Cosgriff in 1960 [6], were also used in previous studies [7–9]. However, these previous studies used the Fourier descriptors to characterize objects in different applications (e.g., object retrieval and recognition), but did not employ them in designing a shape-preserving FCN. On the other hand, *FourierNet* uses the Fourier descriptors to construct a shape-preserving FCN for the first time, and demonstrates that this use is effective for more accurate HFL segmentation.

## 1.2 Outline of the Thesis

The thesis is organized as follows: In Chapter 2, methods proposed for retinal layer segmentation are discussed. In Chapter 3, the details of the methodology regarding the proposed method are explained. The information about the dataset, metrics for performance evaluation, and the methods that the proposed method is compared against are all clarified in Chapter 4. Then, the discussions about the results, both quantitative and visual, are given in Chapter 5. Finally, in Chapter 6, the thesis is concluded with potential uses of the proposed method to segment pathological retinal layers and possible future research directions.

# Chapter 2

# Background

This chapter is based on our manuscript submitted for publication [1].

Optical coherence tomography (OCT) is a noninvasive imaging modality and widely used in clinical ophthalmology to visualize the retinal morphology. OCT uses infrared light to differentiate tissue sections with high resolution within the retina [2]. Since its invention, OCT has become an essential tool for the early diagnosis of retinal pathologies, monitoring diseases, and research. In an OCT scan, there are several retinal layers that can be identified based on the level of reflectivity as seen in Figure 2.1. If the reflectivity of a tissue varies and does not follow a uniform pattern under a single imaging beam, visualization of retinal structures requires directional OCTs, which could be obtained by changing the entry position of the imaging beam. An example standard OCT scan and its corresponding directional scans (inferior, superior, nasal and temporal OCTs) are illustrated in Figure 2.2. Note that pixel intensities differ on each scan. Hence, if a retinal layer is not visible locally on the spectral domain OCT scan, it can be visualized by switching to other scans, which facilitates more accurately defining retinal layer boundaries.

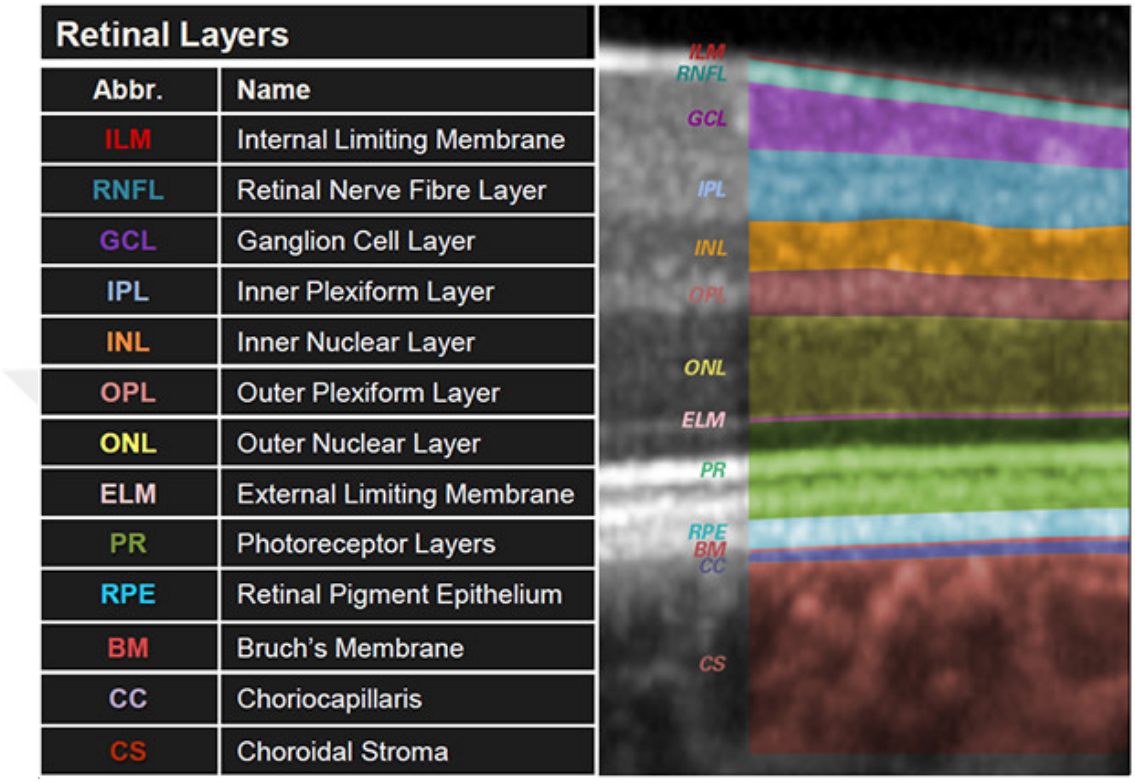



Figure 2.1: Retinal layers in an OCT image. This image is taken from [10].

The interpretation of an OCT image is a challenging task because of the artifacts caused by eye movements, low intensity difference at the boundaries, and deformed retinal structures in the case of retinal pathologies. Hence, manual annotation of the retinal layers is a time consuming process and very subjective. This motivated many research to develop automated methods for accurate and reliable segmentation. With the advances in medical image analysis techniques, various methods have been proposed for automatic segmentation of retinal layers in OCT images. Based on their approaches, these methods could be categorized as traditional methods, classical machine learning methods, and deep learning methods.

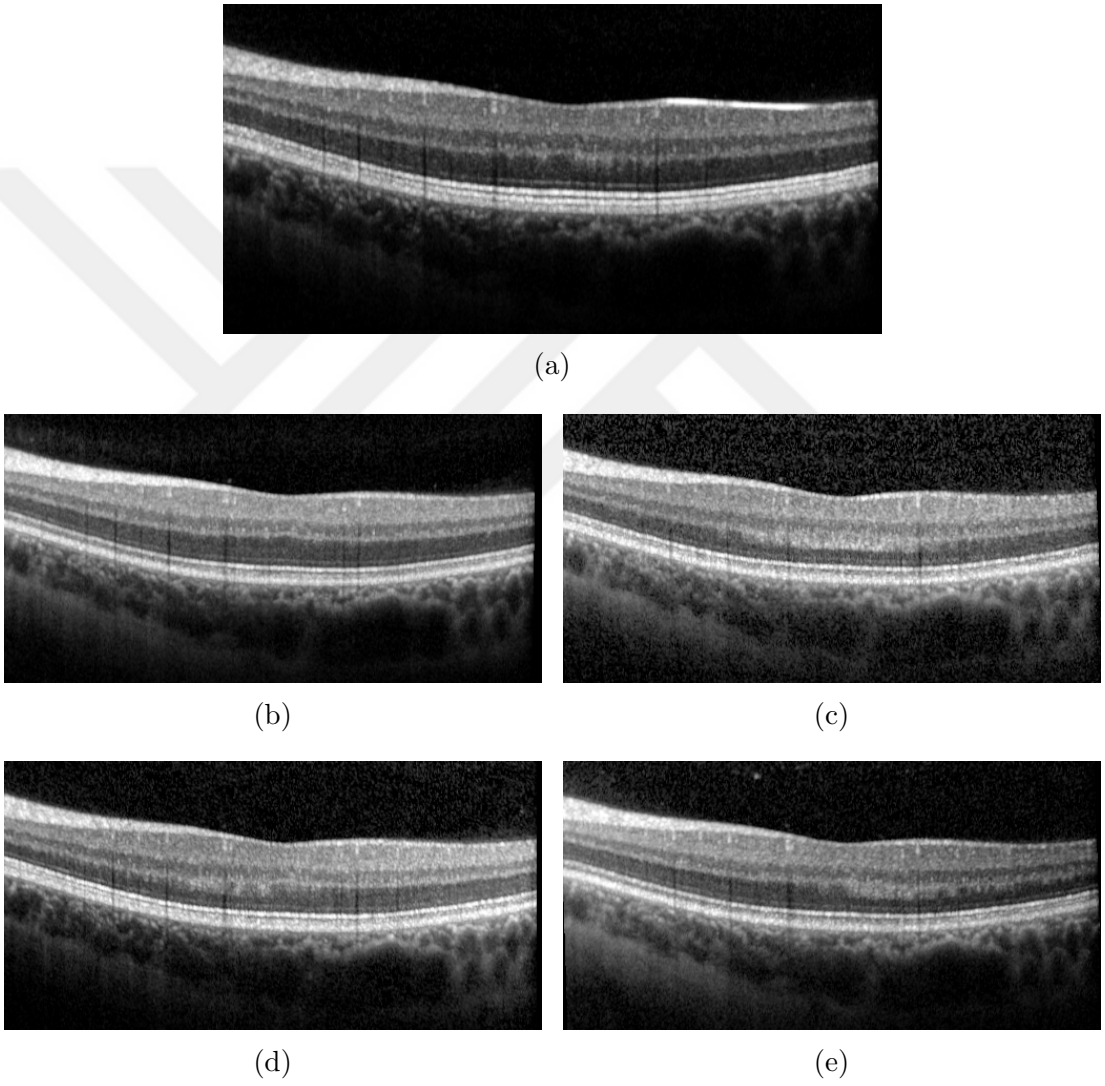


Figure 2.2: An examplary OCT scan with corresponding four directional scans. (a) standard OCT, (b) inferior OCT, (c) superior OCT, (d) nasal OCT, and (e) temporal OCT.

## 2.1 Traditional Methods

Earlier studies typically segment retinal layers by identifying their boundary pixels and refining them with optimization methods. For example, in [11], initial boundaries were obtained using edge detection after denoising the image and these boundaries were fitted through minimization of an energy term that considered vertical gradients and regional smoothness. However, these methods are not generally applicable since mostly there is a small intensity difference between retinal layers. To overcome the limitations of the edge detection algorithms, deformable models and graph-based methods were employed using the order among retinal layers. These approaches were based on some energy function that was minimized to preserve the shape smoothness and obtain the expected segmentation.

Active contour and level set methods are among the widely known deformable models. Extending the traditional active contour model, an adaptive vector-valued kernel function was used in [12] to account for the presence of speckle noise. Alternatively, in [13], retinal layers were segmented using loosely coupled level sets that defined the optimization problem based on boundary smoothness and thickness priors. By the way, the constraints required by these methods increased the complexity of the problem, and thus, led to low efficiency.

In [14–16], segmentation problem was defined as finding minimum volumetric surface cost subject to some regularizing constraints and prior information via a graph-cut technique and adjusting distances between two segmented layers. As an alternative approach to graph-cut segmentation, intensity gradients were used as graph weights and then layer boundaries were obtained by applying dynamic programming on the shortest path problem [17].

## 2.2 Classical Machine Learning Methods

Later, data driven approaches that used labelled data to train classical machine learning models were proposed for retinal layer segmentation. In [18], k-nearest neighbor classifier was applied to voxels for initial segmentation, and then a graph-cut algorithm with probability constraints was used to refine the boundaries. As an another example of three-dimensional retinal layer segmentation, a support vector machine classifier was trained on the interactively marked set of points and employed to define layer boundaries by considering the mean intensity of six neighbors at each voxel [19]. In [20], a graph-cut algorithm and dynamic programming were used together to refine initial layers classified by kernel regression. Apart from these, a trained neural network classifier was used to identify pixels within each layer by taking average minimum and maximum intensities between adjacent layers once the starting points were selected [21]. In [22], a random forest classifier was used to generate segmentation maps. In [23], sparse representations were utilized to improve the initial segmentation obtained using a dynamic programming method.

## 2.3 Deep Learning Methods

More recent studies commonly used UNet architectures [24] and its variants for layer-wise retina segmentation [25]. The UNet architecture consists of symmetric encoding and decoding paths (see Figure 2.3). The encoding path extracts features and learns high level representations. On the other hand, the decoding path uses these features and generates a segmentation mask with the original dimensions. The skip connections between corresponding encoding and decoding blocks are used to recover spatial information lost due to the pooling layers.

In [26], retinal layers were segmented designing an asymmetric U-shaped network that combined residual building blocks with dilated convolutions. In [27], a modified UNet with a context extractor module, which generated high-level

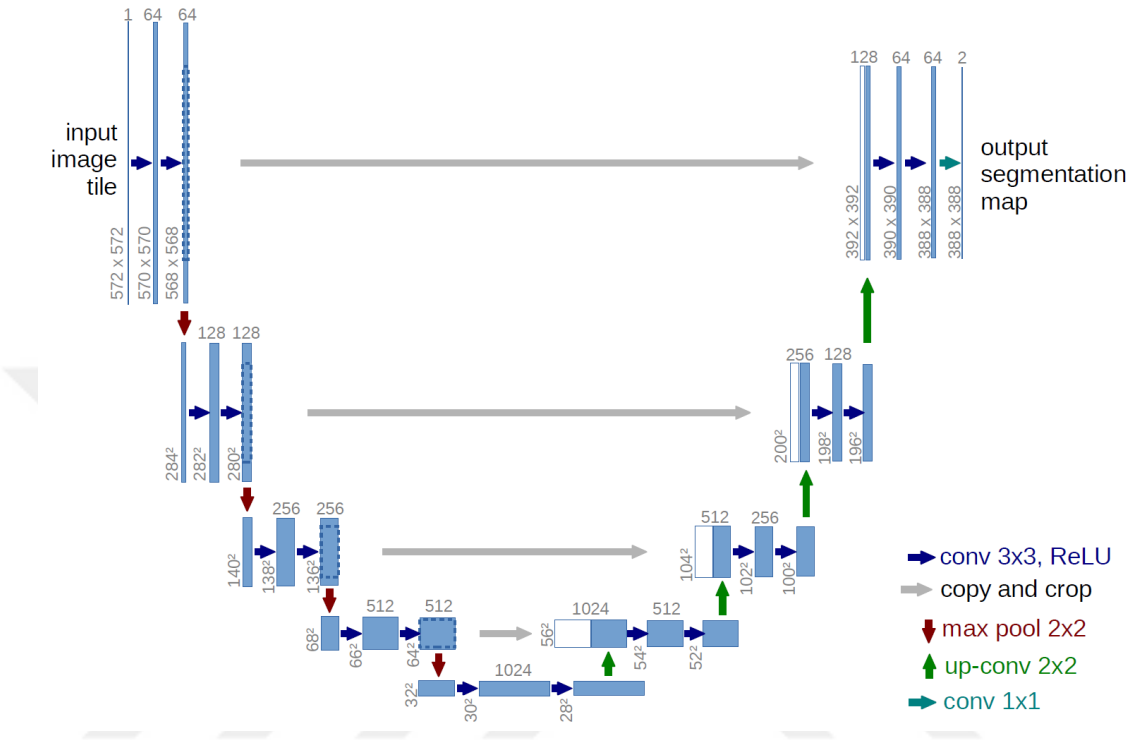



Figure 2.3: The original UNet architecture. This figue is taken from [24].

semantic feature maps, was employed. Adversarial learning technique was also used for segmenting retinal layers in OCT images [28]. Post-processing techniques were used to correct pixel-wise classifications of these networks. For instance, a graph search algorithm was applied on posterior probability maps outputted by a convolution neural network [4] as well as by a recurrent neural network [29] to find the final boundaries. In another study [5], the DenseNet architecture was combined with a Gaussian process regression to smooth the segmentation results. In [30], a random forest was trained on handcrafted features along with those learned by a residual network and this trained classifier obtained the contour probabilities of each retinal layer. Some studies proposed methods that handled topology correction in an end-to end manner. For example, an edge detection module was applied to the output of U-structure network which employed multi-scale inputs [31].

Cascaded network architectures were also used for segmenting OCT images. They typically had two separately trained modules. In [32, 33], these modules

achieved lesion segmentation and classification consecutively. In [34], they classified noise distribution in an image and then removed noise from the image based on the selected distribution. Segmented layers can also be refined in a cascaded framework. The network proposed by [35] first segmented retinal layers and then performed topology correction on the segmented layers. In [36], the network automatically configured itself using 2D or 3D UNets where segmentation and refinement of an output map were achieved subsequently. The modules within this architecture were mainly used for parameter adjustment, preprocessing, and postprocessing. The UNet architecture in [37] employed a feature pyramid module as its bottleneck layer to capture multi-scale features and handled the problems of scale variations and discontinuity of retinal vessels.

Although all these networks yielded promising results for the segmentation of retinal layers, none of them segmented the HFL separately. Moreover, they neither defined Fourier descriptors to represent the shape prior of a retinal layer nor integrated this information with a design of a cascaded architecture. On the other hand, the proposed *FourierNet* model introduces a shape-preserving network for HFL segmentation for the first time and presents an effective way of representing the shape prior of the HFL defining the Fourier descriptors on its contours.

There also exist previous studies that used the Fourier transform to facilitate different types of tasks in neural network classifiers. It was proposed to conduct convolution operations in the Fourier domain in order to gain speed and enable processing of larger images [38, 39], as well as to obtain non-local receptive fields [40]. In [41], all network operations were transformed to the Fourier space to satisfy translation equivariance with respect to rotations in spherical images. In [42], Fourier basis functions characterized universal adversarial perturbations on input images that could change a model's prediction while preserving the image content. As opposed to the proposed *FourierNet* model, none of these previous studies used the Fourier transform to represent the shape prior and defined a shape-preserving network for image segmentation.

# Chapter 3

# Methodology

This chapter is based on our manuscript submitted for publication [1].

The *FourierNet* model relies on characterizing the HFL shape with Fourier descriptors, defining a regression task of estimating the maps of the Fourier descriptors, and learning this regression task together with the main task of HFL segmentation by designing a cascaded FCN. The following subsections give the implementation details.

## 3.1 Fourier Descriptors

We quantify the HFL shape with a function defined along its contour points. This is the distance-to-center function that outputs the distance from an object centroid to the contour point at a given arc length. The object centroid is calculated by averaging the coordinates of all contour points. This function characterizes how subsequent boundary points distribute in the space to form the object's contour, and hence, the shape of the object. It can be defined for any arbitrarily shaped object. For instance, it is a constant function for a circular object since the distance from any boundary point to the centroid (radius) is the same.

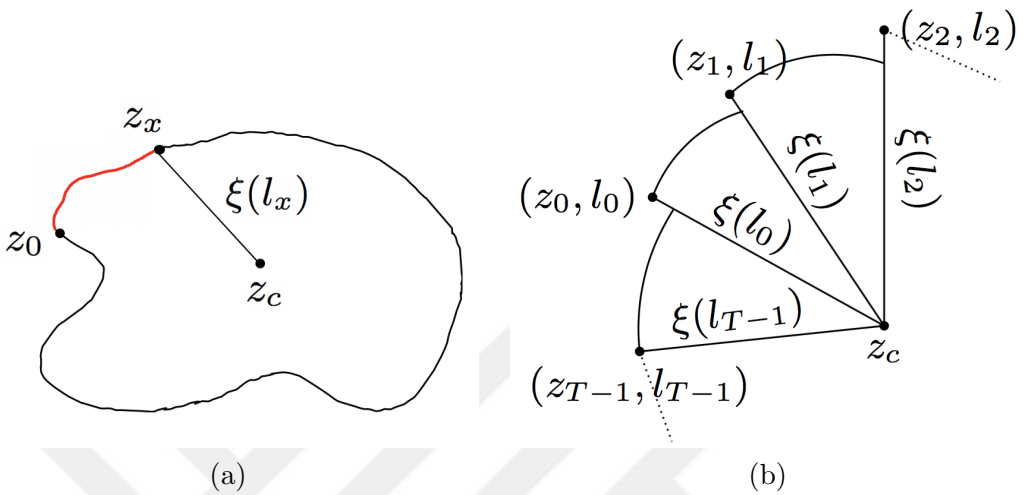



Figure 3.1: (a) The distance-to-center function $\xi(.)$ defined on the domain of length $l_x \in [0, L]$ where $l_x$ denotes the arc length of a section of the continuous curve $\gamma$ from its starting point $z_0$ to the point $z_x$ of the same curve. (b) Illustration of a circular arc interpolation between the discrete points (pixels) of HFL contour $\gamma_h$. This interpolation will be used to define a continuous curve for Fourier descriptor calculations. In this illustration, consecutive pixels are drawn widely separated from each other for demonstration and each pixel is denoted with its coordinate $z_x$ and its corresponding arc length $l_x$.

Let $\gamma$ be a closed continuous curve with a length of $L$. Fourier descriptors are calculated for the distance-to-center function $\xi(.)$ on the domain of length $l_x \in [0, L]$ where $l_x$ denotes the arc length of a section of the curve $\gamma$ from its starting point $z_0$ to the point $z_x$ of the same curve (Figure 3.1a). Since the contour of an object in a digital image does not form a continuous curve but contains finitely many discrete points (pixels), we assume an arc interpolation between these discrete points to define a continuous curve. Thus, the curve $\gamma_h$ that corresponds to the contour of the HFL $h$ becomes the interpolation of $T$ boundary pixels, $\{z_0, z_1, ..., z_{T-1}\}$, each of which lies on the curve $\gamma_h$ at arc length $l_t$ (Figure 3.1b).

The distance-to-center function $\xi(l_x)$ outputs the distance from the centroid $z_c$ to the point $z_x$ for which the arc length is $l_x$. This function is expanded in a

Fourier series as

$$\xi(l_x) = a_0 + \sum_{n=1}^{\infty} \left[ a_n \cos\left(\frac{2\pi n l_x}{L}\right) + b_n \sin\left(\frac{2\pi n l_x}{L}\right) \right], \tag{3.1}$$

where

$$a_n = \frac{2}{L} \int_0^L \xi(l_x) \cos\left(\frac{2\pi n l_x}{L}\right) dl_x \text{ and} \tag{3.2}$$

$$b_n = \frac{2}{L} \int_0^L \xi(l_x) \sin\left(\frac{2\pi n l_x}{L}\right) dl_x. \tag{3.3}$$

For the curve $\gamma_h$, which is an interpolation of $T$ discrete pixels, Equation 3.2 can be divided into $T$ intervals of $[l_{t-1}, l_t)$. Thus, the coefficient $a_n$ can be written as

$$a_n = \frac{2}{L} \sum_{t=1}^{T} \int_{l_{t-1}}^{l_t} \xi(l_x) \cos\left(\frac{2\pi n l_x}{L}\right) dl_x. \tag{3.4}$$

Here we use the circular arc interpolation to estimate the value of $\xi(l_x)$ for all lengths $l_x$ that do not correspond to any boundary pixels. Since the distance from the circle centroid to any point on a circle is the same, this interpolation gives $\xi(l_x) = \xi(l_{t-1}), \forall l_x \in [l_{t-1}, l_t)$, and as a result, allows to take $\xi(l_x)$ outside the integral.

$$\begin{aligned}
a_n =& \frac{2}{L} \sum_{t=1}^{T} \xi(l_{t-1}) \int_{l_{t-1}}^{l_t} \cos\left(\frac{2\pi n l_x}{L}\right) dl_x \\
a_n =& \frac{1}{\pi n} \sum_{t=1}^{T} \xi(l_{t-1}) \left[ \sin\left(\frac{2\pi n l_t}{L}\right) - \sin\left(\frac{2\pi n l_{t-1}}{L}\right) \right] \\
a_n =& \frac{1}{\pi n} \left[ \xi(l_0) \sin\left(\frac{2\pi n l_1}{L}\right) - \xi(l_0) \sin\left(\frac{2\pi n l_0}{L}\right) \right. \\
& + \xi(l_1) \sin\left(\frac{2\pi n l_2}{L}\right) - \xi(l_1) \sin\left(\frac{2\pi n l_1}{L}\right) \\
& \vdots \\
& + \xi(l_{T-2}) \sin\left(\frac{2\pi n l_{T-1}}{L}\right) - \xi(l_{T-1}) \sin\left(\frac{2\pi n l_{T-2}}{L}\right) \\
& \left. + \xi(l_{T-1}) \sin\left(\frac{2\pi n l_T}{L}\right) - \xi(l_T) \sin\left(\frac{2\pi n l_{T-1}}{L}\right) \right].
\end{aligned}$$

Since $\gamma_h$ is a closed curve, the last point $z_T$ is indeed the starting point $z_0$, and hence, $\xi(l_0) = \xi(l_T)$ and $\sin(\frac{2\pi n l_0}{L}) = \sin(\frac{2\pi n l_T}{L})$. By defining $\Delta\xi_t = \xi(l_{t-1}) - \xi(l_t)$

$$a_n = \frac{1}{\pi n} \sum_{t=1}^{T} \Delta\xi_t \ \sin\left(\frac{2\pi n l_t}{L}\right). \tag{3.5}$$

Following similar steps, the coefficient $b_n$ is expressed as

$$b_n = -\frac{1}{\pi n} \sum_{t=1}^{T} \Delta\xi_t \ \cos\left(\frac{2\pi n l_t}{L}\right). \tag{3.6}$$

For the $n$-th Fourier coefficients $(a_n, b_n)$, the polar coordinates are $(A_n, \alpha_n)$ where $A_n = \sqrt{{a_n}^2 + {b_n}^2}$ is the harmonic amplitude and $\alpha_n = \arctan(b_n/a_n)$ is the harmonic phase.

The first $N$ harmonic amplitudes of a truncated expansion of $\xi(l_x)$ are used as a set of Fourier descriptors $\texttt{FD}(\gamma_h) = [A_1, A_2, ..., A_N]$ to characterize the contour $\gamma_h$ of a given HFL $h$, and therefore, its shape. Note that when $N \to \infty$, the curve can be reconstructed using these harmonic amplitudes together with their corresponding harmonic phases. However, we do not use the harmonic phases to define a descriptor set as they provide less shape-related information [9]. Note that this Fourier descriptor calculation was also used by the previous studies [7,8] for object retrieval and recognition. Similar calculations were followed by [9] not on the distance-to-center function but on the cumulative-angle function that outputs a normalized cumulative angle from the starting point $z_0$ of the curve to the point $z_x$ at arc length $l_x$. Nevertheless, none of these studies used the Fourier descriptors in the design of a shape-preserving FCN.

## 3.2 Fourier Descriptor Map Generation

The calculation explained in the previous subsection outputs a set of Fourier descriptors $\texttt{FD}(\gamma_h) = [A_1, A_2, ..., A_N]$ for the contour (outermost pixels) of the HFL. In order to define the maps of intermediate regression tasks, these contour-wise descriptors are mapped onto every pixel, both HFL and background pixels, using an iterative algorithm. In its initial step, this algorithm calculates the Fourier descriptors $\texttt{FD}(\gamma_h)$ on the contour $\gamma_h$ of the HFL $h$ in an image and assigns these descriptors to each pixel $z_t \in \gamma_h$. In the next iteration, the algorithm obtains a new HFL region $h'$ by removing the pixels $z_t \in \gamma_h$ from $h$ and finds the contour $\gamma_{h'}$ of this new region. It then calculates the Fourier descriptors $\texttt{FD}(\gamma_{h'})$ on $\gamma_{h'}$ and assigns them to the pixels $z'_t \in \gamma_{h'}$. The algorithm iteratively continues until there exists no HFL region in the image. The default value of 0 is used as the descriptors of the background pixels. When $N > 1$, a contour pixel has $N$ Fourier descriptors, each of which is represented in a separate channel. This results in generating an $N$-channel map for a given image. In our experiments, we selected $N = 1$, which gave the highest f-score on validation images. This selection did not use any test images. For $N = 1$, the map calculation of a single image implemented in Python took $0.141 \pm 0.044$ s on an Intel Xeon 2.20 GHz CPU. This time increased to $0.213 \pm 0.067$ s when $N = 5$.

For an example OCT image, Figure 3.2 illustrates the map of the first Fourier descriptors calculated with respect to the given HFL annotations. This calculation is possible only for a training image, for which the HFL is manually annotated. For test images, the maps are to be estimated by the intermediate regression tasks of the cascaded FCN.

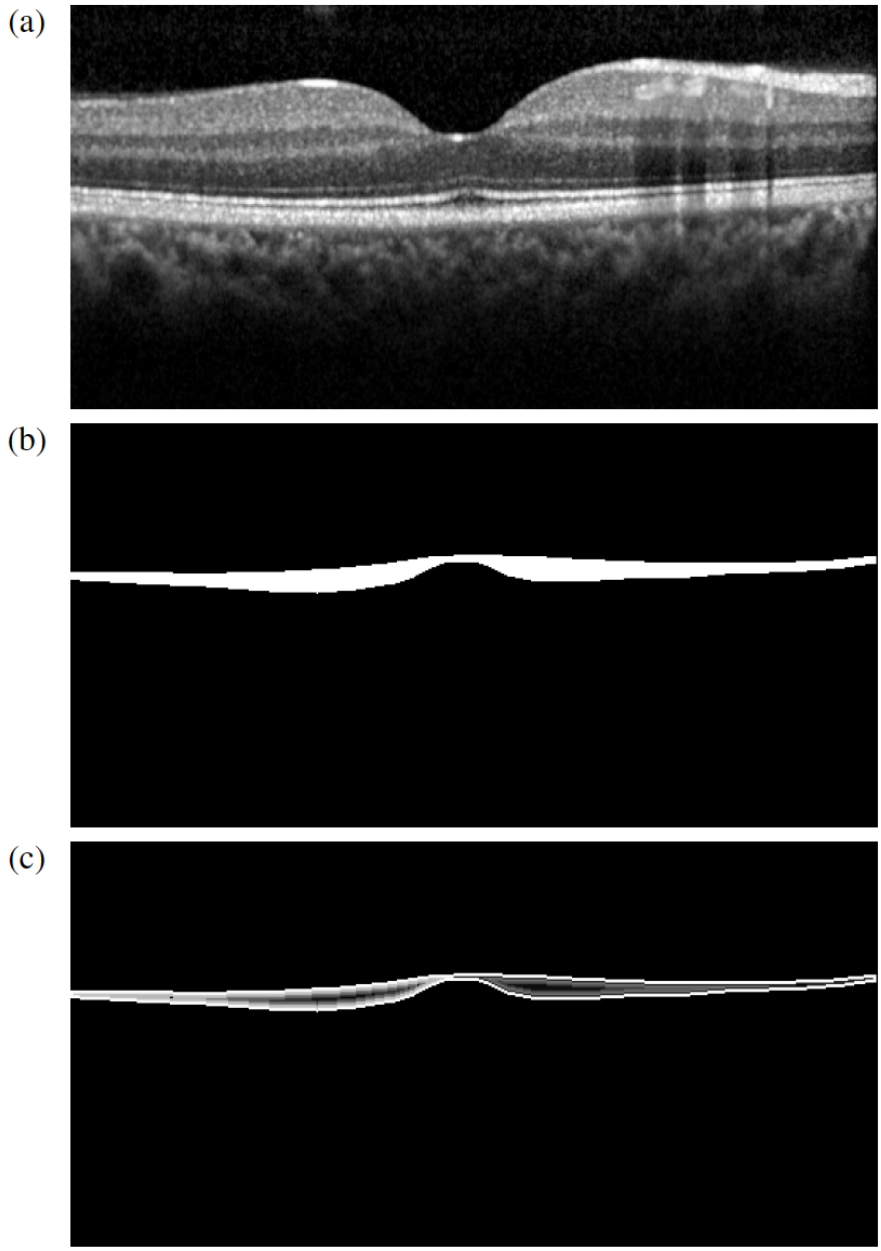



Figure 3.2: (a) An image taken from the standard OCT, (b) its manual HFL annotations, and (c) the map of the first Fourier descriptors calculated with respect to the given HFL annotations. Note that Fourier descriptors' magnitudes in this map are represented using gray tones; the brighter the pixel is, the larger its Fourier descriptor is.

## 3.3 Cascaded Network Architecture and Training

The proposed *FourierNet* model uses a cascaded FCN to concurrently learn the regression and classification tasks. This cascaded FCN uses a multi-task network to estimate $N$ Fourier descriptor maps from an input image in its initial (intermediate) regression tasks. This multi-task network has one encoder to learn shared feature maps and $N$ decoders, each of which learns a different Fourier descriptor map from the shared features. Note that when $N = 1$, the multi-task network has one decoder, which converts it to a single-task UNet architecture. The selected architectures of the encoder and decoders are illustrated in Figure 3.3. The model consecutively uses another UNet that consists of an encoder and a decoder with the same architecture (Figure 3.3) to predict segmentation labels from the estimated Fourier descriptor maps along with the input image in its final classification task.

In all these selected architectures, the convolutional layers use $3 \times 3$ filters and are followed by the rectified linear unit (ReLU) activation function except the output layers. The output layers of the regression and classification tasks use the linear and softmax activations, respectively. The pooling and upsampling layers use $2 \times 2$ filters. Long-skip connections are added between the corresponding layers of the encoder and the decoder. Dropout layers at the rate of 0.2 are used for regularization. The number of layers and the feature maps used in each convolution layer are illustrated in Figure 3.3.

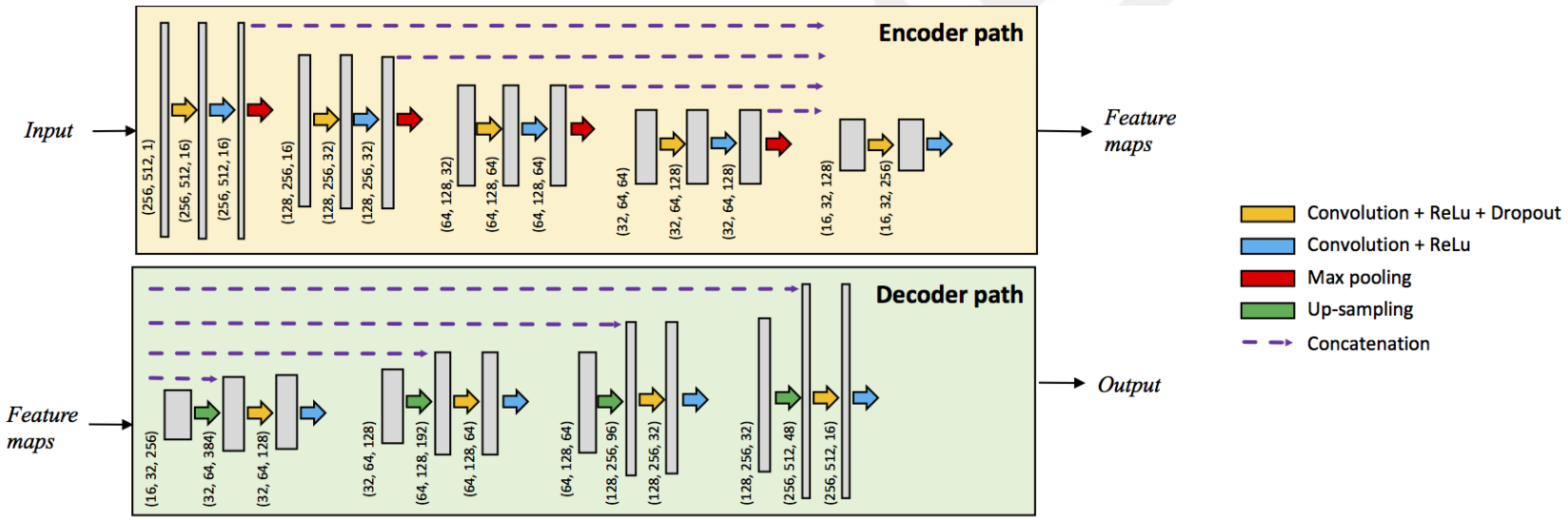



Figure 3.3: Encoder and decoder paths used in the UNet architectures. Each box corresponds to an operation distinguishable by its color. The numbers on the left of each box are of the form $(h, w, c)$ where $h$ and $w$ denote the input's height and width and $c$ denotes the number of input channels.

The overall architecture of the cascaded FCN is shown in Figure 3.4. This cascaded design, which obtains the Fourier descriptor representation in the middle of the entire architecture, enables both the encoder-decoder networks to exploit the feature maps learned for this representation through gradients flowing in the entire network. The cascaded FCN is implemented in Python using the Keras deep learning library. It is end-to-end trained from scratch to concurrently learn the intermediate regression and final classification tasks. The mean square error and the categorical cross-entropy are used as the loss function, respectively. All tasks contribute to the joint loss function equally with the unit weight. AdaDelta is used to adaptively adjust the learning rate and the momentum. The batch size is selected as 1. Early stopping is used in training. The network training stops if there is no improvement on the validation set loss in the last $M$ epochs. Too small $M$ values may cause insufficient training of the network whereas too large values make training unnecessarily long. Considering this tradeoff, we set $M = 50$ epochs in our experiments.

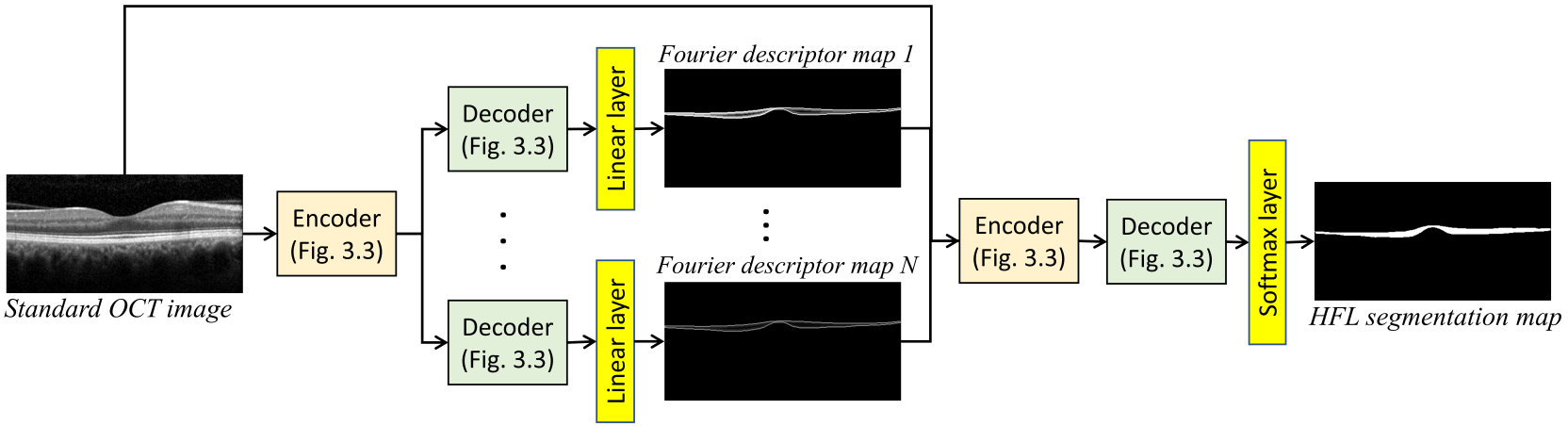



Figure 3.4: Architecture of the cascaded FCN to concurrently learn the regression and classification tasks, along with an example standard OCT image as the input and the estimated Fourier descriptor maps and the segmentation map as the outputs of the regression and classification tasks, respectively. Note that Fourier descriptors' magnitudes in the estimated maps are represented using gray tones; the brighter the pixel is, the larger its Fourier descriptor is.

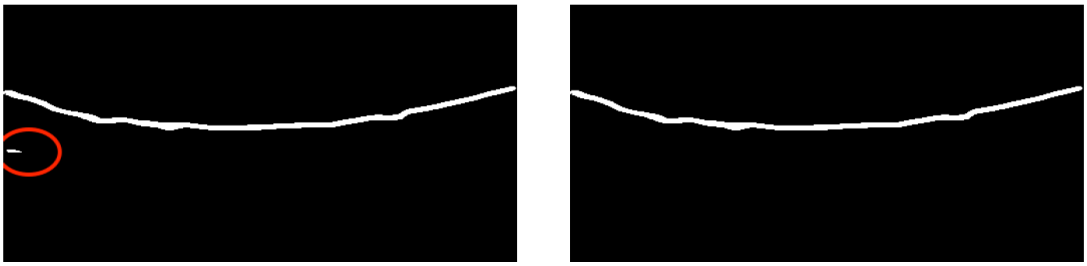



Figure 3.5: (a) Segmentation map estimated by a network and (b) the resulting map after postprocessing. The red oval indicates an examplary noisy region, which need to be eliminated by postprocessing the estimated map.

## 3.4 HFL Segmentation

Pixels are classified as HFL if their estimated posteriors are greater than 0.5. This may yield small noisy regions (see Figure 3.5), which need to be eliminated. For that, a simple postprocessing method is applied separately on each column C in the estimated segmentation map. This method keeps only pixels of the largest connected component of C, removing the other pixels from the column C. Note that this step considers pixels of the other columns to identify the largest component of C. However, while processing C, it does not remove any pixels from the other columns even though these pixels may belong to the non-largest components of C. Although this simple method does not guarantee to remove only non-HFL pixels, it is effective to remove noisy regions for our model as well as for the comparison algorithms. For a single test image, our model achieved HFL segmentation in $0.027 \pm 0.002$ s on a GeForce GTX 1080 Ti GPU and postprocessing in $0.178 \pm 0.025$ s on an Intel Core 3.40 GHz CPU.

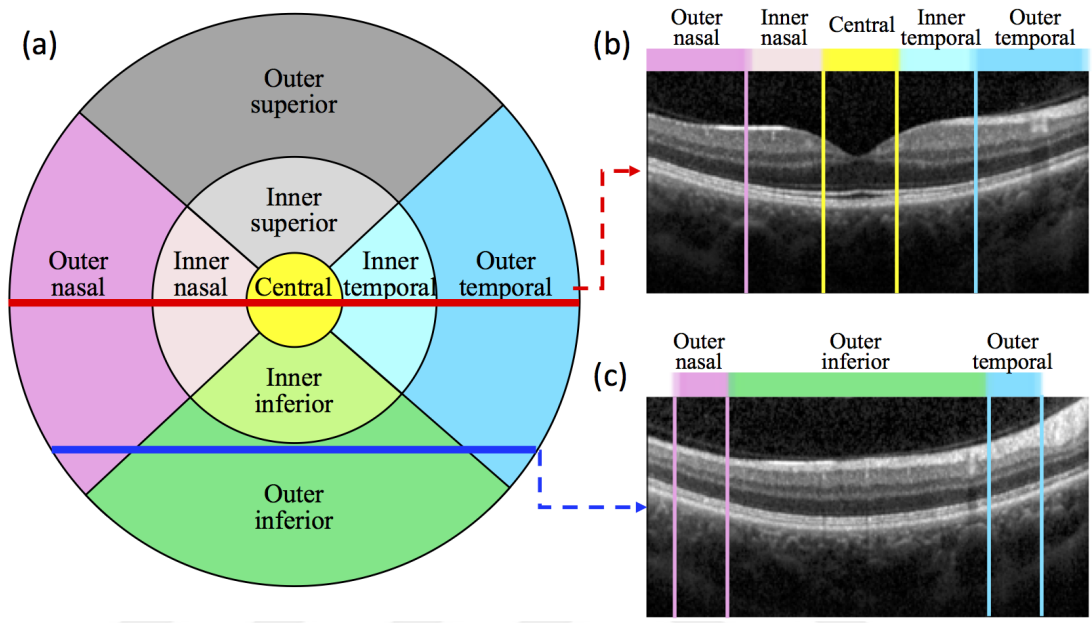



Figure 3.6: (a) Illustration of the ETDRS grid that divides the retina volume into nine macular sectors based on the three concentric circles with 1, 3, and 6 mm diameters. This illustration represents the top view for a retinal OCT volume scan. This ETDRS grid is depicted for the right eye; the nasal and temporal sectors should be the opposite for the left eye. (b)-(c) Two samples of cross-sectional OCT scan slices corresponding to the red and blue lines in the illustrated ETDRS grid.

## 3.5 ETDRS Grid Analysis

In routine clinical practice, Early Treatment of Diabetic Retinopathy Study (ETDRS) grid is used to quantify topographical properties in the central subfield and inner-outer quadrantal areas of macula. This grid divides the retina into nine macular sectors based on the three concentric circles of 1, 3, and 6 mm in diameter (see Figure 3.6). The layers' volume and thickness in the whole ETDRS grid as well as in each sector give valuable information. In our experiments, we will analyze the performance of *FourierNet* for each sector in the ETDRS grid. Additionally, we will conduct volumetric and thickness analyses of the HFL within the ETDRS grid.

# Chapter 4

# Experiments

This chapter is based on our manuscript submitted for publication [1].

## 4.1 Dataset

The *FourierNet* model was tested on a dataset that contains standard OCT scans of 30 eyes belonging to 17 subjects with healthy-looking macula. Study subjects were recruited from healthy participants with no reported ocular or systemic disease. All patients underwent full ophthalmological examination including best corrected visual acuity, slit-lamp biomicroscopy, applanation tonometry, and dilated fundus examination. Patients with more than $\pm 3.0$ diopters of refractive error, any ocular disease including cataract, glaucoma, inflammation, and macular degeneration, and history of ocular surgery were excluded. The data collection was conducted in accordance to the tenets of the Declaration of Helsinki and was approved by Koç University Institutional Review Board (Protocol number: 2020.443.IRB2.117). Informed consent was obtained from all participants of the study.

Eye scans were acquired using Heidelberg SD-OCT imaging equipment (Spectralis®, Heidelberg Engineering GmbH, Heidelberg, Germany). Images were acquired with the macular $20^\circ \times 20^\circ$ dense protocol that included 49 consecutive scans with ART set to 16 images without Enhanced Depth Imaging (EDI) mode.

For each eye, a total of five sets of 49 grayscale OCT scans were acquired. The first set was the *standard* scan acquired using standard settings and the incident light beam directed to the foveal center. In order to acquire the other four directional scans with *superior, inferior, nasal,* and *temporal* tilts from the same retinal locations, the built-in auto-alignment feature of the OCT device was used where it considered the first set as the reference scan and the others as the follow-up scans. Directional scans were acquired by directing the light beam to upper, lower, nasal, and temporal juxtafoveal areas that created an arbitrary $10^\circ$ of inclination of the live OCT image when compared to the standard reference scan. Two ophthalmologists then used all these five scans simultaneously to manually delineate HFL boundaries on the standard scan. An input image with the resolution of $256 \times 512$ pixels was cropped from each of these scans. It is important to note that the OCT images acquired with superior, inferior, nasal, and temporal tilts were used only for annotating the HFL but not for training the *FourierNet* model. The further details about data acquisition, HFL annotation, and directional scan alignment can be found in [43].

## 4.2 Evaluation

Three-fold cross-validation was used in the experiments. For that, 30 eye scans were randomly split into three folds. Note that if the subject had the scans for both of her eyes, these scans were put in the same fold. Then, for each fold, the images in that fold were used as the test set and those in the remaining two folds were used for training the network. For each run, the training set contained 735 OCT images of 15 eyes (49 images from each eye), on which the network weights were learned by backpropagation; the validation set contained 245 OCT images

of five eyes, which were used for early stopping; and the test set contained 490 OCT images of 10 eyes, which were used for final evaluation of the model.

HFL segmentations were evaluated visually and quantitatively on the test set images. For each image, we found the number of true positive (TP), false positive (FP), and false negative (FN) pixels, comparing the estimated maps with manual annotations, and calculated the precision = TP / (TP + FP), recall = TP / (TP + FN), and f-score. These metrics were averaged over the test set images in all folds. Furthermore, for *FourierNet* and the comparison algorithms, the networks were trained three times and the average metrics of these three runs together with their standard deviations were reported.

## 4.3 Comparisons

*FourierNet* was compared against five algorithms: *Standard-Unet*, *Standard-RelayNet*, *Standard-FCNN*, *Directional-Cascaded*, and *Directional-Unet*. First, *Standard-Unet* was the baseline algorithm that also used the same encoder and decoder architectures given in Figure 3.3. It took a standard scan as its input and output an HFL segmentation map. It used neither Fourier descriptors to quantify the HFL shape nor a cascaded design to learn these Fourier descriptors. Second, *Standard-RelayNet* used the RelayNet model proposed in [25]. This network included a repeated application of convolution, batch normalization, ReLU, and max pooling layers in a U-shaped architecture. Its decoder upsampled feature maps using indices passed from the pooling layers and concatenated them with the corresponding feature maps in its encoder. The network was trained by jointly optimizing the weighted cross entropy and the Dice loss. The third comparison algorithm was *Standard-FCNN*. It had the same encoder and decoder architectures given in Figure 3.3 except that all convolution operators were defined in the Fourier domain as explained in [38]. It is worth noting that, as opposed to this algorithm, our model does not perform any of its network operations in the Fourier domain but uses Fourier series to represent the shape prior of the HFL.

Fourth, *Directional-Cascaded*, employed directional scans in a cascaded network design given in Figure 3.4. Similar to *FourierNet*, it took a standard scan as its input and formulated HFL segmentation as concurrent learning of regression and classification tasks. However, in its regression task, instead of estimating the Fourier descriptor maps, it estimated the four directional scans from the standard scan employing a multi-task network. In its classification task, it predicted an HFL segmentation map from these four estimated scans along with the standard scan. This comparison algorithm was used to understand the effectiveness of using an intermediate task that estimates the Fourier descriptors in a cascaded FCN design.

Fifth, *Directional-Unet* took both standard and directional scans as its inputs and output an HFL segmentation map without using a cascaded design. It also used the encoder and decoder architectures given in Figure 3.3. It was used to understand a target performance when directional OCT was available, which is not the case in common practice.

To summarize, *FourierNet* input only the standard scans and needed the directional scans neither as its input nor in its training. Likewise, the first three comparison methods, *Standard-Unet*, *Standard-RelayNet*, *Standard-FCNN*, did not make use of the directional scans. On the other hand, *Directional-Unet* input both the standard and directional scans. *Directional-Cascaded* did not take the directional scans as its input but used them as the outputs of its intermediate regression task, and thus, needed them for the network training. The last two comparison methods used the scans aligned with the built-in auto-alignment feature of the OCT device.

# Chapter 5

# Results and Discussion

This chapter is based on our manuscript submitted for publication [1].

The quantitative test set results obtained by the algorithms were reported in Table 5.1. This table revealed that although *Standard-Unet* and *Standard-RelayNet* led to the highest precision and recall, respectively, they could not reach the highest f-score. On the other hand, *FourierNet*, which used only the standard scan as its input, yielded the highest f-score compared to the four algorithms that also took the standard scans as their inputs. These results were statistically significant with $p < 0.05$ when the paired-sample t-test was applied. This indicated the effectiveness of representing the prior knowledge on the HFL shape in a cascaded network design. Moreover, *FourierNet* gave an f-score as high as that of *Directional-Unet*, which used four more directional scans in addition to the standard one as its inputs. The t-test revealed that there was no statistically significant difference between the f-scores of *FourierNet* and *Directional-Unet* with $p = 0.05$. It is worth noting that the *Directional-Unet* comparison algorithm was used to understand a target performance when directional OCT was available. Hence, the proposed *FourierNet* model aimed to achieve similar segmentation results as *Directional-Unet*, and did not target to outperform this comparison algorithm. *Directional-Cascaded* also used a cascaded design similar to ours but with different intermediate regression tasks. Table 5.1 also revealed that this way

Table 5.1: Cross-validation test set results with postprocessing. Averages and standard deviations across three runs. Significantly best metrics ($p = 0.05$) are indicated in bold.

| | **With postprocessing** | | |
|---|---|---|---|
| | Precision | Recall | F-score |
| *FourierNet* | 82.92 ± 0.44 | 87.74 ± 0.29 | **84.93 ± 0.20** |
| Standard-Unet | **84.71 ± 0.68** | 83.65 ± 1.12 | 83.82 ± 0.27 |
| Standard-RelayNet | 76.33 ± 0.41 | **93.00 ± 0.76** | 83.49 ± 0.11 |
| Standard-FCNN | 82.91 ± 0.77 | 85.80 ± 0.88 | 83.97 ± 0.13 |
| Directional-Cascaded | 83.75 ± 0.73 | 85.34 ± 1.16 | 84.02 ± 0.41 |
| Directional-Unet | 83.62 ± 0.17 | 86.99 ± 0.41 | **84.91 ± 0.13** |

Table 5.2: Cross-validation test set results without postprocessing. Averages and standard deviations across three runs. Significantly best metrics ($p = 0.05$) are indicated in bold.

| | **Without postprocessing** | | |
|---|---|---|---|
| | Precision | Recall | F-score |
| *FourierNet* | 82.70 ± 0.44 | 87.74 ± 0.29 | **84.80 ± 0.20** |
| Standard-Unet | **84.39 ± 0.72** | 83.66 ± 1.11 | 83.63 ± 0.25 |
| Standard-RelayNet | 76.05 ± 0.53 | **93.00 ± 0.76** | 83.29 ± 0.08 |
| Standard-FCNN | 82.43 ± 0.90 | 86.25 ± 1.10 | 83.79 ± 0.16 |
| Directional-Cascaded | 83.48 ± 0.60 | 85.34 ± 1.16 | 83.86 ± 0.36 |
| Directional-Unet | 83.31 ± 0.26 | 87.00 ± 0.42 | **84.73 ± 0.08** |

of defining an intermediate task was less effective than using the intermediate task of Fourier descriptor estimation. This might be attributed to the following: Estimating an image from another one might require more complex models for this particular task. In contrast, the Fourier descriptor representation, which summarized the shape prior of the HFL, facilitated defining a more effective and more compact learning task.

Table 5.3: Standard deviations calculated on the cross-validation test set f-scores across runs, subjects, and images.

| | Runs | Subjects | Images |
|---|---|---|---|
| *FourierNet* | 0.20 | 3.28 | 2.97 |
| Standard-Unet | 0.27 | 3.54 | 3.21 |
| Standard-RelayNet | 0.11 | 2.72 | 2.50 |
| Standard-FCNN | 0.13 | 3.71 | 3.50 |
| Directional-Cascaded | 0.41 | 3.39 | 3.16 |
| Directional-Unet | 0.13 | 3.26 | 2.99 |

Table 5.2 reported the metrics calculated on the estimated segmentation maps before postprocessing. Eliminating noisy regions similar to the one shown in Figure 3.5 slightly decreased the number of FPs, which in turn slightly increased the precision. No decrease in the recall indicated that no TPs are eliminated by this postprocessing (note that the only way to decrease the recall was to decrease the number of TPs since this postprocessing could not decrease the number of FNs). The slight increase was observed for all methods when postprocessing was used. Table 5.1 and Table 5.2 reported the standard deviations across three runs. Table 5.3 presented those calculated on the cross-validation test set f-scores across 30 subjects and all individual images. As seen in this table, difficulty variations existed among images and subjects, but for all algorithms.

The visual results obtained on four exemplary test images were shown in Figures 5.1-5.4. They were also consistent with the quantitative results. In our experiments, we observed that *FourierNet* led to enhanced performance at the fovea (see the center zone of the two OCT images marked with red and yellow in Figure 5.1 and Figure 5.2, respectively) as well as at the outer regions. It also tended to generate continuous segmentation maps of the HFL compared to its counterparts (see the OCT image in Figure 5.3 marked with green). Additionally, compared to the comparison algorithms, the HFL thickness in the estimated maps of *FourierNet* was more consistent with the HFL thickness calculated with reference to the manual HFL annotations (see Figure 5.4).

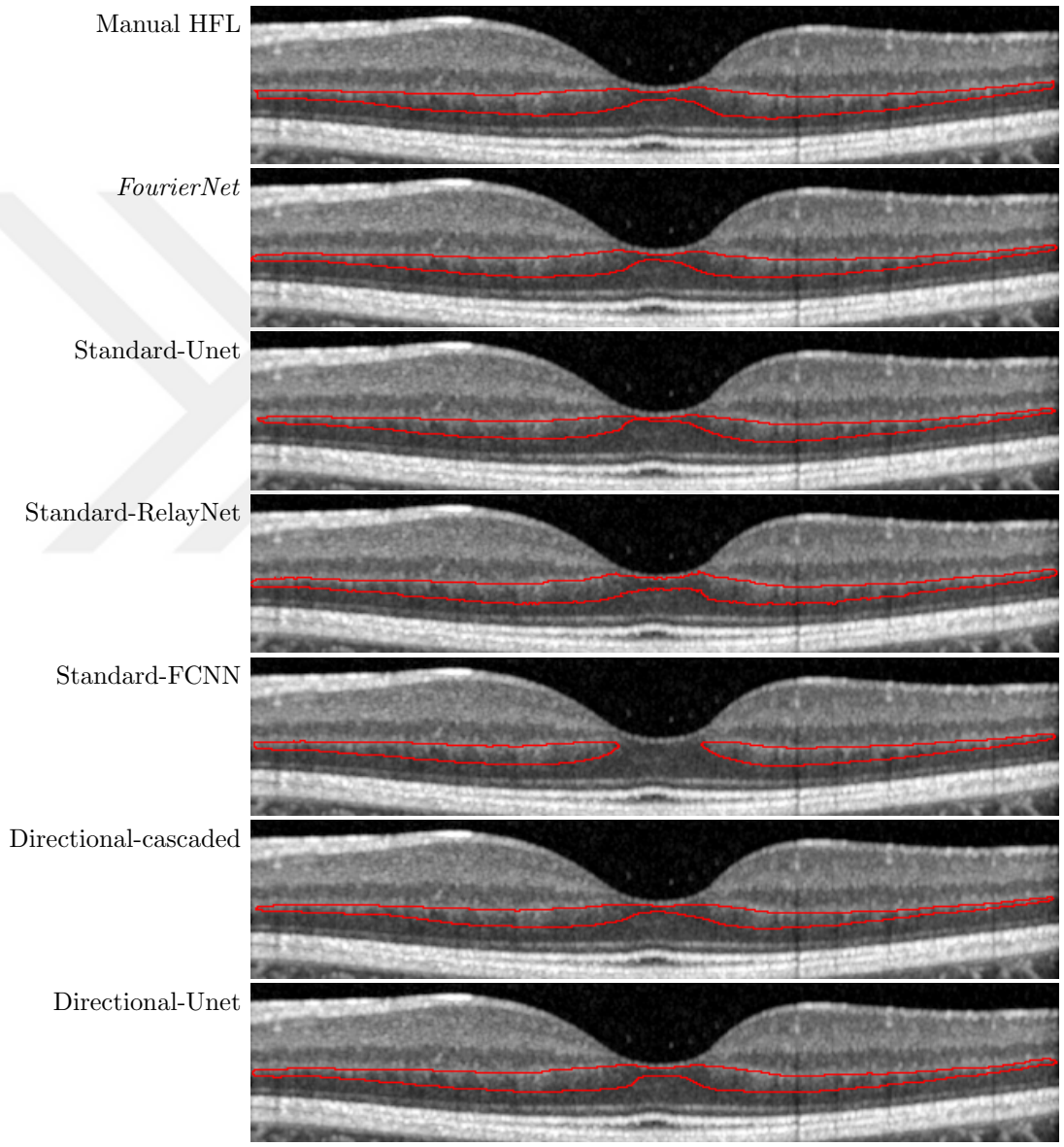



Figure 5.1: Exemplary test set image showing the manual HFL annotations and the results of all algorithms. All results are embedded on the standard OCT images.

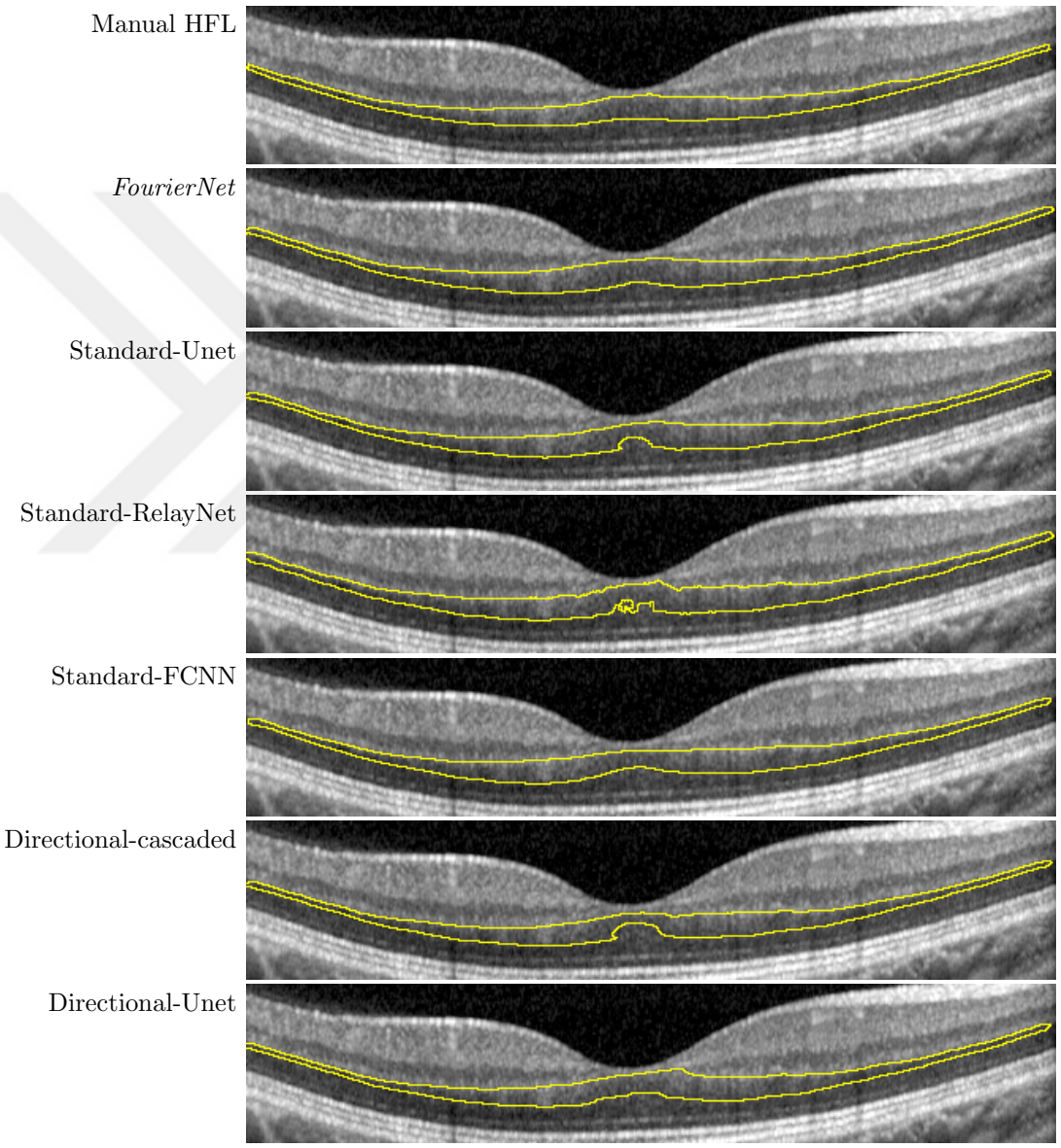



Figure 5.2: Exemplary test set image showing the manual HFL annotations and the results of all algorithms. All results are embedded on the standard OCT images.

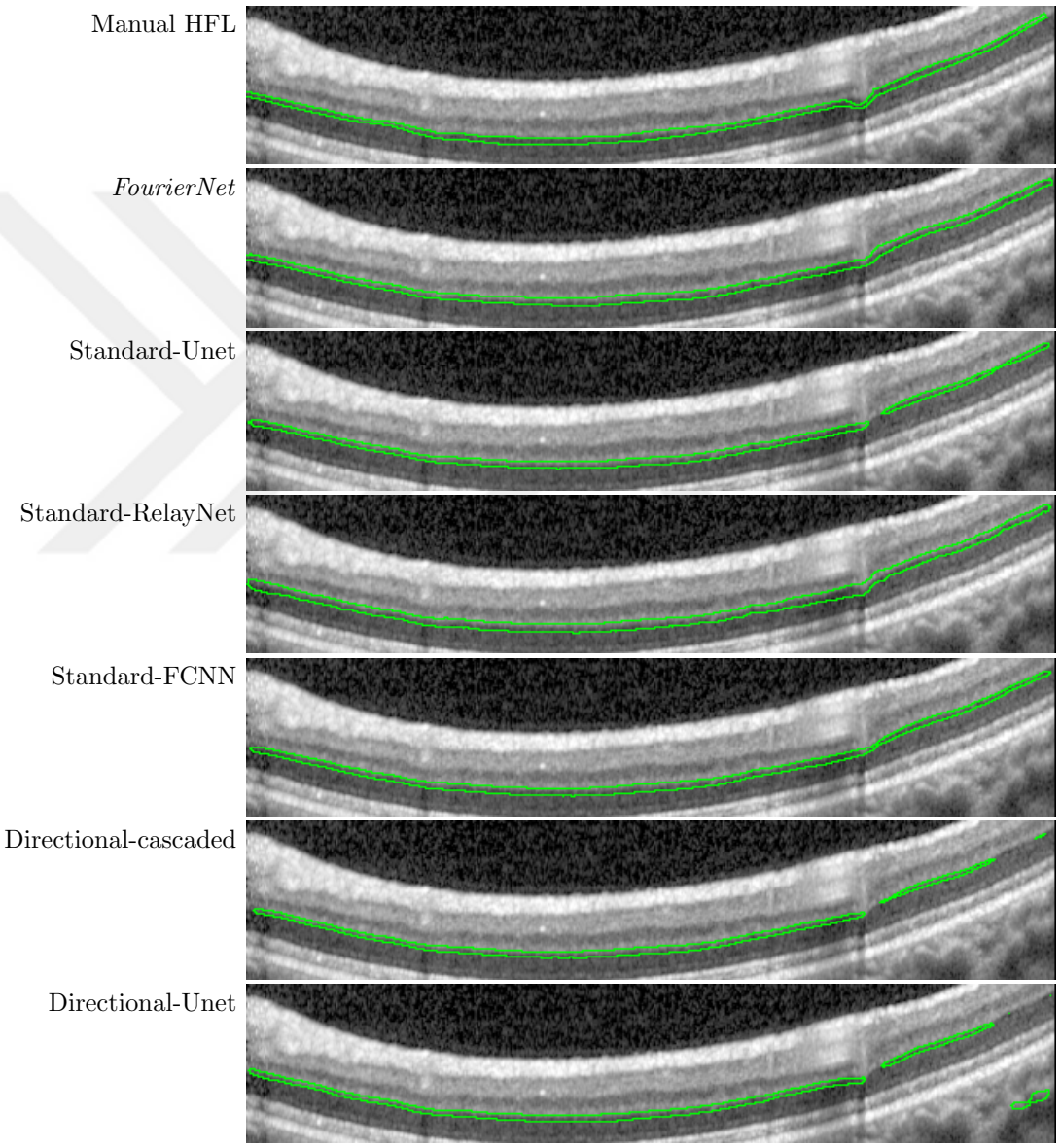



Figure 5.3: Exemplary test set image showing the manual HFL annotations and the results of all algorithms. All results are embedded on the standard OCT images.

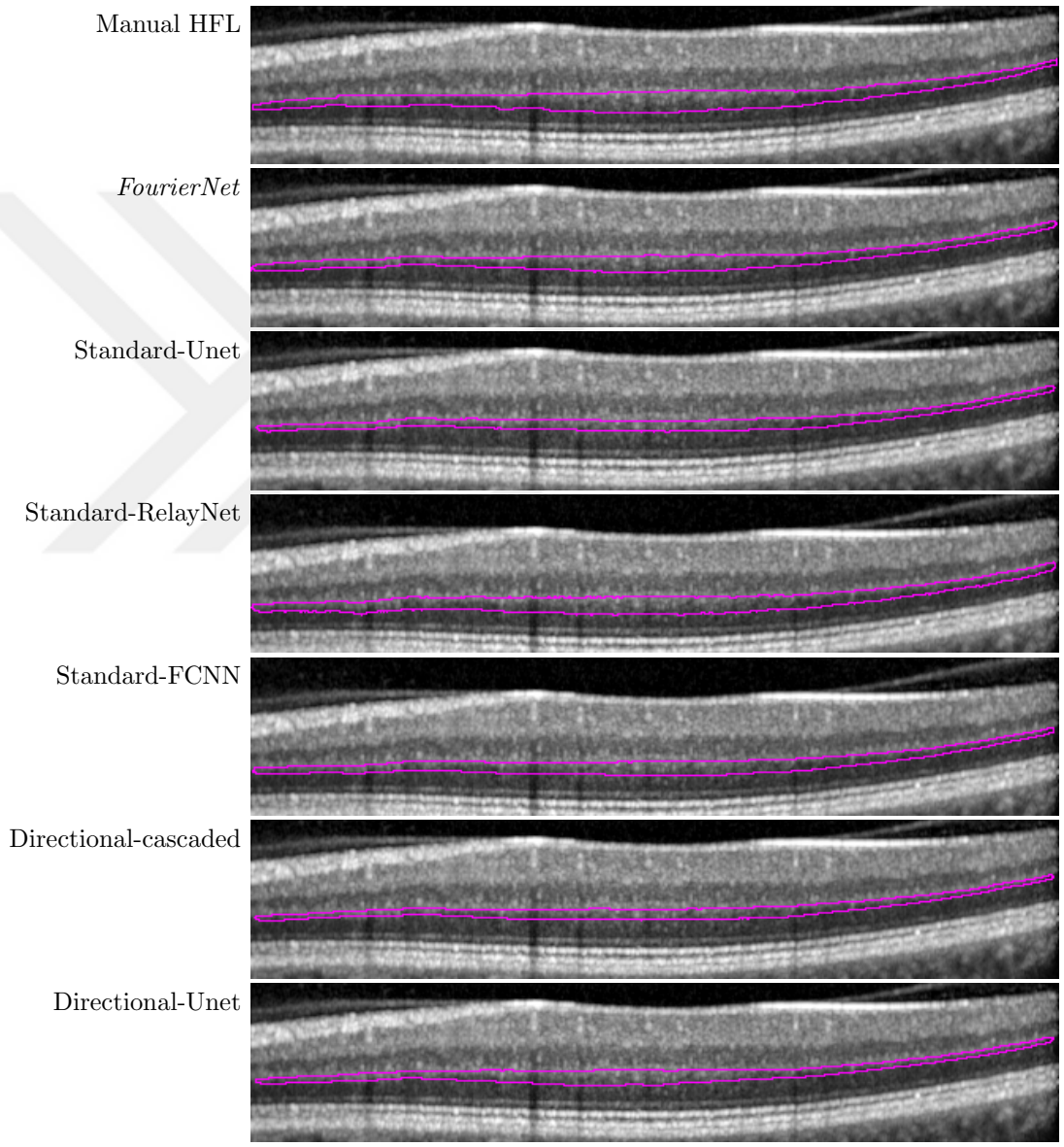



Figure 5.4: Exemplary test set image showing the manual HFL annotations and the results of all algorithms. All results are embedded on the standard OCT images.

The quantitative metrics were separately calculated for each sector in the ETDRS grid by considering only the pixels falling in the corresponding sector. The sector-wise average test set f-scores were reported in Table 5.4. This demonstrated that compared to *Standard-UNet* and *Directional-Cascaded*, *FourierNet* yielded statistically better f-scores ($p < 0.05$) in all sectors. *Standard-RelayNet* estimated thicker HFL regions, which caused it to find more TPs in the central field and the inner quadrantal areas at the expense of finding many FPs in the other regions. This was reflected by the high f-scores for the central field and the inner quadrantal areas but the low f-scores for the outer quadrantal areas and the lower precision in the overall f-score (Tables 5.1, 5.2, and 5.4). Although *FourierNet* and *Standard-FCNN* resulted in comparable results for the inner quadrantal areas, our model yielded statistically better f-scores for the outer quadrantal areas. Compared to *Directional-Unet*, which also used directional scans, *FourierNet* led to statistically better f-scores for two, similar f-scores for three, and worse f-scores for four sectors but without using any extra directional scans ($p < 0.05$). Table 5.5 presented the total volume and average thickness of the HFL calculated with reference to manual annotations. It also reported the absolute errors between the actual values and those calculated with reference to the estimated segmentation maps. This table showed that the proposed *FourierNet* model and the targeted *Directional-Unet* algorithm yielded similar estimation errors. There was no statistically significant difference with $p = 0.05$.

Table 5.4: Cross-validation test set results. Sector-wise average f-scores and their standard deviations across three runs. For each sector, the bold font indicates the significantly best methods among the first five algorithms ($p < 0.05$). The font color is used to identify whether these best methods are statistically better (red), similar (blue), or worse (black) than Directional-Unet.

| | Central | Inner superior | Inner inferior | Inner nasal | Inner temporal |
|---|---|---|---|---|---|
| *FourierNet* | 85.3 ± 0.1 | **90.1 ± 0.3** | **90.5 ± 0.2** | **90.8 ± 0.1** | **90.6 ± 0.2** |
| Standard-Unet | 84.6 ± 0.6 | 89.2 ± 0.6 | 90.0 ± 0.4 | 90.5 ± 0.2 | 89.8 ± 0.3 |
| Standard-RelayNet | **86.1 ± 0.4** | 89.4 ± 0.2 | **90.6 ± 0.1** | 90.2 ± 0.3 | **90.8 ± 0.1** |
| Standard-FCNN | 85.4 ± 0.3 | **89.9 ± 0.2** | **90.4 ± 0.3** | **90.7 ± 0.1** | **90.4 ± 0.4** |
| Directional-Cascaded | 83.9 ± 0.8 | 88.7 ± 0.3 | 89.6 ± 0.6 | 90.0 ± 0.4 | 89.9 ± 0.3 |
| Directional-Unet | 86.6 ± 0.9 | 90.0 ± 0.1 | 91.2 ± 0.2 | 90.9 ± 0.0 | 91.3 ± 0.0 |

| | Outer superior | Outer inferior | Outer nasal | Outer temporal |
|---|---|---|---|---|
| *FourierNet* | **84.7 ± 0.4** | **84.4 ± 0.2** | **85.0 ± 0.1** | **84.6 ± 0.3** |
| Standard-Unet | 83.4 ± 0.3 | 83.1 ± 0.1 | 84.3 ± 0.1 | 83.5 ± 0.4 |
| Standard-RelayNet | 82.1 ± 0.2 | 83.2 ± 0.1 | 83.6 ± 0.1 | 83.2 ± 0.3 |
| Standard-FCNN | 84.0 ± 0.3 | 84.0 ± 0.7 | 84.5 ± 0.1 | 84.0 ± 0.3 |
| Directional-Cascaded | 83.6 ± 0.3 | 84.0 ± 0.6 | 84.3 ± 0.4 | 83.6 ± 0.4 |
| Directional-Unet | 84.3 ± 0.7 | 84.8 ± 0.1 | 84.8 ± 0.2 | 84.0 ± 0.1 |

Table 5.5: First row: Total volume and average thickness of the HFL within the ETDRS grid. They were calculated with reference to the manual annotations. Other rows: Absolute errors between the actual values and those calculated with reference to the estimated segmentation maps for cross-validation test set images.

| | Total volume ($mm^3$) | Average thickness ($\mu$m) |
|---|---|---|
| Manual annotations | 0.743 ± 0.082 | 26.464 ± 2.908 |
| | Estimation error | Estimation error |
| *FourierNet* | 0.066 ± 0.053 | 2.345 ± 1.887 |
| Standard-Unet | 0.073 ± 0.048 | 2.596 ± 1.706 |
| Standard-RelayNet | 0.159 ± 0.084 | 5.660 ± 3.002 |
| Standard-FCNN | 0.068 ± 0.061 | 2.423 ± 2.156 |
| Directional-Cascaded | 0.096 ± 0.065 | 3.411 ± 2.303 |
| Directional-Unet | 0.066 ± 0.058 | 2.339 ± 2.055 |

Table 5.6: Cross-validation test set results for different $N$ values. Averages and standard deviations across three runs.

| | Precision | Recall | F-score |
|---|---|---|---|
| $N = 1$ | 82.92 ± 0.44 | 87.74 ± 0.29 | 84.93 ± 0.20 |
| $N = 2$ | 83.38 ± 0.98 | 86.27 ± 0.71 | 84.48 ± 0.16 |
| $N = 3$ | 82.46 ± 0.41 | 87.60 ± 0.23 | 84.60 ± 0.11 |
| $N = 4$ | 83.01 ± 0.52 | 86.87 ± 0.79 | 84.54 ± 0.27 |
| $N = 5$ | 82.00 ± 0.10 | 87.71 ± 0.66 | 84.40 ± 0.31 |

*FourierNet* has one external parameter $N$, which is the number of Fourier coefficients in the truncated expansion of the distance-to-center function $\xi(l_x)$. First Fourier descriptors contain more general information about a function whereas latter ones contain its finer details. Table 5.6 showed the test set performance for different $N$ values. This table revealed that the first Fourier descriptor carried sufficient information to capture the HFL contour shape and the latter dimensions did not bring about additional information.

# Chapter 6

# Conclusion

This chapter is based on our manuscript submitted for publication [1].

This thesis proposed and showed a shape-preserving network for automated HFL segmentation in standard OCT scans. This network, which we named *FourierNet*, relied on benefiting the shape prior of the HFL in its training. To this end, it proposed to quantify the shape prior by extracting Fourier descriptors on the HFL contours and introduced a new cascaded network design that learned these descriptors concurrently with the segmentation task. We tested the *FourierNet* model for HFL segmentation on 1470 OCT images of 30 eye scans. Our experiments revealed that *FourierNet* achieved HFL segmentation on standard scans at least with the performance that would be obtained when directional OCT was used. As a result, this model proved to be a useful tool for HFL segmentation by reducing the necessity to perform directional OCT imaging.

This work focused on segmenting the HFL in healthy macula that is devoid of any retinal disease. To understand the potential use of the proposed segmentation model in the analysis of HFL-affected retinal pathologies, we further conducted a preliminary experiment on four exemplary images showing retinal involvement (Figures 6.1 and 6.2). The two images in Figure 6.1 were acquired from patients with diabetic macular edema where edematous regions disturbed the normal

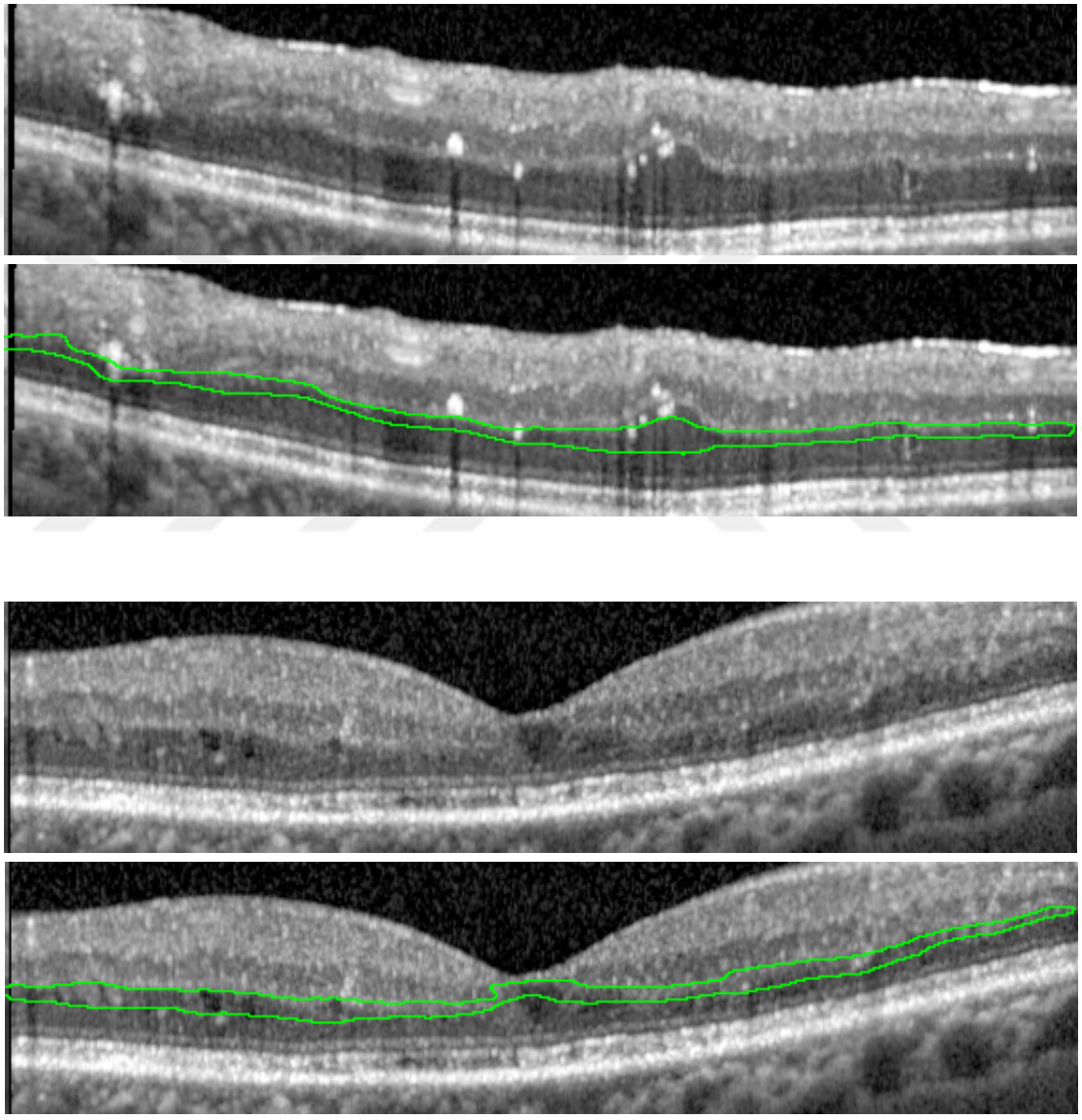


Figure 6.1: Visual results obtained on exemplary images showing retinal involvement. Original images with diabetic macular edema and HFL regions estimated by *FourierNet*.

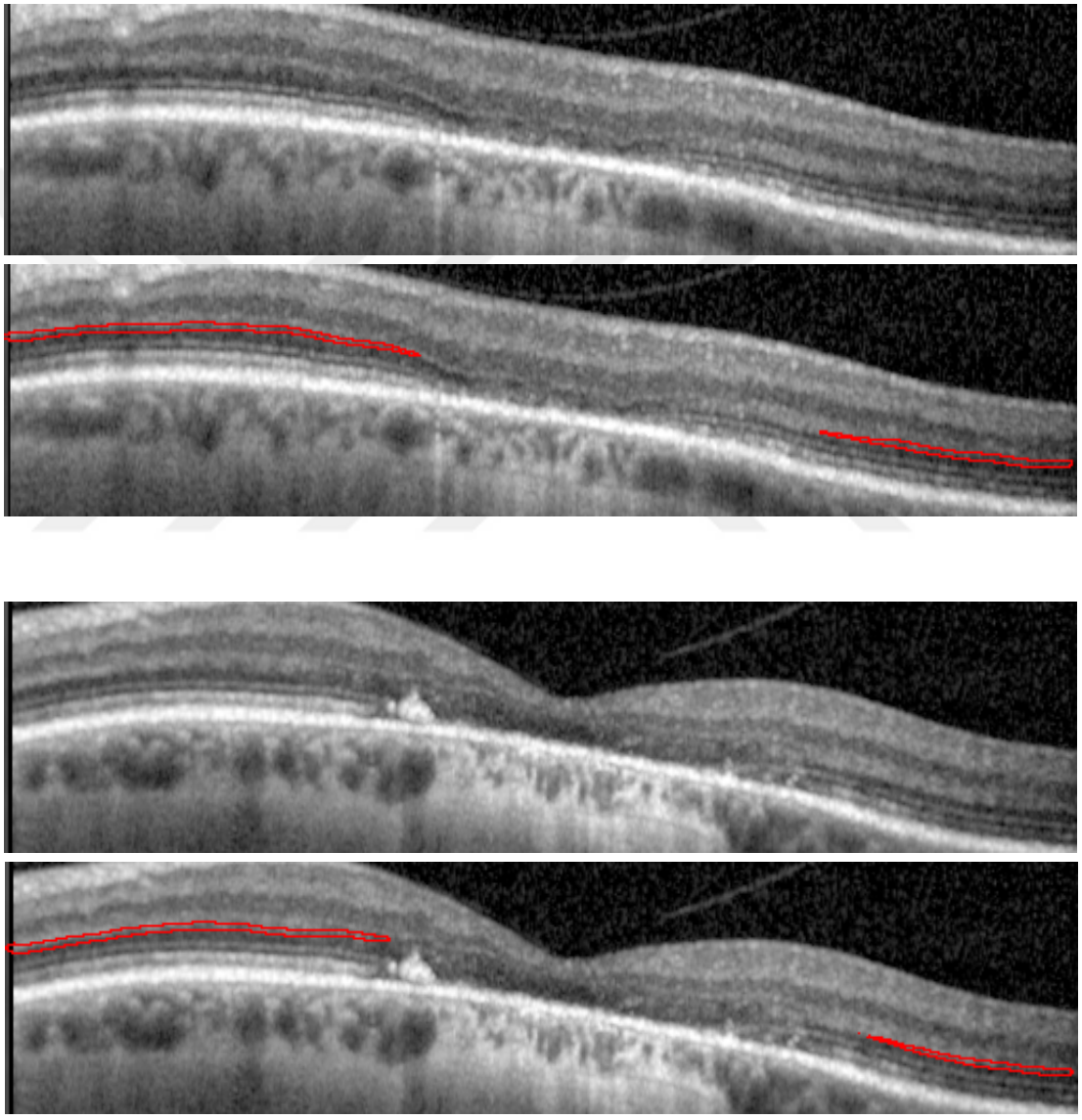


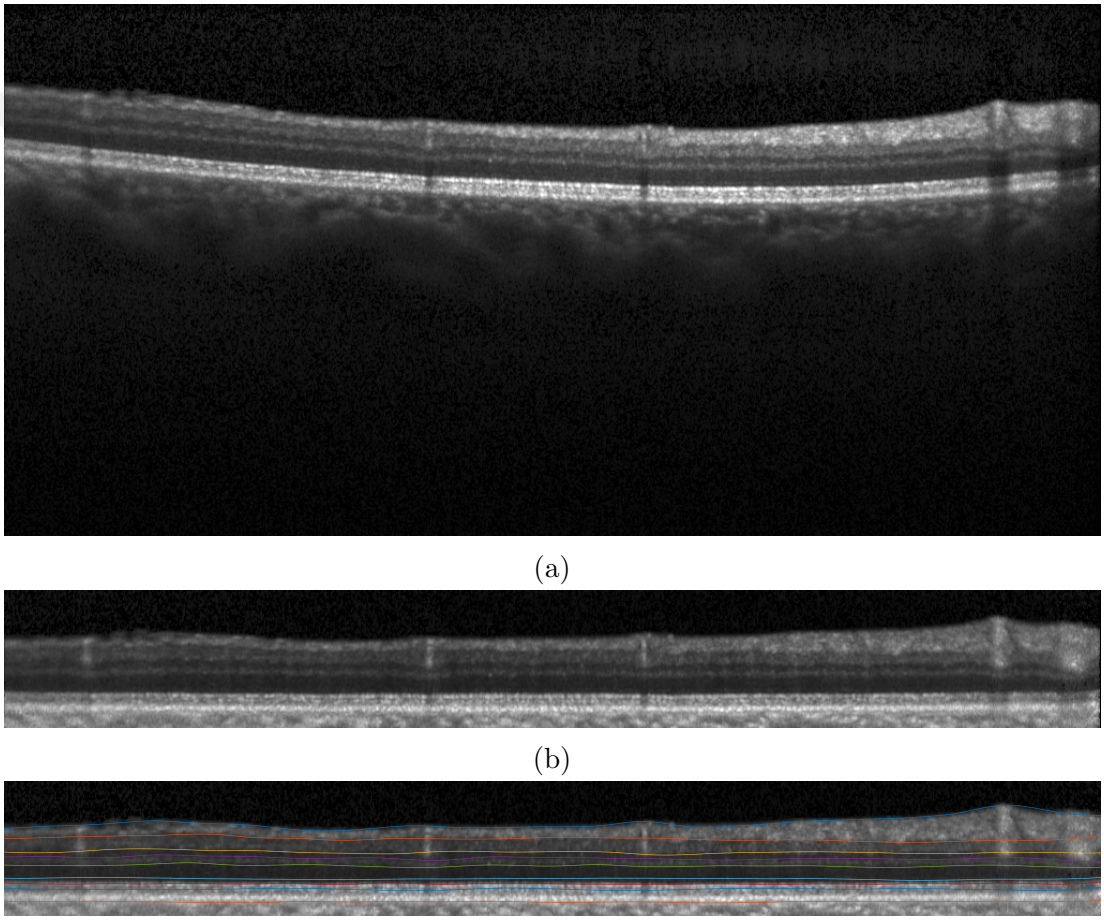
Figure 6.2: Visual results obtained on exemplary images showing retinal involvement. Original images with hereditary retinal dystrophy and HFL regions estimated by *FourierNet*.

structure of the HFL. The other two images in Figure 6.2 were taken from a patient within the spectrum of hereditary retinal dystrophy where the HFL is presumed to be defective in the fovea and thinner than normal in the outer regions. As seen in Figure 6.1, *FourierNet* was able to include edematous regions in the segmented HFL to some extent. For the images of retinal dystrophy in Figure 6.2, *FourierNet* led to thinner segmentations for outer regions and did not identify any HFL in the fovea. Although these results are promising, especially considering the fact that the model did not see any retinal pathologies in its training, it deserves a systematic and larger-scale experimentation. This requires systematic data collection for the selected retinal pathology and detailed analyses for correlating the identified HFL regions and the severity of these pathologies. Further investigation of this possibility is considered as future work.

The *FourierNet* model presented a shape-preserving network designed to model the shape prior of the HFL in healthy macula. To understand the effectiveness of this shape-preserving network for segmenting other retinal layers, we also conducted experiments on the dataset including scans of healthy subjects and those diagnosed with multiple sclerosis (MS) [44]. The dataset consists of 35 standard OCT volumes (14 healthy subjects and 21 abnormal subjects with MS) obtained using Heidelberg Spectralis OCT device. Each volume includes 49 scans with the size of 496×1024. The eight retinal layers on each scan were manually annotated with nine boundaries. The names of these layers are retinal nerve fiber layer (RNFL), ganglion cell layer and inner plexiform layer (GCL+IPL), inner nuclear layer (INL), out plexiform layer (OPL), outer nuclear layer (ONL), inner segment (IS), outer segment (OS), and retinal pigment epithelium (RPE), respectively. To prepare the dataset for training, all scans were flattened along the RPE layer and cropped with the size of 128×1024. An example OCT image, its flattened and crop version along with annotations are shown in Figure 6.3. The retinal layers in patients with MS tend to be thinner [45], and thus, segmentation of them becomes harder. Note that the retinal layers of MS subjects typically had less severe damage when compared to the HFL disturbance in the presence of diabetic macular edema or retinal dystrophy.

For this dataset, we trained multiple *FourierNet* models to segment the layers one by one and compared its results with those of the baseline (*Standard-Unet*), which did not involve any shape-preserving property in its design. Table 6.1 presented the average f-scores of three runs. This table revealed that *FourierNet* yielded significantly better results for six layers, similar results for one layer, and worse results for another layer ($p < 0.05$). We also analyzed the visual results obtained on two exemplary images from heathy subjects and patients with MS. In order to obtain visual maps, the layerwise predictions of the algorithms were combined using a postprocessing method. Basically, the upper and lower retinal boundaries were detected using the posterior probabilities greater than 0.5. Then, the pixels remaining between the upper and lower boundaries were classified with the layer having maximum posterior probability for each pixel. Finally, intertwined regions were smoothed topologically. In the results, we observed that *FourierNet* defined the boundaries better, especially in the regions where the contrast difference between adjacent layers was not visible well (see Figure 6.4). It also generated better predictions when these was a significant change in retinal morphology, as in the case of patients with MS (see Figure 6.5).

*FourierNet* led to larger improvements on the baseline results for HFL segmentation. However, the performance of the proposed model was not significantly better for two retinal layers in this dataset. Here it is worth noting two issues: First, this might be attributed to the fact that the layers in this new dataset can be delineated on standard scans by human annotators whereas it is challenging for them to delineate the HFL using only the standard scans when directional OCT is not available. This might indicate the effectiveness of using the shape prior in the network design, as the proposed *FourierNet* model does, when incomplete information is present.



(a)

(b)

(c)

Figure 6.3: Processing stages on an example of OCT image. (a) The original OCT image, (b) flattened and cropped version of the original OCT image, and (c) boundary annotations of eight layers.

Table 6.1: Test set results on the dataset of [44]. Significantly better metrics ($p = 0.05$) are indicated in bold.

| | RNFL | GCL+IPL | INL | OPL |
|---|---|---|---|---|
| *FourierNet* | **92.93 ± 0.02** | **94.36 ± 0.06** | **87.65 ± 0.03** | **90.09 ± 0.06** |
| Standard-Unet | 92.64 ± 0.25 | 93.50 ± 0.29 | 86.87 ± 0.37 | 89.95 ± 0.08 |
| With topology correction | | | | |
| *FourierNet* | **93.16 ± 0.05** | **94.58 ± 0.09** | **87.76 ± 0.07** | **90.18 ± 0.12** |
| Standard-Unet | 92.75 ± 0.07 | 93.96 ± 0.15 | 87.22 ± 0.19 | 90.10 ± 0.02 |

| | ONL | IS | OS | RPE |
|---|---|---|---|---|
| *FourierNet* | **94.83 ± 0.02** | 86.81 ± 0.07 | **87.43 ± 0.33** | 91.45 ± 0.13 |
| Standard-Unet | 94.76 ± 0.11 | **87.09 ± 0.11** | 86.79 ± 0.23 | 91.46 ± 0.15 |
| With topology correction | | | | |
| *FourierNet* | **94.97 ± 0.02** | 87.22 ± 0.10 | 87.71 ± 0.25 | **91.59 ± 0.12** |
| Standard-Unet | 94.86 ± 0.09 | **87.31 ± 0.10** | **87.88 ± 0.19** | 91.50 ± 0.16 |

Second, for both *FourierNet* and the baseline, we used simple encoder and decoder architectures in the network design and followed simple training and testing procedures. However, it is usually possible to increase a model's performance with data augmentation, additional training regularizations, more complex backbone architectures, and additional postprocessing such as topology correction. For the new dataset, we exploited a possibility of using topology correction to show that higher performances could be obtained for our model and also for the baseline. This topology correction combined the segmentation maps predicted by the layer-wise networks and corrected ambiguous predictions as follows: First, a mask was created combining all layer predictions and including pixels in between them. Within this mask, unambiguous pixels, for which only one network made a layer prediction, were identified as layer seeds. Then, these seeds were iteratively grown on the other pixels within the mask. Each iteration selected a non-seed pixel with the maximum number of adjacent seed pixels belonging to a single layer type, and included this selected pixel to the corresponding layer seed. This region growing continued until no non-seeds pixels were left. As expected, topology correction increased the performance of both models but did not change the conclusion that *FourierNet* resulted in a significantly better performance for more layers than the corresponding baseline (Table 6.1). Exploring the shape-preserving network designs with more complex architectures and using more complex training procedures is considered as another future work.

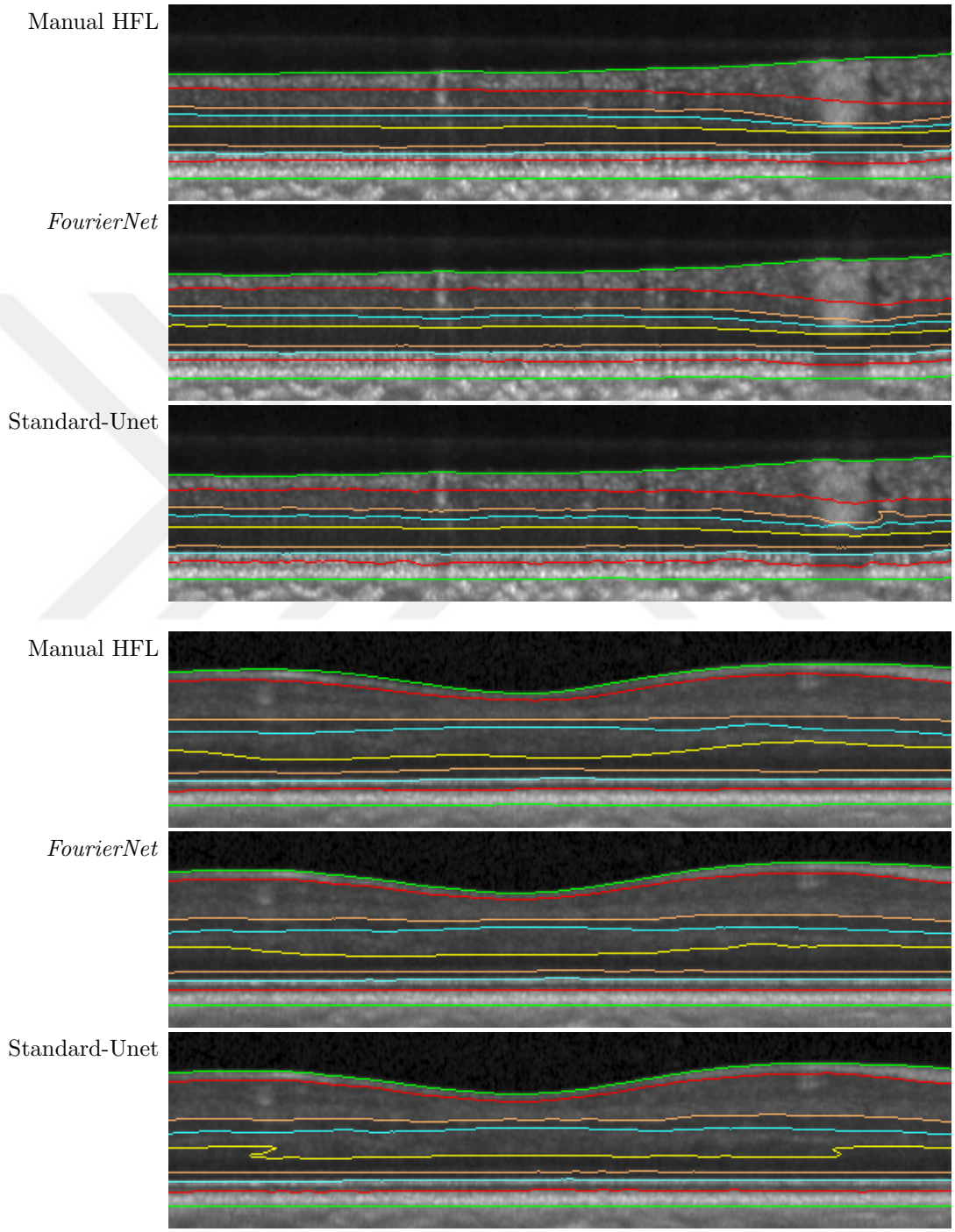



Figure 6.4: Two exemplary test set images obtained from healthy subjects. For each image, the manual HFL annotations and the results of the proposed and the baseline algorithms are shown. All results are embedded on the flattened and cropped OCT images. Different colors are used to define the boundaries between retinal layers.

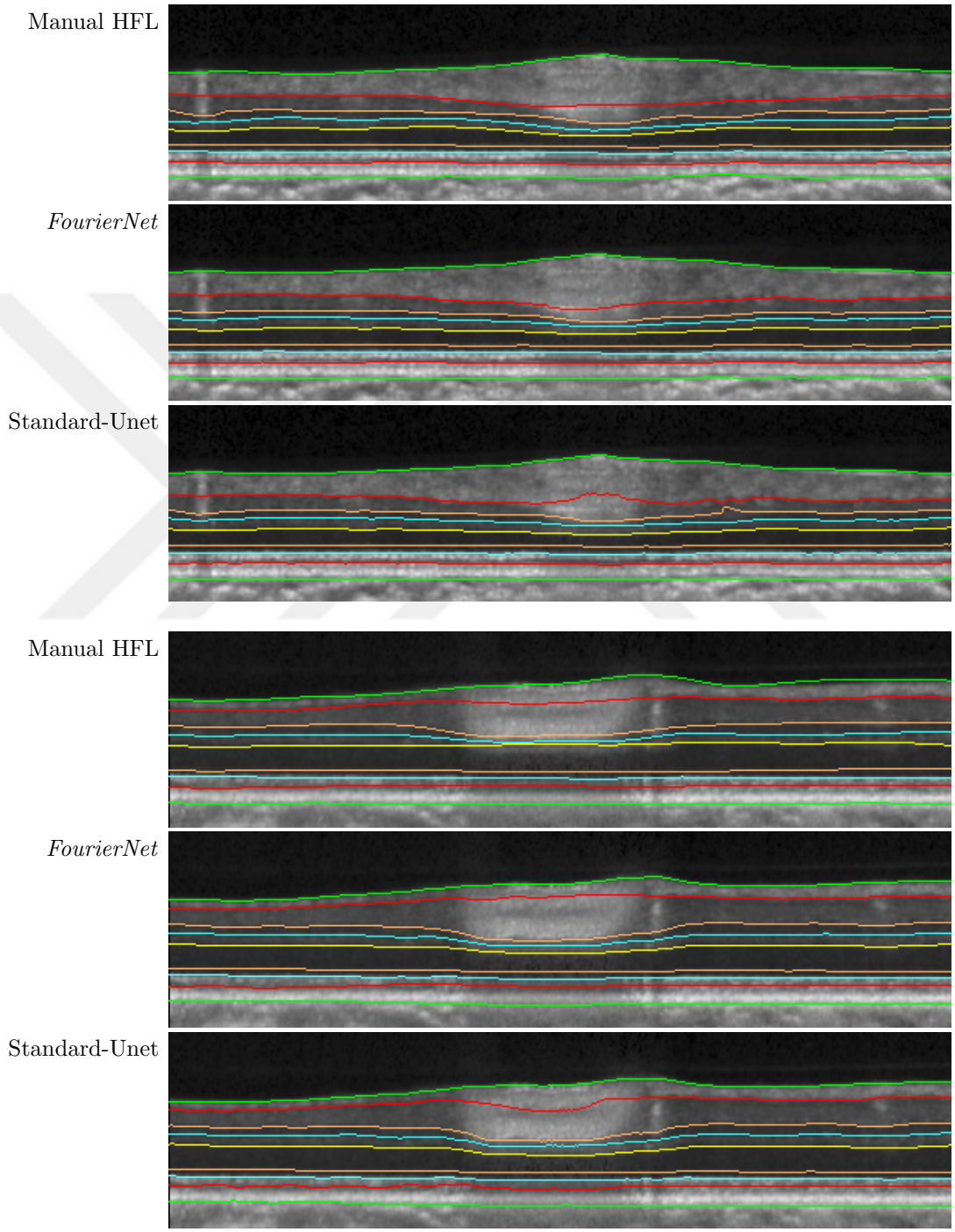



Figure 6.5: Two exemplary test set images obtained from subjects with MS. For each image, the manual HFL annotations and the results of the proposed and the baseline algorithms are shown. All results are embedded on the flattened and cropped OCT images. Different colors are used to define the boundaries between retinal layers.

# Bibliography


[1] S. Cansiz, C. Kesim, S.N. Bektas, Z. Kulali, M. Hasanreisoglu, and C. Gunduz-Demir, "FourierNet: Shape-Preserving Network for Henle's Fiber Layer Segmentation in Optical Coherence Tomography Images," 2022, *arXiv:2201.06435*. [Online]. Available: https://arxiv.org/abs/2201.06435.

[2] B. J. Lujan, A. Roorda, R. W. Knighton, and J. Carroll, "Revealing Henles fiber layer using spectral domain optical coherence tomography," *Invest. Ophthalmol. Vis. Sci.*, vol. 52, no. 3, pp. 1486–1492, 2011.

[3] J. Tian *et al.*, "Performance evaluation of automated segmentation software on optical coherence tomography volume data," *J. Biophotonics*, vol. 9, no. 5, pp. 478–489, 2016.

[4] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Express*, vol. 8, no. 5, pp. 2732–2744, 2017.

[5] M. Pekala, N. Joshi, T. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina, "Deep learning based retinal OCT segmentation," *Comput. Biol. Med.*, vol. 114, p. 103445, 2019.

[6] R. L. Cosgriff, "Identification of shape," Ohio State Univ. Res. Foundation, Columbus Rep. 820-11 (39–57), 1960.

[7] D. Zhang and G. Lu, "A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval," *J. Vis. Comm. Image Repr*, vol. 14, pp. 39–57, 2003.

[8] F. Larsson and M. Felsberg, "Using Fourier descriptors and spatial models for traffic sign recognition," in *Scandinavian Conf. Image Analysis*, 2011, pp. 238–249.

[9] C. T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Computers*, vol. 100, pp. 269–281, 1972.

[10] Retinal-Layers, digital photograph, Heidelberg Engineering, accessed 22 August 2022, <https://business-lounge.heidelbergengineering.com/vg/en/news/news/know-your-retinal-layers-33401465>.

[11] M. A. Mayer, J. Hornegger, C. Y. Mardin, and R. P. Tornow, "Retinal nerve fiber layer segmentation on FD-OCT scans of normal subjects and glaucoma patients," *Biomed. Opt. Express*, vol. 1, no. 5, pp. 1358–1383, 2010.

[12] A. Mishra, A. Wong, K. Bizheva, and D.A. Clausi, "Intra-retinal layer segmentation in optical coherence tomography images," *Biomed. Opt. Express*, vol. 17, no. 26, pp. 23719–23728, 2009.

[13] J. Novosel, G. Thepass, H. G. Lemij, J. F. de Boer, K. A. Vermeer, and L. J. van Vliet, "Loosely coupled level sets for simultaneous 3D retinal layer segmentation in optical coherence tomography," *Med. Image Anal.*, vol. 26, no. 1, pp. 146–158, 2015.

[14] P. A. Dufour *et al.*, "Graph-based multi-surface segmentation of OCT data using trained hard and soft constraints," *IEEE Trans. Med. Imaging*, vol. 32, no. 3, pp. 531–543, 2012.

[15] J. Oliveira, S. Pereira, L. Gonçalves, M. Ferreira, and C.A. Silva, "Multi-surface segmentation of OCT images with AMD using sparse high order potentials," *Biomed. Opt. Express*, vol. 8, no. 1, pp. 281–297, 2017.

[16] M.K. Garvin, M.D. Abramoff, X. Wu, S.R. Russell, T.L. Burns, and M. Sonka, "Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images," *IEEE Trans. Med. Imaging*, vol. 28, no. 9, pp. 1436–1447, 2009.

[17] S.J. Chiu, X.T. Li, P. Nicholas, C.A. Toth, J.A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Biomed. Opt. Express*, vol. 18, no. 18, pp. 19413–19428, 2010.

[18] X. Chen, M. Niemeijer, L. Zhang, K. Lee, M. D. Abramoff, and M. Sonka, "Three-dimensional segmentation of fluid-associated abnormalities in retinal OCT: Probability constrained graph-search-graph-cut," *IEEE Trans. Med. Imaging*, vol. 31, no. 8, pp. 1521–1531, 2012.

[19] A. Fuller, R. Zawadzki, S. Choi, D. Wiley, J. Werner, and B. Hamann, "Segmentation of three-dimensional retinal image data," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1719–1726, 2007.

[20] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomed. Opt. Express*, vol. 6, no. 4, pp. 1172–1194, 2015.

[21] K. McDonough, I. Kolmanovsky, and I. V. Glybina, "A neural network approach to retinal layer boundary identification from optical coherence tomography images," in *Proc. IEEE Conf. Comput. Intell. Bioinform. Comput. Biol.*, 2015, pp. 1–8.

[22] A. Lang *et al.*, "Retinal layer segmentation of macular OCT images using boundary classification," *Biomed. Opt. Express*, vol. 4, no. 7, pp. 1133–1152, 2013.

[23] L. Fang, S. Li, D. Cunefare, and S. Farsiu, "Segmentation based sparse reconstruction of optical coherence tomography images," in *IEEE Trans. Med. Imaging*, vol. 36, no. 2, pp. 407–421, 2016.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2015, pp. 234–241.

[25] A. G. Roy *et al.*, "Relaynet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Express*, vol. 8, no. 8, pp. 3627–3642, 2017.

[26] S. Apostolopoulos, S. De Zanet, C. Ciller, S. Wolf, and R. Sznitman, "Pathological OCT retinal layer segmentation using branch residual U-shape networks," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2017, pp. 294–301.

[27] Z. Gu *et al.*, "Ce-net: Context encoder network for 2-D medical image segmentation," *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[28] X. Liu, J. Cao, T. Fu, Z. Pan, W. Hu, K. Zhang, and J. Liu, "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning," in *IEEE Access*, vol. 7, pp. 3046–3061, 2018.

[29] J. Kugelman, D. Alonso-Caneiro, S.A. Read, S.J. Vincent, and M.J. Collins, "Automatic segmentation of OCT retinal boundaries using recurrent neural networks and graph search," in *Biomed. Opt. Express*, vol. 9, no. 11, pp. 5759–5777, 2018.

[30] X. Liu *et al.*, "Automated layer segmentation of retinal optical coherence tomography images using a deep feature enhanced structured random forests classifier," *IEEE J. Biomed. Health. Inform.*, vol. 23, no. 4, pp. 1404–1416, 2018.

[31] B. Wang, W. Wei, S. Qiu, S. Wang, D. Li, and H. He, "Boundary aware U-Net for retinal layers segmentation in optical coherence tomography images," in *IEEE J. Biomed. Health. Inform.*, vol. 25, no. 8, pp. 3029–3040, 2021.

[32] B. Hassan *et al.*, "CDC-Net: Cascaded decoupled convolutional network for lesion-assisted detection and grading of retinopathy using optical coherence tomography (OCT) scans," *Biomed. Signal Process. Control*, vol. 70, pp. 103030, 2021.

[33] P. Zhong, J. Wang, Y. Guo, X. Fu, and R. Wang, "Multiclass retinal disease classification and lesion segmentation in OCT B-scan images using cascaded convolutional networks," *Appl. Opt.*, vol. 59, no.33, pp. 10312–10320, 2020.

[34] B. N. Anoop *et al.*, "A cascaded convolutional neural network architecture for despeckling OCT images," *Biomed. Signal Process. Control*, vol. 66, pp. 102463, 2021.

[35] Y. He *et al.*, "Towards topological correct segmentation of macular OCT from cascaded FCNs," *Fetal, Infant Opht. Med. Image Anal.*, pp. 202–209, 2017.

[36] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[37] W. Wang, J. Zhong, H. Wu, Z. Wen, and J. Qin, "RVSeg-Net: An efficient feature pyramid cascade network for retinal vessel segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2020, pp. 796–805.

[38] H. Pratt, B. M. Williams, F. Coenen, and Y. Zheng, "FCNN: Fourier convolutional neural networks," in *Proc. ECML/PKDD*, 2017, pp. 786–798.

[39] T. Highlander and A. Rodriguez, "Very efficient training of convolutional neural networks using fast Fourier transform and overlap-and-add," 2016, *arXiv:1601.06815*. [Online]. Available: https://arxiv.org/abs/1601.06815.

[40] L. Ch, B. Jiang, and Y. Mu, "Fast Fourier convolution," in *Proc. Adv. Neural Inf. Proc. Syst.*, 2020, pp. 4479–4488.

[41] R. Kondor, Z. Lin, and S. Trivedi, "Clebsch-Gordan nets: A fully Fourier space spherical convolutional neural network," in *Proc. Adv. Neural Inf. Proc. Syst.*, 2018, pp. 10138–10147.

[42] Y. Tsuzuku and I. Sato, "On the structural sensitivity of deep convolutional networks to the directions of Fourier basis functions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 51–60.

[43] C. Kesim *et al.*, "Henle fiber layer mapping with directional optical coherence tomography," *RETINA J. Ret. Vit. Dis*, 2022.

[44] Y. He *et al.*, "Retinal layer parcellation of optical coherence tomography images: Data resource for multiple sclerosis and healthy controls," *Data in Brief*, vol. 22, pp. 601–604, 2022.

[45] E. Garcia-Martin, V. Polo, J.M. Larrosa, M.L. Marques, R. Herrero, J. Martin, J.R. Ara, J. Fernandez, and L.E. Pablo, "Retinal layer segmentation in patients with multiple sclerosis using spectral domain optical coherence tomography," *Ophthalmology*, vol. 121, no. 2, pp. 573–579, 2014.