DEVELOPMENT OF AUTOMATED METHODS FOR RETINAL OPTICAL COHERENCE TOMOGRAPHY IMAGE ANALYSIS

A THESIS

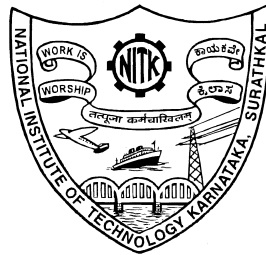Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

ANOOP B N

(177150)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA
SURATHKAL, MANGALORE - 575025, INDIA

November, 2021

DECLARATION

I hereby *declare* that the Research Thesis entitled **DEVELOPMENT OF AUTO-MATED METHODS FOR RETINAL OPTICAL COHERENCE TOMOGRAPHY IMAGE ANALYSIS** which is being submitted to the *National Institute of Technology Karnataka, Surathkal* in partial fulfillment of the requirements for the award of the Degree of *Doctor of Philosophy* is a *bona fide report of the research work carried out by me*. The material contained in this thesis has not been submitted to any University or Institution for the award of any degree.

ANOOP B N
Register No.: 177150
Department of Computer Science and Engineering
National Institute of Technology Karnataka
Surathkal-575025

Place: NITK - Surathkal

Date: 03 NOVEMBER 2021

CERTIFICATE

This is to *certify* that the Research Thesis entitled **DEVELOPMENT OF AUTO-MATED METHODS FOR RETINAL OPTICAL COHERENCE TOMOGRAPHY IMAGE ANALYSIS**, submitted by **ANOOP B N** (Register Number: 177150) as the record of the research work carried out by him, is *accepted* as the *Research Thesis submission* in partial fulfillment of the requirements for the award of degree of ***Doctor of Philosophy***.

03/11/2021

Dr. Jeny Rajan
Research Guide
Assistant Professor
Department of Computer Science and Engineering
National Institute of Technology Karnataka
Surathkal-575025

Chairman - DRPC
Department of Computer Science and Engineering
National Institute of Technology Karnataka
Surathkal-575025
(Signature with Date and Seal)

To My Beloved Wife, Parents, Teachers
and
Friends

ACKNOWLEDGEMENTS

ABSTRACT

Cystoid Macular Edema (CME) is a retinal abnormality causing fluid accumulations in the retina due to various retinal diseases. The early diagnosis of CME and its quantification is vital for treatment planning. Optical Coherence Tomography (OCT) is a non-invasive imaging technique used to visualize the human retina and the retinal abnormalities. The OCT images have corrupted with speckle noise due to the coherence detection, which will degrade the quality of the OCT images. Also, the human retina is a layered structure. The segmentation of retinal layers helps in diagnose various retinal diseases and finding the locations of retinal pathologies. This thesis focused on the development of automated methods for retinal OCT image analysis and tried to provide deep learning-based solutions for each of the stages.

OCT is an imaging technique widely used for medical imaging. Noise in an OCT image generally degrades its quality, thereby obscuring clinical features and making the automated segmentation task suboptimal. Obtaining higher quality images requires sophisticated equipment and technology, available only in selected research settings, and is expensive to acquire. Developing effective denoising methods to improve the quality of the images acquired on systems currently in use has the potential for vastly improving image quality and automated quantitative analysis. Noise characteristics in images acquired from machines of different makes and models may vary. Our experiments show that any single state-of-the-art method for noise reduction fails to perform equally well on images from various sources. Therefore, detailed analysis is required to determine the exact noise type in images acquired using different OCT machines. In the second chapter, we studied noise characteristics in the publicly available DUKE and OPTIMA datasets to build a more efficient model for noise reduction. These datasets have OCT images acquired using machines of different manufacturers. We further pro-

pose a patch-wise training methodology to build a system to effectively denoise OCT images. We have performed an extensive range of experiments to show that the proposed method performs superior to other state-of-the-art methods.

Segmentation of retinal layers is a vital step in computerized processing and the study of retinal OCT images. However, automatic segmentation of retinal layers is challenging due to the presence of noise, widely varying reflectivity of image components, variations in morphology and alignment of layers in the presence of retinal diseases. In the third chapter, we propose a Fully Convolutional Network (FCN) termed as DelNet based on a deep ensemble learning approach to selectively segment retinal layers from OCT scans. The proposed model is tested on a publicly available DUKE DME dataset. Comparative analysis with other state-of-the-art methods on a benchmark dataset shows that the performance of DelNet is superior to other methods.

In the fourth chapter, we propose attention assisted convolutional neural network-based architecture to detect and quantify three types of retinal cysts namely the intra-retinal cyst, sub-retinal cyst and pigmented epithelial detachment from the OCT images of the human retina. The proposed architecture has an encoder-decoder structure with an attention and a multi-scale module. The qualitative and quantitative performance of the model is evaluated on the publicly available RETOUCH retinal OCT fluid detection challenge data set. The proposed model outperforms the state-of-the-art methods in terms of precision, recall, and dice coefficient. Furthermore, the proposed model is computationally efficient due to its less number of model parameters.

Keywords: Retinal cysts; Image Segmentation; Retinal Layer Segmentation; Optical Coherence Tomography; Speckle Noise; Denoising; Deep Learning; Ensemble Learning; Convolutional Neural Networks; Fully Convolutional Networks; Patch-wise Training; Attention module; Multi-scale features.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xiii

ABBREVIATIONS

AG	Amsler Grid test
ASPP	Atrous Spatial Pyramid Pooling
CFP	Color fundus photography
CME	Cystoid Macular Edema
DME	Diabetic Macular Edema
CNN	Convolutional Neural Network
FA	Fundus fluorescein angiography
FCN	Fully Convolutional Network
GCL	Ganglion Cell Layer
GCDS	Gated Convolution Deconvolution Structure
Grad-CAM	Gradient-weighted Class Activation Mapping
ILM	Internal Limiting Membrane
IPL	Inner Plexiform Layer
INL	Inner Nuclear Layer
ILM	Internal Limiting Membrane
IRF	Intra-Retinal cyst
ML	Maximum likelihood
MS	Model Selector
NFL	Nerve Fiber Layer
NLM	Non local Means
OCT	Optical Coherence Tomography
OPL	Outer Plexiform Layer
ONL	Outer Nuclear Layer
OLM	Outer Limiting Membrane

PED	Pigmented Epithelial Detachment
PSNR	Peak Signal to Noise Ratio
RPE	Retinal Pigment Epithelium
SNR	signal-to-noise ratio
SSIM	Structural Similarity Index Metric
SRF	Sub-Retinal cyst
TV	Total Variation
VA	Visual Acuity

CHAPTER 1

INTRODUCTION

1.1 Human Eye and Retina

Second, to the brain, the eyes are the most complex organ in human body. The eye consists of complex mechanisms and pathways and with the help of several structures like the cornea, retina, iris, pupils, optic nerves and the ciliary muscles. All these parts work together to help the eye function properly and make our vision perfect. The anatomy of the human eye is shown in Figure. 1.1.

The eye's anatomical structures are arranged in three layers, namely outer fibrous tunic, intermediate vascular tunic and the innermost retina. The outer protective fibrous
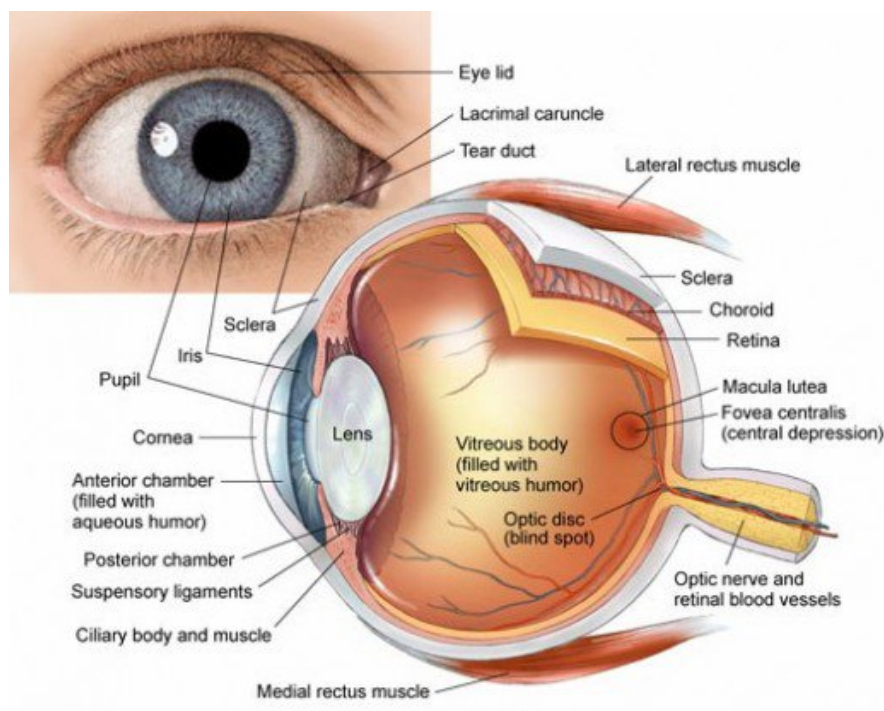
Figure 1.1: Anatomy of the human eye. (Image Courtesy: (TheEye, 2021))

tunic is composed of cornea and sclera. Cornea transmits and focuses lights into the eye. Sclera is continuous with the cornea and forms the supporting wall of the eyeball. The middle layer, or the vascular tunic, consists of choroid, ciliary body, pigmented epithelium and iris. Iris is visible as the coloured part of the eye that functions like a diaphragm and helps in regulating the amount of light that enters the eye. The anterior chamber between the cornea and iris as well as the posterior chamber between the lens and the iris is occupied by the aqueous humour, a fluid that nourishes the anterior structures.

The lens is a transparent structure that focuses light rays onto the retina. Lens is attached to the ciliary body by suspensory ligament, which transmit muscular forces for changing the lens shape to focus on objects at various distances. The vitreous body is a clear, jelly-like substance that fills the eye. The innermost layer or retina, gets oxygenation from choroidal blood vessels and retinal vessels. The retina is a multilayered neural tissue that lines the back of the eye, senses light and creates electrical impulses that travel through the optic nerve to the visual cortex, which is the part of the brain that controls our senses of sight.

The retina is a thin layered tissue. It is situated near to the optic nerve. Retina collects the light focused by the lens and converts the luminance into nerve impulse. Nerve impulses are transmitted through the optic nerve to the visual cortex to establish the vision. The layered structure of human retina (de Campos *et al.*, 2020) is shown in Figure. 1.2.

Retinal layers are organized into two distinct functional components namely neurosensory retina and retinal pigment epithelium. Neurosensory layers of the retina contains total 9 layers. Starting from the vitreous surface, they are the internal limiting membrane (ILM), nerve fiber layer (NFL), ganglion cell layer (GCL), inner plexiform layer (IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), outer limiting membrane (OLM) and the layer of photoreceptors. Retinal pigment epithelium (RPE) is the outermost layer of the retina, this layer provides metabolic
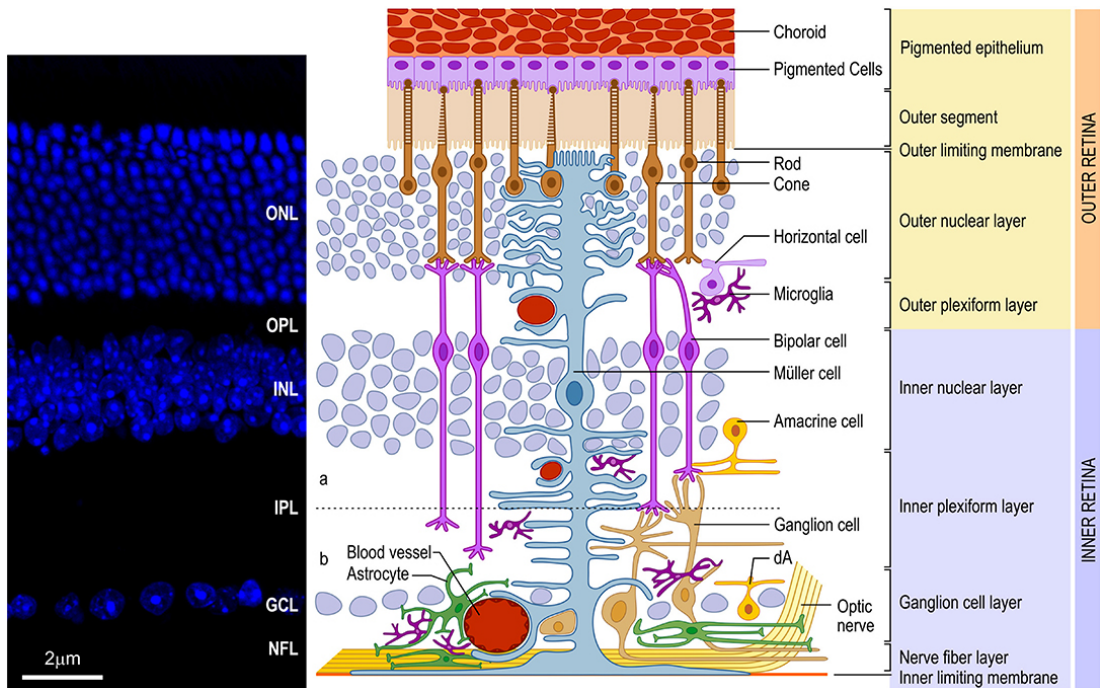
Figure 1.2: Cross sectional view of the retina. (Image courtesy: (de Campos *et al.*, 2020))

support to the nuerosensory layers of the retina and it absorbs incident scattered light.

1.2   Retinal Cyst and Cystoid Macular Edema

A retinal cyst is a fluid-filled space in the retina, and the presence of edema and cysts increases the thickness of the involved retina. Retinal cysts have pathological significance and are present in several eye disorders like age related macular degeneration, diabetic retinopathy, retinal vein occlusion, ocular inflammation, diabetic macular edema (DME) etc.

Cystoid macular edema (CME) develops when excess fluid accumulates within the macular retina. This is thought to occur following disruption of the blood retinal barrier because of the aforementioned eye disorders. This fluid accumulation in retina

reduces macular retinal function. Medical studies shows that visual acuity can be reasonably correlated from the volume of retinal cystic fluids and their location in the retina (Wilkins *et al.*, 2012).

1.3   Diagnosis of CME

Primary symptoms of CME are decreased, blurry or wavy vision, faded colors, and distortion near or at the center of vision. Ocular diseases can be characterized using several techniques. Initial Visual Acuity (VA) test is used to measure central vision according to the size of letters or symbols viewed on Snellen chart or E chart respectively. Figure. 1.3 shows an example of Snellen and E chart.

Amsler Grid test (AG) is used to detect visual distortion caused in macular disorders. Amlser Grid contains horizontal and vertical lines with a small dot at the center of the grid and is shown in Figure. 1.4(a). Patient is asked to view the AG from each eye separately. If the patient has macular pathology, he/she may notice distortion or they may be unable to see some lines in the AG. Both VA and AG tests are psychophysical tests and needs patient cooperation. For clinical anatomical evaluation of the macula, ophthalmoscopy or slit lamp biomicroscopy test (with a 78 or 90 diopter aspheric lens) is used. Figure. 1.4(b) shows slit lamp system.

Imaging techniques can also be used for the detection of cystic fluids in the retina. Color fundus photography (CFP) is used for imaging of the macula. With stereoscopic view of the retina using CFP, ophthalmologist can identify thickening of retina that is caused due to intra-retinal fluid accumulation. Fundus fluorescein angiography (FA) is widely available and used for dynamic evaluation of the retinal vasculature. In FA, the amount of fluorescein leakage depends on the dysfunction of the retinal vascular endothelium. Apart from being a significant diagnostic modality, Fundus Fluorescein Angiography also improves the accuracy for planning the treatment of CME. Figure. 1.5 shows CFP and FA images of a retina with CME. Figure. 1.5(b) shows fluorescein leak-

1	20/200	
2	20/100	
3	20/70	
4	20/50	
5	20/40	
6	20/30	
7	20/25	
8	20/20	
9		
10		
11		
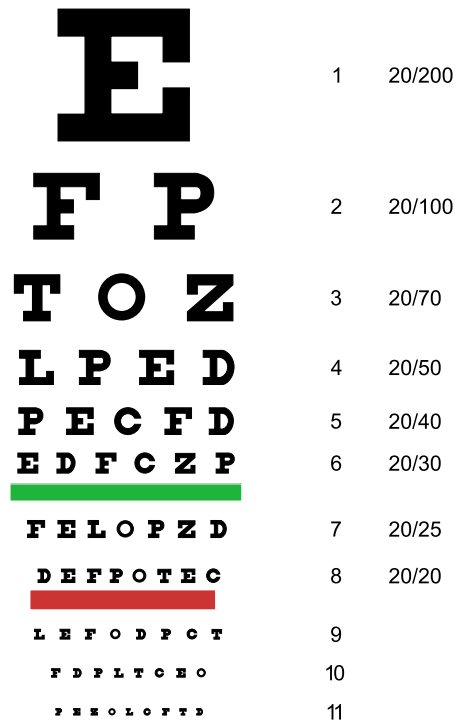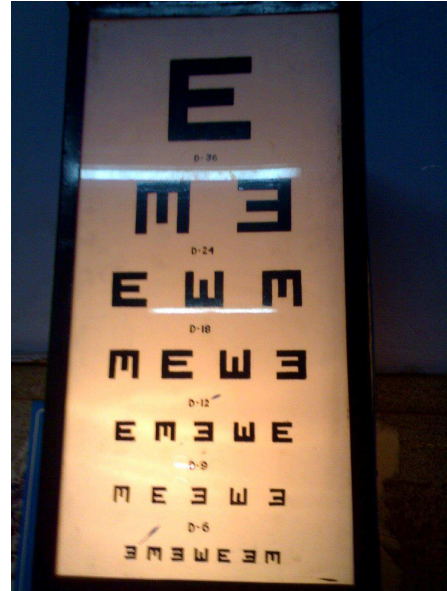
(a) Snellen chart (b) E chart

Figure 1.3: Different type of charts used in visual acuity test.
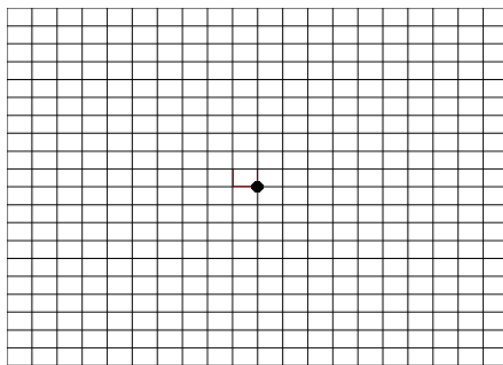
(a) (b)

Figure 1.4: Amsler grid and slit lamp machine.
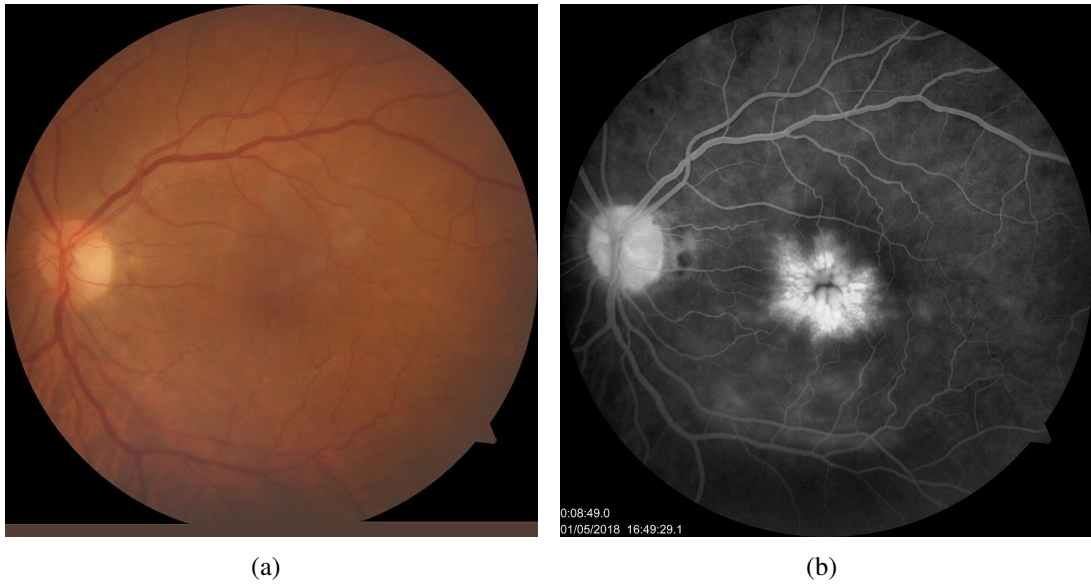
Figure 1.5: Color fundus photography and fundus fluorescein angiography scan of a retina with CME.

age into the cystoid spaces is distributed radially in retina forming the classic honey-comb like structure appearance near fovea. Slit-lamp bio-microscopy, fundus photography and FA techniques gives 2D *en face* view of the retina and are prone to differences in subjective assessment to quantify and diagnose retinal pathology. OCT was introduced in clinical practice for the purpose of high resolution cross-sectional imaging (as opposed to *en face*) and for objective measurement of retinal structures.

1.4 Optical Coherence Tomography

OCT is a non-invasive imaging technique that uses low coherence light for resolving internal structures of biological tissues and obtains cross-sectional, high-resolution images of the retina and can detect retinal thickness (Huang *et al.*, 1991). It is a widely used diagnostic technique for several ocular diseases.

Figure. 1.6 shows a schematic diagram of OCT Scanner. A low coherence light source (LCS) (Eg:superluminescent light diode) emits light towards beam splitter (BS).
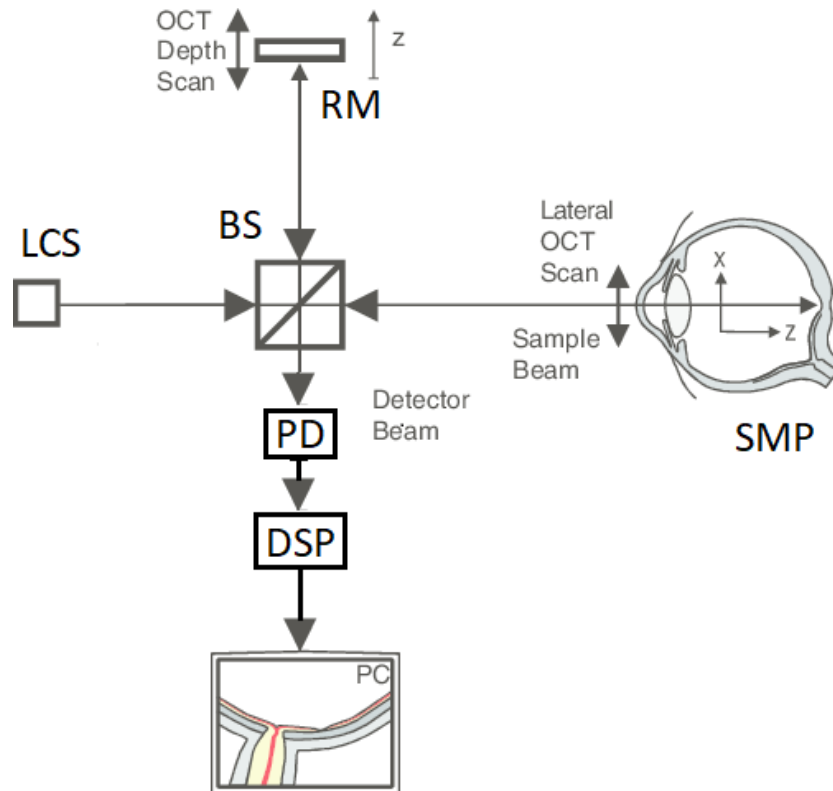
Figure 1.6: Schematic of the Conventional OCT system based on a low time-coherence Michelson interferometer. LCS = low time-coherence light source; PC = personal computer. (Image courtesy: (Fercher *et al.*, 2003))

Part of this beam is projected on to reference mirror (RM) and the rest to the sample (SMP) to be scanned. Reflected light from both RM and SMP are combined at BS, creating interference patterns based on the interferometry concept. When the reflected light beams from RM and SMP are *in phase*, this generates constructive inference. At other times, it results in destructive inference. Photo detector receives combined reflected light with interference pattern and forward the signal to a digital signal processor unit (DSP). DSP unit uses Fourier transform to create a linear A-scan of the point scanned. Several adjacent A scans along a linear path of scanning are combined to give a B-scan image. An OCT scan typically consists of several B scan images taken over a fixed area, with a fixed protocol. The B scans will be saved in a personal computer (PC).

Figure. 1.7 shows a commercially available OCT Scanner. OCT provides excellent

Figure 1.7: Cirrus optical coherence tomography scanner.

visualization of several retinal structures like cysts, exudates and retinal layers. Figure. 1.8 shows OCT B-scan of the retina with *en face* macular projection. The dark green line with arrow head shown on the *en face* macular image (left side of the figure) corresponds to the projection line of the OCT B-scan shown on right side of the figure. Figure. 1.9 shows OCT scans of the retina with normal and cystic fluids obtained using Cirrus and Spectralis vendor OCT machines.

In principle, OCT works similar to ultrasound (US) imaging systems, based on interferometric techniques. However, in practice, OCT make use of light beams in position of sound profiles in US system (Wong *et al.*, 2005; Pagliara *et al.*, 2008). As in US imaging, speckle noise hampers the quality of OCT and limits its clinical utility. Speckle arises because of the coherence of optical waves (Schmitt *et al.*, 1999) and destroys the edges of the retinal layers and the cystic boundaries. Hence, despeckling of OCT images is a significant preprocessing stage for the proper diagnosis of the retinal disorders and its quantification.
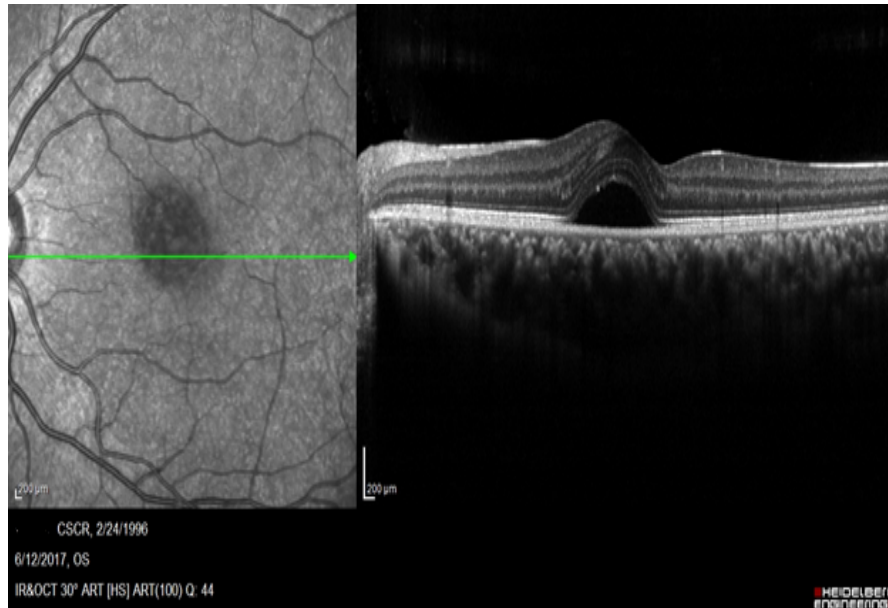
Figure 1.8: *En face* macular projection and its corresponding OCT B-Scan of a retina.

(a)

(b)

(c)

(d)
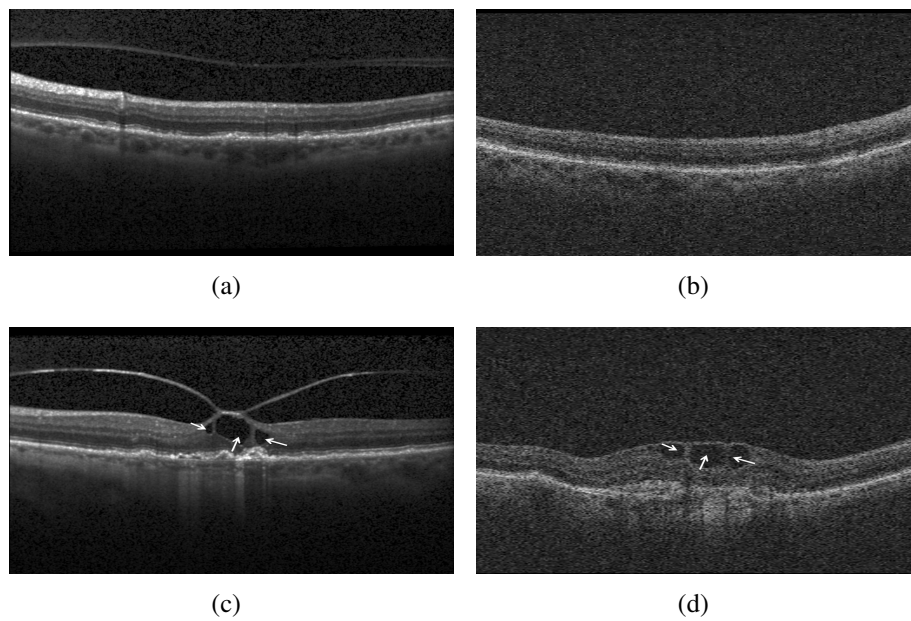
Figure 1.9: Optical coherence tomography scans of retina from two different vendors: (a), (b) normal retina; (c), (d) retina with cystoid macular edema. (White colored arrow shows cystic fluids). (a) and (c) obtained from Spectralis imaging system, (b) and (d) obtained from Cirrus imaging system. (Image courtesy: OPTIMA cyst challenge dataset (OCSC, 2015))

.

1.5   Motivation and Challenges

According to the Vision Loss Expert Group of the Global Burden of Disease Study, "the leading global causes of blindness in those aged 50 years and older in 2020 were cataract (152 million cases), followed by glaucoma (36 million cases), age-related macular degeneration (18 million cases), and diabetic retinopathy (086 million cases)" (Steinmetz *et al.*, 2021). One of the key anatomical derangements caused by the aforementioned diseases is CME. This is a common painless end result condition in which multiple cystic fluid filled spaces form in the retina. Visual morbidity from CME is a major public health problem globally. If detected early and treated effectively, most of this visual impairment from CME can be prevented or cured.

Manually analyzing OCT data for the detection of cysts and measuring their volume is time consuming and needs expertise, and currently is not performed routinely, leading to qualitative assessment of the scan alone. Automated cyst segmentation and quantification methods enable accurate quantification for faster diagnosis and treatment. The development of automated cyst segmentation techniques for OCT scans pose several challenges such as, speckle noise in the scans, poor signal-to-noise ratio (SNR), retinal vessel shadows, scan intensity difference across the vendors, and the location of the cyst based on the retinal layers.

Even though numerous methods have been proposed for denoising OCT images, the methods failed to produce better texture preservation, edge preservation, and noise removal. Another drawback is that the methods available in the literature are suitable only for specific datasets. With the advent of deep neural network architectures such as deep convolutional neural networks and stacked denoising autoencoders, promising results in noise removal from natural images and medical images are available in the literature. Also from the literature, it is evident that the deep convolutional neural networks, can automate the object segmentation and image classification tasks accurately.

1.5.1 Problem Statement

Noise in an OCT image degrades its quality, thereby obscuring clinical features and making the automated segmentation task suboptimal. Developing effective denoising methods to improve the quality of the images acquired on systems currently in use is vital for automated quantitative analysis. Noise characteristics in images acquired from machines of different makes and models may vary. Also, any single state-of-the-art method for noise reduction fails to perform equally well on images from various sources. Therefore, a detailed analysis is required to determine the exact noise type in images acquired using different OCT machines and also, to propose a solution for eliminating these noise types.

Segmentation of retinal layers is an essential step in computerized processing and the study of retinal OCT images. Even though the automated retinal layer segmentation algorithms are present in the literature, they haven't considered the characteristics of the retinal layers and hence, failed in achieving better segmentation accuracy. Therefore, an automated retinal layer segmentation algorithm, which can understand the characteristics of the retinal layers is vital for the accurate segmentation of the retinal layers.

The early diagnosis of CME and its quantification is requisite for treatment planning. The automated methods proposed for the segmentation of all the three kinds of retinal cysts in the literature are computationally expensive. Hence, developing an automated retinal cyst segmentation algorithm, which can segment all the three types of retinal cysts and which is computationally less expensive is required for the early diagnosis and treatment planning of CME.

Research Objectives:

1. To study the data distribution in the OCT images.

2. To develop an automated despeckling technique such that it can provide better edge-preservation and texture preservation in the OCT images.

3. To develop an automated algorithm for segmenting various layers of retina from OCT scans of retina.

4. To develop an automated algorithm to segment the three types of retinal cysts named as an intra-retinal cyst (IRF), sub-retinal cyst (SRF) and the pigmented epithelial detachment (PED) from OCT images.

1.6 Major Contributions

The main focus of this thesis is to develop novel automated image analysis solutions for the segmentation of retinal cyst from the retinal OCT scans. The main contributions of this thesis can be summarized as follows:

- Speckle in OCT images generally degrades the quality of the OCT images and makes the clinical diagnosis tedious. To improve the quality of the OCT images, we have proposed an automated despeckling scheme called Gated Convolution Deconvolution Structure (GCDS). We demonstrate experimentally that the distribution of speckle in OCT images acquired with machines of different manufacturers are different. A single model or distribution is unable to account for the behaviour of noise in OCT images acquired from different makes of the equipment. Based on the aforementioned experiment, we designed a DenseNet based architecture for identifying the type of noise in the input images. Also, we have developed a novel patch-wise methodology to properly train the existing GCDS model. This methodology enables the GCDS model to isolate and learn the noise characteristic of the OCT image. The proposed method denoises the image while maintaining the structural and chromatic integrity of the underlying ground truth. Hence, we have used an array of GCDS architectures to despeckle the input images.

- Segmentation of retinal layers is a vital and important step in computerized processing and the study of retinal OCT images. We have developed an ensemble architecture for the effective segmentation of the most prominent retinal layers. The proposed ensemble-based architecture helped the model to learn different layers better when compared to the stand-alone single models. This is demonstrated with the help of experiments on a standard benchmark dataset and also through Gradient-weighted Class Activation Mapping (Grad-CAM), visualization.

- Accurate estimation of the type of retinal cyst and its volume is vital for effective treatment planning of retinal abnormalities. We have proposed attention assisted convolutional neural network-based architecture for detecting and quantifying three types of retinal cysts such as the intra-retinal cyst, sub-retinal cyst and pigmented epithelial detachment from the retinal OCT images. The proposed model is computationally efficient due to its less number of model parameters.

1.7   Organization of this Thesis

This dissertation is organized into five chapters as follows.

Chapter 1 presents the motivation, clinical significance of retinal OCT image analysis and its challenges, followed by the proposed approach to the problem and its contributions.

Chapter 2 discuss the statistical models of the speckle present in the OCT images acquired from different OCT machines and presents a novel deep learning approach to remove speckle from the OCT images using a cascaded patch-wise gated convolution deconvolution structure.

Chapter 3 presents a fully convolutional network for selectively segment retinal layers from the retinal OCT scans termed DelNet, based on deep ensemble learning.

Chapter 4 introduces attention assisted convolutional neural network-based architecture for detecting and quantifying three types of retinal cysts, namely the intra-retinal cyst, sub-retinal cyst and pigmented epithelial detachment using the OCT images of the human retina.

Chapter 5 concludes this thesis by providing a general summary of the presented research work and its future scope.

CHAPTER 2

A CASCADED CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE FOR DESPECKLING OCT IMAGES

In this chapter, we have studied the noise characteristics in the publicly available DUKE and OPTIMA datasets to build a more efficient model for noise reduction. These datasets have OCT images acquired using machines of different manufacturers. We further propose a patch-wise training methodology to build a system to effectively denoise OCT images in this chapter [1]. We have performed an extensive range of experiments to show that the proposed method performs superior to other state-of-the-art-methods.

2.1 Introduction

Optical Coherence Tomography (OCT) is an imaging technology widely used for the diagnosis of retinal disorders. OCT uses the principle of coherent detection to capture the information present in the retina. The coherent detection also introduces multiplicative noise called speckle in the acquired images. Speckle degrades the quality of images

---

[1]The work described in this Chapter has been published in: **Anoop, B. N.**, Kaushik S. Kalmady, Akhil Udathu, V. Siddharth, G. N. Girish, Abhishek R. Kothari, and Jeny Rajan, "A cascaded convolutional neural network architecture for despeckling OCT images." *Biomedical Signal Processing and Control*, 66 (2021): 102463.

Menon, Sandeep N., VB Vineeth Reddy, A. Yeshwanth, **B. N. Anoop**, and Jeny Rajan, "A novel deep learning approach for the removal of speckle noise from optical coherence tomography images using gated convolutiondeconvolution structure." *In Proceedings of 3rd International Conference on Computer Vision and Image Processing*, pp. 115-126. Springer, Singapore, 2020.

Anoop, B. N., G. N. Girish, P. V. Sudeep, and Jeny Rajan, "Despeckling Algorithms for Optical Coherence Tomography Images: A Review." *Advanced Classification Techniques for Healthcare Analysis* (2019): 286-310.

and makes visual as well as automated analysis of OCT images difficult. Despeckling is used as a preprocessing step in many OCT image analysis tools. Despeckling improves the signal-to-noise ratio of images and as a result, retinal structures and boundaries between the layers can be seen more clearly (Girish *et al.*, 2018*a*). It also improves visualization of pathological conditions like retinal edema, cystoid macular edema, pigment epithelial detachment and other kinds of lesions (Gopinath and Sivaswamy, 2018; Girish *et al.*, 2019; Rao *et al.*, 2019).

The statistics of noise present in OCT images has an important role in developing despeckling algorithms. Different data distributions were suggested in the literature for modeling speckle in OCT images. According to Schmitt (Schmitt *et al.*, 1999), the speckle in OCT follows a decaying exponential distribution. Bashkansky and Reintjes (Bashkansky and Reintjes, 2000) suggested that speckle follows a Gaussian distribution. Pircher et al. (Pircher *et al.*, 2003) and Karamata et al. (Karamata *et al.*, 2005) used Rayleigh distribution to model the statistical behavior of the speckle. Sudeep et al. (Sudeep *et al.*, 2016) proposed a denoising method based on Gamma distribution for the speckle reduction in OCT images.

In the last few years, several methods were published in the literature to remove speckle from OCT images. Among them, Non local Means (NLM) based approaches are popular (Aum *et al.*, 2015; Yu *et al.*, 2016; Tang *et al.*, 2017; Cuartas-Vélez *et al.*, 2018; Girish *et al.*, 2018*b*). The NLM method computes the true underlying intensity by replacing every noisy pixel with the weighted average of the non-local neighboring pixels. The weights are assigned based on the similarity of the patches (neighboring pixels around the pixel of interest and the neighboring pixels around non-local pixel). Since the noise is non-Gaussian, the bias in the mean also has to be taken care of. Even though NLM based methods are relatively easy to implement, their performance on non-Gaussian noisy images is suboptimal. These methods also have relatively high computational complexity.

Thapa et al. (Thapa *et al.*, 2015) proposed a weighted nuclear norm minimization

16

method for denoising OCT images. An adaptive singular value shrinking method was the proposed by Chen et al. (Chen *et al.*, 2018a), where the authors applied adaptive singular value shrinking method for removing the speckle. Iterative bilateral filter (Sudeep *et al.*, 2016) has also been proposed to denoise OCT images. In (Sudeep *et al.*, 2016), the authors assumed that speckle follows Gamma distribution and maximum likelihood (ML) method was employed to estimate the parameters of the Gamma model. These parameters (shape and scale) are used to compute the unbiased image. Total Variation (TV) minimization methods were proposed in (Gong *et al.*, 2015; Duan *et al.*, 2016) for despeckling OCT images. Our own experiments and experimental results reported in (Anoop *et al.*, 2019) show that this filter introduces artifacts in the denoised OCT images. Wang et al. (Wang *et al.*, 2018) categorized OCT speckle into additive and multiple portions and presented a two step iteration mechanism to suppress the noise.

Wavelet-based approaches have also been explored (Zaki *et al.*, 2017; Kafieh *et al.*, 2019) to denoise OCT images. In this procedure, wavelet coefficients in the selected sub-bands with magnitude greater than an arbitrary threshold are attenuated. Wavelet thresholding fails if the coefficients are biased and leads to degradation of fine details in high noise conditions. Jian et al. (Jian *et al.*, 2010) proposed a 3D despeckling approach to despeckle OCT images. They used the volumetric information and the Curvelet transform for speckle attenuation. Xu et al. (Xu *et al.*, 2013) proposed a despeckling method based on shrinkage in the contourlet domain. Zhang et al. (Zhang *et al.*, 2017a) exploited the phase information of the complex OCT data to denoise the Fourier domain OCT images. Other popular methods mentioned in the literature for OCT despeckling include Principal Component Analysis (PCA) based despeckling (Lv *et al.*, 2018), dictionary-based denoising or K-SVD (Kafieh *et al.*, 2014), Bayesian denoising (Rajabi and Zirak, 2016), diffusion potential based despeckling (Paul *et al.*, 2018), ensemble framework of Multi-Layer Perceptron (MLP) (Adabi *et al.*, 2016), and spatially constrained Gaussian mixture model (Amini and Rabbani, 2017).

Recently, deep learning-based methods were also proposed for image denoising

(Zhang *et al.*, 2017*b*, 2018; Lefkimmiatis, 2018; Tian *et al.*, 2020; Yuan *et al.*, 2018; Chen *et al.*, 2017). The aforementioned methods were mainly proposed for images corrupted with Gaussian noise. Very recently, a Fully Convolutional Neural Network (FCNN) model called Gated Convolution Deconvolution Structure (GCDS) for despeckling OCT images was proposed by Sandeep et al. (Menon *et al.*, 2020*a*). Though effective, this model failed to perform uniformly on OCT images acquired with different machines, especially when the resolution differed.

Most of the despeckling methods proposed in the literature were built on the assumption that the probability density function (pdf) of the speckle in the OCT follows some known family of distributions (e.g. Gamma or Rayleigh). If the pdf of the noise is different from the assumptions made in the model, then its performance will be suboptimal. Additionally, most methods proposed in the literature were tested against data acquired from a single machine and these methods may not perform well on images acquired from machines of other manufacturers. Our experiments with images acquired using machines of different vendors show that speckle data distributions vary significantly in these images. The aforementioned drawbacks of the existing methods are addressed in this chapter. Major highlights of our work are:

1. We demonstrate experimentally that the distribution of speckle in OCT images acquired with machines of different manufacturers are different. A single model or distribution is unable to account for the behavior of noise in OCT images acquired from different makes of the equipment.

2. We used an array of GCDS architectures to despeckle the input images. These architectures take care of different data distributions.

3. Based on the aforementioned experiment, we designed a DenseNet (Huang *et al.*, 2017) based architecture for identifying the type of noise in the input images. This architecture acts also as a switch to select the appropriate despeckling model.

4. We have developed a novel patch-wise methodology to properly train the existing GCDS (Menon *et al.*, 2020*a*) model. This methodology enables the GCDS model to isolate and learn the noise characteristic of the OCT image. It denoises the image while maintaining the structural and chromatic integrity of the underlying ground truth.
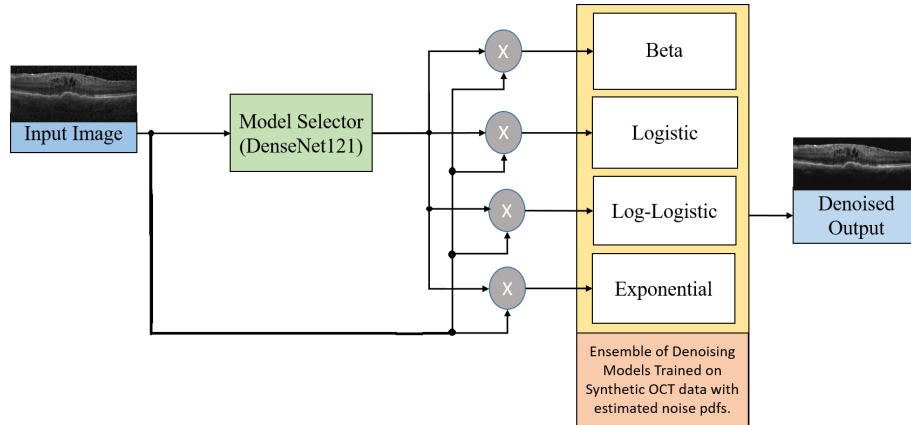
Figure 2.1: Workflow of the proposed methodology.

The chapter is structured as follows. The model, synthetic eye dataset, proposed methodology and experiments are discussed in section 2.2. Experimental results are presented in section 2.3, and finally the conclusions and remarks are drawn in section 2.4.

2.2 Methods and Data

An overview of the proposed CNN based cascaded OCT image despeckling architecture is shown in Figure 2.1. The architecture consists of two stages. The first stage is a fully connected deep convolutional neural network called Model Selector (MS). The second stage consists of an array of fully convolutional autoencoder models that have each been trained exclusively for denoising specific types of noises. The details of the experimental setup, data sets used for experiments, and a study of noise in the OCT data acquired using different equipments are presented in this section. We also provide a detailed description of the model selector, FCNN model used for denoising, and the challenges faced as well as solutions for effectively training an FCNN model for denoising.
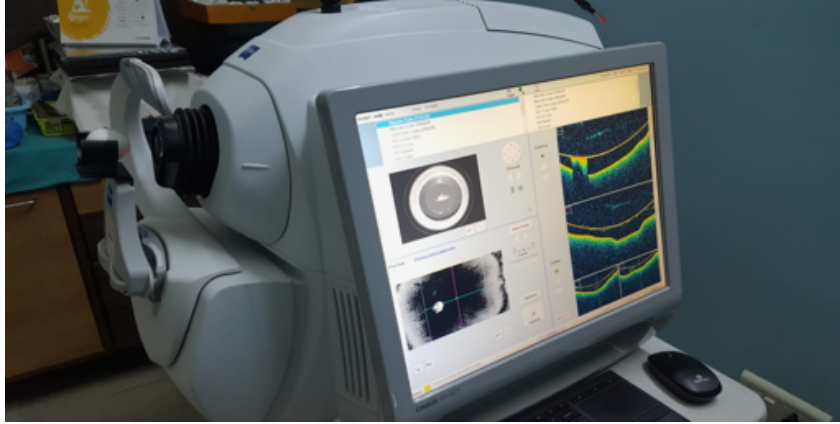
Figure 2.2: Experimental setup to capture images from the synthetic eye.

2.2.1 Experimental setup

In order to develop effective denoising models, we studied the noise characteristics in OCT images using three datasets. These included the two publicly available DUKE (Fang *et al.*, 2013) and OPTIMA (OCSC, 2015) datasets, and one privately created synthetic eye dataset. The details about the spatial dimensions and number of volumes of images in each dataset is depicted in Table 2.1.

The synthetic images used for the experiments are obtained from an artificial eye fixed onto the OCT acquisition system (Figure 2.2). The artificial training eye model (Training Eye Zeiss Ref.64389, Carl Zeiss Meditec, Inc., Dublin, CA and RetiEye, Aurolab, Madurai, India) are plastic eyes with an anterior segment having optical power close to the human eye and a fixed dilated entry pupil. Zeiss Cirrus 500 OCT machine was used for acquiring the images. The posterior part of these eyes is made of a thin translucent plastic layer painted similar to that of the human retina. These eyes are used primarily for practice of ophthalmoscopy and laser application by trainee ophthalmologists. The eye was fixed to the OCT apparatus and 25 consecutive scans (volumes) were taken for one position of the eye without altering the alignment. This constituted a single dataset with complete elimination of artifacts from saccadic movements that are universally present in the eyes of human subjects which give rise to varying levels of motion artifacts despite eye tracking. The dimensions of each volume

20

Table 2.1: The details of the dataset used for this study with total number of B - scans per volume (TNBPV).

Vendor \Dataset	OPTIMA(OCSC, 2015)	Duke(Fang et al., 2013)	Synthetic eye
Cirrus	No. of Vol.: 8 Size: 1024512 TNBPV: 128 - 200	-	No. of Vol.: 25 Size: 1024512 No. of GT: 128 TNBPV: 128
Spectralis	No. of Vol.: 8 Size: 496512 TNBPV: 7 - 49	-	-
Nidek	No. of Vol.: 6 Size: 512512 TNBPV: 7 - 128	-	-
Topcon	No. of Vol.: 8 Size: 885512 TNBPV: 128	-	-
Bioptigen	-	No. of Vol.: 18 Size: 450900 No. of GT: 18 TNBPV: 6	-

is $1024 \times 512 \times 128$. We used these 25 scans to create a high SNR image by averaging them.

2.2.2 Modelling Speckle Noise

We conducted experiments to study the noise characteristics and the distribution of speckle in OCT images acquired with equipments of different manufacturers. This also enabled us to verify experimentally, the claims laid out in earlier works regarding the nature of speckle noise in OCT images. This experiment was motivated by a simi-

lar study put forth by Vegas et al. (Vegas-Sanchez-Ferrero, 2010) to study the noise characteristics in ultrasound images. Homogeneous background regions were cropped from the images and the distribution of the intensity values were analyzed statistically. We used 1500 samples (homogeneous patches) each from Duke dataset, four vendors from OPTIMA dataset (1500×4) and 1500 samples from the synthetic eye dataset. The distributions considered for evaluation were: (i) Gamma, (ii) Log-normal, (iii) Rayleigh, (iv) Normal, (v) Beta, (vi) Log-Logistic (Fisk), (vii) Rice, (viii) Exponential, (ix) Weibull-min, (x) Weibull-max and (xi) Logistic. These are the popular distributions mentioned in the literature regarding speckle in OCT (Kirillin *et al.*, 2014; Bashkansky and Reintjes, 2000; Schmitt *et al.*, 1999; Karamata *et al.*, 2005; Vegas-Sanchez-Ferrero, 2010). However, some of these distributions are related.

For every input patch, the histogram of pixel intensities was computed and each of the earlier mentioned distributions were fit and their respective χ^2 statistic was computed. For implementation we used python 3.7 and SciPy library's Statistical functions (SciPy.stats). For every noisy patch considered, SciPys *stats.fit()* method which uses the Downhill Simplex algorithm for optimization was applied on the chosen distributions. The distribution with the lowest χ^2 statistic value was chosen as the best fit distribution for the provided input patch. The χ^2 test was used to determine the significance of differences between the expected frequencies and the observed frequencies and was also used to estimate goodness-of-fit measure (Vegas-Sanchez-Ferrero, 2010). The overall results for all the patches were then consolidated. Table 2.2 summarizes the results of the aforementioned experiment. Most prominent noise distributions exhibited in OCT images of different manufacturers are listed.

The study on noise characteristics in OCT images shows its diverse nature and vendor dependency. Most modern OCT scanning devices incorporate post-processing operations either at the software or hardware level in an attempt to produce better images. This post-processing operations may change the noise characteristics. This could be one of the reasons for the observed variation. The obtained values also show that prior

Table 2.2: Consolidated results of the experiments performed on different OCT datasets to find the characteristics of speckle present in OCT images.

SI.No.	Dataset and OCT machine used	Data distribution
1	OPTIMA (Cirrus)	Log-logistic
2	OPTIMA (Nidek)	Exponential
3	OPTIMA (Topcon)	Logistic
4	OPTIMA (Spectralis)	Beta
5	Duke (Bioptigen)	Beta
6	Synthetic Eye (Cirrus)	Log-logistic

assumptions regarding the nature of speckle noise in OCT images cannot be considered as standard for all OCT systems, and a detailed study would be required to determine the system-specific true noise characteristics.

2.2.3 GCDS Model

For denoising OCT images, we used the recently proposed Gated Convolution Deconvolution Structure (GCDS) (Menon *et al.*, 2020*a*) model. The GCDS model consists of two phases - an encoding phase and a decoding phase. The encoding phase consists of 5 convolution layers, that create a representation that encapsulates all fundamental features but leaves out the noise. The decoding phase consists of 5 deconvolution layers, which upsample the compressed image to restore the full image. Since the intermediate representation does not contain noise information, the restored image obtained is noise free. The model also contains skip connections between corresponding convolution and deconvolution layers. These skip connections reduce the number of weights to be trained in the neural network, which leads to quicker convergence of the model. They also enable more effective upsampling (decoding). Each skip connection is associated with a gating factor which determines the ratio of split of the information between the next convolution layer and the corresponding deconvolution layer. Max pooling layers with filter size 2×2 are added after each convolution and deconvolution. A 3×3 ker-
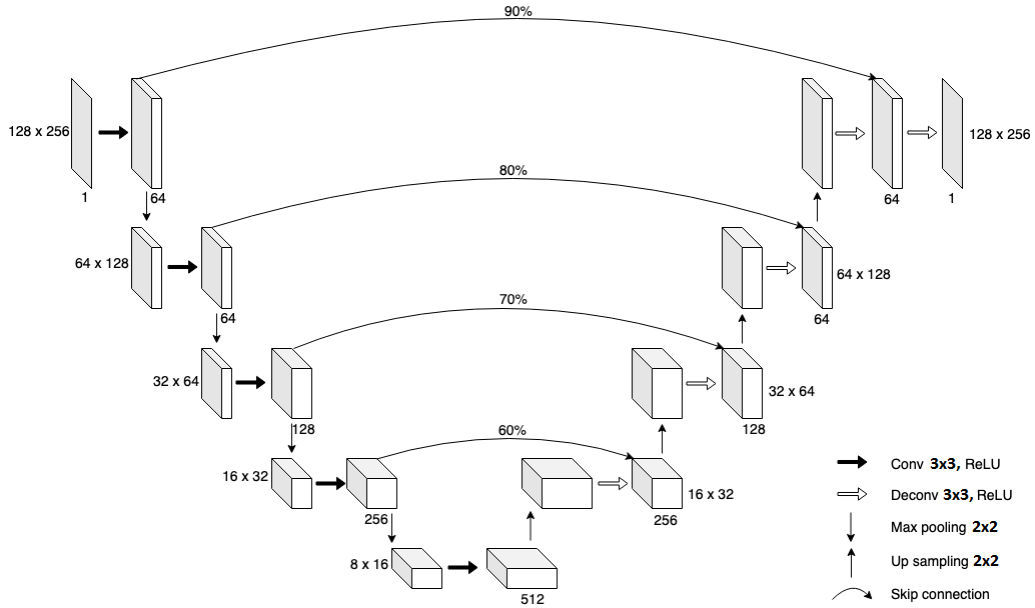
Figure 2.3: Architecture of the 10 layer GCDS model used in (Menon *et al.*, 2020*a*).

nel with ReLU activation function was used in all convolution and deconvolution layers. The GCDS architecture proposed in (Menon *et al.*, 2020*a*) is shown in Figure 2.3.

2.2.4 Limitations of GCDS model

While obtaining accurate noise characteristics allows us to generate a dataset that closely resembles the actual images that the model has to denoise, the methodology followed in (Menon *et al.*, 2020*a*) presents a few drawbacks. The model assumes that noise in all OCT scans follows Gamma distribution. The training samples were generated by corrupting the 18 ground truth images (from DUKE dataset) with Gamma noise (of difference scale and shape parameters). The model is then trained using the resized version (96 × 192) of these artificially generated training samples. Even though a large number of noisy images are constructed, the intended output still belong to the same pool of 18 images. This inadvertently creates a highly skewed dataset where the model, instead of learning the intended task of denoising, will also learn the spatial and structural features required to reconstruct those 18 images.

The limitations of the GCDS model are demonstrated in Figure 2.4. Figure 2.4(a) is the noisy image in full size and Figure 2.4(b) shows the output of the GCDS model when the patches from the full image are given as input. Figure 2.4(c) shows the output of the GCDS model when the full image is given as input. Note that the model is trained on resized images of size 96×192. Resizing the image may also change the noise characteristics (because of interpolation operations).

(a)	(b)
(c)	(d)
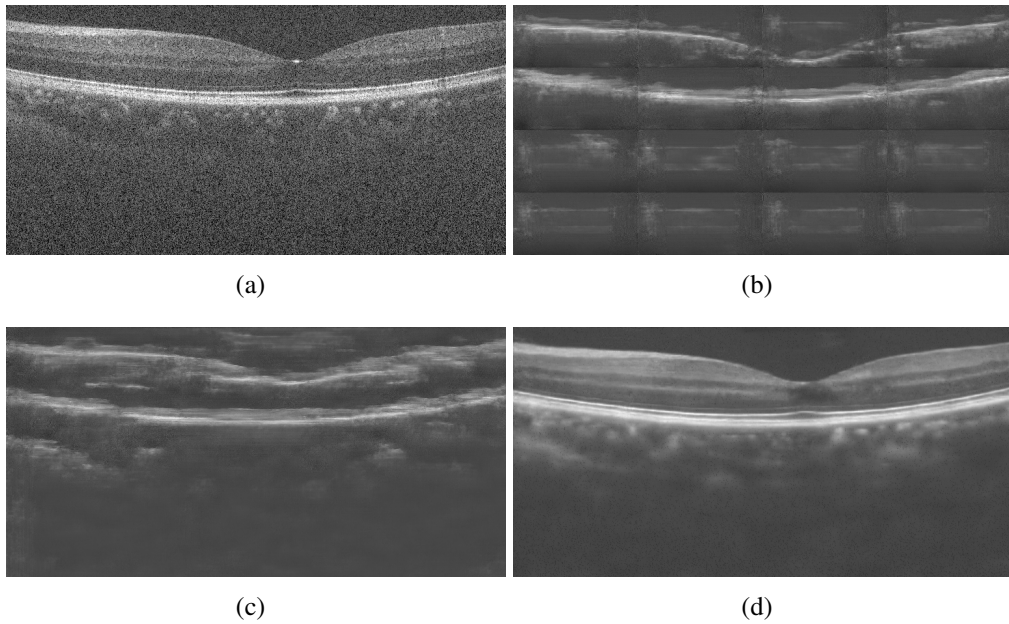
Figure 2.4: OCT denoising using the conventional GCDS and the proposed modified GCDS. (a) Sample full-size noisy image from DUKE dataset, (b) Output of the GCDS model when the patches from the full image are given as input, (c) Output of the GCDS model when the full image is given as input (the model is trained on resized images of size 96×192), (d) Image denoised by proposed model trained using patchwise method.

For applications such as medical image denoising, where training data is scarce, such overfitting of the model towards the available data is undesirable. We now present a novel approach for training denoising models in a data scarce setup in order to achieve state of the art denoising capabilities.
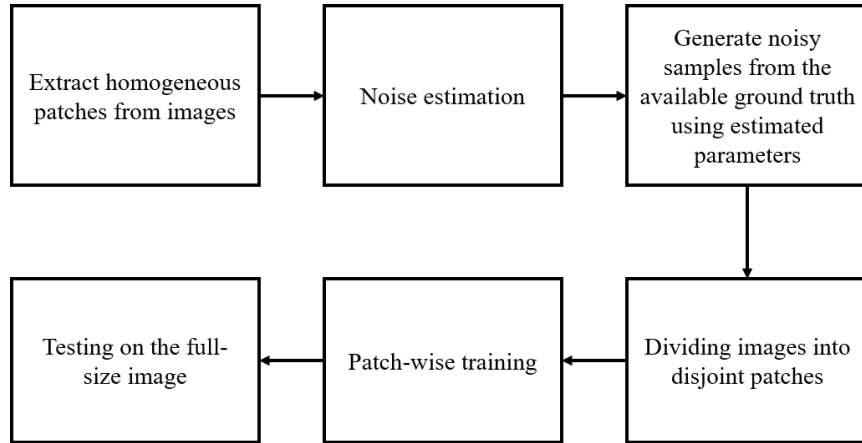
Figure 2.5: Pipeline of the proposed patch-wise training method.

2.2.5 Patch-wise training

The entire pipeline of the proposed patch-wise training method is outlined in Figure 2.5. The first step is to estimate the noise type and noise level in the target images we wish to denoise. We then generate an artificial training set by corrupting the available ground truth images with different levels of noise. The range of noise levels in the training set is based on the observations that we made in the noise study mentioned earlier. Next, we propose the patch-wise method of learning to mitigate the drawbacks mentioned in the previous section. Instead of downsampling the original image of dimension M × N to a lower dimension m × n, the original full-sized noisy image is divided into disjoint smaller patches of dimension m × n. This is demonstrated in Figure 2.6. In a highly data constrained setup, this allows to achieve the following advantages:

1. Since the model is fed with patches of the overall image as input, the training samples now consist of a highly diverse set of images. This prevents the model from overfitting to any particular type of spatial or structural feature. At the same time, this also extends the size of the training data by a considerable factor, since one image is now replaced with a batch of (M × N)/(m × n) patches.

2. The model is only shown disjoint patches from the OCT image. When we test the model by providing an entire OCT image as input - which it has never seen in its entirety, if the model is able to effectively denoise it, it is conclusive evidence of the model's generalized ability to denoise unseen images. This would prove that earlier issues of overfitting to spatial and structural features in the training set is
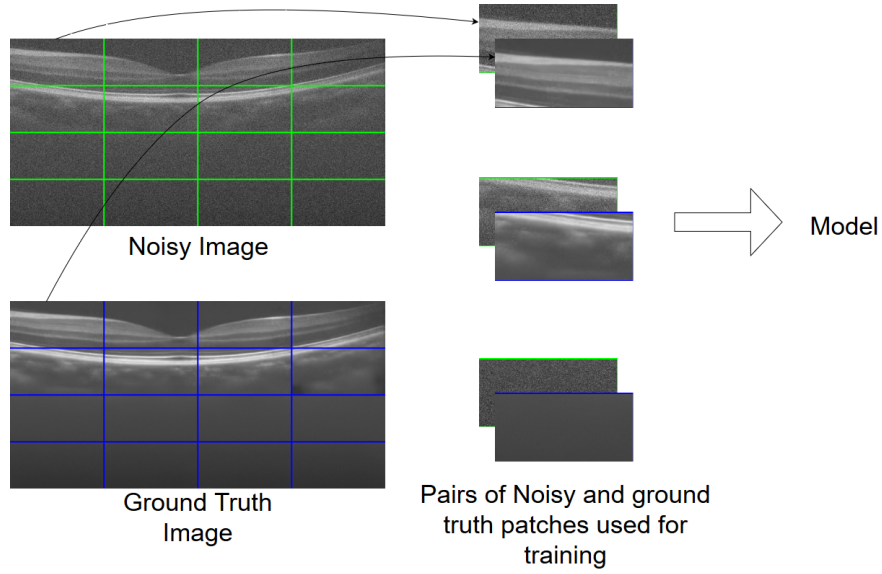
Figure 2.6: Full-sized sample training image and the corresponding ground-truth image of size 384×768 split into disjoint patches of size 96×192. The patches are then passed as a batch to the model.

mitigated. We observed that the model was able to do so, thereby confirming our hypothesis.

The experimental output of patch-wise trained model is shown in Figure 2.4(d). The model was tested on full size noisy image of dimension 450×900. It is important to note that the model was trained only on patches from the full images. Hence, this experiment also served as a good indicator of it's generalising capacity. In our experiment we used patches of size 64×64. The patch size 64×64 was selected after empirically training the models with different patch sizes. For a fair comparison with conventional GCDS model, this experiment was conducted on images corrupted with Gamma noise.

Since the images acquired with machines of different manufacturers containing different kinds of noise (or can be modeled better with different kinds of distributions), we observed that a model trained on a single noise type is not able to discern the noise characteristics and thereby unable to universally denoise the noisy OCT images from multiple vendors. Hence, we created an array of GCDS architectures and trained them with the most prominent noise types observed in Table 2.2 namely Beta, Logistic, Log-

Logistic, and Exponential. Experiments with this array of GCDS architectures gave us superior results compared to a single GCDS model. Since we had four trained GCDS models for denoising, we needed a model selector to route the input image to the correct model. For this purpose we developed a model selector.

The goal of the model selector is to automatically route the input image to the appropriately trained GCDS model. We used DenseNet 121 (Huang *et al.*, 2017) architecture for designing the model selector. It was trained to classify the input image into one of the four aforementioned classes (Beta, Logistic, Log-Logistic and Exponential). The model selector is trained on patches (of size 200×200) extracted from OCT images in the OPTIMA dataset. The training set consists of 1600 samples (400 samples each for Beta, Logistic, Log-Logistic, and Exponential) and the test set consists of 400 samples, by picking 100 samples each from four classes. The model selector achieved an accuracy of nearly 100% on the test patches; it correctly classified the input images from the OPTIMA dataset into one of the four noise models. The input to the trained model selector is the full OCT image and the model selector determines the noise type by selecting a 200×200 region from the bottom left side of the image. We observed that for most of the images this region consists of only noise.

The model was trained for 25 epochs with a batch size of 8, Adam (Kingma and Ba, 2014) as the optimization method and categorical cross-entropy as the loss function. In cases when an input image with a different noise type (other than the four noise types considered by the model selector) is provided, it selects the noise type (from among the four) that is closest match to the noise in the input image.

2.3 Results and Discussion

We used the datasets mentioned in Table 2.1 for evaluating the performance of the proposed cascaded despeckling architecture and other state-of-the-art methods. The proposed method was implemented in Keras with Tensorflow as the backend. All the

Table 2.3: Range of shape and scale parameters used to generate noisy training samples.

	shape	scale	loc	alpha
Beta	1.62 - 21.62	2.29 - 6.29	-	-
Logistic	-	5 - 8	26.5 - 29.5	-
Log-logistic	13 - 16	90 - 110	-	-
Exponential	-	-	-	6 - 12

experiments were evaluated and performed on a 64-bit workstation with an Ubuntu 18.04 operating system, Intel Xeon(R) Gold 5120 CPU @ 2.20 GHz x 28, solid state hard drive, 64 GB of RAM and NVIDIA Quadro P5000 GPU with 16 GB dedicated memory. The other conventional despeckling algorithms used for the analysis were implemented on the same machine using MATLAB R2019a.

The first stage of the proposed cascaded architecture is the model selector. The four different GCDS models designed to learn different noise types were trained separately. Every model was trained with 2176 images (2,13,248 patches) and validated on 544 images (53,312 patches). Various parameters of the noise model used to create noisy samples are mentioned in Table 2.3. The models were trained for 50 epochs with Structural dissimilarity metric (DSSIM)(Menon *et al.*, 2020*a*) as the loss metric.

2.3.1   Experiments on Synthetic Eye Images

This section discusses the experiments that we conducted on the synthetic eye image to validate the performance of the proposed method. We used two images for this experiment. One image without averaging and the other one was created by averaging 25 acquisitions. The image created by averaging 25 acquisitions was considered the ground truth. The image acquired without averaging was given as an input to the proposed method and all other methods it was compared against. The outputs of these methods were compared with the ground truth. The values of various parameters used

while implementing other filters (against which we compared) are listed below.

1. Curvature Anisotropic Diffusion (CAD) (Zhang *et al.*, 2015*a*): No. of Iterations=5, Time step=0.0625 and Conductance=3.

2. Optimized Bayesian non-local means filter (OBNLM) (Coupé *et al.*, 2009): Search area size, M=5, Smoothing parameter, h=computed as the standard deviation of a homogeneous region in the image.

3. Total Generalized Variation filter (TVG) (Duan *et al.*, 2016): alpha=[10 30 5], Theta=[5 5 5], No.0f iterations=1000.

4. Wavelet (Stankiewicz *et al.*, 2017) : Maximum wavelet decomposition level=50, Threshold=17, Wavelet basis function='Haar'.

5. K-SVD (Kafieh *et al.*, 2015*b*): Dictionary size = 256, No. of training iteration = 4000, No. of iteration = 20, blocksize = 8.

6. DnCNN (Zhang *et al.*, 2017*b*): We used the trained DnCNN provided by Zhang et al. in (Zhang *et al.*, 2017*b*).

The results of this experiment are depicted in Figure 2.7 and Table 2.4. Visual analysis shows that the image despeckled with proposed method is closer to the averaged image than the other methods against which it was compared. This is also evident from the quantitative analysis given in Table 2.4. The values (PSNR and SSIM of denoised Synthetic Eye) shown in the Table 2.4 are average of five experiments. We randomly picked 5 volumes out of 25 volumes and denoised the same frame from the 5 volumes. The mean and standard deviation of the five experiments can be studied from Table 2.4. It can be observed that the Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Metric (SSIM) of the image denoised with the proposed method is superior to those of images denoised with other methods. To demonstrate how existing methods behave on speckle noise with different distributions, we have also conducted additional experiments on the synthetic eye image. We corrupted the averaged image with four different noise types (the prominent noise types that we observed in real OCT images) and denoised the noisy images using the methods that we considered for comparative analysis. This experiment was also repeated for DUKE dataset. The results of these experiments are given in Table 2.5 and Table 2.6. These experiments also demonstrate the superiority of the proposed method over state-of-the-art methods.

Table 2.4: The quantitative analysis of the performance of the proposed denoising method with other state-of-the-art denoising methods on the synthetic eye image and Duke image in terms of PSNR and SSIM.

SI.No	Method	Synthetic EYE		Duke dataset	
		PSNR	SSIM	PSNR	SSIM
1.	Noisy	26.57(0.3218)	0.67(0.0097)	17.78(0.0493)	0.10(0.0010)
2	Wavelet (Stankiewicz et al., 2017)	28.87(0.3720)	0.81(0.0076)	20.26(0.0704)	0.16(0.0013)
3	TGV (Duan et al., 2016)	28.22(0.1417)	0.83(0.0023)	22.63(0.2357)	0.66(0.0034)
4	OBNLM (Coupé et al., 2009)	27.94(0.0968)	0.81(0.0016)	22.58(0.1381)	0.64(0.0025)
5	CAD (Zhang et al., 2015a)	29.44(0.3734)	0.78(0.0076)	22.68(0.1480)	0.39(0.0014)
6	KSVD (Kafieh et al., 2015b)	30.67(0.3762)	0.89(0.0042)	20.11(0.0825)	0.17(0.0014)
7	DnCNN (Zhang et al., 2017b)	26.76(0.3289)	0.68(0.0097)	18.03(0.0518)	0.11(0.0011)
8	Proposed Method	**31.01(0.2009)**	**0.91(0.0020)**	**23.97(0.1738)**	**0.66(0.0020)**

Table 2.5: The quantitative analysis of the performance of the proposed denoising method with other state-of-the-art denoising methods on the synthetic eye image by adding different prominent noise types present in the OCT machines (Beta distribution: shape - 8, scale- 3; Exponential distribution: alpha - 8; Logistic distribution: scale - 27, location - 6;Log-logistic distribution: shape - 13.5, scale - 95).

SI. No	Method	Noise Type							
		Beta		Exponential		Logistic		Log logistic	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	Noisy	11.27	0.33	26.06	0.89	16.59	0.62	21.82	0.72
2	Wavelet (Stankiewicz et al., 2017)	11.84	0.41	27.36	0.90	16.96	0.70	22.26	0.78
3	TGV (Duan et al., 2016)	12.01	0.48	26.40	0.81	16.87	0.64	22.14	0.73
4	OBNLM (Coupé et al., 2009)	11.52	0.47	28.48	0.84	16.93	0.69	22.15	0.80
5	CAD (Zhang et al., 2015a)	11.92	0.44	28.32	0.92	17.04	0.72	22.72	0.80
6	KSVD (Kafieh et al., 2015b)	11.96	0.45	27.02	0.84	16.95	0.66	22.36	0.75
7	DnCNN (Zhang et al., 2017b)	11.66	0.34	28.23	0.90	16.94	0.69	22.27	0.75
8	Proposed Method	**28.39**	**0.92**	**33.64**	**0.94**	**30.88**	**0.93**	**29.77**	**0.93**

2.3.2 Experiments on real retinal OCT Images

This section elaborates the experiments that we performed on real retinal OCT images. For these experiments we used the images from DUKE and OPTIMA datasets. The performance of the proposed method on the OPTIMA datasets and DUKE dataset is shown in Figure 2.8, Figure 2.9 and Figure 2.10. Careful visual inspection on the frames in Figure 2.8, Figure 2.9, Figure 2.10 and Figure 2.11 shows that the proposed method preserved structural details in the images better than other methods. Figure 2.8, Figure 2.9 and Figure 2.10 shows the experimental results on images acquired with four different machines in OPTIMA dataset. Figure 2.10 shows the results on images in DUKE

Table 2.6: The quantitative analysis of the performance of the proposed denoising method with other state-of-the-art denoising methods on the Duke dataset by adding different prominent noise types present in the OCT machines (Beta distribution: shape - 8, scale - 3; Exponential distribution: alpha - 8; Logistic distribution: scale - 27, location - 6; Log-logistic distribution: shape - 13.5, scale - 95).

SI. No	Method	Noise Type							
		Beta		Exponential		Logistic		Log logistic	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	Noisy	15.98	0.18	18.43	0.64	27.77	0.60	21.97	0.54
2	Wavelet (Stankiewicz *et al.*, 2017)	16.55	0.29	18.55	0.73	29.63	0.73	22.38	0.68
3	TGV (Duan *et al.*, 2016)	17.38	0.65	18.54	0.68	29.81	0.72	22.52	0.70
4	OBNLM (Coupé *et al.*, 2009)	17.62	0.62	18.76	0.67	28.50	0.69	22.62	0.69
5	CAD (Zhang *et al.*, 2015*a*)	17.14	0.52	18.48	0.74	29.97	0.78	22.64	0.74
6	KSVD (Kafieh *et al.*, 2015*b*)	17.09	0.40	18.60	0.70	29.76	0.75	22.69	0.73
7	DnCNN (Zhang *et al.*, 2017*b*)	16.11	0.20	18.47	0.66	28.28	0.62	22.11	0.57
8	Proposed Method	**25.32**	**0.72**	**23.70**	**0.74**	**30.02**	**0.75**	**23.70**	**0.74**

Table 2.7: The quantitative analysis of the performance of the proposed denoising method with other state-of-the-art denoising methods on the Duke dataset by adding Exponential noise of varying noise levels (alpha - 6, alpha - 7, alpha - 8, alpha - 10, alpha - 12).

SI. No	Method	Noise Type : Exponential									
		alpha - 6		alpha - 7		alpha - 8		alpha - 10		alpha - 12	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	Noisy	20.97	0.72	18.52	0.69	18.43	0.64	17.84	0.56	17.59	0.49
2	Wavelet (Stankiewicz *et al.*, 2017)	21.56	0.79	18.58	0.75	18.55	0.73	17.99	0.68	17.82	0.63
3	TGV (Duan *et al.*, 2016)	21.08	0.71	18.56	0.68	18.54	0.68	18.03	0.67	17.09	0.67
4	OBNLM (Coupé *et al.*, 2009)	21.38	0.69	18.78	0.66	18.76	0.76	18.24	0.66	18.11	0.65
5	CAD (Zhang *et al.*, 2015*a*)	21.09	0.79	18.55	0.75	18.48	0.74	17.93	0.71	17.75	0.68
6	KSVD (Kafieh *et al.*, 2015*b*)	21.16	0.73	18.62	0.70	18.60	0.70	18.09	0.69	17.97	0.69
7	DnCNN (Zhang *et al.*, 2017*b*)	21.04	0.74	18.56	0.70	18.47	0.66	17.88	0.58	17.65	0.52
8	Proposed Method	**24.23**	**0.80**	**23.95**	**0.78**	**23.70**	**0.74**	**23.19**	**0.73**	**22.31**	**0.72**

dataset. Figure 2.11 shows the zoomed view of the images for the coordinate region marked in red in Figure 2.8 and Figure 2.9. We also performed quantitative analysis on images in DUKE dataset since it had averaged images (average of 4 acquisitions). For quantitative analysis, we denoised the same frame from all four acquisitions and computed the mean PSNR and standard deviation. The results are depicted in Table 2.4. To study how the denoising methods perform on images with different noise types, we corrupted the averaged image with four prominent noise that we mentioned earlier. The results of this experiment is shown in Table 2.6. Table 2.7 shows the experimental results of the denoising methods on the Duke dataset with varying levels of Exponential noise. These experiments on the real OCT dataset also demonstrate the superior

Table 2.8: Expert Evaluation on Retinal Layer Segmentation.

Method	Score
Noisy Image	2.5
CAD	2.5
OBNLM	2.5
TGV	3
Wavelet	2
KSVD	3
DnCNN	2.5
Proposed Method	3

performance of the proposed method over other methods considered.

As a further stratagem for evaluation, we performed retinal layer segmentation (Duan *et al.*, 2017; Reddy *et al.*, 2020; Anoop *et al.*, 2020) using OCTSEG tool (Mayer *et al.*, 2011). Layer segmentation was performed on the noisy as well as the denoised images and the results are shown in Fig. 2.12. We have obtained scoring from an expert retinal surgeon on the quality of retinal layer segmentation and is given in Table. 2.8. The ophthalmologist ranked the outputs from different methods on a scale of 1 to 5, where 1 and 5 are used to indicate the worst and best results, respectively. All these experiments on synthetic and real OCT scans shows that the proposed cascaded method is a viable choice for despeckling OCT images.

2.4   Summary

A cascaded CNN based architecture to denoise OCT images is presented in this chapter. The proposed model identifies the noise type in an input image and classifies it into one of the four different noise types. It subsequently denoises the image effectively using a GCDS array that has models catering to specific noise types. To the best of our knowledge, a single model with composite architecture capable of dealing with multiple noise types has not hitherto been reported in the literature. Models reported in the literature are trained on one noise type alone and this makes them vulnerable to

errors when input images have noise different from that assumed.

Another unique feature proposed in our work is the use of the synthetic eye to create datasets for simulations and experiments. This dataset was useful in doing quantitative evaluation of different OCT denoising methods. Other important contribution in this chapter is a patch-wise training approach to improve the performance of the existing GCDS model. This methodology prevents the model from learning the structural details and also provides a way of dealing with limited ground truth data. Experiments on synthetic and real OCT datasets show that the proposed method performs better than other reported methods.
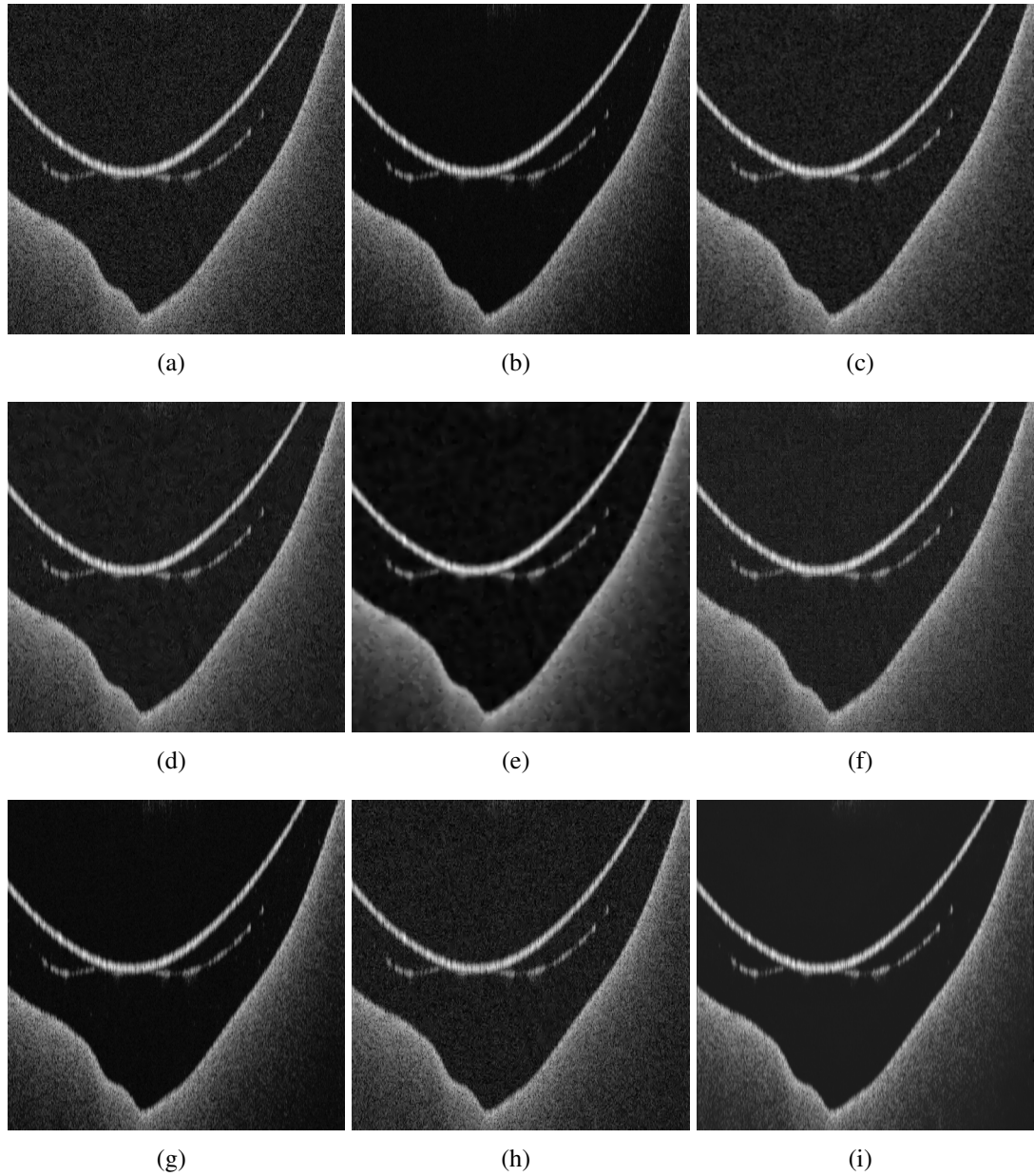
Figure 2.7: The visual analysis of the performance of the proposed denoising method with other state-of-the-art denoising methods on the synthetic eye dataset. (a) Noisy synthetic eye image, (b) Ground truth image, (c) CAD, (d) OBNLM, (e) TGV, (f) Wavelet, (g) KSVD, (h) DnCNN, and (i) Proposed method. All images are displayed in the range 0-255.

35
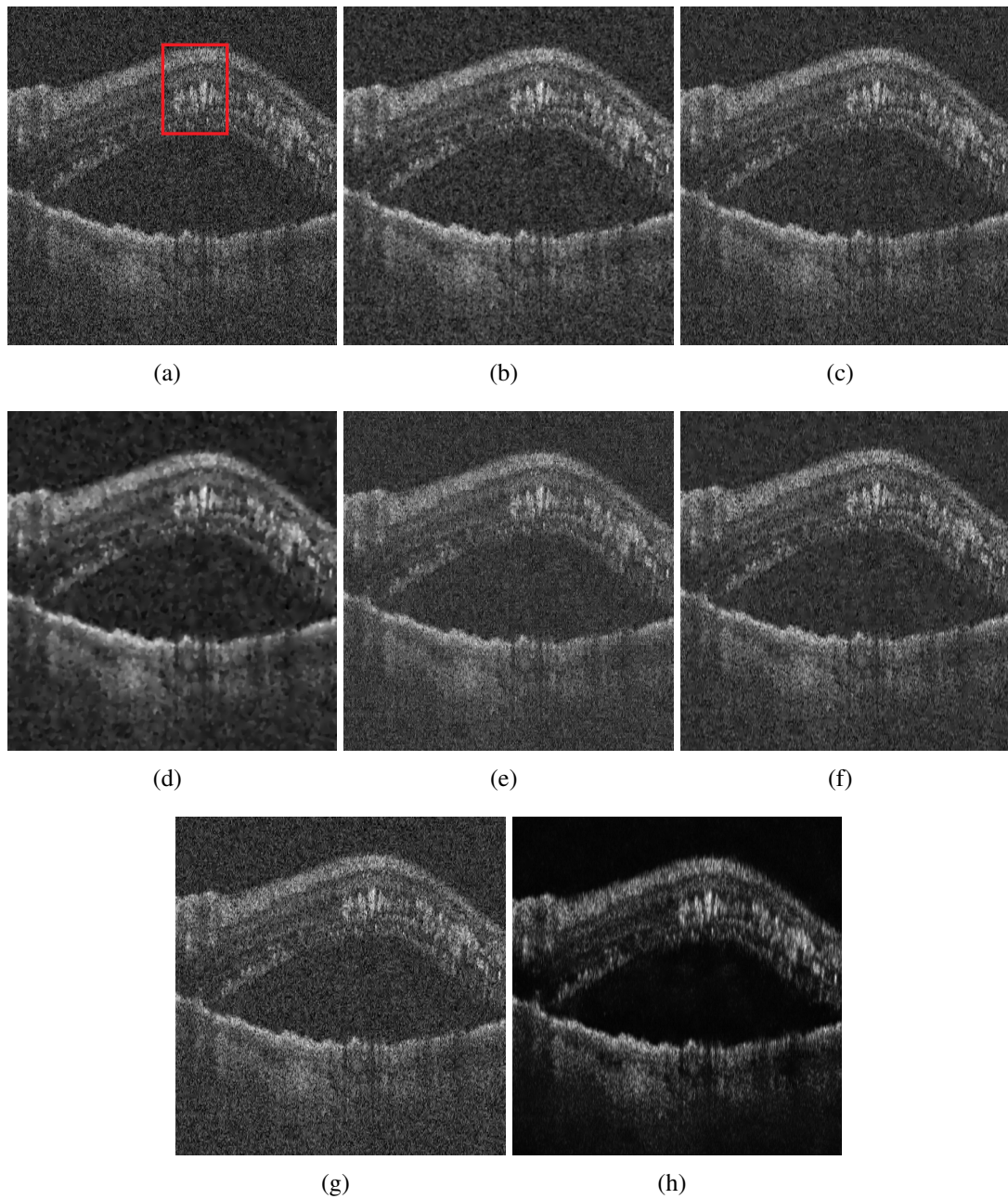
Figure 2.8: The visual analysis of the performance of the proposed denoising method with other state-of-the-art denoising methods on the Cirrus data from the OPTIMA dataset. (a) Noisy synthetic eye image, (b) CAD, (c) OBNLM, (d) TGV, (e) Wavelet, (f) KSVD, (g) DnCNN, and (h) Proposed method. The intensity values of all the images are displayed in the range of 0-255.
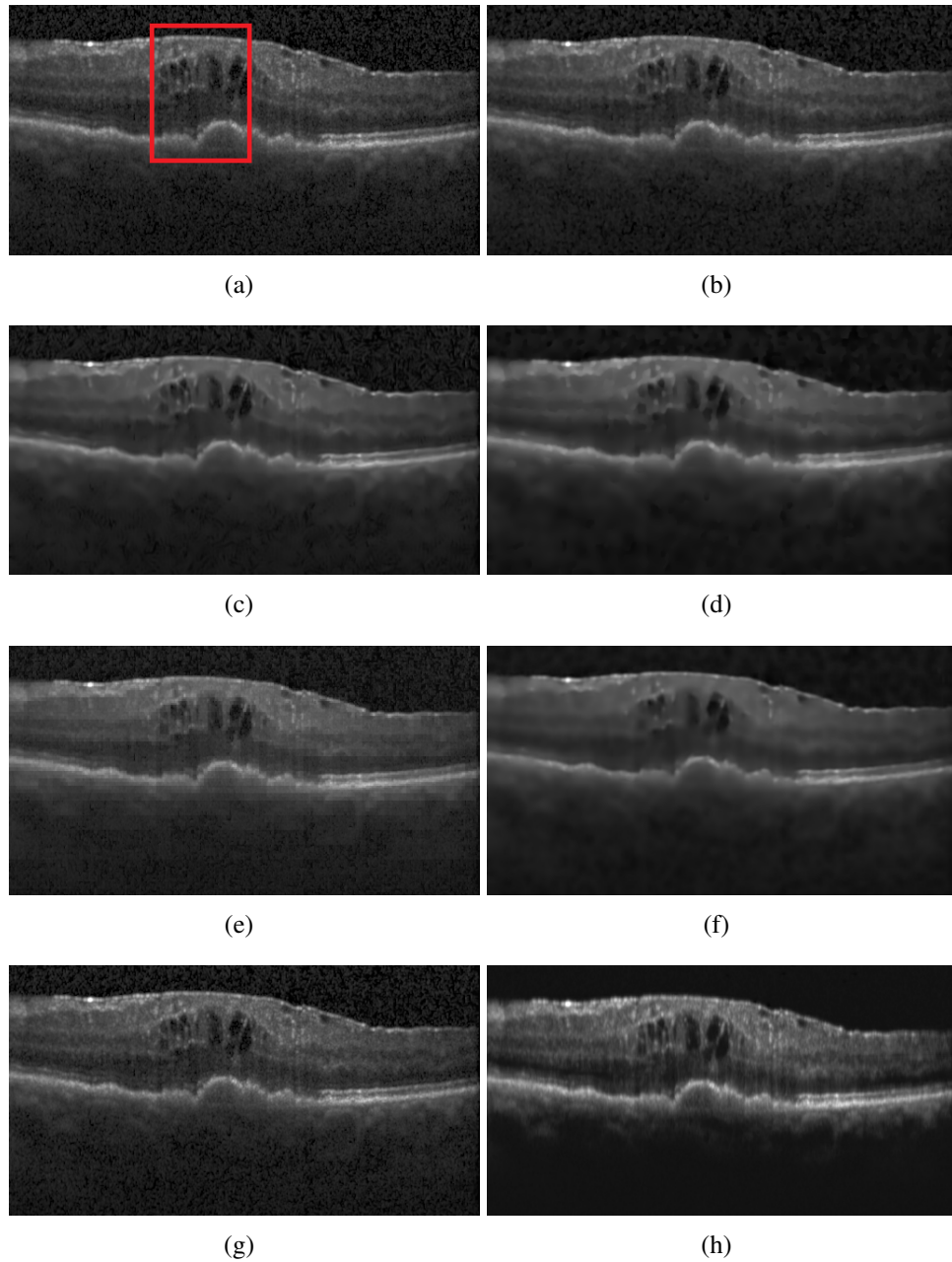
Figure 2.9: The visual analysis of the performance of the proposed denoising method with other state-of-the-art denoising methods on the Spectralis data from the OPTIMA dataset. (a) Noisy synthetic eye image, (b) CAD, (c) OBNLM, (d) TGV, (e) Wavelet, (f) KSVD, (g) DnCNN, and (h) Proposed method. The intensity values of all the images are displayed in the range of 0-255.
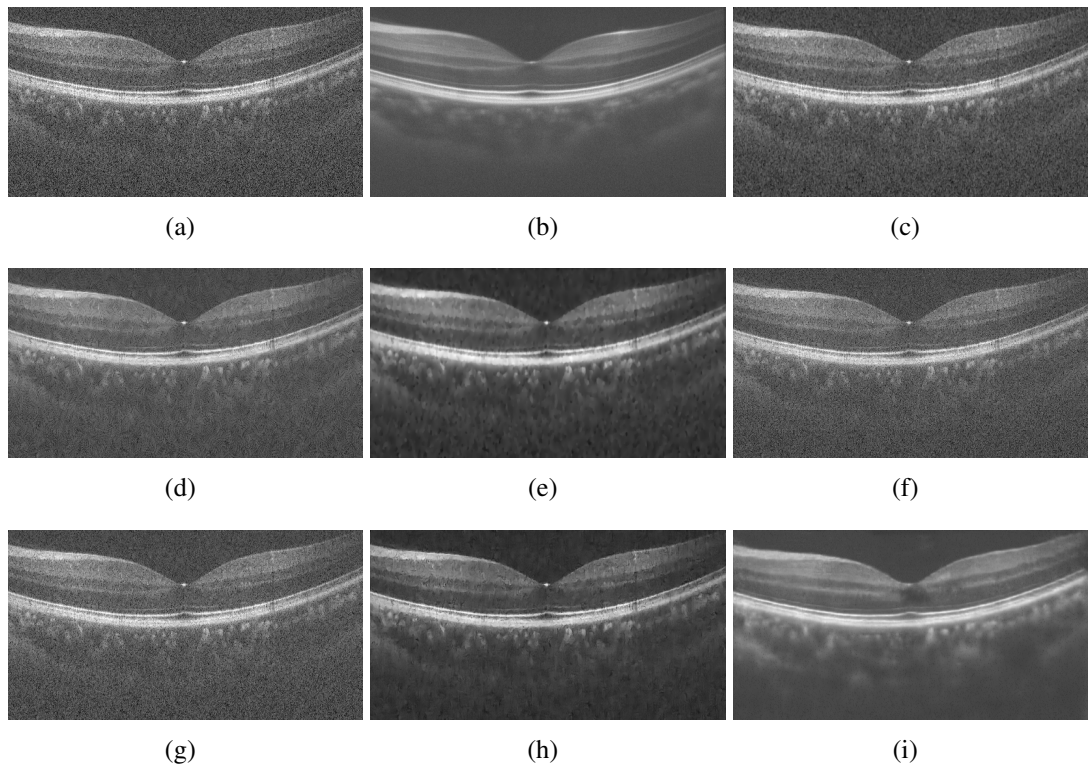
Figure 2.10: The visual analysis of the performance of the proposed denoising method with other state-of-the-art denoising methods on the Duke dataset. (a) Noisy image, (b) Ground truth image, (c) CAD, (d) OBNLM, (e) TGV, (f) Wavelet, (g) KSVD, (h) DnCNN, and (i) Proposed method. All images are displayed in the range 0-255.
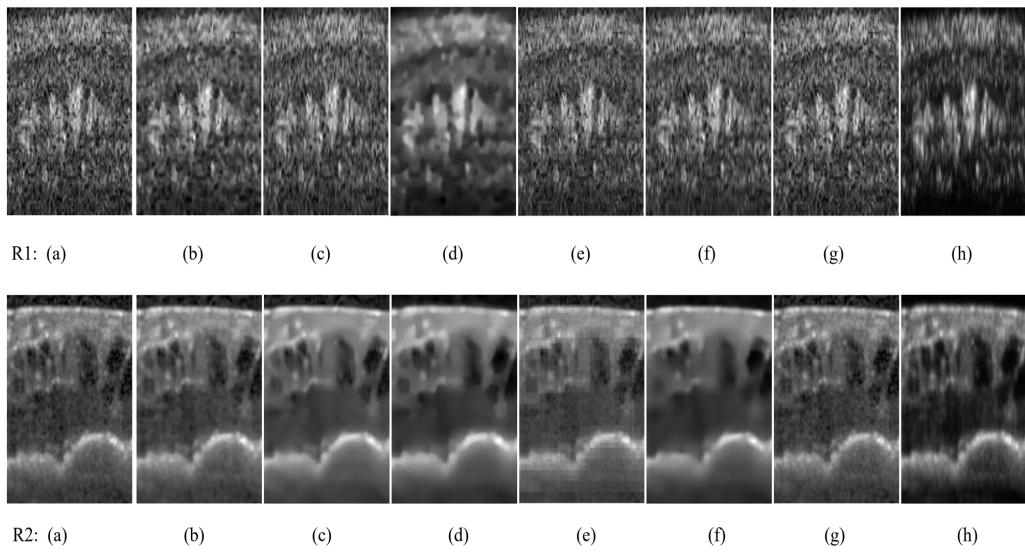
Figure 2.11: The qualitative analysis of the performance of the proposed denoising method with other state-of-the-art denoising methods (with a zoomed region) on Cirrus (R1) and Spectralis (R2) vendor data of OPTIMA dataset. (a) Noisy input image, (b) CAD, (c) OBNLM, (d) TGV, (e) Wavelet, (f) KSVD, (g) DnCNN, and (h) Proposed method. All images are displayed in the range 0-255.
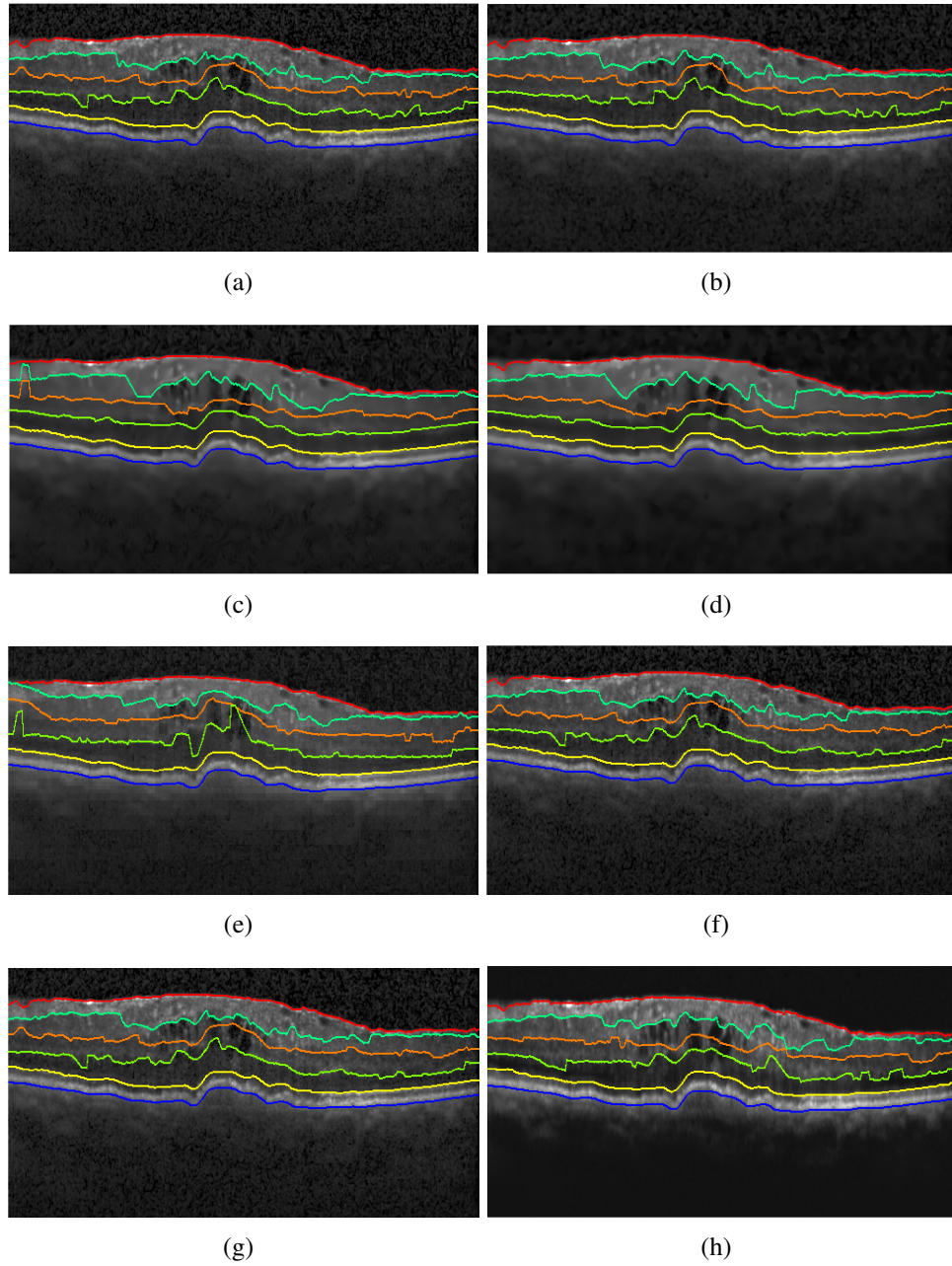
Figure 2.12: The qualitative analysis of the retinal layer segmentation performed on the proposed denoised algorithm with respect to the state-of-art denoising algorithms on Spectralis vendor data of OPTIMA dataset. (a) Noisy input image, (b) CAD, (c) OBNLM, (d) TGV, and (e) Wavelet, (f) KSVD, (g) DnCNN, (h) Proposed method. All images are displayed in the range 0-255.

CHAPTER 3

STACK GENERALIZED DEEP ENSEMBLE LEARNING FOR RETINAL LAYER SEGMENTATION IN OPTICAL COHERENCE TOMOGRAPHY IMAGES

A Fully Convolutional Network (FCN) termed as DelNet based on a deep ensemble learning approach to selectively segment retinal layers from OCT scans is proposed in this Chapter[2]. The proposed model is tested on a publicly available DUKE DME dataset. Comparative analysis with other state-of-the-art methods on a benchmark dataset shows that the performance of DelNet is superior to other methods.

3.1   Introduction

The retina is a thin multi-layered tissue at the back of the eye which converts photic signals to electrical nerve impulses. Due to the retina's important role in vision, retinal diseases may cause temporary or permanent blindness. The retina has 10 anatomical layers that are seen on OCT images as alternating bands with hyper- and hypo-reflectivity. Layers primarily composed of axons and their plexiform connections appear hyper-reflective. Layers with cell bodies and nuclei appear hypo-reflective. Major retinal disorders cause fluid accumulation within and under the retina, whereas others

cause loss of retinal tissue, extra tissue over or under the sensorineural retina (Girish *et al.*, 2016, 2018*b*).

OCT is one of the most popular non-invasive imaging procedures used in ophthalmology (Drexler *et al.*, 2001). OCT devices allow cross-sectional viewing of retina with localization of level of the lesion within the retina (Girish *et al.*, 2018*a*). Most OCT device software allows precise detection of retinal boundaries in normal or minimally distorted retina (Drexler *et al.*, 2001). Their reliability suffers in images with retinal disorders where moderate to severe retinal distortion is present. Layer boundaries in these cases are frequently identified incorrectly. Another major challenge while segmenting retinal layers in OCT images is the presence of speckle noise (Anoop *et al.*, 2019; Duan *et al.*, 2016). Speckle degrades the quality of OCT images and it may influence the performance of automated methods. Despeckling is required in this case to enhance the quality of OCT images (Girish *et al.*, 2018*b*). The role of OCT retinal layer thickness measurements in identifying and predicting ocular and even systemic disease (Glaucoma (Schuman *et al.*, 1995), alzheimers diseases (Paquet *et al.*, 2007) and Multiple sclerosis (Gordon-Lipkin *et al.*, 2007), has gained widespread attention. Hence, there is a need to refine the performance of retinal layer segmentation algorithms to enable better prediction and diagnoses of these diseases, and similar emerging indications (Parkinsons (Inzelberg *et al.*, 2004), ataxia (Michalik *et al.*, 2004) and other neurodegenerative disorders (Lamirel *et al.*, 2010).

Several semi-automatic methods and automatic methods are proposed in the literature for retinal layer segmentation. In (Fernández *et al.*, 2005), Fernandez et al. showed that retinal layers can be automatically and/or interactively determined from retinal OCT images by inferring the local characteristics of retinal layers. Yin et al. (Yin *et al.*) proposed a user-guided segmentation method using likelihood estimation. Mayer et al. (Mayer *et al.*, 2008) proposed an algorithm based on fuzzy c-means approach for the segmentation of the retinal layers. Experiments shows that the performance of this method is influenced by outliers and it may settle in the local minima. Xiang et al. (Xi-

ang *et al.*, 2019) suggested a random forest classifier based segmentation algorithm for the retinal layer segmentation. A live wire algorithm (the lowest cost path algorithm) is employed in (Xiang *et al.*, 2019) for the accurate prediction of the retinal layer boundaries. This algorithm uses a set of handcrafted features for the layer segmentation. Even though this algorithm gave better accuracy in the custom dataset, similar performance is not guaranteed in other datasets since the handcrafted features used is computed from the custom data set that they used. Other drawbacks of this method are the number of trees required for the optimal results, the interpretability and overfitting.

Garvin et al. (Garvin *et al.*, 2008) studied the feasibility of graph theory and energy minimization constraints (Li *et al.*, 2005) and proposed an automated algorithm for the segmentation of the five layers from retinal OCT images. They also suggested a modified version (Garvin *et al.*, 2009) of this approach by employing feasibility constraints and regional information. Lu et al. (Lu *et al.*, 2010) and Yang et al. (Yang *et al.*, 2011) studied the feasibility of gradient-based approaches for retinal layer segmentation using an optimal graph search method. Duan et al. (Duan *et al.*, 2018*b*) proposed a generative retinal layer segmentation method based on group wise curve alignment. Ghorbel et al. (Ghorbel *et al.*, 2011) proposed a segmentation algorithm using the active contour method and random Markov fields. Yazdanpanah et al. (Yazdanpanah *et al.*, 2011) proposed an active contour-based segmentation algorithm with shape constraint and contextual weights. Duan et al. proposed a geodesic distance-based retinal layer segmentation (Duan *et al.*, 2017) and an advanced version of it by employing shape constraint to intensity-based Mumford-Shah variational functional algorithm (Duan *et al.*, 2018*a*) for the segmentation of nine retinal layers. The limitations of these algorithms are its inability to segment the layers properly when the quality of the images are not good or when the layer boundaries are not that visible in the images. The graph cut optimization, kernel regression and machine learning technique such as Support Vector Machines (SVM) are also used for the retinal layer segmentation (Chiu *et al.*, 2010, 2015; Dufour *et al.*, 2013; Ehnes *et al.*, 2014; Shi *et al.*, 2015; Vermeer *et al.*, 2011; Srinivasan *et al.*, 2014).

The aforementioned methods proposed for retinal layer segmentation are not end-to-end models. Usually, heuristics and hand-crafted features are used. To address these problems, recently convolutional neural network (CNN) based methods were proposed to segment the retinal layers. In CNN, the handcrafted features are replaced with features computed using convolution operations and also CNN based methods are the state-of-the-art. In this category, Fang et al. (Fang *et al.*, 2017) experimented a combination of CNN and graph search methods for automatic segmentation of layer boundaries from retinal OCT images. Hamwood et al. (Hamwood *et al.*, 2018) also used a combination of CNN and graph search based segmentation algorithm, and studied the effect of patch sizes in training the model and the behavior of parameter variations in the CNN models for the layer segmentation.

Roy et al. (Roy *et al.*, 2017) proposed a fully convolutional neural network architecture termed as ReLayNet, for end-to-end segmentation of retinal layers. Kugelman et al. (Kugelman *et al.*, 2018) used a recurrent neural network as the feature extractor and graph search for classification. Shah et al. (Shah *et al.*, 2018) showed that a single CNN based framework can be used to segment multiple surfaces simultaneously. Kiaee et al. (Kiaee *et al.*, 2018) developed a 3D fully convolutional encoder-decoder structure for automated segmentation of retinal layers in OCT scans. Reddy et al. (Reddy *et al.*, 2020) proposed a CNN architecture named as DilatedReLayNet for the segmentation of retinal layers by extending the approach proposed in (Roy *et al.*, 2017) by replacing normal convolution operations with dilated convolutions. One of the biggest challenges in developing the CNN based method for retinal layer segmentation is the unavailability of sufficient labeled data.

Apart from this, there are software packages like OCTExplorer (Li *et al.*, 2005; Garvin *et al.*, 2009; Abràmoff *et al.*, 2010; Quellec *et al.*, 2010; Antony *et al.*, 2011; MeindertNiemeijer *et al.*, 2012) and OCTSEG tool (Mayer), available for automated segmentation of retinal layers from OCT images. OCTExplorer is developed by the Retinal Image Analysis Lab (Iowa Institute for Biomedical Imaging, Iowa City, IA).

This publicly available, standalone tool segments 10 layers of the retina from the volumetric OCT data by employing graph theory-based segmentation approach. It can also read most formats of the volumetric OCT scans acquired with various OCT machines. The OCTSEG is also a publically available retinal layer segmentation tool (created by Markus Mayer). Six prominent retinal layers can be automatically segmented using OCTSEG tool. Unlike the OCTExplorer, segmentation from 2D circular and linear scans, as well as volume scans, are possible with OCTSEG tool.

Literature on the biopsy of the human retina reports 18 layers (Staurenghi *et al.*, 2014). Out of this only 11 layers can be distinguished by humans with naked eyes (Kafieh *et al.*, 2015*a*). There are algorithms proposed in the literature to segment 5 to 11 layers from retinal OCT images (Mayer). Among them, the methods that segment more number of layers usually perform well on the normal retina and fail to perform equally well on retinal images with pathologies. This is mainly because the pathologies distort the structure of the layers present in the retina. Also, the clinical importance of segmenting more numbers of layers are not studied well. Hence, in this study, we considered segmenting only those layers which are clinically significant.

Our major contribution in this work is the development of an ensemble architecture for the effective segmentation of most prominent retinal layers. The proposed ensemble based architecture helped the model to learn different layers better when compared to the stand-alone single models. This is demonstrated in this chapter with the help of experiments on standard benchmark dataset and also through Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju *et al.*, 2017) visualization. Various experiments are conducted to optimize the model and we evaluated the performance of the model using standard evaluation metrics. Experiments were also conducted to study how the model performs on normal retinal scans and subjects with pathologies. We also studied the influence of despeckling in the retinal layer segmentation process by repeating the experiments on the raw dataset as well as the denoised dataset.

The following section (section 3.2) provides detailed description of the proposed methodology (including information about the data set, overall pipeline, and experimental settings). Section 3.3 discusses various experiments conducted for evaluating the proposed method. Finally, Section 3.4 summarizes and concludes the chapter.

3.2 Methodology

The proposed CNN model is an ensemble learning based approach and contains four base models followed by a predictor block. The objective of the proposed model is to map the pixels in the given retinal OCT image $I(x, y)$ to a particular label in the label space $Q=\{q\}=\{1, ..., q\}$ for q classes. The proposed model addressed the segmentation task as $q = 8$ class classification problem which includes the tissue classes of 7 retinal layers and a background class.

3.2.1 Proposed DelNet Architecture

The proposed DelNet architecture is an ensemble based stack generalization (Wolpert, 1992) of deep learning models (Xiao *et al.*, 2018) for computerized segmentation of retinal layers from OCT scans. In the literature, ensemble-based approaches have been used in classification and regression tasks (Ren *et al.*, 2016). The basic idea of ensemble based model is to generate a set of weak base-learners and to combine their output instead of trying to create one single optimal learner. Stack generalization (Mitchell *et al.*, 1986) refers to a special type of ensemble learning, where we use another model (or convolutional block) to combine the outputs from the base-learners. Here we use a predictor model to combine the learners feature map from the base models to compute the output.

The main motivation behind using ensemble learning is due to the fact that the characteristics of each layer is different from other layers in the retina (i.e, presence

46
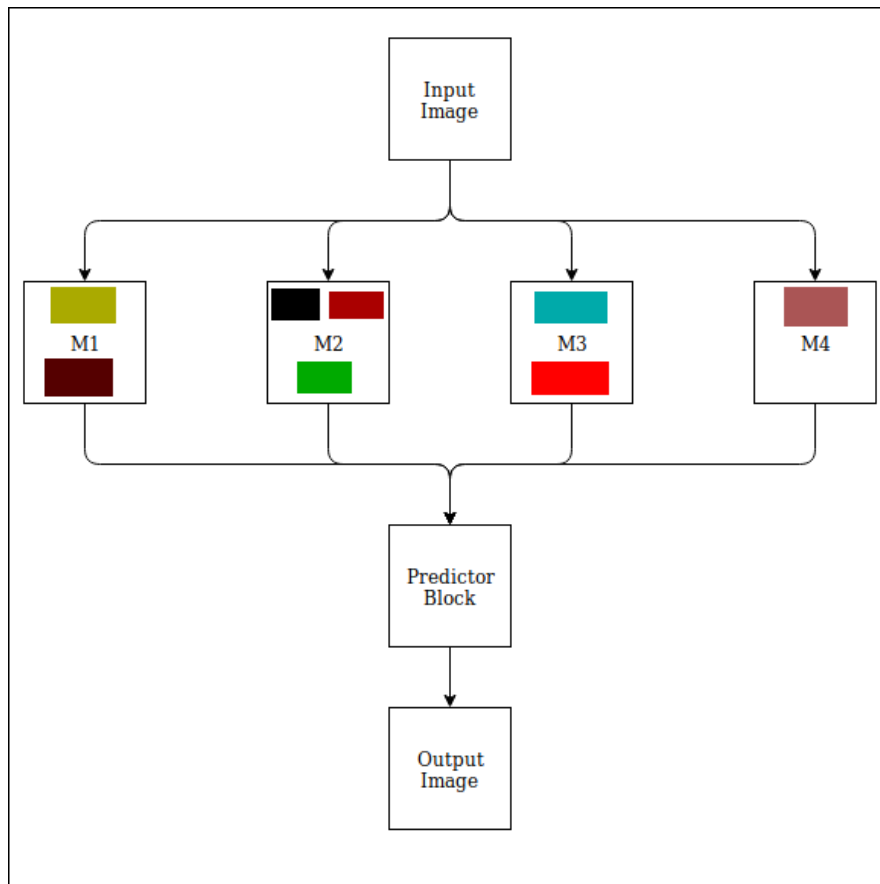
Figure 3.1: The Proposed DelNet architecture. The input image is fed to the ensemble model. All the 4 models make independent predictions on the same input image and the predictions from all the 4 models are processed by the predictor block. The color codes shows the layer that a particular models focuses. The colour coding of the retinal layers is depicted in Fig. 3.5.

of hyper and hypo reflective layers, as well as variations in the spatial structure and alignment of the layers) (Hee *et al.*, 1995). This difference is very apparent between the thicker and thinner layers, and using a fixed model for segmentation might not be a good choice here because the model will only learn the features that are common to all the layers and as a result the performance will be sub-optimal. Hence, we use a set of four different models, each of which is fine tuned to meet the requirement of a particular class of similar layers. By training each of these models separately on the data and using another trained convolutional block to get a combined output, the model can do a better prediction compared to a fixed model that learns all the layers together.

In the proposed DelNet architecture we used four base models and each of them is fine tuned to meet the requirement of the layers that have common features. Every base model predicts eight classes (7 layers + background) but with different levels of accuracy. The output of the base models is fed to the predictor block which finally classifies each and every pixel in the image to one of the eight classes. The number of models to use as base-learners is again a hyper-parameter, and based on our experimental analysis we selected four base-learners. Using more than four models did not improved the result and increases the computational complexity.

The overall learning pipeline is as follows: first train each of the base-learners individually on the dataset. The input is an image of dimension 216×64 and the output is the prediction of dimension $216 \times 64 \times 8$ against the one hot encoded ground truth which is also of dimension $216 \times 64 \times 8$. The eight output layers contains eight different classes. The next step is to fix the base-learners and train the predictor block. The predictor block takes input as the output of all four base-learners stacked on top of each other (total dimension $216 \times 64 \times 32$) and produce an output of dimension $216 \times 64 \times 8$. The loss is computed against the ground truth (of dimension $216 \times 64 \times 8$). Fig. 3.1 shows the proposed DelNet architecture. All models M_i ($1 \leq i \leq 4$) have the same base architecture (with different hyper parameters) followed by a predictor model. Out of the four base learners, Model M1 will focus on the Inner Nuclear Layer (INL) and Outer

Nuclear Layer - Inner Segments of Myeloids (ONL-ISM). Model M1 predicts these two layers with high accuracy when compared to other base learners. Model M2 focuses on three layers such as Background, Internal Limiting Membrane (ILM) and Nerve Fibre Layer - Inner Plexiform Layer (NFL-IPL). Model M3 has tuned to segment the Outer Plexiform Layer (OPL) and Inner Segment Ellipsoid (ISE). Model M4 focuses on the Outer Segment - Retinal Pigmented Epithelium (OS-RPE) layer. The colour coding of the retinal layers is depicted in Fig. 3.5. The base model and the predictor model architecture are explained in the following subsections.

3.2.1.1 Base Model

The base model of the proposed DelNet architecture is depicted in Fig. 3.2. The base model follows DilatedReLayNet architecture (Reddy *et al.*, 2020), which follows an encoder-decoder structure with dilated convolutions, skip connections, and classification layer. In (Reddy *et al.*, 2020), dilated convolution filters are used to enhance the receptive field without increasing the parameters in the network. Skip connections are used to traverse information faster through concatenation layers for better learning capabilities. The final classification layer - a softmax activation layer which produces the probability score for each class. Even though all the base learners are based on DilatedReLayNet architecture, the hyper parameters used in different ensemble models are different. The hyper parameters for different base learners are fixed in such a way to give maximum performance in segmenting the layers that each base learner is intended for.

3.2.2 Prediction Model

The predictor block is a CNN model which combines all four base learners. This model receives the output prediction map of four base models as input and produces a final segmentation output. The predictor model contains four fully convolutional layers in
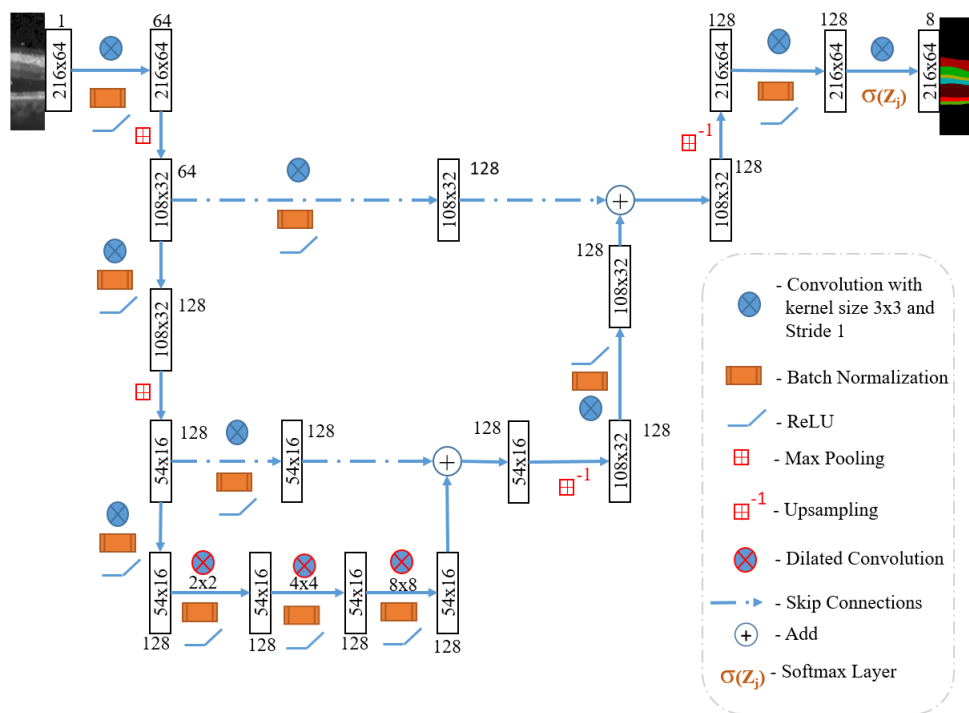
Figure 3.2: The architecture of the base model used in all the 4 rank ensemble models M1 to M4. The base model follows DilatedReLayNet architecture (Reddy *et al.*, 2020) trained with different set of hyperparameters.
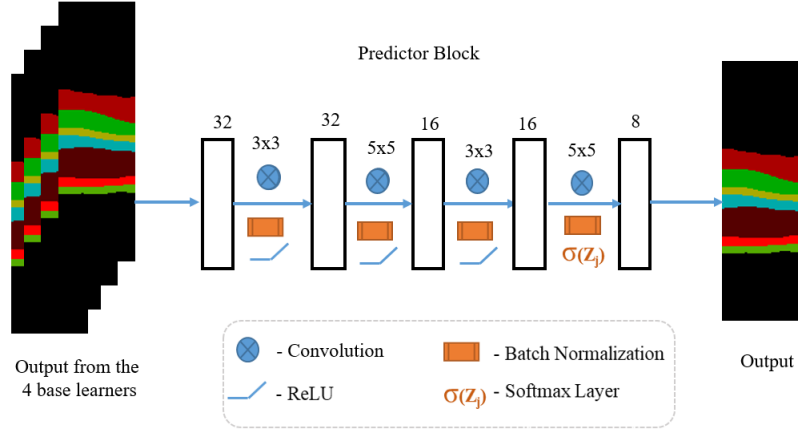
Figure 3.3: The architecture of the predictor block, which is used to fuse the outputs of the base-learners.

which the initial layer input is a stacked version of four base layers (each of size $216 \times 64 \times 8$ yielding $216 \times 64 \times 32$). The predictor model architecture is shown in Fig. 3.3.

3.2.3 Training

3.2.3.1 Cost Function

The cost function considered in our model is a combination of multiclass weighted cross entropy loss and Dice loss and is given below (Roy *et al.*, 2017):

$$loss = \lambda_1 MCWCE + \lambda_2 DL + \lambda_3 \left\| W^{(.)} \right\|_F^2 \qquad (3.1)$$

where $loss$ is the training loss, $MCWCE$ is the multiclass weighted cross entropy, DL is the Dice loss and the quantity $\left\| W^{(.)} \right\|_F$ is a regularisation term which represents the Frobenius norm of the weights matrix. The λ parameters determine the trade-off between different components in the loss function. We used the multiclass weighted cross entropy method suggested in Reddy et al. (Reddy *et al.*, 2020). They used a weighting scheme based on the median frequency method and was found effective in segmenting the retinal layers. This weighing scheme assures that the minority class pixels carry

a weight than the value of weights designated to any of the predominant classes. Interested readers can refer (Reddy *et al.*, 2020) for more information. The motivation behind using the combined loss function is to trade off between the advantages and disadvantages of the respective loss functions.

The reason for including Dice loss component in the loss function is to maximize the Dice coefficient. Dice coefficient performs better at class imbalanced problems and segmentation of finer small structures in the image. But it leads to unstable optimization and gradients blowing up and thus the training process convergence is delayed. On the contrary, cross entropy is just a proxy loss which is easier to maximize using backpropagation. cross entropy loss function leads to better (faster and stable) convergence due to smooth form and nicer gradients for multi-class problems. The gradients of cross entropy is of the form $Q - P$, where Q is the softmax outputs and P is the target. The class imbalance can be taken care of by simply assigning loss multipliers to each class, such that the network is more sensitive to not ignore a particular class pixel which is infrequent. The regularisation term $\left\| W^{(.)} \right\|_F$ ensures that the network doesn't over-fit to the training sample and improves its ability to generalize better. More analysis of the loss function and its parameter sensitivity can be found in papers (Wong *et al.*, 2018; Sudre *et al.*, 2017; Zhu *et al.*, 2019).

3.2.4 Data Set and Preparation

The experiments are conducted on the publicly available Duke DME SD-OCT dataset (Chiu *et al.*, 2015). A short description of the DUKE DME dataset is provided in the Table 4.1. The dataset consists of OCT scans acquired from 10 subjects with DME. Out of the 10 volumes, 110 images (11 B-scans per subject) were annotated for the retinal layers by two expert ophthalmologists. Each image is having a resolution of 496×768. Out of these 110 images, 57 images are the B-scans of the normal retina and the remaining 53 are the B-scans of retina having pathology.

Table 3.1: The description of the DUKE DME dataset (Chiu *et al.*, 2015)

No. of Volume	No. of annotated B-scans / Volume	Total No. of B-Scans with annotations	The dimensions of each B-scans	Make of machine employed for acquiring the OCT scans
10	11	110	496×768	Spectralis SD-OCT (Heidelberg Engineering Inc., Heidelberg, Germany)
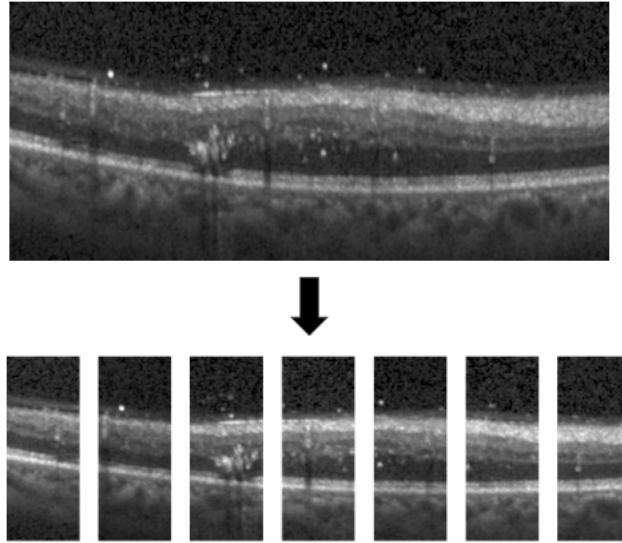
Figure 3.4: Vertically slicing of the input OCT B-scan into 7 equal parts.

To avoid class imbalance problems (due to background), background cropping is employed and the dataset now comprises 110 images of size 216×500. The number of training images plays a vital role in training fully convolutional networks. Hence, the images are vertically sliced as in Reddy et al. (Reddy *et al.*, 2020) to increase the number of samples to 770 images. Each image is now of size 216×64. The dataset preparation process is depicted in Fig. 3.4 and Fig. 3.5. The eight classes of the retinal layers considered in this work with the segmentation mask are shown in Fig. 3.5.

3.2.4.1   Experimental Settings

Thus the samples available for conducting experiments are 770 images. Out of these 770 images, 500 images were used for training and 40 images were used for validation
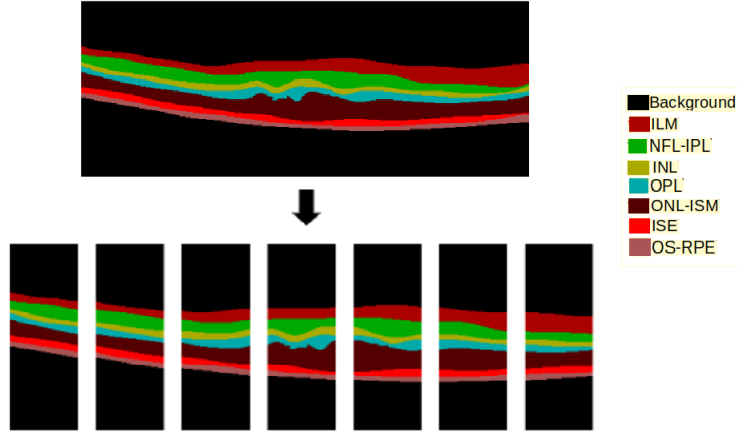
Figure 3.5: The segmentation masks (ground truth) are also sliced vertically in similar manner to the test images.

and the remaining 230 images were employed for testing. To improve the models capacity to generalize, the train, validation and test images were disseminated in an equal way over the data. The values of $\lambda 1$, $\lambda 2$ and $\lambda 3$ in the cost function are set as 1, 0.5 and 1×10^{-4} respectively (chosen experimentally, other combinations yielded inferior results).

All model weights were initialized randomly and the optimization was carried out using the Stochastic Gradient Descent (SGD) optimizer with Nesterov Momentum (Zhang *et al.*, 2015*b*). A learning rate reducer was used to monitor the learning of the model which reduces the learning rate by a factor of 0.5 on the plateau where the loss stops decreasing for more than 6 epochs is used. The minimum learning rate for training the model is set as 5×10^{-6}. The number of epochs and batch sizes for different models were fixed based on the experimental analysis. The details of the hyper-parameters used in all four base learners used in the proposed DelNet model is given in Table 3.2. The training and validation errors of the 4 models used in the proposed method are shown in Fig. 3.6. The training and validation errors of the prediction block used in the proposed method is shown in Fig. 3.7. We implemented all the modules of the ReLayNet (Roy *et al.*, 2017), DilatedReLayNet (Reddy *et al.*, 2020), DelNet architectures using Python

Table 3.2: The details of the hyper-parameters used in all four base learners used in the proposed DelNet model.

Model	No.of epochs	Batch size	Initial learning rate	Momentum
M1	75	32	0.005	0.9
M2	60	20	0.01	0.92
M3	57	20	0.005	0.9
M4	60	128	0.01	0.95
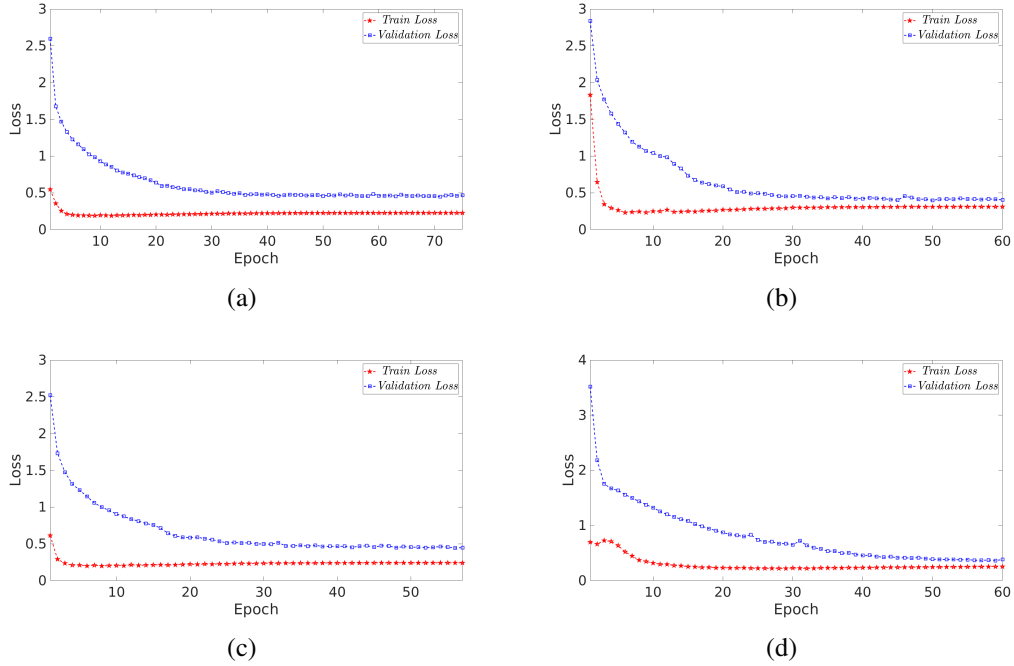
Figure 3.6: The training and validation errors of the 4 models used in the proposed method. (a) Model 1, (b) Model 2, (c) Model 3, (d) Model 4.

3 with the help of Keras TensoFlow API (Chollet *et al.*, 2015) on Intel Xeon CPU with 64GB RAM and Nvidia Tesla K40 GPU with 10 GB dedicated memory. The other conventional layer segmentation algorithms used for the analysis are implemented on a workstation with a 64-bit Ubuntu 18.04 OS, Intel Xeon(R) Gold 5120 CPU @ 2.20 GHz x 28, solid-state hard drive, 64 GB of RAM and NVIDIA Quadro P5000 GPU with 16 GB dedicated memory using MATLAB R2019a.

From the loss curve graphs in Fig. 3.6 we can observe that the models converge at different points and the loss follows different curve paths for different hyper-parameter settings. This is a useful behaviour we exploit in ensemble learning by making use of
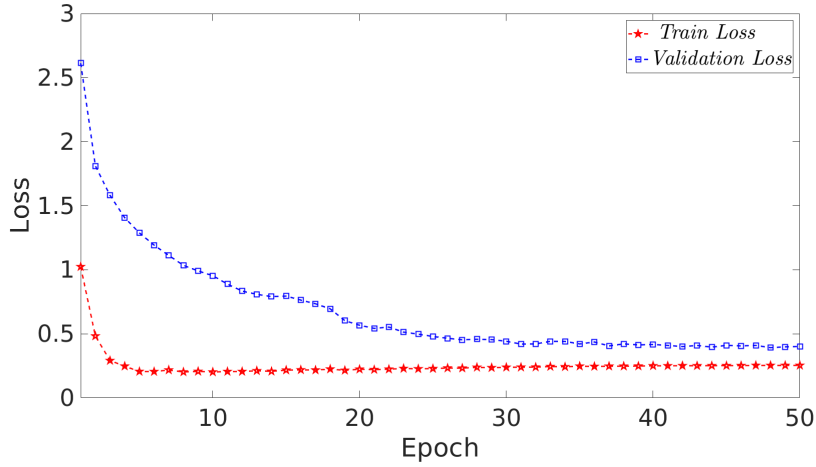
Figure 3.7: The training and validation errors of the prediction block used in the proposed method.

all the 4 base-learners which have learned different information and features from the same training data.

3.3 Experimental Results and Discussions

In this section, we explain the set of experiments conducted to evaluate our method and also validated its overall robustness and adaptability. The proposed algorithm was tested on the publicly available Duke DME OCT Dataset (Chiu *et al.*, 2015). The performance of the proposed method is evaluated (on the test data) both qualitatively and quantitatively. We have used the standard quality metrics such as Precision (Powers, 2011), Recall (Powers, 2011) and Dice coefficient (DC) (Dice, 1945) for quantitative analysis. To justify the results, we have also presented Grad-CAM visualizations of the model gradient maps.

To fix the number of base learners, we tried using different combinations of base learners, and observed that 4 models produced the optimal results. The performance using different number of base learners is summarised in Table 3.3. The individual performance of all the 4 base learners is given Table 3.4.

Table 3.3: The performance comparison using different number of base learners (best results are highlighted using boldface).

No.of Base Models	Layers							
	Background	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE
2	0.99	0.89	0.94	0.88	0.87	0.96	0.94	0.90
3	0.99	0.89	0.94	0.89	0.88	0.96	0.95	0.90
4	**0.99**	**0.91**	**0.96**	**0.92**	**0.90**	**0.97**	**0.95**	**0.91**
5	0.99	0.91	0.96	0.92	0.90	0.97	0.95	0.90
6	0.99	0.91	0.95	0.92	0.90	0.96	0.95	0.90

Table 3.4: The individual performance of the base learners on test data (best results are highlighted using boldface).

Model No.	Layers							
	Background	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE
M1	0.98	0.89	0.92	**0.89**	0.85	**0.96**	0.93	0.89
M2	**0.99**	**0.90**	**0.95**	0.87	0.84	0.93	0.93	0.88
M3	**0.99**	0.88	0.94	0.87	**0.89**	0.94	**0.95**	0.90
M4	0.98	0.87	0.92	0.88	0.88	0.94	0.94	**0.91**

OCT scans carry varying degrees of noise (Anoop *et al.*, 2019; Sudeep *et al.*, 2016)). The presence of speckle noise in OCT degrades the quality of the acquired images and may produce poor segmentation results. To see how the model performs on noisy and denoised data, we trained and tested the model on both noisy and denoised data. For denoising, we have used the model developed in Chapter 2. In the first experiment, we directly fed the noisy images for training and testing and calculated the results and metrics. In our second set of experiment, we fed the denoised images for training and testing. The results are depicted in Table 3.5. It can be observed from the Table 3.5 that by denoising the performance of segmentation can be improved. The experiments that we discussed earlier (i.e. to fix the number of base learners) are also conducted on the denoised images.

To evaluate the segmentation accuracy of the proposed model and other state-of-the-art methods, the performance in terms of mean Dice coefficient is computed. The comparison is mostly restricted to CNN based methods because it has been previously established (Roy *et al.*, 2017) that CNN based methods outperform all the traditional

Table 3.5: Performance of the proposed DelNet model on raw OCT images and denoised OCT images measured in terms of precision, recall and F_beta (best results are highlighted using boldface).

Layers \Metric	Precision		Recall		F_beta	
	Noisy	Denoised	Noisy	Denoised	Noisy	denoised
Background	0.99	**0.99**	0.98	**0.98**	0.99	**0.99**
ILM	0.92	**0.88**	0.82	**0.97**	0.87	**0.92**
NFL-IPL	0.89	**0.96**	0.91	**0.95**	0.91	**0.96**
INL	0.78	**0.95**	0.83	**0.93**	0.81	**0.93**
OPL	0.72	**0.91**	0.80	**0.95**	0.78	**0.91**
ONL-ISM	0.94	**0.96**	0.91	**0.96**	0.93	**0.97**
ISE	0.89	**0.94**	0.91	**0.95**	0.9	**0.96**
OS-RPE	0.81	**0.88**	0.92	**0.95**	0.87	**0.92**

segmentation algorithms. Table 3.6 shows the performance evaluation of the proposed method with two CNN-based techniques named as ReLayNet (Roy *et al.*, 2017), DilatedRelayNet (Reddy *et al.*, 2020) and two traditional approaches such as CM-GDP (Chiu *et al.*, 2010; Roy *et al.*, 2017), CM-KR Chiu *et al.* (2015); Roy *et al.* (2017), in terms of the mean Dice coefficient on expert annotations. It can be observed from the Table 3.6 that the performance of the proposed method is superior to other methods considered. OPL layer was the most challenging retinal layer to segment and with the proposed method we got a mean DC of 0.90 for the segmentation of this layer. We also conducted experiments by dividing the test dataset into normal and cases with pathologies. The results are reported in Table 3.7. It can be seen that the proposed method outperforms other methods in segmenting layers from OCT scans of normal eye and cases with pathologies.

Also, we have computed the width of the confidence interval using mean and standard deviation of the Dice coefficient on different confidence levels (50%, 75%, 90%, 95%). The obtained results are compared with the ReLayNet (Roy *et al.*, 2017) and DilatedReLayNet (Reddy *et al.*, 2020) models (Table 3.7 and Table 3.8). We observed that the DelNet showed a lower standard deviation and more consistent predictions. Table 3.8 summarises the results of the width of the interval (on either side of the mean) at different confidence levels for each layer.

Qualitative analysis of the proposed method with other methods considered is shown in Figure 3.8. An OCT image of the normal retina is shown in Figure 3.8 (a). The cor-

Table 3.6: The performance evaluation of DelNet with respect to the ReLayNet, DilatedReLayNet models and the conventional approaches in terms of the mean Dice coefficient on expert 2 annotations (best results are highlighted using boldface).

Methods	Layers							
	Background	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE
CM-GDP (Chiu *et al.*, 2010; Roy *et al.*, 2017)	NA	0.77	0.77	0.65	0.67	0.86	0.87	0.82
CM-KR (Chiu *et al.*, 2015; Roy *et al.*, 2017)	NA	0.85	0.89	0.75	0.74	0.93	0.87	0.82
ReLayNet(Roy *et al.*, 2017)	0.99	0.90	0.94	0.88	0.85	0.93	0.92	0.90
DilatedReLayNet (Reddy *et al.*, 2020)	0.99	0.89	0.94	0.89	0.88	0.96	0.94	0.90
DelNet MC	0.99	0.91	0.95	0.91	0.90	0.96	0.95	0.90
DelNet	**0.99**	**0.92**	**0.96**	**0.93**	**0.91**	**0.97**	**0.96**	**0.92**

responding labelled image from the Duke data set is shown in Fig. 3.8 (b). Fig. 3.8 (c), Fig. 3.8 (d), Fig. 3.8 (e), Fig. 3.8 (f) show the prediction map of RelayNet, DilatedReLayNet, DelNet with MC and the proposed DelNet architectures respectively. It can be observed from the images that the proposed method outperforms other methods. Comparing Fig. 3.8 (c), Fig. 3.8 (d), Fig. 3.8 (e), Fig. 3.8 (f) with respect to the ground truth (Fig. 3.8 (b)), it is evident that there are misclassifications in the outer layer boundaries of Fig. 3.8 (c) and Fig. 3.8 (d). The thickness of the layers like ILM, NFL-IPL plays a vital role in diagnosing diseases like glaucoma. Our proposed approach gave better predictions and it is evident from Fig. 3.8 (e) and Fig. 3.8 (f).
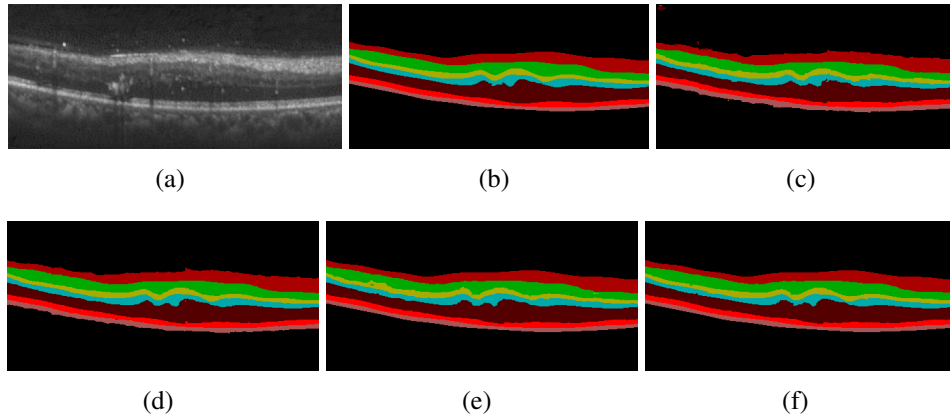
Figure 3.8: Qualitative comparison of the proposed DelNet model with existing CNN based methods. (a) An OCT image of the normal retina, (b) the corresponding labeled image of expert 2 from Duke data set, (c) the predicted results of the ReLayNet model, (d) the result of the DilatedReLayNet model, (e) the output of the DelNet model with manual combining, (f) the output of the proposed DelNet model.

Fig. 3.9 shows the edge map of retinal layers generated with different methods. The

Table 3.7: The performance evaluation on normal retinal scans and retinal scans with the pathology of DUKE DME dataset in terms of the mean (standard deviation) Dice coefficient on expert 2 annotations (best results are highlighted using boldface).

Methods	Observations	Layers							
		Background	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE
ReLayNet (Roy et al., 2017)	Normal Scans	0.99 (0.0451)	0.90 (0.9723)	0.93 (0.7725)	0.88 (1.6838)	0.88 (1.0462)	0.95 (0.4710)	0.93 (0.7325)	0.89 (0.6269)
	Scans with Pathology	0.99 (0.0882)	0.88 (1.3514)	0.94 (1.0827)	0.89 (1.8493)	0.85 (1.4517)	0.96 (0.6470)	0.93 (1.0386)	0.89 (0.7052)
	Overall	0.99 (0.0589)	0.90 (1.3224)	0.94 (0.9111)	0.88 (1.8494)	0.85 (1.3265)	0.93 (0.5305)	0.92 (1.0525)	0.90 (0.6371)
DilatedReLayNet (Reddy et al., 2020)	Normal Scans	0.99 (0.0402)	0.90 (0.9713)	0.94 (0.3305)	0.88 (0.8819)	0.88 (1.0432)	0.95 (0.3425)	0.93 (0.5116)	0.89 (0.6924)
	Scans with Pathology	0.99 (0.0980)	0.88 (1.3514)	0.94 (0.5052)	0.89 (0.9790)	0.85 (1.4317)	0.96 (0.3861)	0.93 (0.7590)	0.89 (0.7836)
	Overall	0.99 (0.0766)	0.89 (1.3124)	0.94 (0.4344)	0.89 (0.9791)	0.88 (1.3255)	0.96 (0.4451)	0.94 (0.7033)	0.90 (0.7524)
DelNet	Normal Scans	**0.99 (0.0251)**	**0.93 (0.4047)**	**0.96 (0.2575)**	**0.95 (0.4955)**	**0.93 (0.5226)**	**0.97 (0.1385)**	**0.97 (0.2508)**	**0.93 (0.3477)**
	Scans with Pathology	**0.99 (0.0490)**	**0.89 (0.5631)**	**0.95 (0.3709)**	**0.89 (0.5539)**	**0.89 (0.7168)**	**0.96 (0.1608)**	**0.95 (0.3795)**	**0.91 (0.3717)**
	Overall	**0.99 (0.0384)**	**0.92 (0.5448)**	**0.96 (0.3133)**	**0.93 (0.5739)**	**0.91 (0.6727)**	**0.97 (0.1854)**	**0.96 (0.3412)**	**0.92 (0.4772)**

Table 3.8: The performance evaluation on the retinal scans of DUKE DME dataset in terms of the width of interval (on either side mean Dice) for different confidence levels on expert 2 annotations.

Methods	Confidence Level	Width of the interval on either side of mean Dice of retinal layers							
		Background	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE
ReLayNet (Roy et al., 2017)	50 %	0.0044	0.0846	0.0600	0.1193	0.0855	0.0406	0.0679	0.0436
	75 %	0.0076	0.1447	0.1026	0.2039	0.1461	0.0695	0.1161	0.0746
	90 %	0.0109	0.2075	0.1472	0.2925	0.2096	0.0997	0.1666	0.1071
	95 %	0.0130	0.2480	0.1759	0.3495	0.2505	0.1191	0.1990	0.1279
DilatedReLayNet (Reddy et al., 2020)	50 %	0.0049	0.0846	0.0280	0.0631	0.0855	0.0287	0.0453	0.0485
	75 %	0.0084	0.1447	0.0479	0.1079	0.1461	0.0490	0.0774	0.0829
	90 %	0.0121	0.2075	0.0687	0.1548	0.2096	0.0704	0.1110	0.1190
	95 %	0.0144	0.2480	0.0821	0.1850	0.2505	0.0841	0.1327	0.1421
DelNet	50 %	0.0024	0.0352	0.0200	0.0350	0.0427	0.0119	0.0226	0.0242
	75 %	0.0042	0.0603	0.0342	0.0599	0.0730	0.0204	0.0387	0.0414
	90 %	0.0060	0.0864	0.0490	0.0860	0.1048	0.0293	0.0555	0.05595
	95 %	0.0072	0.1033	0.0586	0.1027	0.1252	0.0350	0.0663	0.0710
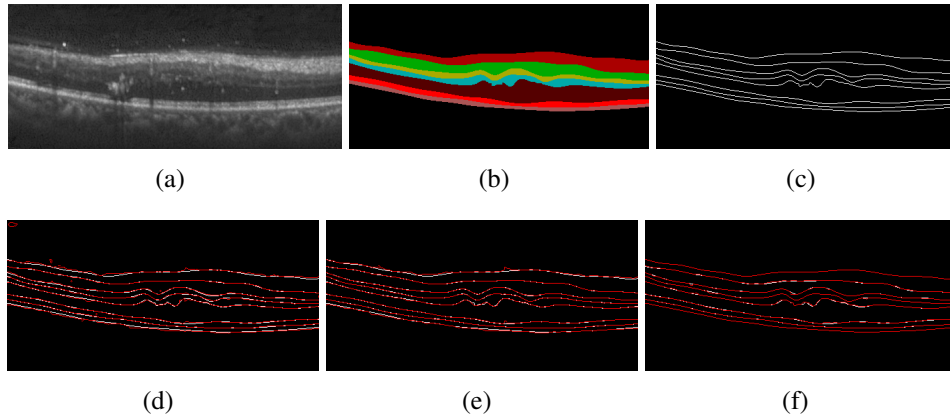
Figure 3.9: Qualitative comparison of the proposed DelNet model with existing CNN based methods. (a) An OCT image of the normal retina. (b) The corresponding labeled image of expert 2 from Duke data set. (c) Retinal layer boundaries of ground truth. (d) Retinal layer boundaries of ReLayNet predictions. (e) Retinal layer boundaries of DilatedReLayNet Prediction. (f) Retinal layer boundaries of DelNet Prediction.

edge maps clearly visualize the accuracy of correct segmentation with respect to the ground truth. Here, the edges of the ground truth are provided in white colour, corresponding predictions are depicted in red colour. Fig. 3.9 (d) and Fig. 3.9 (e) shows the predictions of ReLayNet and DilatedReLayNet, it is clear that there are misclassifications in the retinal boundaries. But the predictions of the proposed method shown in Fig. 3.9 (f) is very close to the ground truth. From the superimposed images (Fig. 3.9 (d), (e) and (f)), it can be seen that the layer boundary of the image generated using the proposed method fits better with the ground truth. Fig. 3.10 shows how the proposed model performs on images with pathology. Fig. 3.10 (a), Fig. 3.10 (b) and Fig. 3.10 (c) shows the OCT image of a normal retina, retina with pathology, image with fovea region respectively. The corresponding ground truth and predictions are shown in the second column and third column of Fig. 3.10. This experiment additionally shows the robustness of the proposed method in segmenting retinal layers from images with pathology.

Grad-CAM (Selvaraju *et al.*, 2017) help us to understand what the model is actually learning. Grad-CAM is a class-discriminative localization technique and unlike CAM,
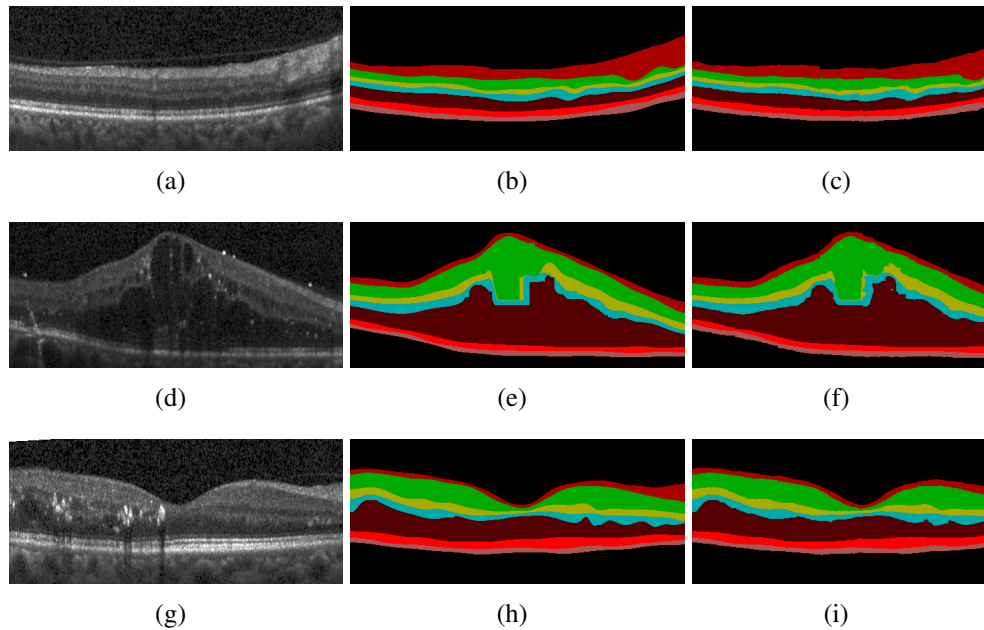
Figure 3.10: Qualitative Comparison of the proposed DelNet model with three sets of images. Images (a), (d) and (g) are OCT image of the normal retina, the retina having pathology and the retina with fovea region respectively. Images (b), (e) and (h) shows the corresponding ground truth images. The corresponding predictions are depicted in images (c), (f) and (i).

it uses the gradient information flowing into the last convolutional layer of the CNN to understand each neuron for a decision of interest. As an additional demonstration of the superiority of the proposed method, GradCAM visualization of the filter gradient for all the eight retinal layers from the proposed DelNet architecture compared with DilatedReLayNet model is depicted in Fig. 3.11. Here, Fig. 3.11 (R1), (a) represents the input image, (b) ground truth, (c) Prediction map from DilatedReLayNet, and (d-k) GradCAM visualization of filter gradient for all the eight retinal layers from the predictor block of DilatedReLayNet Model. Also, in Fig. 3.11 (R2), (a) indicates the input image, (b) ground truth, (c) Prediction map from the proposed model, and (d-k) GradCAM visualization of filter gradient for all the eight retinal layers of the proposed Model. Also, (d-k) shows a heat map where the red indicates higher value and blue indicates lower values. Comparing the Grad-CAM visualization of DilatedReLayNet model depicted in Fig. 3.11 (R1) with the Grad-CAM visualization of the
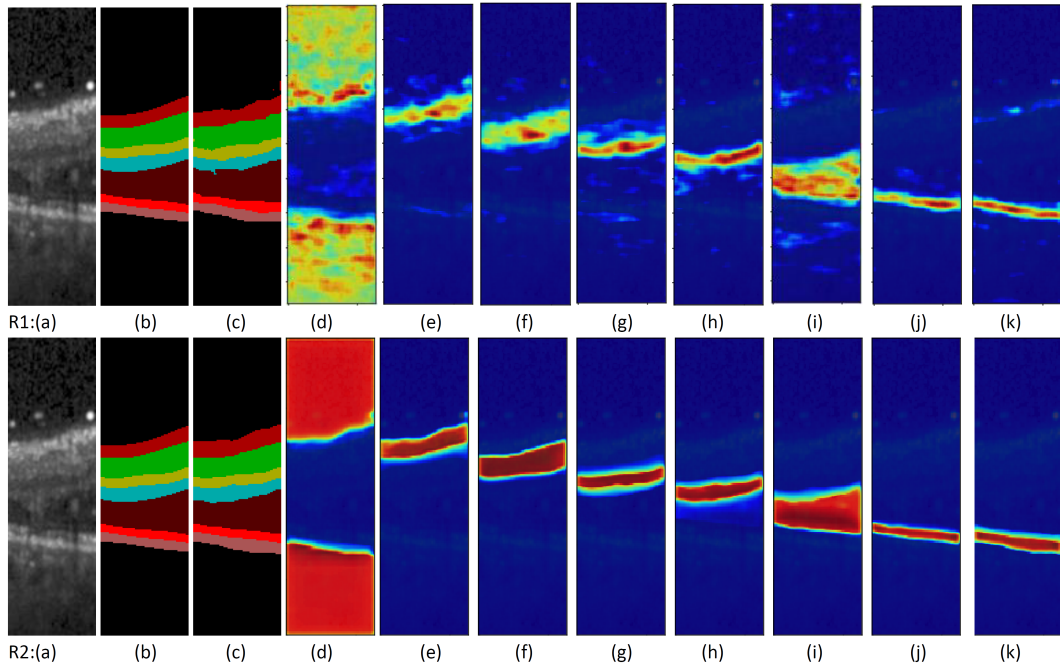
Figure 3.11: GradCAM visualizations on retinal layers of proposed DelNet architecture compared with DilatedReLayNet model. R1: (a) Input, (b) ground truth, (c) Prediction map from DilatedReLayNet and (d-k) GradCAM visualization of filter gradient for all the eight retinal layers from the predictor block of DilatedReLayNet Model. R2: (a) Input, (b) ground truth, (c) Prediction map from the proposed model and (d-k) GradCAM visualization of filter gradient for all the eight retinal layers of the proposed Model. Also, (d-k) shows a heat map where the red indicates higher value and blue indicates lower values.

DelNet (ensemble) model given in Fig. 3.11 (R2). This visualization clearly shows that the proposed DelNet model learns the layers better than the single model. The better performance of the proposed method is because of the ensemble approach. Different base models are trained and fine tuned to learn certain layers properly. As a result this approach performs better than the stand-alone single model.

The time complexity analysis of the proposed and other methods considered are shown in Table 3.9. It has been reported in the literature that graph-based methods are faster compared to other conventional retinal layer segmentation algorithms (Kafieh et al., 2013). Also from our analysis, one of the graph-based approaches consider in this comparison (Chiu et al., 2010) taking 9.7 seconds per image for yielding the result. Hence, from Table 3.9 it is clear that the CNN-based methods are faster compared

Table 3.9: Time Complexity analysis.

Method	Total train time (seconds)	Avg. prediction time per image (seconds)
CM-GDP (Chiu *et al.*, 2010)	-	9.74s
ReLayNet (Roy *et al.* (2017))	1375s	0.032s
DilatedReLayNet (Reddy *et al.*, 2020)	1441s	0.030s
M1	1596s	0.032s
M2	1501s	0.030s
M3	1472s	0.030s
M4	1527s	0.031s
DelNet (proposed method)	7082s	0.1524s

to the conventional approaches. As expected, it can be observed from the table that the ensemble method takes more time (almost 5 times) to train and test compared to other CNN-based methods. The memory requirements also will be more for ensemble approaches. This is one drawback of the proposed approach.

3.3.1 DelNet Manual Combining Algorithm (DelNet MC)

In this experiment, we have tried a traditional ensemble approach which is established in machine learning algorithms. Here, we replaced the predictor block with a layer-by-layer manual combing algorithm, which used a weighted ensemble for each layer (weighted mean of the base-learner outputs).

The developed step by step procedure using weighted ensemble for an optimal 8 class segmentation from the 4 model outputs obtained from the models M_i ($1 \leq i \leq 4$) is described below:

Let P^m be the output prediction of model M_m ($1 \leq m \leq 4$). Clearly P^m has dimensions $216 \times 64 \times 8$, we see that $P^m_{i,j}$ gives the 8 class softmax score vector for the pixel i, j ($1 \leq i \leq 216$, $1 \leq j \leq 64$).

The procedure follows a general algorithm to obtain segmentation mask for each layer l ($1 \leq l \leq 8$) one by one in the following order : OPL, INL, ILM, OS-RPE, ISE, NFL-IPL, ONL-ISM and finally the Background. We first define a quantity G as the

weighted mean of P:

$$G = \frac{\sum_{m=1}^{4} A_{l,m} P^m}{\sum_{m=1}^{4} A_{l,m}} \tag{3.2}$$

where the weights tensor A is a scalar, it's also a hyper-parameter which is intuitively chosen for an optimal performance. The values of $A_{l,m}$ for $1 \leq l \leq 8$ and $1 \leq m \leq 4$ are given in Table 3.10.

We see that G has dimensions $216 \times 64 \times 8$, and $G_{i,j}^l$ gives the softmax score/probability for the pixel i, j ($1 \leq i \leq 216$, $1 \leq j \leq 64$) belonging to layer l ($1 \leq l \leq 8$). Now for each layer $1 \leq l \leq 8$, and each previously unlabelled pixel i, j ($1 \leq i \leq 216$, $1 \leq j \leq 64$), we label pixel i, j as layer l only if it satisfies the following condition:

$$G_{i,j}^l > \beta_l$$

Here, β_l ($1 \leq l \leq 8$) are scalars (and hyper-parameters) chosen carefully to produce optimal results. The optimal value of β is given in Table 3.11.

For the remaining unlabelled pixels, we first define G as the arithmetic mean of P:

$$G = \frac{\sum_{m=1}^{4} P^m}{4} \tag{3.3}$$

and then label each unlabelled pixel to the class corresponding to the maximum softmax score for that pixel in G^u (the normal way of classification). Even though the manual combining algorithm is an option for combining the base learners, our experiments shows that the CNN based predictor block is superior in terms of performance. This is

Table 3.10: The values of $A_{l,m}$ used in each layer and models.

Layer (l)	OPL	INL	ILM	OS-RPE	ISE	NFL-IPL	ONL-ISM	Background
$A_{l,1}$	0.08	0.10	0.10	0.80	0.80	0.02	0.02	0.02
$A_{l,2}$	0.02	0.02	0.02	0.02	0.02	0.80	0.80	0.80
$A_{l,3}$	0.10	0.80	0.80	0.10	0.10	0.10	0.10	0.10
$A_{l,4}$	0.80	0.08	0.08	0.08	0.08	0.08	0.08	0.08

Table 3.11: The values of Beta (β_l) used in each layer for optimal result.

Layer	OPL	INL	ILM	OS-RPE	ISE	NFL-IPL	ONL-ISM	Background
Beta (β_l)	0.75	0.75	0.80	0.85	0.85	0.90	0.90	0.95

evident from Fig. 3.8.

Algorithm 1: DelNet Manual Combining Algorithm (DelNet MC).

Result: Gives the class label for each pixel i, j in the image.

for $1 \leq i \leq 216$ *and* $1 \leq j \leq 64$ **do**
   | $label_{i,j} = 0$;
end

for $1 \leq l \leq 8$ **do**

$$G = \sum_{m=1}^{4} A_{l,m} P^m;$$

   for $1 \leq i \leq 216$ *and* $1 \leq j \leq 64$ **do**
      if $label_{i,j} = 0$ *and* $G^l_{i,j} \geq \beta_l$ **then**
         | $label_{i,j} = l$;
      end
   end

end

$$G = 0.25 \times \sum_{m=1}^{4} P^m;$$

for $1 \leq i \leq 216$ *and* $1 \leq j \leq 64$ **do**

   if $label_{i,j} = 0$ **then**
      | $label_{i,j} = \underset{1 \leq l \leq 8}{\arg\max}\{G_{i,j}\}$;
   end

end

3.4 Summary

This chapter proposes an automated method for segmenting retinal layers from OCT images using a deep ensemble learning based technique. To create the ensemble based architecture, we used 4 base models which follows DilatedReLayNet architecture and a predictor block. Experimental results on a standard benchmark dataset shows that the proposed architecture improved the segmentation accuracy when compared to the stand-alone single DilatedReLayNet model and other state-of-the-art methods.

To demonstrate the robustness of the proposed method we conducted experiments on retinal scans with pathology and compared it with the performance of the model on normal eye scans. The results show that there is not much difference in the performance of the model in both cases. Also, the statistical hypothesis testing showed lower standard deviation and more consistent predictions for the proposed model. The Grad-CAM visualization of the proposed model also shows that the model is learning different layers much better than the single DilatedReLayNet model.

Since the proposed model follows an ensemble approach, the training and testing time is more when compared to the stand-alone single model. Also the memory requirements for training the model is high. This is a drawback of the proposed model. However, the test time is still less than 0.25 seconds (on our experimental settings).

CHAPTER 4

ATTENTION ASSISTED PATCH-WISE CNN FOR THE SEGMENTATION OF FLUIDS FROM THE RETINAL OPTICAL COHERENCE TOMOGRAPHY IMAGES

An attention assisted convolutional neural network-based architecture to detect and quantify three types of retinal cysts namely the intra-retinal cyst, sub-retinal cyst and pigmented epithelial detachment from the OCT images of the human retina is proposed in this chapter [3]. The proposed architecture has an encoder-decoder structure with an attention and a multi-scale module. The qualitative and quantitative performance of the model is evaluated on the publicly available RETOUCH retinal OCT fluid detection challenge data set. The proposed model outperforms the state-of-the-art methods in terms of precision, recall, and dice coefficient. Furthermore, the proposed model is computationally efficient due to its less number of model parameters.

4.1 Introduction

Eyes provide a gateway to the outward world due to their ability to perceive light and process it into visual information. Eyes receive reflected light from the environment which is converted to neural electrochemical signals. These are then transmitted to the brain and processed in the visual cortex. The retina is responsible for the conversion

---

[3]The work described in this Chapter has been published in:**Anoop, B. N.**, Saswat Parida, Ajith B, G N Girish, Abhishek R. Kothari, Muthu Subash Kavitha, and Jeny Rajan, "Attention Assisted Patch-wise CNN for the Segmentation of Fluids from the Retinal Optical Coherence Tomography Images." *International Conference on Pattern Recognition and Machine Intelligence*. Springer, 2021. (Submitted).

of optical signals to electrochemical signals, and its central part, called the macula, is responsible for fine visual discrimination. Due to this vital role, abnormalities in the macula can lead to serious visual impairment.

Macular edema is a painless disorder caused by the collection of fluid from leaking retinal capillaries in the retina (Roy *et al.*, 2017; Anoop *et al.*, 2020). The fluid may be present in the retina (Reddy *et al.*, 2020) (intraretinal cysts, IRC), under the retina (subretinal fluid, SRF) or under the retinal pigment epithelium (Pigment epithelial detachment, PED). This causes loss of vision proportionate to its degree and extent. This condition occurs in several retinal disease such as age-related macular degeneration (AMD), retinal vein occlusion (RVO) and diabetic macular edema (DME). These three conditions affect a large number of people in the middle and older age groups, and together, constitute majority of the cases of vision loss due to retinal disease.

OCT is a quick and non-invasive imaging modality commonly used to investigate and characterize retinal disorders. The quality of OCT output volumes from different instruments varies due to the presence of speckle and other artefacts (Menon *et al.*, 2020*b*; Anoop *et al.*, 2021) (Fig. 4.1). This poses a challenge in identification, interpretation and reproducibility of morphological features of disease, like intraretinal cysts, subretinal fluid, etc. Also, manual segmentation and subsequent quantification of these features on OCT images is tedious and has significant inter-observer variability. It has been a focus of effort for the medical imaging community in recent years to develop automated methods for retinal OCT fluid segmentation and quantification which could augment the clinical characterization and quantification, thereby resulting in better decision making.

Several methods have been proposed in the literature to automate the segmentation of either IRC (Girish *et al.*, 2018*b*, 2019), SRF (Rao *et al.*, 2019) or PED (Wu *et al.*, 2017) individually. However, methods proposed for segmenting all three retinal fluid components simultaneously are rare in the literature. Limitations in the algorithms proposed and unavailability of quality annotated ground truth datasets have been major
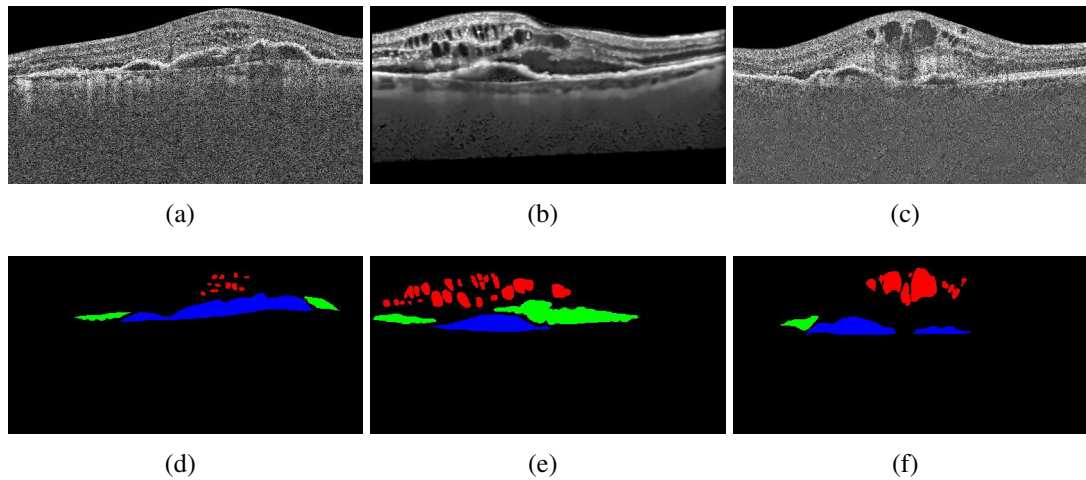
Figure 4.1: The visual quality differences of the OCT scans due to the presence of speckle is depicted here. A retina with macular edema imaged with three different OCT scanners namely: (a) Cirrus, (b) Spectralis, (c) Topcon and (d), (e), and (f) are the annotations of the three fluid types on 2D B-scans: IRF (red), SRF (green), PED (blue).

impediments in the development of such a comprehensive algorithm. The RETOUCH challenge (Bogunović *et al.*, 2019) (April 2017) was aimed at creating a benchmark that could be used to evaluate future algorithms for detection and segmentation of all three fluid components across different retinal disorders and OCT vendors. The algorithm presented by Lu et al. (Lu *et al.*, 2017) performed superior to others in this challenge. They used U-net architecture as the segmentation network with an additional channel to provide relative- information in the form of a distance map and reported better performance for the model. Subsequently, Alsaih et al (Alsaih *et al.*, 2020) proposed an ensemble based deep learning approach for end-to-end segmentation of retinal cysts from OCT images by employing the well-known segmentation architecture, DeepLabv3+. The authors demonstrated the effectiveness of patch-wise training approach for the segmentation of retinal cysts. Ten different models were trained for a specific task and predicted the best results using majority voting. Though the method gave good results, the complexity of the model was its major drawback. In this chapter, we propose an attention assisted CNN model that addresses some of the issues of existing methods. The main contributions in this chapter are as follows-

- Proposed a modified DeepLabv3+ architecture by adjusting the field of view of the convolutional kernels in the initial two stages of the encoder part, which can extract more predominant features.

- Unlike the conventional DeepLabv3+, we have employed a modified atrous spatial pyramid pooling (ASPP) module with weight sharing concept. The dilation rates of the ASPP module is fixed experimentally.

- We have employed a self attention-based mechanism in the skip connections, which helped the model learn more reliable features.

- Employed relative layer information of the retinal layers as an additional feature for the segmentation of retinal cyst.

- The proposed approach uses fewer network parameters compared to the state-of-the-art techniques.

The rest of the chapter is structured as follows, Section 4.2 describes the proposed method which includes the data set, pre-processing, and model architecture. Section 4.3 discusses the results and comparative analysis with existing methods. Finally, conclusions and remarks are drawn in section 4.4.

4.2 Methodology

4.2.1 Dataset and data preprocessing

The RETOUCH retinal OCT fluid challenge (Bogunović *et al.*, 2019) dataset is used to analyze the performance of the proposed method. RETOUCH challenge dataset consists of OCT scans having all the three kinds of retinal cysts such as IRF, SRF and PED acquired from three different OCT machines namely Cirrus, Spectralis, and Topcon. The dataset contains 112 volumes of OCT scans, out of which 70 volumes are kept for training and the remaining 42 volumes are reserved for the testing. Since the test set was not available we split the training set of the RETOUCH dataset into train, validation, and test set. Details of the RETOUCH dataset is given in Table 4.1.

Table 4.1: RETOUCH dataset details.

Vendor (Dimension)	No. of volumes for Training	No. of volumes for Testing
Cirrus (512 x 1024 x 128)	24	14
Spectralis (512 x 496 x 49)	24	14
Topcon (512 x 885 x 128)	22	14
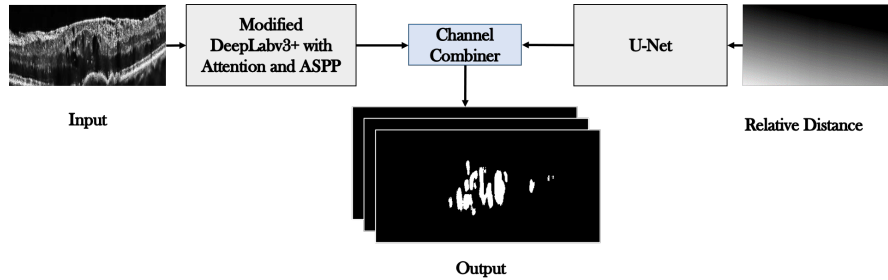
Figure 4.2: Block diagram of the proposed network.

OCT images are generally corrupted with speckle (Anoop *et al.*, 2019). The level and distribution of the noise depends on the equipment used to acquire the images. We have used the model developed in Chapter 2 for denoising the OCT images. Contrast Limited Adaptive Histogram Equalization (CLAHE) (Zuiderveld, 1994) is employed over the denoised image to increase the intensity difference between cystic and non-cystic regions. The contrast enhancement helped the deep learning model to extract more relevant features compared to the raw images.

4.2.2   Network Architecture

The architecture of the proposed method is depicted in Fig. 4.2. The proposed approach mainly consists of an encoder-decoder structure and is improvement over the popular DeepLabv3+ architecture (Fig. 4.3). We added a multi-scale module and an attention module (Fig. 4.4(a)) to the DeepLabv3+ architecture to improve the performance. The multi-scale module extracts features of different scales and the attention module provide attention to relevant features from the encoder side during the decoding stage. The combined effect of the multi-scale module and attention module with relative distance approach(Fig. 4.4(b)) yielded an improved performance.
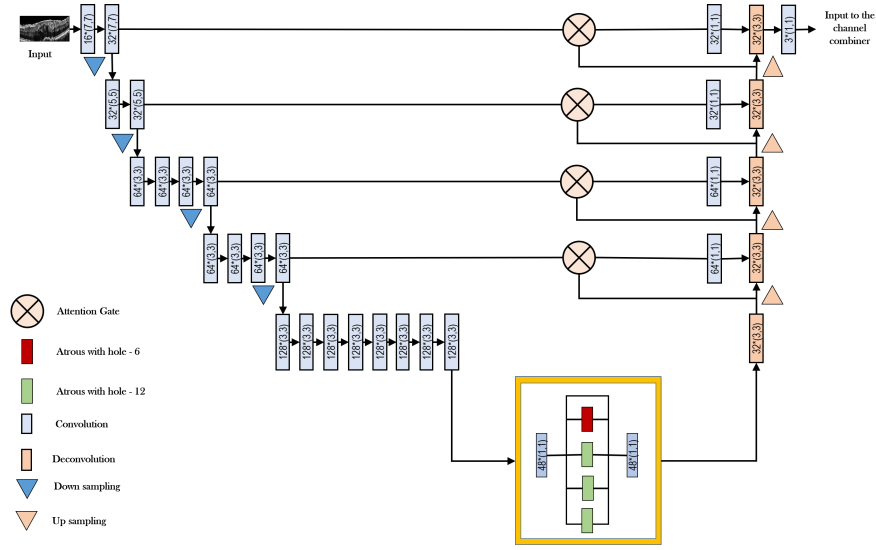
Figure 4.3: A detailed view of the modified DeepLabv3+ architecture with a self-attention mechanism and ASPP module.

The proposed architecture is shown in Fig 4.3. The model is designed to take an input OCT image of size 256×512 and generate a probability map of the same size. The kernel size of the encoder varies from 7×7 to 3×3. The first and second stages of the encoder use 7×7 and 5×5 kernels respectively which can extract more prominent features from the initial stages compared to the kernel of size 3×3. Batch normalization and ReLU activation are used after each convolution. The decoder has a fixed kernel size of 3×3. The encoder and decoder are connected with skip connections that enable to pass the features extracted by the encoder to the decoder in the same stage. The skip connection is passed through the attention module that provides attention to certain feature maps thereby the total performance can be increased. The multi-scale module is attached to extract the multi-scale features from the high dimensional space. It can extract features without the need of deep layer network.

The attention module in the proposed model uses the self-attention (Oktay *et al.*, 2018) approach to extract both spatial and channel information, that gave importance to the most relevant features to do the classification of the target task. A channel-wise $1 \times 1 \times 1$ convolution is carried out on both the inputs followed by a sigmoid activation function and it is used to normalize the attention coefficients. The attention coefficients

74
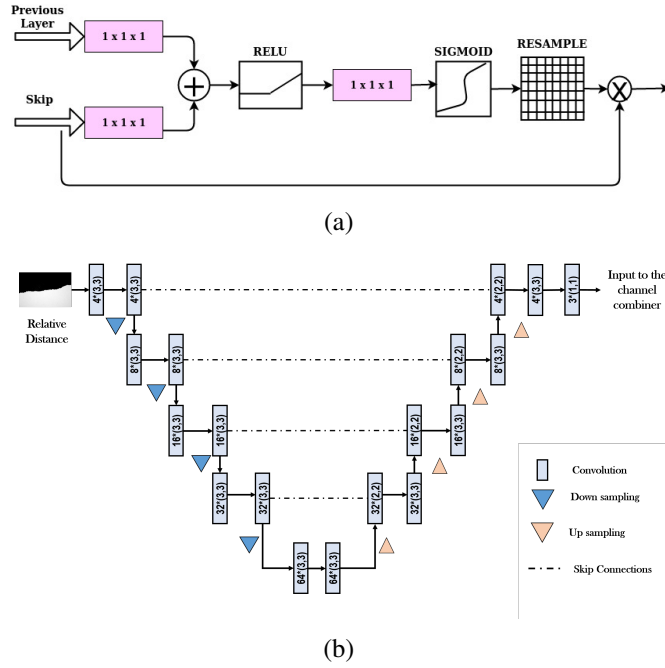
Figure 4.4: (a) Block Diagram of the Attention Module, (b) The proposed network architecture by incorporating relative distance of the data as an additional information to aid the training of the model.

give importance to the relevant features that passes from the encoder to the decoder. Fig. 4.4(a) describes the attention module used in the proposed architecture. The attention module assist the decoder to extract the prominent feature from the symmetric scale encoder.

The multi-scale (ASPP) module is used to extract different scale features from the high dimensional level. In (Chen *et al.*, 2018*b*), the spatial pyramid pooling is utilized to extract multi-scale features from the high dimensional features. A kernel for different scales in the spatial pyramid pooling module to extract certain common features of different scales are proposed in (Huang *et al.*, 2019). In our experiment, we have observed that the network is not picking up a common feature with a common kernel of four different dilation rates in the spatial pyramid pooling module. In the proposed multi-scale module the features are processed from particular dilation rates (dilation rates - 6 and 12) with more convolutional filters.

We also included additional information regarding the location of the cyst to im-

prove the segmentation. To do this, we incorporated relative layer distance information as a feature, which helps in distinguishing the type of cysts based on its location (different types of cysts present between different retinal layers). Relative distance describes the ratio between the pixel distance from the top layer to the distance from the top and bottom layers. If the value is 1, then it indicates the pixel is in the bottom layer and 0 indicates the pixel is in the top layer. Furthermore, the value increases from 0 to 1 as we move from top to bottom. We used the U-Net architecture to extract features from relative distance map which is further used along with the output of the modified DeepLabv3+ model to predict the output.

4.3   Results and Discussion

To avoid any possible bias towards the test set, we did three-fold cross-validation. The mean outcome of all three experiments are shown in Table 4.2 and Table 4.4. Each split consists of separate train validation and test sets. From the total 70 volumes in the RETOUCH training set, 22 volumes are used for testing, 6 volumes are used for validation and the remaining volumes are utilized for training the model. Patch-wise training is adopted for the proposed model. We experimentally fixed the optimal patch size as 128×128 with 0.72 percentage overlap. All model weights were initialized with $'he\_normal'$ initializer and the optimization was carried out using the Adam optimizer with a learning rate of 0.0003. The model is trained for 200 epochs, with a batch sizes of 8, and the objective function used to train the proposed model is the categorical cross-entropy loss. All experiments were performed on an NVIDIA DGX station with 64-bit Ubuntu operating system, NVIDIA Tesla V100 GPU with 32 GB dedicated memory and implemented in Keras with Tensorflow as backend.

The performance of the proposed approach is evaluated both qualitatively and quantitatively. The performance of the proposed architecture is compared with the state-of-the-art methods such as SFU (Lu *et al.*, 2017) and patch-DeepLabv3+ (Alsaih *et al.*,

2020). In Alsaih *et al.* (2020) an ensemble approach is also proposed. Since it is difficult to implement the ensemble approach we were not able to compare with it. The performance comparisons are based on the well-known evaluation metrics such as precision, recall, and mean dice scores. The quantitative analysis of the proposed model on the RETOUCH dataset is given in Table 4.2. It can be observed that the proposed method outperformed other methods in terms of precision, recall, and the mean Dice coefficient. A vendor-wise and cyst-wise analysis of the proposed method compared to the other state-of-the-art methods is listed in Table 4.3. It can be noticed that the proposed method outperforms the other recently proposed methods.

Qualitative comparison of the proposed model over the state-of-the-arts are depicted in Figure 4.5, Figure 4.6, and Figure 4.7. Figure 4.5(a) is one of the input image from cirrus vendor, Figure 4.6(a) is one of the input image from spectralis vendor, and Figure 4.7(a) is one of the input image from topcon vendor, used for testing. Figure 4.5(b), Figure 4.6(b), and Figure. 4.7(b) are the Ground Truth image marked by the expert ophthalmologist. Figure 4.5(c), Figure 4.5(d), and Figure 4.5(e) are the predicted images from the SFU (Lu *et al.*, 2017) model, the patch-DeepLabv3+ (Alsaih *et al.*, 2020) model, and the proposed model respectively on an image from the cirrus vendor. Figure 4.6(c), Figure 4.6(d), and Figure 4.6(e) are the predicted images from the SFU (Lu *et al.*, 2017) model, the patch-DeepLabv3+ (Alsaih *et al.*, 2020) model, and the proposed model respectively on an image from the spectralis vendor. Figure 4.7(c), Figure 4.7(d), and Figure 4.7(e) are the predicted images from the SFU (Lu *et al.*, 2017) model, the patch-DeepLabv3+ (Alsaih *et al.*, 2020) model, and the proposed model respectively on an image from the topcon vendor. From all these images (Figure 4.5, Figure 4.6, and Figure 4.7), it is clear that the predictions of the proposed model shows fewer false positives compared to the other two. These results additionally highlight the efficacy of the proposed method in segmenting retinal cysts from the retinal OCT images.

The comparison of the complexity of the proposed model in terms of the number of

Table 4.2: Quantitative comparison of the proposed model with the SFU and Patch-DeepLabv3+ in terms of complexity (number of parameters), precision, recall and mean dice scores on the RETOUCH dataset.

Method	Total no. of parameters	Vendor	IRF			SRF			PED		
			Precision	Recall	Dice	Pricision	Recall	Dice	Precision	Recall	Dice
SFU (Lu *et al.*, 2017)	18,511,491	Cirrus	74.43	65.67	63.71	75.65	68.35	68.46	59.67	60.75	56.33
		Spectralis	71.23	65.34	67.32	79.45	82.58	79.57	69.17	60.32	62.82
		Topcon	77.92	57.87	64.12	67.43	54.95	59.43	65.21	67.34	58.39
Patch-DeepLabv3+	2,599,363	Cirrus	69.95	56.03	58.38	79.93	58.35	64.49	57.91	48.18	48.68
		Spectralis	68.10	55.31	58.52	83.63	75.45	77.68	68.53	57.40	59.80
		Topcon	75.70	50.10	58.87	64.93	42.92	50.03	71.44	56.36	59.48
Proposed	9,105,391	Cirrus	**74.83**	62.92	**63.89**	**80.20**	64.98	68.53	**63.59**	54.36	55.30
		spectralis	**77.63**	**65.83**	**69.78**	**84.98**	79.96	**81.68**	68.73	**66.65**	**66.92**
		Topcon	76.59	53.80	60.89	67.56	49.89	56.79	**71.93**	66.75	**66.93**

Table 4.3: Quantitative evaluation of the proposed model, SFU and Patch-DeepLabv3+ over the three individual data splits in terms of dice scores.

Model	Split-1 (Dice)			Split-2 (Dice)			Split-3 (Dice)		
	IRF	SRF	PED	IRF	SRF	PED	IRF	SRF	PED
SFU (Lu *et al.*, 2017)	71.58	72.72	63.60	70.40	72.53	49.59	52.32	70.70	61.00
Patch-DeepLabv3+	59.53	54.67	69.53	63.17	71.64	47.95	51.13	66.03	50.20
Proposed	70.58	62.69	74.61	70.12	74.03	53.30	53.95	69.99	61.54

parameters with respect to the SFU and Patch-DeepLabv3+ is presented in Table 4.2. Compared to the SFU (Lu *et al.*, 2017) model the proposed model used almost 50% of fewer parameters. Also, we have conducted an ablation study to evaluate the performance of the proposed system by removing patch-wise training, relative layer information, attention module and multi-resolution module to understand the contribution of each of the components to the proposed architecture. The results of the ablation study is given in Table 4.4.

To obtain the optimal architecture, three sets of experiments have been conducted. The first set of experiments were to fix the depth and width of the model.The second set of experiments was based on the dilation rates of the multi-scale module. We tried with different dilation rates (3, 4, 5, 6,12, 18 and 24) and 6 and 12 gave us maximum performance. The third set of experiments is conducted by varying the filter size. we observed that the usage of filter size 7×7 and 5×5 at the initial stage of encoders helps to increase the performance of the architecture compared to the architecture with a fixed filter size of 3×3.
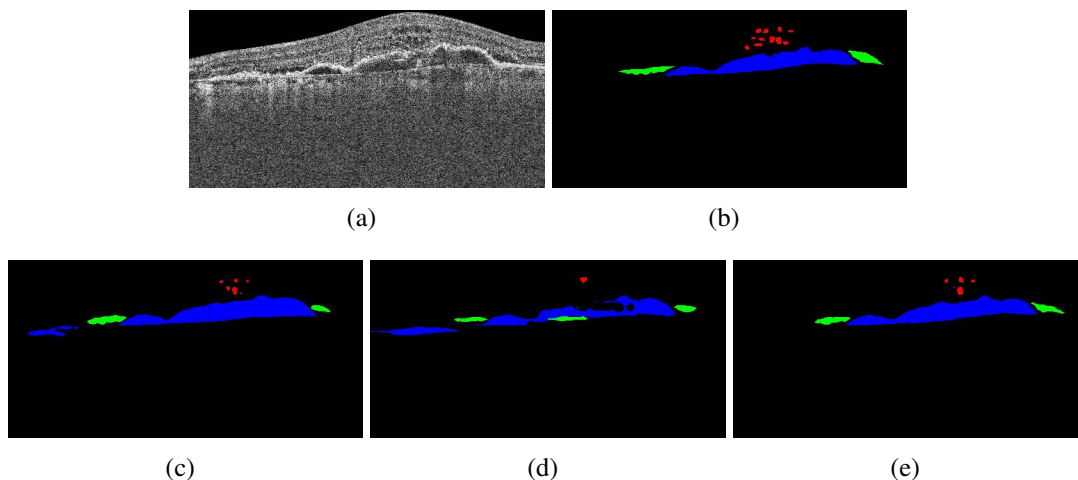
Figure 4.5: Qualitative comparison of the proposed model with existing CNN based methods on the cirrus data of the RETOUCH dataset. (a) shows an OCT image of the cirrus scan with all three kinds of cysts (IRF, SRF, and PED). (b) is the corresponding labeled image of experts from RETOUCH data set, (c) indicate the predicted results of the SFU model, (d) indicate the results of the patch-DeepLabv3+ model, and (e) indicate the prediction of the proposed model.

4.4 Summary

In this chapter, we proposed a CNN based model for segmenting three prominent types of retinal cyst from OCT images. The proposed model uses the attention modules to provide importance to certain features, which improved the performance of the model. The proposed model also comes up with the observations regarding extracting multi-scale features that can improve its performance. The overlapped patch-wise training and the relative layer information helped the model to learn the features of the retinal

Table 4.4: Ablation study.

Method	Vendor	IRF			SRF			PED		
		Precision	Recall	Dice	Pricision	Recall	Dice	Precision	Recall	Dice
DeepLabv3+	Cirrus	65.34	50.35	53.56	68.79	43.50	47.33	58.69	34.36	38.84
	Spectralis	61.52	51.34	53.97	80.78	66.53	69.49	71.93	43.52	49.99
	Topcon	72.94	47.45	54.23	44.24	39.34	36.96	83.34	39.90	50.23
Patch-DeepLabv3+	Cirrus	69.95	56.03	58.38	79.93	58.35	64.49	57.91	48.18	48.68
	Spectralis	68.10	55.31	58.52	83.63	75.45	77.68	68.53	57.40	59.80
	Topcon	75.70	50.10	58.87	64.93	42.92	50.03	71.44	56.36	59.48
Proposed	Cirrus	**74.37**	**62.85**	**63.75**	**79.99**	**64.90**	**68.23**	**63.07**	**54.06**	**54.97**
	spectralis	**77.03**	**65.12**	**69.36**	**84.44**	**79.76**	**81.16**	68.50	**66.31**	**66.09**
	Topcon	**76.29**	**53.30**	**60.69**	**67.16**	**49.71**	**56.39**	71.64	**66.49**	**66.68**
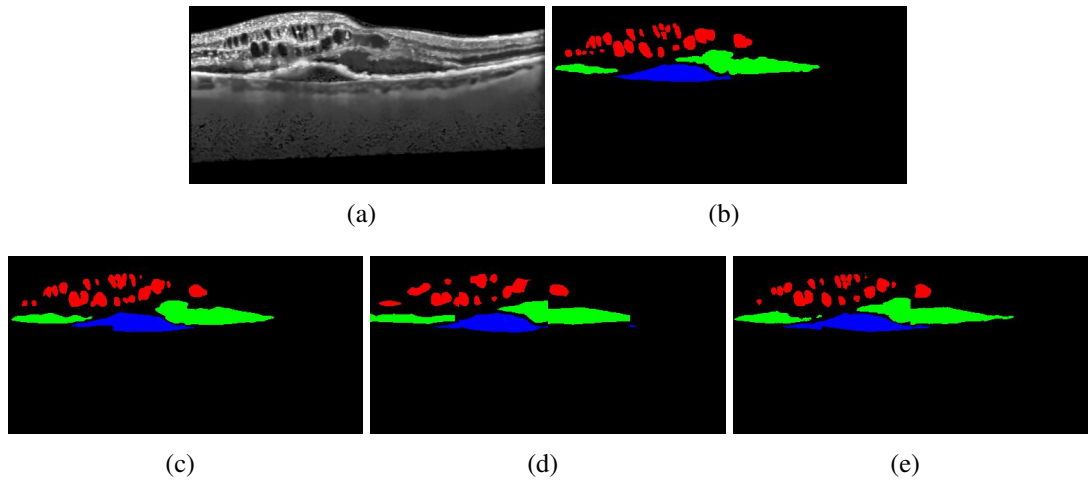
Figure 4.6: Qualitative comparison of the proposed model with existing CNN based methods on the spectralis data of the RETOUCH dataset. (a) shows an OCT image of the spectralis scan with all three kinds of cysts (IRF, SRF, and PED). (b) is the corresponding labeled image of experts from RETOUCH data set, (c) indicate the predicted results of the SFU model, (d) indicate the results of the patch-DeepLabv3+ model, and (e) indicate the prediction of the proposed model.

cysts much better. The less number of false positives prediction of our model shows its higher ability in predicting retinal cysts compared to the other competitive methods used in this study.

(a)   (b)

(c)   (d)   (e)
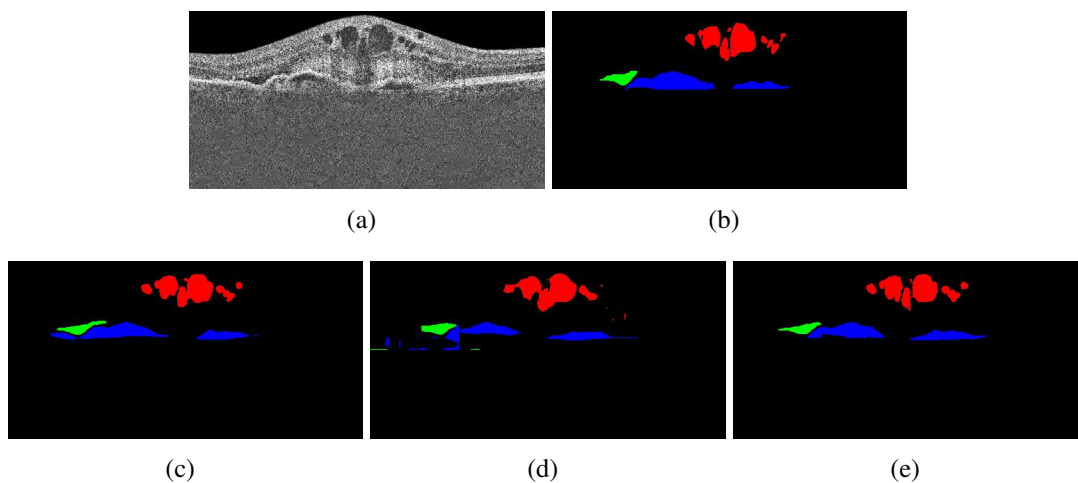
Figure 4.7: Qualitative comparison of the proposed model with existing CNN based methods on the topcon data of the RETOUCH dataset. (a) shows an OCT image of the topcon scan with all three kinds of cysts (IRF, SRF, and PED). (b) is the corresponding labeled image of experts from RETOUCH data set, (c) indicate the predicted results of the SFU model, (d) indicate the results of the patch-DeepLabv3+ model, and (e) indicate the prediction of the proposed model.

CHAPTER 5

CONCLUSIONS AND FUTURE SCOPE

This thesis discusses the steps and challenges for developing automated methods based on deep learning techniques for the analysis of OCT images such as despeckling, retinal layer segmentation and retinal cyst segmentation. OCT is an imaging technique mainly used to analyse the abnormalities present in the human retina. The presence of speckle in the OCT images hampers its visual quality and leads to an inaccurate clinical diagnosis. The initial part of this thesis studied the statistical characteristics of the noise present in the OCT images and proposed a method to improve the quality of the OCT images by eliminating the speckle. The proposed model identifies the noise type in an input image and classifies it into one of the four different noise types. It subsequently denoises the image effectively using a GCDS array that has models catering to specific noise types. Another unique feature proposed in our work is the use of the synthetic eye to create datasets for simulations and experiments. This dataset was useful in doing the quantitative evaluation of different OCT denoising methods. Another vital contribution in this chapter was the patch-wise training approach, which helped to attain an improved performance of the existing GCDS model. This methodology prevents the model from learning the structural details and also provides a way of dealing with limited ground truth data. The experimental results on synthetic and real OCT image datasets showed that the proposed method performed better than other reported methods.

Apart from this, an automated retinal layer segmentation algorithm based on ensemble learning, proposed for the selective segmentation of retinal layers. To create the ensemble-based architecture, we used 4 base models which follows DilatedReLayNet architecture and a predictor block. To demonstrate the robustness of the proposed method we conducted experiments on retinal scans with pathology and compared it

with the performance of the model on normal eye scans. The results show that there is not much difference in the performance of the model in both cases. Also, the statistical hypothesis testing showed lower standard deviation and more consistent predictions for the proposed model. The Grad-CAM visualization of the proposed model also shows that the model is learning different layers much better than the single DilatedReLayNet model.

Also, we proposed a CNN based model for segmenting three types of the retinal cyst from OCT images. The proposed algorithms can help clinicians with the early detection and quantification of abnormalities in the human retina. The proposed model uses the attention modules to provide importance to certain features, which improved the performance of the model. The proposed model also comes up with observations regarding extracting multiscale features that can improve its performance. The overlapped patch-wise training and the relative layer information helped the model to learn the features of the retinal cysts much better. The less number of false positives prediction of our model shows its higher ability in predicting retinal cysts compared to the other competitive methods used in this study.

The CNN despeckling methods proposed in this thesis are based on the noise types. We trained separate models for different kinds of noises. One drawback of this approach is the introduction of bias in the denoised images, when the noise doesnt belong to the four categories that we considered. One solution to this problem is to develop a noise independent denoising model that can take care of any type of noise. Future works should focus on developing such models. The retinal layer segmentation and cyst segmentation techniques proposed in this thesis can segment the retinal layers and cysts in OCT scan up to their true boundary. However, proposed CNN methods may fail to delineate the cysts accurately when the quality of the scans is poor. Hence, future work may be directed towards developing end to end models that incorporate the noise information and edge enhancement.

REFERENCES

Abràmoff, M. D., **M. K. Garvin**, and **M. Sonka** (2010). Retinal imaging and image analysis. *IEEE reviews in biomedical engineering*, **3**, 169–208. 44

Adabi, S., **S. Conforto**, **A. Clayton**, **A. G. Podoleanu**, **A. Hojjat**, and **M. R. Avanaki**, An intelligent speckle reduction algorithm for optical coherence tomography images. *In 2016 4th International Conference on Photonics, Optics and Laser Technology (PHO-TOPTICS)*. IEEE, 2016. 17

Alsaih, K., **M. Yusoff**, **T. Tang**, **I. Faye**, and **F. MÚriaudeau** (2020). Deep learning architectures analysis for age-related macular degeneration segmentation on optical coherence tomography scans. *Computer Methods and Programs in Biomedicine*, 105566. 71, 76, 77

Amini, Z. and **H. Rabbani** (2017). Optical coherence tomography image denoising using gaussianization transform. *Journal of Biomedical Optics*, **22**(8), 086011. 17

Anoop, B., **G. Girish**, **P. Sudeep**, and **J. Rajan**, Despeckling algorithms for optical coherence tomography images: A review. *In Advanced Classification Techniques for Healthcare Analysis*. IGI Global, 2019, 286–310. 17, 42, 57, 73

Anoop, B., **K. S. Kalmady**, **A. Udathu**, **V. Siddharth**, **G. Girish**, **A. R. Kothari**, and **J. Rajan** (2021). A cascaded convolutional neural network architecture for despeckling oct images. *Biomedical Signal Processing and Control*, **66**, 102463. 70

Anoop, B., **R. Pavan**, **G. Girish**, **A. R. Kothari**, and **J. Rajan** (2020). Stack generalized deep ensemble learning for retinal layer segmentation in optical coherence tomography images. *Biocybernetics and Biomedical Engineering*, **40**(4), 1343–1358. 33, 70

Antony, B., **M. D. Abramoff**, **L. Tang**, **W. D. Ramdas**, **J. R. Vingerling**, **N. M. Jansonius**, **K. Lee**, **Y. H. Kwon**, **M. Sonka**, and **M. K. Garvin** (2011). Automated 3-d method for the correction of axial artifacts in spectral-domain optical coherence tomography images. *Biomedical optics express*, **2**(8), 2403–2416. 44

Aum, J., **J.-h. Kim**, and **J. Jeong** (2015). Effective speckle noise suppression in optical coherence tomography images using nonlocal means denoising filter with double gaussian anisotropic kernels. *Applied Optics*, **54**(13), D43–D50. 16

Bashkansky, M. and **J. Reintjes** (2000). Statistics and reduction of speckle in optical coherence tomography. *Optics Letters*, **25**(8), 545–547. 16, 22

Bogunović, H., **F. Venhuizen**, **S. Klimscha**, **S. Apostolopoulos**, **A. Bab-Hadiashar**, **U. Bagci**, **M. F. Beg**, **L. Bekalo**, **Q. Chen**, **C. Ciller**, *et al.* (2019). Retouch: The retinal oct fluid detection and segmentation benchmark and challenge. *IEEE transactions on medical imaging*, **38**(8), 1858–1874. 71, 72

Chen, H., **S. Fu**, **H. Wang**, **H. Lv**, and **C. Zhang** (2018*a*). Speckle attenuation by adaptive singular value shrinking with generalized likelihood matching in optical coherence tomography. *Journal of biomedical optics*, **23**(3), 036014. 17

Chen, H., **Y. Zhang**, **M. K. Kalra**, **F. Lin**, **Y. Chen**, **P. Liao**, **J. Zhou**, and **G. Wang** (2017). Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, **36**(12), 2524–2535. 18

Chen, L.-C., **Y. Zhu**, **G. Papandreou**, **F. Schroff**, and **H. Adam**, Encoder-decoder with atrous separable convolution for semantic image segmentation. *In Proceedings of the European conference on computer vision (ECCV)*. 2018*b*. 75

Chiu, S. J., **M. J. Allingham**, **P. S. Mettu**, **S. W. Cousins**, **J. A. Izatt**, and **S. Farsiu** (2015). Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomedical optics express*, **6**(4), 1172–1194. xi, 43, 52, 53, 56, 58, 59

Chiu, S. J., **X. T. Li**, **P. Nicholas**, **C. A. Toth**, **J. A. Izatt**, and **S. Farsiu** (2010). Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation. *Optics express*, **18**(18), 19413–19428. 43, 58, 59, 64, 65

Chollet, F. *et al.* (2015). Keras. `https://github.com/fchollet/keras`. 55

Coupé, P., **P. Hellier**, **C. Kervrann**, and **C. Barillot** (2009). Nonlocal means-based speckle filtering for ultrasound images. *IEEE transactions on image processing*, **18**(10), 2221–2229. 30, 31, 32

Cuartas-Vélez, C., **R. Restrepo**, **B. E. Bouma**, and **N. Uribe-Patarroyo** (2018). Volumetric non-local-means based speckle reduction for optical coherence tomography. *Biomedical optics express*, **9**(7), 3354–3372. 16

de Campos, V. S., **K. C. Calaza**, and **D. Adesse** (2020). Implications of torch diseases in retinal developmentspecial focus on congenital toxoplasmosis. *Frontiers in Cellular and Infection Microbiology*, **10**. xiii, 2, 3

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, **26**(3), 297–302. 56

Drexler, W., **U. Morgner**, **R. K. Ghanta**, **F. X. Kärtner**, **J. S. Schuman**, and **J. G. Fujimoto** (2001). Ultrahigh-resolution ophthalmic optical coherence tomography. *Nature medicine*, **7**(4), 502. 42

Duan, J., **W. Lu**, **C. Tench**, **I. Gottlob**, **F. Proudlock**, **N. N. Samani**, and **L. Bai** (2016). Denoising optical coherence tomography using second order total generalized variation decomposition. *Biomedical Signal Processing and Control*, **24**, 120–127. 17, 30, 31, 32, 42

Duan, J., **C. Tench**, **I. Gottlob**, **F. Proudlock**, and **L. Bai** (2017). Automated segmentation of retinal layers from optical coherence tomography images using geodesic distance. *Pattern Recognition*, **72**, 158–175. 33, 43

Duan, J., **W. Xie**, **R. W. Liu**, **C. Tench**, **I. Gottlob**, **F. Proudlock**, and **L. Bai**, Oct segmentation: Integrating open parametric contour model of the retinal layers and shape constraint to the mumford-shah functional. *In International Workshop on Shape in Medical Imaging*. Springer, 2018*a*. 43

Duan, W., **Y. Zheng**, **Y. Ding**, **S. Hou**, **Y. Tang**, **Y. Xu**, **M. Qin**, **J. Wu**, **D. Shen**, and **H. Bi** (2018*b*). A generative model for oct retinal layer segmentation by groupwise curve alignment. *IEEE Access*, **6**, 25130–25141. 43

Dufour, P. A., **L. Ceklic**, **H. Abdillahi**, **S. Schroder**, **S. De Dzanet**, **U. Wolf-Schnurrbusch**, and **J. Kowal** (2013). Graph-based multi-surface segmentation of oct data using trained hard and soft constraints. *IEEE transactions on medical imaging*, **32**(3), 531–543. 43

Ehnes, A., **Y. Wenner**, **C. Friedburg**, **M. N. Preising**, **W. Bowl**, **W. Sekundo**, **E. M. zu Bexten**, **K. Stieger**, and **B. Lorenz** (2014). Optical coherence tomography (oct) device independent intraretinal layer segmentation. *Translational vision science & technology*, **3**(1), 1–1. 43

Fang, L., **D. Cunefare**, **C. Wang**, **R. H. Guymer**, **S. Li**, and **S. Farsiu** (2017). Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search. *Biomedical optics express*, **8**(5), 2732–2744. 44

Fang, L., **S. Li**, **R. P. McNabb**, **Q. Nie**, **A. N. Kuo**, **C. A. Toth**, **J. A. Izatt**, and **S. Farsiu** (2013). Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. *IEEE transactions on medical imaging*, **32**(11), 2034–2049. 20, 21

Fercher, A. F., **W. Drexler**, **C. K. Hitzenberger**, and **T. Lasser** (2003). Optical coherence tomography-principles and applications. *Reports on progress in physics*, **66**(2), 239. xiii, 7

Fernández, D. C., **H. M. Salinas**, and **C. A. Puliafito** (2005). Automated detection of retinal layer structures on optical coherence tomography images. *Optics express*, **13**(25), 10200–10216. 42

Garvin, M. K., **M. D. Abràmoff**, **R. Kardon**, **S. R. Russell**, **X. Wu**, and **M. Sonka** (2008). Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-d graph search. *IEEE transactions on medical imaging*, **27**(10), 1495–1505. 43

Garvin, M. K., **M. D. Abramoff**, **X. Wu**, **S. R. Russell**, **T. L. Burns**, and **M. Sonka** (2009). Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE transactions on medical imaging*, **28**(9), 1436–1447. 43, 44

Ghorbel, I., **F. Rossant**, **I. Bloch**, **S. Tick**, and **M. Paques** (2011). Automated segmentation of macular layers in oct images and quantitative evaluation of performances. *Pattern Recognition*, **44**(8), 1590–1603. 43

Girish, G., **V. Anima**, **A. R. Kothari**, **P. Sudeep**, **S. Roychowdhury**, and **J. Rajan** (2018*a*). A benchmark study of automated intra-retinal cyst segmentation algorithms using optical coherence tomography b-scans. *Computer methods and programs in biomedicine*, **153**, 105–114. 16, 42

Girish, G., **A. R. Kothari**, and **J. Rajan**, Automated segmentation of intra-retinal cysts from optical coherence tomography scans using marker controlled watershed transform. *In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016. 42

Girish, G., **B. Saikumar**, **S. Roychowdhury**, **A. R. Kothari**, and **J. Rajan**, Depthwise separable convolutional neural network model for intra-retinal cyst segmentation. *In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019. 16, 70

Girish, G., **B. Thakur**, **S. R. Chowdhury**, **A. R. Kothari**, and **J. Rajan** (2018*b*). Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model. *IEEE journal of biomedical and health informatics*, **23**(1), 296–304. 16, 42, 70

Gong, G., **H. Zhang**, and **M. Yao** (2015). Speckle noise reduction algorithm with total variation regularization in optical coherence tomography. *Optics express*, **23**(19), 24699–24712. 17

Gopinath, K. and **J. Sivaswamy** (2018). Segmentation of retinal cysts from optical coherence tomography volumes via selective enhancement. *IEEE journal of biomedical and health informatics*, **23**(1), 273–282. 16

Gordon-Lipkin, E., **B. Chodkowski**, **D. Reich**, **S. Smith**, **M. Pulicken**, **L. Balcer**, **E. Frohman**, **G. Cutter**, and **P. Calabresi** (2007). Retinal nerve fiber layer is associated with brain atrophy in multiple sclerosis. *Neurology*, **69**(16), 1603–1609. 42

Hamwood, J., **D. Alonso-Caneiro**, **S. A. Read**, **S. J. Vincent**, and **M. J. Collins** (2018). Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of oct retinal layers. *Biomedical optics express*, **9**(7), 3049–3066. 44

Hee, M. R., **J. A. Izatt**, **E. A. Swanson**, **D. Huang**, **J. S. Schuman**, **C. P. Lin**, **C. A. Puliafito**, and **J. G. Fujimoto** (1995). Optical coherence tomography of the human retina. *Archives of ophthalmology*, **113**(3), 325–332. 48

Huang, D., **E. A. Swanson**, **C. P. Lin**, **J. S. Schuman**, **W. G. Stinson**, **W. Chang**, **M. R. Hee**, **T. Flotte**, **K. Gregory**, **C. A. Puliafito**, *et al.* (1991). Optical coherence tomography. *science*, **254**(5035), 1178–1181. 6

Huang, G., **Z. Liu**, **L. Van Der Maaten**, and **K. Q. Weinberger**, Densely connected convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. 18, 28

Huang, Y., **Q. Wang**, **W. Jia**, and **X. He** (2019). See more than once–kernel-sharing atrous convolution for semantic segmentation. *arXiv preprint arXiv:1908.09443*. 75

Inzelberg, R., **J. A. Ramirez**, **P. Nisipeanu**, and **A. Ophir** (2004). Retinal nerve fiber layer thinning in parkinson disease. *Vision research*, **44**(24), 2793–2797. 42

Jian, Z., **L. Yu**, **B. Rao**, **B. J. Tromberg**, and **Z. Chen** (2010). Three-dimensional speckle suppression in optical coherence tomography based on the curvelet transform. *Optics express*, **18**(2), 1024–1032. 17

Kafieh, R., **H. Rabbani**, **F. Hajizadeh**, **M. D. Abramoff**, and **M. Sonka** (2015*a*). Thickness mapping of eleven retinal layers segmented using the diffusion maps method in normal eyes. *Journal of ophthalmology*, **2015**. 45

Kafieh, R., **H. Rabbani**, and **S. Kermani** (2013). A review of algorithms for segmentation of optical coherence tomography from retina. *Journal of medical signals and sensors*, **3**(1), 45. 64

Kafieh, R., **H. Rabbani**, and **I. Selesnick** (2014). Three dimensional data-driven multi scale atomic representation of optical coherence tomography. *IEEE transactions on medical imaging*, **34**(5), 1042–1062. 17

Kafieh, R., **H. Rabbani**, and **I. Selesnick** (2015*b*). Three dimensional data-driven multi scale atomic representation of optical coherence tomography. *IEEE transactions on medical imaging*, **34**(5), 1042–1062. 30, 31, 32

Kafieh, R., **H. Rabbani**, and **G. Unal** (2019). Bandlets on oriented graphs: Application to medical image enhancement. *IEEE Access*, **7**, 32589–32601. 17

Karamata, B., **K. Hassler**, **M. Laubscher**, and **T. Lasser** (2005). Speckle statistics in optical coherence tomography. *JOSA A*, **22**(4), 593–596. 16, 22

Kiaee, F., **H. Fahimi**, and **H. Rabbani**, Intra-retinal layer segmentation of optical coherence tomography using 3d fully convolutional networks. *In 2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018. 44

Kingma, D. P. and **J. Ba** (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 28

Kirillin, M. Y., **G. Farhat**, **E. A. Sergeeva**, **M. C. Kolios**, and **A. Vitkin** (2014). Speckle statistics in oct images: Monte carlo simulations and experimental studies. *Optics letters*, **39**(12), 3472–3475. 22

Kugelman, J., **D. Alonso-Caneiro**, **S. A. Read**, **S. J. Vincent**, and **M. J. Collins** (2018). Automatic segmentation of oct retinal boundaries using recurrent neural networks and graph search. *Biomedical optics express*, **9**(11), 5759–5777. 44

Lamirel, C., **N. J. Newman**, and **V. Biousse** (2010). Optical coherence tomography (oct) in optic neuritis and multiple sclerosis. *Revue neurologique*, **166**(12), 978–986. 42

Lefkimmiatis, S., Universal denoising networks: a novel cnn architecture for image denoising. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. 18

Li, K., **X. Wu**, **D. Z. Chen**, and **M. Sonka** (2005). Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE transactions on pattern analysis and machine intelligence*, **28**(1), 119–134. 43, 44

Lu, D., **M. Heisler**, **S. Lee**, **G. Ding**, **M. V. Sarunic**, and **M. F. Beg** (2017). Retinal fluid segmentation and detection in optical coherence tomography images using fully convolutional neural network. *arXiv preprint arXiv:1710.04778*. 71, 76, 77, 78

Lu, S., **C. Y.-l. Cheung**, **J. Liu**, **J. H. Lim**, **C. K.-s. Leung**, and **T. Y. Wong** (2010). Automated layer segmentation of optical coherence tomography images. *IEEE Transactions on biomedical engineering*, **57**(10), 2605–2608. 43

Lv, H., **S. Fu**, **C. Zhang**, and **L. Zhai** (2018). Speckle noise reduction of multi-frame optical coherence tomography data using multi-linear principal component analysis. *Optics express*, **26**(9), 11804–11818. 17

Mayer, M. (). Octseg. *Friedrich-Alexander-Universität Erlangen-Nürnberg*. 44, 45

Mayer, M. A., **J. Hornegger**, **C. Y. Mardin**, and **R. P. Tornow** (2011). Retinal layer segmentation on oct-volume scans of normal and glaucomatous eyes. *Investigative Ophthalmology & Visual Science*, **52**(14), 3669–3669. 33

Mayer, M. A., **R. P. Tornow**, **J. Hornegger**, and **F. E. Kruse** (2008). Fuzzy c-means clustering for retinal layer segmentation on high resolution oct images. *Analysis of Biomedical Signals and Images*, 160–166. 42

MeindertNiemeijer, X. C., **L. Z. K. Lee**, **M. D. Abràmoff**, and **M. Sonka** (2012). 3d segmentation of fluid-associated abnormalities in retinal oct: Probability constrained graph-search-graph-cut. *IEEE Transactions on Medical Imaging*, **31**(8), 1521–1531. 44

Menon, S. N., **V. V. Reddy**, **A. Yeshwanth**, **B. Anoop**, and **J. Rajan**, A novel deep learning approach for the removal of speckle noise from optical coherence tomography images using gated convolution–deconvolution structure. *In Proceedings of 3rd International Conference on Computer Vision and Image Processing*. Springer, 2020*a*. xiii, 18, 23, 24, 29

Menon, S. N., **V. V. Reddy**, **A. Yeshwanth**, **B. Anoop**, and **J. Rajan**, A novel deep learning approach for the removal of speckle noise from optical coherence tomography images using gated convolution–deconvolution structure. *In Proceedings of 3rd International Conference on Computer Vision and Image Processing*. Springer, 2020*b*. 70

Michalik, A., **J. Martin**, and **C. Van Broeckhoven** (2004). Spinocerebellar ataxia type 7 associated with pigmentary retinal dystrophy. *European journal of human genetics*, **12**(1), 2. 42

Mitchell, T. M., **R. M. Keller**, and **S. T. Kedar-Cabelli** (1986). Explanation-based generalization: A unifying view. *Machine learning*, **1**(1), 47–80. 46

OCSC (2015). Ocsc. https://optima.meduniwien.ac.at/research/challenges/. xiii, 9, 20, 21

Oktay, O., **J. Schlemper**, **L. L. Folgoc**, **M. Lee**, **M. Heinrich**, **K. Misawa**, **K. Mori**, **S. McDonagh**, **N. Y. Hammerla**, **B. Kainz**, *et al.* (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*. 74

Pagliara, M. M., **D. Lepore**, and **E. Balestrazzi** (2008). The role of oct in glaucoma management. *Progress in brain research*, **173**, 139–148. 8

Paquet, C., **M. Boissonnot**, **F. Roger**, **P. Dighiero**, **R. Gil**, and **J. Hugon** (2007). Abnormal retinal thickness in patients with mild cognitive impairment and alzheimer's disease. *Neuroscience letters*, **420**(2), 97–99. 42

Paul, A., **D. P. Mukherjee**, and **S. T. Acton** (2018). Speckle removal using diffusion potential for optical coherence tomography images. *IEEE journal of biomedical and health informatics*, **23**(1), 264–272. 17

Pircher, M., **E. Götzinger**, **R. A. Leitgeb**, **A. F. Fercher**, and **C. K. Hitzenberger** (2003). Speckle reduction in optical coherence tomography by frequency compounding. *Journal of biomedical optics*, **8**(3), 565–570. 16

Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 56

Quellec, G., **K. Lee**, **M. Dolejsi**, **M. K. Garvin**, **M. D. Abramoff**, and **M. Sonka** (2010). Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in sd-oct of the macula. *IEEE transactions on medical imaging*, **29**(6), 1321–1330. 44

Rajabi, H. and **A. Zirak** (2016). Speckle noise reduction and motion artifact correction based on modified statistical parameters estimation in oct images. *Biomedical Physics & Engineering Express*, **2**(3), 035012. 17

Rao, T. N., **G. Girish**, **A. R. Kothari**, and **J. Rajan**, Deep learning based sub-retinal fluid segmentation in central serous chorioretinopathy optical coherence tomography scans. *In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019. 16, 70

Reddy, T. G. P., **K. S. Ashritha**, **T. Prajwala**, **G. Girish**, **A. R. Kothari**, **S. G. Koolagudi**, and **J. Rajan**, Retinal-layer segmentation using dilated convolutions. *In Proceedings of 3rd International Conference on Computer Vision and Image Processing*. Springer, 2020. xv, 33, 44, 49, 50, 51, 52, 53, 54, 58, 59, 60, 61, 65, 70

Ren, Y., **L. Zhang**, and **P. N. Suganthan** (2016). Ensemble classification and regression-recent developments, applications and future directions. *IEEE Computational Intelligence Magazine*, **11**(1), 41–53. 46

Roy, A. G., **S. Conjeti**, **S. P. K. Karri**, **D. Sheet**, **A. Katouzian**, **C. Wachinger**, and **N. Navab** (2017). Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomedical optics express*, **8**(8), 3627–3642. 44, 51, 54, 57, 58, 59, 60, 61, 65, 70

Schmitt, J. M., **S. Xiang**, and **K. M. Yung** (1999). Speckle in optical coherence tomography. *Journal of biomedical optics*, **4**(1), 95–106. 8, 16, 22

Schuman, J. S., **M. R. Hee**, **A. V. Arya**, **T. Pedut-Kloizman**, **C. A. Puliafito**, **J. G. Fujimoto**, and **E. A. Swanson** (1995). Optical coherence tomography: a new tool for glaucoma diagnosis. *Current opinion in ophthalmology*, **6**(2), 89–95. 42

Selvaraju, R. R., **M. Cogswell**, **A. Das**, **R. Vedantam**, **D. Parikh**, and **D. Batra**, Grad-cam: Visual explanations from deep networks via gradient-based localization. *In Proceedings of the IEEE International Conference on Computer Vision*. 2017. 45, 62

Shah, A., **L. Zhou**, **M. D. Abrámoff**, and **X. Wu** (2018). Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in oct images. *Biomedical optics express*, **9**(9), 4509–4526. 44

Shi, F., **X. Chen**, **H. Zhao**, **W. Zhu**, **D. Xiang**, **E. Gao**, **M. Sonka**, and **H. Chen** (2015). Automated 3-d retinal layer segmentation of macular optical coherence tomography images with serous pigment epithelial detachments. *IEEE transactions on medical imaging*, **34**(2), 441–452. 43

Srinivasan, P. P., **S. J. Heflin**, **J. A. Izatt**, **V. Y. Arshavsky**, and **S. Farsiu** (2014). Automatic segmentation of up to ten layer boundaries in sd-oct images of the mouse retina with and without missing layers due to pathology. *Biomedical optics express*, **5**(2), 348–365. 43

Stankiewicz, A., **T. Marciniak**, **A. Dbrowski**, **M. Stopa**, **P. Rakowicz**, and **E. Marciniak** (2017). Denoising methods for improving automatic segmentation in oct images of human eye. *Bulletin of the Polish Academy of Sciences Technical Sciences*, **65**(1), 71–78. 30, 31, 32

Staurenghi, G., **S. Sadda**, **U. Chakravarthy**, **R. F. Spaide**, *et al.* (2014). Proposed lexicon for anatomic landmarks in normal posterior segment spectral-domain optical coherence tomography: the in oct consensus. *Ophthalmology*, **121**(8), 1572–1578. 45

Steinmetz, J. D., **R. R. Bourne**, **P. S. Briant**, **S. R. Flaxman**, **H. R. Taylor**, **J. B. Jonas**, **A. A. Abdoli**, **W. A. Abrha**, **A. Abualhasan**, **E. G. Abu-Gharbieh**, *et al.* (2021). Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: the right to sight: an analysis for the global burden of disease study. *The Lancet Global Health*, **9**(2), e144–e160. 10

Sudeep, P., **S. I. Niwas**, **P. Palanisamy**, **J. Rajan**, **Y. Xiaojun**, **X. Wang**, **Y. Luo**, and **L. Liu** (2016). Enhancement and bias removal of optical coherence tomography images: an iterative approach with adaptive bilateral filtering. *Computers in biology and medicine*, **71**, 97–107. 16, 17, 57

Sudre, C. H., **W. Li**, **T. Vercauteren**, **S. Ourselin**, and **M. J. Cardoso**, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *In Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2017, 240–248. 52

Tang, C., **L. Cao**, **J. Chen**, and **X. Zheng** (2017). Speckle noise reduction for optical coherence tomography images via non-local weighted group low-rank representation. *Laser Physics Letters*, **14**(5), 056002. 16

Thapa, D., **K. Raahemifar**, and **V. Lakshminarayanan** (2015). Reduction of speckle noise from optical coherence tomography images using multi-frame weighted nuclear norm minimization method. *Journal of Modern Optics*, **62**(21), 1856–1864. 16

TheEye (2021). How the eye works. http://www.pceyeglasses.com/eye-anatomy.html. xiii, 1

Tian, C., **Y. Xu**, and **W. Zuo** (2020). Image denoising using deep cnn with batch renormalization. *Neural Networks*, **121**, 461–473. 18

Vegas-Sanchez-Ferrero, G. (2010). "on the influence of interpolation on probabilistic models for ultrasonic images.". *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 22

Vermeer, K., **J. Van der Schoot**, **H. Lemij**, and **J. De Boer** (2011). Automated segmentation by pixel classification of retinal layers in ophthalmic oct images. *Biomedical optics express*, **2**(6), 1743–1756. 43

Wang, X., **X. Yu**, **X. Liu**, **S. Chen**, **S. Chen**, **N. Wang**, and **L. Liu** (2018). A two-step iteration mechanism for speckle reduction in optical coherence tomography. *Biomedical Signal Processing and Control*, **43**, 86–95. 17

Wilkins, G. R., **O. M. Houghton**, and **A. L. Oldenburg** (2012). Automated segmentation of intraretinal cystoid fluid in optical coherence tomography. *IEEE Transactions on Biomedical Engineering*, **59**(4), 1109–1114. 4

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, **5**(2), 241–259. 46

Wong, B. J., **R. P. Jackson**, **S. Guo**, **J. M. Ridgway**, **U. Mahmood**, **J. Su**, **T. Y. Shibuya**, **R. L. Crumley**, **M. Gu**, **W. B. Armstrong**, *et al.* (2005). In vivo optical coherence tomography of the human larynx: normative and benign pathology in 82 patients. *The Laryngoscope*, **115**(11), 1904–1911. 8

Wong, K. C., **M. Moradi**, **H. Tang**, and **T. Syeda-Mahmood**, 3d segmentation with exponential logarithmic loss for highly unbalanced object sizes. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018. 52

Wu, M., **W. Fan**, **Q. Chen**, **Z. Du**, **X. Li**, **S. Yuan**, and **H. Park** (2017). Three-dimensional continuous max flow optimization-based serous retinal detachment segmentation in sd-oct for central serous chorioretinopathy. *Biomedical optics express*, **8**(9), 4257–4274. 70

Xiang, D., **G. Chen**, **F. Shi**, **W. Zhu**, **Q. Liu**, **S. Yuan**, and **X. Chen** (2019). Automatic retinal layer segmentation of oct images with central serous retinopathy. *IEEE journal of biomedical and health informatics*, **23**(1), 283–295. 42, 43

Xiao, Y., **J. Wu**, **Z. Lin**, and **X. Zhao** (2018). A deep learning-based multi-model ensemble method for cancer prediction. *Computer methods and programs in biomedicine*, **153**, 1–9. 46

Xu, J., **H. Ou**, **E. Y. Lam**, **P. Chui**, and **K. K. Wong** (2013). Speckle reduction of retinal optical coherence tomography based on contourlet shrinkage. *Optics letters*, **38**(15), 2900–2903. 17

Yang, Q., **C. A. Reisman**, **K. Chan**, **R. Ramachandran**, **A. Raza**, and **D. C. Hood** (2011). Automated segmentation of outer retinal layers in macular oct images of patients with retinitis pigmentosa. *Biomedical optics express*, **2**(9), 2493–2503. 43

Yazdanpanah, A., **G. Hamarneh**, **B. R. Smith**, and **M. V. Sarunic** (2011). Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach. *IEEE transactions on medical imaging*, **30**(2), 484–496. 43

Yin, X., **J. R. Chao**, and **R. K. Wang** (). User-guided segmentation for volumetric retinal optical coherence tomography images. *Journal of biomedical optics*, **19**(8). 42

Yu, H., **J. Gao**, and **A. Li** (2016). Probability-based non-local means filter for speckle noise suppression in optical coherence tomography images. *Optics letters*, **41**(5), 994–997. 16

Yuan, Q., **Q. Zhang**, **J. Li**, **H. Shen**, and **L. Zhang** (2018). Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, **57**(2), 1205–1218. 18

Zaki, F., **Y. Wang**, **H. Su**, **X. Yuan**, and **X. Liu** (2017). Noise adaptive wavelet thresholding for speckle noise removal in optical coherence tomography. *Biomedical optics express*, **8**(5), 2720–2731. 17

Zhang, A., **J. Xi**, **J. Sun**, and **X. Li** (2017*a*). Pixel-based speckle adjustment for noise reduction in fourier-domain oct images. *Biomedical optics express*, **8**(3), 1721–1730. 17

Zhang, K., **W. Zuo**, **Y. Chen**, **D. Meng**, and **L. Zhang** (2017*b*). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, **26**(7), 3142–3155. 18, 30, 31, 32

Zhang, K., **W. Zuo**, and **L. Zhang** (2018). Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, **27**(9), 4608–4622. 18

Zhang, L., **W. Zhu**, **F. Shi**, **H. Chen**, and **X. Chen**, Automated segmentation of intraretinal cystoid macular edema for retinal 3d oct images with macular hole. *In 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015*a*. 30, 31, 32

Zhang, S., **A. E. Choromanska**, and **Y. LeCun**, Deep learning with elastic averaging sgd. *In Advances in Neural Information Processing Systems*. 2015*b*. 54

Zhu, W., **Y. Huang**, **L. Zeng**, **X. Chen**, **Y. Liu**, **Z. Qian**, **N. Du**, **W. Fan**, and **X. Xie** (2019). Anatomynet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical physics*, **46**(2), 576–589. 52

Zuiderveld, K., Graphics gems iv. chapter Contrast Limited Adaptive Histogram Equalization. Academic Press Professional, Inc., San Diego, CA, USA, 1994. ISBN 0-12-336155-9, 474–485. URL http://dl.acm.org/citation.cfm?id= 180895.180940. 73

PUBLICATIONS BASED ON THIS THESIS

Journal Papers

1. **Anoop, B. N.**, Kaushik S. Kalmady, Akhil Udathu, V. Siddharth, G. N. Girish, Abhishek R. Kothari, and Jeny Rajan, "A cascaded convolutional neural network architecture for despeckling OCT images." *Biomedical Signal Processing and Control*, 66 (2021): 102463.

2. **Anoop, B. N.**, Rakesh Pavan, G. N. Girish, Abhishek R. Kothari, and Jeny Rajan, "Stack generalized deep ensemble learning for retinal layer segmentation in Optical Coherence Tomography images", *Biocybernetics and Biomedical Engineering*, 40, no. 4 (2020): 1343-1358.

Conference Papers

1. **Anoop, B. N.**, Saswat Parida, Ajith B, G N Girish, Abhishek R. Kothari, Muthu Subash Kavitha, and Jeny Rajan, "Attention Assisted Patch-wise CNN for the Segmentation of Fluids from the Retinal Optical Coherence Tomography Images." *International Conference on Pattern Recognition and Machine Intelligence*. Springer, 2021. (Submitted).

2. Menon, Sandeep N., VB Vineeth Reddy, A. Yeshwanth, **B. N. Anoop**, and Jeny Rajan, "A novel deep learning approach for the removal of speckle noise from optical coherence tomography images using gated convolution-deconvolution structure." *In Proceedings of 3rd International Conference on Computer Vision and Image Processing*, pp. 115-126. Springer, Singapore, 2020.

Book Chapter

1. **Anoop, B. N.**, G. N. Girish, P. V. Sudeep, and Jeny Rajan, "Despeckling Algorithms for Optical Coherence Tomography Images: A Review." *Advanced Classification Techniques for Healthcare Analysis* (2019): 286-310.

ANOOP B N

Anoop Bhavan, Vlathankara P O, Thiruvananthapuram, Kerala - 695 134

(+91)9071897997 ⋄ anoopcem@gmail.com

EDUCATION

National Institute of Technology Karnataka - Surathkal.   *December 2017 - July 2021*
Doctor of Philosophy (Ph.D)
Department of Computer Science and Engineering

Thesis Title: Analysis and quantification of retinal pathologies from the retinal OCT images using Deep Learning techniques.
Advisor: Dr. Jeny Rajan, Department of CSE, NITK Surathkal.
Summary: The Research mainly aims at developing Preprocessing, Layer segmentation and Cyst segmentation techniques from the retinal OCT images using Deep Learning methods.

National Institute of Technology Calicut, Kerala   *July 2011 - May 2013*
Master of Technology
Department of Electronics and Communication Engineering (Signal Processing)

Thesis Title: Secure Image and Video Transcoders.
Advisor: Dr. Sudhish N George, Department of ECE, NIT Calicut.
Summary: In this project work, developed a secure transcoder for images and videos by utilizing the JPEG, MPEG and H.264 standards.

College of Engineering Munnar, Kerala   *2003 - 2007*
Bachelor of Technology
Department of Electronics and Communication Engineering

Thesis Title: GSM SCADA with Robotics..
Advisor: Dr. Ramesh P, Department of ECE, College of Engineering Munnar.
Summary: In this project work, developed a SCADA system with the help of GSM and Robotics.

CAREER OBJECTIVE

To work for an organization which provides me the opportunity to improve my skills and knowledge to grow along with the organization objective.

WORK EXPERIENCE: 10 YEARS 9 MONTHS

NITK Surathkal, Karnataka.   May 2019 - March 2020
Senior Research Fellow   *(10 Months)*

NITK Surathkal, Karnataka.   May 2017 - May 2019
Junior Research Fellow   *(2 Years)*

St. Joseph's College of Engineering and Technology, Palai, Kerala.   May 2013 - May 2017
Assistant Professor   *(4 Years)*

St. Joseph's College of Engineering and Technology, Palai, Kerala.   January 2008 - July 2011
Lecturer   *(3 Years 5 Months)*

College of Engineering Munnar, Kerala.   August 2007 - January 2008
Lecturer   *(6 Months)*

PUBLICATIONS

SCI Journals

1. Anoop, B. N., et al. "A cascaded convolutional neural network architecture for despeckling OCT images." Biomedical Signal Processing and Control 66 (2021): 102463. (Impact factor. 3.88).

2. Pawan, S. J., Rahul Sankar, Anubhav Jain, Mahir Jain, D. V. Darshan, B. N. Anoop, Abhishek R. Kothari, M. Venkatesan, and Jeny Rajan. "Capsule Networkbased architectures for the segmentation of sub-retinal serous fluid in optical coherence tomography images of central serous chorioretinopathy." Medical Biological Engineering Computing (2021): 1-15.(Impact factor. 2.61)

3. Anoop, B. N., et al. "Stack generalized deep ensemble learning for retinal layer segmentation in Optical Coherence Tomography images." Biocybernetics and Biomedical Engineering (2020). (Impact factor. 4.314).

4. Joseph, Justin, B. N. Anoop, and Joseph Williams. "A modified unsharp masking with adaptive threshold and objectively defined amount based on saturation constraints." Multimedia Tools and Applications 78.8 (2019): 11073-11089. (Impact factor. 2.313).

5. Anoop, B. N., et al. "A prospective case study of high boost, high frequency emphasis and two-way diffusion filters on MR images of glioblastoma multiforme." Australasian physical engineering sciences in medicine 41.2 (2018): 415-427. (Impact factor. 1.161).

6. Babu, Jalumedi, et al. "Development of a comprehensive delamination assessment factor and its evaluation with high-speed drilling of composite laminates using a twist drill." Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 232.12 (2018): 2109-2121. (Impact factor. 0.66).

Journals

1. Anoop, B. N., Sudhish N. George, and P. P. Deepthi. "Secure Image Transcoding technique using chaotic key based algorithm." International Journal of Advanced Computer Research 2.4 (2012): 84.

2. Jacob, Pooja, and B. N. Anoop. "Design and implementation of polyphase decimation filter." International Journal of Computer Networks and Wireless Communications (IJCNWC), ISSN (2014): 2250-3501.

Book Chapter

1. Anoop, B. N., et al. "Despeckling Algorithms for Optical Coherence Tomography Images: A Review." Advanced Classification Techniques for Healthcare Analysis. IGI Global, 2019. 286-310.

International Conferences

1. Anoop, B. N., Saswat Parida, Ajith B, G N Girish, Abhishek R. Kothari, Muthu Subash Kavitha, and Jeny Rajan. "Attention Assisted Patch-wise CNN for the Segmentation of Fluids from the Retinal Optical Coherence Tomography Images." International Conference on Pattern Recognition and Machine Intelligence. Springer, 2021. (Submitted).

2. Menon, Sandeep N., et al. "A Novel Deep Learning Approach for the Removal of Speckle Noise from Optical Coherence Tomography Images Using Gated ConvolutionDeconvolution Structure." Proceedings of 3rd International Conference on Computer Vision and Image Processing. Springer, Singapore, 2020.

3. Anoop, B. N., P. E. Ameenudeen, and Justine Joseph. "A Meta-Analysis of Contrast Measures Used for the Performance Evaluation of Histogram Equalization Based Image Enhancement Techniques." 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2018.

4. Anoop, B. N., Sudhish N. George, and P. P. Deepthi. "Secure video transcoders based on correlation preserving sorting algorithm." 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). IEEE, 2014.

5. Anoop, B. N., Sudhish N. George, and P. P. Deepthi. "Secure image transcoding technique using chaotic key based algorithm with improved security." 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s). IEEE, 2013.

National Conferences

1. Anoop, B. N. "Lane and Car Traffic Controller." National conference VLSI Design and Information Technologies (NVDIT-2011) at METs School of Engineering, Mala.

2. James et al. " Character Recognition and Conversion to Sound - A Simplified Approach." Proceedings of National Conference on Signal Processing, Communication and Nanotechnology 2014.

GATE 2009

Score	341.
Rank	5932.

TECHNICAL STRENGTHS

Skills	Deep Learning, Convolutional Neural Networks, Auto-encorders, GANs.
Software & Tools	MATLAB, Python, Tensorflow, Keras.

MEMBERSHIP IN PROFESSIONAL BODIES

IEEE, IEEE Signal Processing Society.

Internet Society

POSITIONS HELD WITH RESPONSIBILITIES

Reviewer - IEEE Transactions on Medical Imaging.

Reviewer - Biocybernetics and Biomedical Engineering.

Reviewer - Biomedical Signal Processing and Control.

Reviewer - International Journal for Light and Electron Optics.

IEEE Student Branch Counselor, SJCET - Palai.

INVITED TALKS

- A talk on "Introduction to Deep Learning ", in connection with AICTE Sponsored Faculty Development Progrmme on Intelligent Big Data Analytics Organized by department of CSE, M.A.M.CE, Tiruchirappalli, on November 2019.

- A talk on " Introduction to GANs", in connection with Summer School on Deep Learning Organized by department of CSE, NITK Surathkal, on June 2019.

- A talk on " Introduction to Auto-encorders and GANs", in connection with Industry 4.0 Organized by department of ECE, MITS, Kerala, on May 2019.

- A talk on "Medical Imaging Modalities", in connection with TEQIP-II Sponsored Faculty Development Progrmme on Signal Processing and its Applications Organized by department of EI, College of Engineering Kidangoor, Kerala, on November 2016.

SHORT TERM COURSES ATTENDED

- Workshop on Deep Learning, of five days duration, organized by National Institute of Technology (NIT), Calicut during 18th December to 22nd December, 2018.

- Workshop on Computer Vision and Image Processing, of three days duration, organized by Indian Institute of Technology (IIT), Roorkee during 09th March to 11th March, 2018.

- Faculty Development Programme on Digital Filter Design-Theory and Practice, of one week duration, organized by College of engineering Trivandrum (CET), during 25th July to 30th July, 2016.

- Workshop on Technical Communication of two weeks duration, sponsored by Indian Society of Technical Education (ISTE) conducted by IIT-Bombay during 8th October to 5th December, 2015.

- Workshop on signals and systems of two weeks duration, sponsored by Indian Society of Technical Education (ISTE) conducted by IIT-Kharagpur during 2nd January to 12th January, 2014.

- MHRD- TEQIP II sponsored Faculty Development Programme on Estimation and Coding for Multimedia Processing, of one week duration, organized by National Institute of Technology (NIT), Calicut during 30th June to 05th July, 2014.

REFERENCES

Dr. Jeny Rajan
Assistant Professor
Department of CSE
NITK Surathkal
Mangalore - 575 025
Email: jenyrajan@nitk.edu.in

Dr. Sudhish N George
Assistant Professor
Department of ECE
NIT Calicut
Kerala - 673 601
Email: sudhish@nitc.ac.in

Dr. Girish G N
Assistant Professor
Department of CSE
IIIT Sri City
Chittoor 517 646
Email: girish.anit@gmail.com

Dr. Abhishek Kothari
Vitreoretinal Surgeon
Director and Chief Consultant
Pink City Eye & Retina Center
Jaipur, Rajasthan - 302 015
Email: dr.a.kothari@gmail.com

DECLARATION

I hereby declare that the information furnished above is true to the best of my knowledge.

ANOOP B N