Retinal Image Quality Improvement via Learning

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Electronics and Communication Engineering by Research

by

Sukesh Adiga V
20162017
sukesh.adigav@research.iiit.ac.in

INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY
HYDERABAD

International Institute of Information Technology
Hyderabad - 500 032, INDIA
August 2019

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Retinal Image Quality Improvement via Learning" by Sukesh Adiga V, has been carried out under my supervision and is not submitted elsewhere for a degree.

____
Date

____
Adviser: Prof. Jayanthi Sivaswamy

To Harry,

one who supported and motivated me in every stage

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisor, Prof. Jayanthi Sivaswamy for the continuous support of my master study and research, for her patience to all my silly mistakes and doubts, time to time motivation and suggestion, enthusiasm, and immense knowledge. Coming from an industry with no experience of research, she had taught me an essential aspect of research. I will remember for her advice in my failure time: "Research is not all about success stories, the failure is also part of it. You have to analyse and learn from both". This statement is well applicable to any life situation. Apart from academics, she has imparted knowledge of seeing things differently. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor for my master study. I would also like to thank Dr. Anand Sivaraman and Dr. Ramachandran Rajalakshmi for providing medical data required for my thesis work.

At the start of my journey in IIIT-Hyderabad, I received a lot of help from Sudarshan and Surya Sir from getting accustomed in college to choosing a guide. The discussion with them clarified many things and directed me onto the right path. Samrudhdhi helped my initial days of research. She used to involve discussing deeply in my research, and I was fortunate to get such a senior. She along with Karthik and Chetan were go to friends for any advice and suggestions for any aspect of research, especially during tough times. I am lucky and thankful to earned their friendship and be in touch with them. I want to thank my MIP lab mates Arunava Sir, Raghav, Aabhas, Jahnavi, Pujitha, Alphin, Abhinav for fruitful group meetings. I also like to thank all CVITians for making this learning experience a memorable one.

Every day in college would be incomplete without lunch/dinner/walking time discussion with Krishna, Zakir, Kishor, Jayanth, Josyula and few others. Many activities such as food exploration (especially sweets), badminton and swimming times with Govinda, Joyneel and Venkat; late-night discussion with Suman, Prateek, Vijay; table tennis & movie times and Cricket discussion are some memories of my hostel stay. I want to thank them all and many other friends in IIIT Hyderabad.

Finally, I would like to wholeheartedly thank my parents, who have supported me in every stage of my life. I am fortunate to have an elder brother Nagaraj, who showed me the gist of research and motivated me in every step of my life, and sister-in-law, who encouraged me for every decision.

Abstract

Retinal images are widely used to detect and diagnose many diseases such as Diabetic Retinopathy (DR), glaucoma, Age-related Macular Degeneration, Cystoid Macular Edema, coronary heart disease, and so on. These diseases affect vision and lead to irreversible blindness. Early image-based screening and monitoring of the patient is a solution. Imaging of retina is commonly done either through Optical Coherence Tomography (OCT) or Fundus photography. OCT captures cross-sectional information about the retinal layers in a 3D volume, whereas fundus imaging projects retinal tissues onto the 2D imaging plane. Recently smartphone camera-based fundus imaging is being explored with a relatively low-cost. Imaging retina with these technologies pose challenges due to physical properties of the light source, or quality of optics and sensors used or low and uneven light condition. In this thesis, we look at learning based approaches, namely neural network techniques to improve the quality of retinal images to aid diagnosis.

The first part of this thesis aims at denoising OCT images, which are corrupted by speckle noise due to underlying coherence-based imaging technique. We propose a new method for denoising OCT images based on Convolutional Neural Network by learning common features from unpaired noisy and clean OCT images in an unsupervised, end-to-end manner. The proposed method consists of a combination of two autoencoders with shared encoder layers, which we call as Shared Encoder (SE) architecture. The SE is trained to reconstruct noisy and clean OCT images with respective autoencoders, and denoised OCT image is obtained using a cross-model prediction. The proposed method can be used for denoising OCT images with or without pathology from any scanner. The SE architecture was assessed using public datasets and found to perform better than baseline methods exhibiting a good balance of retaining anatomical integrity and speckle reduction. The second problem we focus on is the enhancement of fundus images acquired with a Smartphone camera (SC). SC image is a cost-effective solution for the assessment of retina, especially in screening. However, imaging at high magnification and low light levels results in loss of details, uneven illumination, noise particularly in the peripheral region and flash-induced artefacts. We address these problems by matching the characteristics of images from SC to those from a regular fundus camera (FC) using either unpaired or paired data. Two mapping solutions are designed using deep learning technique in an unsupervised and supervised manner. The unsupervised architecture called ResCycleGAN is based on the CycleGAN with two significant changes: A residual connection is introduced to aid learning only the correction required; A structure similarity based loss function is used to improve the clarity of anatomical structures and pathologies. This method can handle variations seen in normal and pathological images, acquired even without mydriasis, which is attractive in screening. The method produces consistently balanced results, outperforms CycleGAN both qualitatively and quantitatively, and has more pleasing results. Next, a new architecture is proposed called SupEnh, which handles noise removal using paired data. The proposed method enhances the quality of SC images along with denoising in an end-to-end, supervised manner. Obtaining paired data is challenging; however, it is feasible in fixed clinical settings or commercial product as it is required once for learning. The proposed SupEnh method based on U-net consists of an encoder and two decoders. The network simplifies the task by learning denoising and mapping to FC separately with two decoders. The method handles images with/without pathologies as well as images acquired even without mydriasis. The SupEnh was assessed using private datasets and found to performs better than U-net. The cross-validation results show method is robust to change in image quality. The enhancement using SupEnh method achieves 5% higher AUC for early stage DR detection when compared with original images.

Contents

Chapter Page

1 Introduction . . . 1
- 1.1 Retina and Retinal Imaging . . . 1
- 1.2 Retinal Imaging Modalities . . . 2
  - 1.2.1 Fundus Imaging . . . 3
  - 1.2.2 Optical Coherence Tomography . . . 4
- 1.3 Image Quality . . . 5
  - 1.3.1 Degradation in OCT Image . . . 6
  - 1.3.2 Degradation in Smartphone Camera based Fundus Image . . . 6
  - 1.3.3 Other Challenges . . . 7
- 1.4 Thesis Focus . . . 8
- 1.5 Contributions . . . 8
- 1.6 Outline of the Thesis . . . 9

2 Denoising of Retinal OCT Images: A Semi-supervised Approach . . . 10
- 2.1 Introduction . . . 10
- 2.2 Related Work . . . 11
- 2.3 Proposed Method . . . 12
  - 2.3.1 Shared Encoder Architecture . . . 12
  - 2.3.2 Network Training and Testing . . . 14
    - 2.3.2.1 Backtranslation Training: . . . 14
  - 2.3.3 Loss Function . . . 15
- 2.4 Dataset and Experiments Details . . . 16
  - 2.4.1 Dataset and Parameters . . . 16
  - 2.4.2 Experimental Setting . . . 16
    - 2.4.2.1 Unsupervised Shared Encoder . . . 17
    - 2.4.2.2 Stacked Shared Encoder . . . 17
    - 2.4.2.3 Semi-supervised SE . . . 17
- 2.5 Performance Measures and Comparisons . . . 18
  - 2.5.1 Qualitative Evaluation on SBSDI Dataset . . . 18
  - 2.5.2 Qualitative Evaluation on Chiu's Dataset . . . 20
  - 2.5.3 Denoising Performance with Scanner Variability . . . 21
  - 2.5.4 Quantitative Evaluation . . . 22
- 2.6 Conclusions . . . 23

3 Matching the characteristics of Fundus & Smartphone camera images: An Unsupervised Approach 25
3.1 Introduction 25
3.2 Method 26
3.2.1 Proposed Architecture 27
3.2.2 Loss functions 28
3.2.2.1 Adversarial Loss 28
3.2.2.2 Cycle-Consistency Loss 28
3.2.2.3 Identity Loss 29
3.2.2.4 Overall training Loss 29
3.3 Implementation and Dataset Details 29
3.3.1 Network Architecture 29
3.3.2 Training Details 29
3.3.3 Dataset Details 30
3.4 Performance Analysis 30
3.4.1 Qualitative Assessment 30
3.4.2 Comparison with CycleGAN [75] 30
3.4.3 Quantitative Evaluation 31
3.4.4 Comparison against Standard Retinal Image Enhancement Method [74] 32
3.5 Conclusion 33

4 Enhancement of Smartphone camera based Fundus Images: A Supervised Approach 34
4.1 Introduction 34
4.2 Method 35
4.2.1 Proposed architecture 35
4.2.2 Loss function 36
4.3 Implementation and Dataset Details 37
4.3.1 Network Architecture 37
4.3.2 Training Details 37
4.3.3 Dataset Description 38
4.3.4 Data Preparation 38
4.4 Results 38
4.4.1 Qualitative analysis 39
4.4.1.1 Comparison with U-net and ResCycleGAN 40
4.4.1.2 Cross-validation with iPhone images 43
4.4.2 Quantitative analysis 44
4.4.2.1 Evaluation using Full-Reference based Quality Metric 44
4.4.2.2 Evaluation using No-reference based Quality Metric 45
4.4.2.3 Performance of Early Stage Diabetic Retinopathy Detection 45
4.5 Conclusions 47

5 Conclusions 48
5.1 Future Work 49

Bibliography 51

List of Figures

Figure Page

1.1 Illustration of inner anatomy of eye and retinal layers (Source: Junqueira's Basic Histology: Text and Atlas 12^{th} edition by A. Mescher) 2

1.2 Fundus imaging. (a) First known image of human retina [64] (b) Standard fundus camera image (c) Smartphone based fundus image 3

1.3 Fundus cameras. (a) Standard fundus camera (Zeiss FF450 device) (b) Fundus on Phone (a product of Remidio Innovative Solutions Pvt. Ltd.) 4

1.4 Illustration of OCT scans. Left: A depth profile (A-scan) of backscattered intensity. Middle: The beam is scanned in a transverse direction to obtain 2D imaging (B-scan). Right: Multiple B-scans are acquired to obtain 3D volume 5

1.5 Illustration of OCT image. Inset image from left to right shows degradation of retinal structures in choroid, macula and blood vessel regions 6

1.6 A sample retinal image captured by a standard fundus camera (left) and a smartphone camera (right) 7

1.7 Challenges in fundus images using smartphone camera 7

2.1 OCT B-scan images from different scanner and an averaged OCT image 11

2.2 (a) The proposed Shared Encoder (SE) architecture (b) Schematic representation of the encoder and decoder layers in SE. Solid boxes represent multi-channel feature maps. Dashed boxes represent copied feature maps. Number of channels is denoted on the top of the box 13

2.3 The block diagram of Stacked Shared Encoder (StSE) 17

2.4 Illustration of denoising performance with different method for an example of OCT data from SBSDI [10] 18

2.5 A OCT denoising performance comparison with DP [50] method 19

2.6 Illustration of denoising performance with different method on Chiu's normal OCT data [7] 20

2.7 Illustration of denoising performance with different method on Chiu's DME OCT data [8] 21

2.8 Illustration of denoising results on OCT images from different scanner. Column 1 is noisy image and Column 2 is corresponding denoised image from our best method SSE1 22

3.1 Comparison of images of a scene taken by a standard camera (left) and an SC (right) . 26

3.2 Schematic representation of the proposed architecture. The red and blue color trapezoid represent encoder and decoder layer, respectively. The green color circle represent residual connection 27

3.3 Sample results for ResCycleGAN for images without (top) and with pathologies (bottom) 31

3.4 Comparison of ResCycleGAN with CycleGAN outputs for images without (Row 1) and with pathologies (Row 2) 32
3.5 Comparison of standard retinal image enhancement with the proposed mapping. Left to right: SC image, results of our method and enhancement [74] 33

4.1 Schematic of the proposed architecture 36
4.2 Sample results for SupEnh for images without (top) and with pathologies (bottom) . . 39
4.3 Comparison of mapped results for a sample image (a) with other methods: (b) U-net1, (c) U-net2, (d) R-U-net, (e) Proposed method, and (f) ResCycleGAN 40
4.4 Line profile comparison of mapped results for a sub-region (shown in inset image) . . 41
4.5 Comparison of mapped results for a pathological image (a) with other methods: (b) U-net1, (c) U-net2, (d) R-U-net, (e) Proposed method, and (f) ResCycleGAN 42
4.6 Comparison of mapped results for a sample image from iPhone (a) with other methods: (b) U-net1, (c) U-net2, (d) R-U-net, (e) Proposed method, and (f) ResCycleGAN . . . 43
4.7 ROC curve for early stage DR detection for per case 46

List of Tables

Table Page

2.1 Quantitative comparison of performance on the synthetic dataset 23
2.2 Quantitative comparison of performance on the real dataset 23

3.1 Quantitative comparison of performance using Q_v and D_b on SC images 33

4.1 Dataset description 38
4.2 Full-reference based quantitative comparison of performance on the SC images obtained using Micromax 44
4.3 No-reference based quantitative comparison of performance on the SC images obtained using Micromax and iPhone 45

Chapter 1

Introduction

Medical Imaging is a process of visual representation of the interior body parts or function of some organs or tissues which are used for clinical analysis and diagnosis. Images are generated by reconstruction of garnering measurements through advanced sensors and computer technology. There are several ways of imaging which utilize a broad spectrum of electromagnetic waves such as radio frequency in MRI; visible range in endoscopy, Optical Coherence Tomography (OCT), fundus photography; sound in ultrasound scan; X-rays in radiography, CT scans; gamma-rays in nuclear SPECT, PET imaging. These different techniques, referred to as modalities, lead to different visualizations which are used to assess the condition of an organ or tissue and used in monitoring a patient for diagnostic and treatment evaluation. Medical image processing involves the development of computational methods and algorithms to analyze, enhance and quantify medical images. This thesis focuses on image enhancement methods for retina images to make better visualization for clinicians.

1.1 Retina and Retinal Imaging

Retina is the light sensory layered membrane that lies on the inner surface of the eye (Figure 1.1). It consists of photoreceptor cells which are responsible for visual phototransduction. Photoreceptor cells in retina consist of three primary cells in eyes: rods, cones and photosensitive ganglion cell. The rods are responsible for peripheral, low light and grey vision, whereas cones produce the colour vision. The photosensitive ganglion cell plays other minor roles in human vision. The major visual acuity and colour vision occur in the small central area of the retina called macula, where cones cells are concentrated. The vision/image formed in retina is translated as electrical neural impulses to the brain using optic nerve through the optic disc (OD). It helps to create a visual perception in brain, which is analogous to that of the camera or image sensors. The OD is also the entry point for blood vessels in retina, and it corresponds to a small blind spot as there are no rods or cones cell exists. Retina is considered part of the central nervous system (CNS) and is the only region which can be visualized non-invasively in CNS.

There are a wide variety of retinal diseases and conditions, which can affect any part of the retina and can cause total blindness. The major diseases in the retina are Diabetic Retinopathy, Age-related Macular Degeneration (AMD), Cystoid Macular Edema (CME), Glaucoma and Hypertensive Retinopathy [1]. Apart from this, research have shown several major neurodegenerative disorders [38, 14, 45],

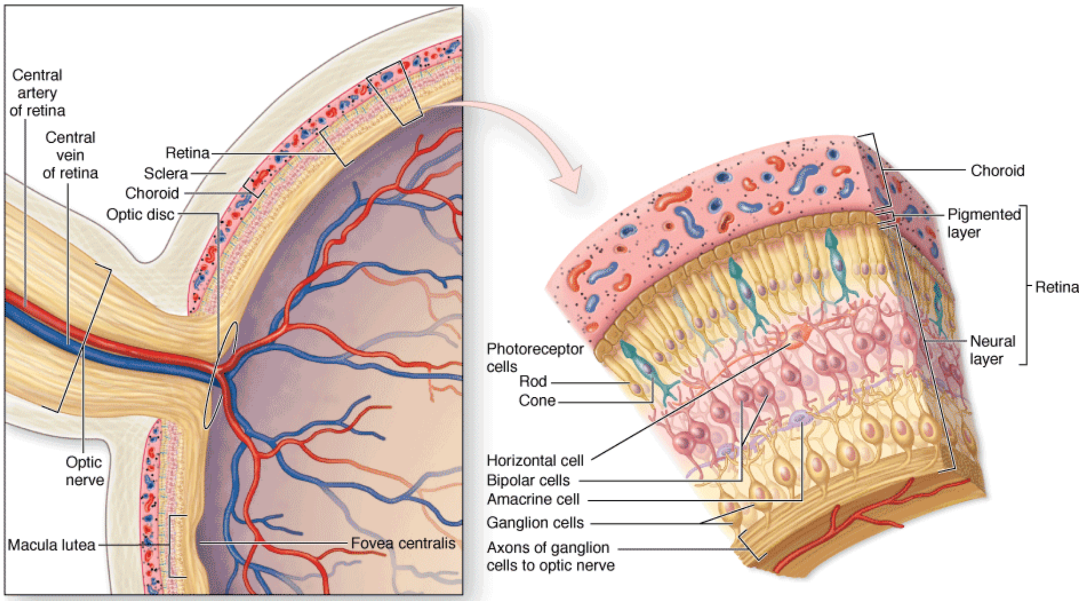


Figure 1.1: Illustration of inner anatomy of eye and retinal layers (Source: Junqueira's Basic Histology: Text and Atlas 12^{th} edition by A. Mescher)

heart disease [36, 12, 51], and chronic kidney disease [17, 70] have manifestations in the retina, suggesting that the eye is a window to the major body parts. The accessibility and the advancement in retinal imaging techniques support effective aid in non-invasive diagnosis of these diseases. To understand the cause and effect of these diseases, the visualization of retina plays an important role in analysis and diagnosis. The retinal imaging gives structural and functional information in a non-invasive manner. The accessibility of the imaging makes it convenient for research studies and can lead to the development of potential new approaches for effective treatment and screening strategies.

1.2 Retinal Imaging Modalities

The optical properties of the eye help in visualization and imaging of retina. The earliest attempt to image the retina was made by immersing live cat in the water and was unfeasible for the human eye. This lead to the invention of ophthalmoscope by J.E. Purkyn (Purkinje) [52]. Several reinvention [13, 28] of ophthalmoscope led to routine use by ophthalmologists for the inspection of retina. Figure 1.2a shows a the first image of retina published in 1853 by Van Trigt [64].

The ophthalmoscope is popular and still used, but it requires the optometrist to come closer to the face of a patient who may sometimes have infectious diseases, which motivates the photography of retina. The first photograph of the retina was captured by Gerloff in 1891, which shows blood vessels [15]. Gullstrand developed later fundus camera in 1910, and the concept is still used to image the retina [18]. Over the years several imaging modalities are developed to image the retina, among them fundus photography and Optical Coherence Tomography (OCT) are most commonly used.

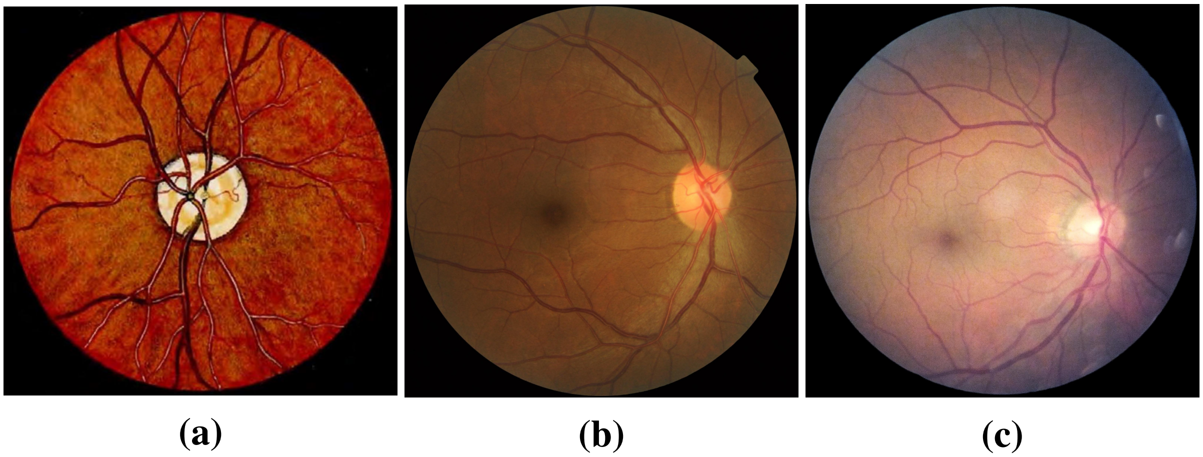


Figure 1.2: Fundus imaging. (a) First known image of human retina [64] (b) Standard fundus camera image (c) Smartphone based fundus image

1.2.1 Fundus Imaging

Fundus imaging (fundus photography) projects 3D retinal tissues onto the 2D imaging plane by the reflected light. It consists of a specialized fundus camera with complex optics of a low power microscope attachment to enable high quality and high-resolution imaging of the fundus (or retina). The intensities in colour fundus image are the amount of reflected red, blue, and green waveband by the retinal surface. Since the optical properties of the eye allow to illuminate the retina by visible light, the modality is cost-effective, non-invasive and safe. Hence it remains as a primary method of imaging retina. Fundus imaging is a popular imaging technique used to document the appearance of the vessel, optic disk, macula and retinal abnormalities such as diabetic retinopathy, glaucoma, age-related macular degeneration. It is also widely used for large scale screening purpose. An illustration of color fundus image is shown in Figure 1.2b.

Angiographic imaging is the other important retinal imaging technique which uses fundus camera with additional filters for imaging vasculature [48]. The image intensities are due to the amount of emitted photons from the fluorescein or indocyanine dye injected for circulation. This method is widely used to assess the damaged retinal vessels, but the technique is limited due to invasive nature.

Recently retinal imaging with a smartphone camera is being explored with a relatively low-cost lens attachment [39, 40, 62, 54]. A sample fundus on a smartphone device is shown in Figure 1.3b. It is becoming a more powerful clinical tool, especially in resource-constrained (medical experts and funds) settings as it offers a scalable, cost-effective solution for prevention of diseases leading to blindness [44]. Since the device is portable, it can reach a remote location and aid for teleophthalmology. A sample of colour fundus image captured from a smartphone camera is shown in Figure 1.2c.

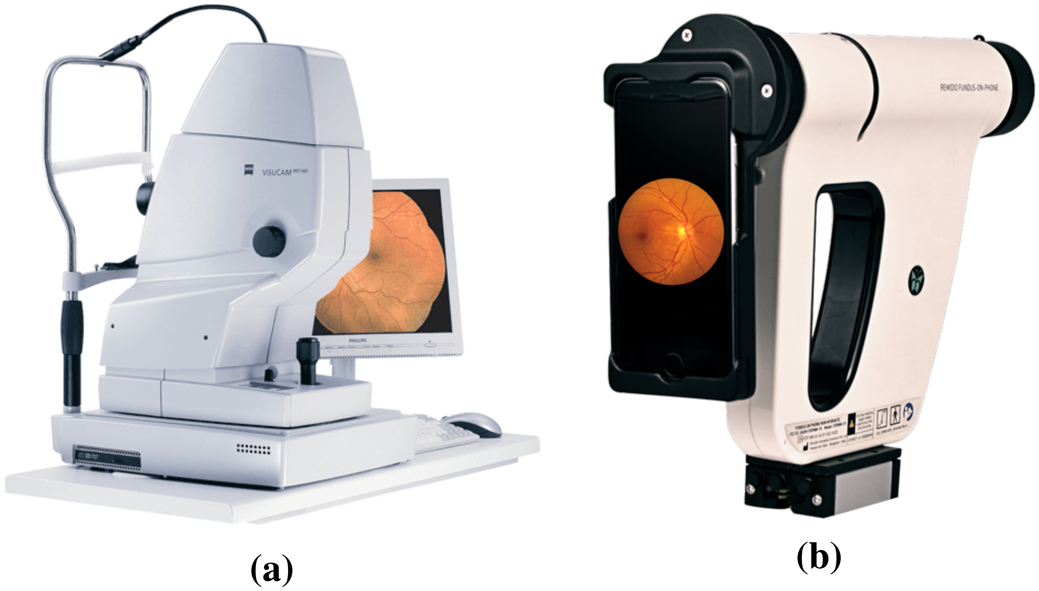


Figure 1.3: Fundus cameras. (a) Standard fundus camera (Zeiss FF450 device) (b) Fundus on Phone (a product of Remidio Innovative Solutions Pvt. Ltd.)

1.2.2 Optical Coherence Tomography

A major limitation of fundus imaging is that it captures a 2D image of 3D retinal tissue. Several methods are developed, such as Stereo fundus photography [3] and confocal scanning laser ophthalmoscopy [68] to obtain a 3D shape of retina. However, these methods result in low resolution due to limits of optics of the eye. Tomography imaging [21] of retina overcomes resolution problem and become common in clinical use as Optical Coherence Tomography (OCT) which allows 3-D cross-sectional image of retina [65].

OCT is a non-invasive imaging technique that used to take cross-section pictures of biological tissue. It is based on interferometry, uses white or low coherence light (typically near-infrared light) to capture micrometre resolution from scattering media (Ex: retina tissue, skin tissue). It works by the principle of depth estimation of particular backscatter originated by calculating the time taken to travel inside the eye. Backscatters are due to the difference in the refractive index of tissue layers. The deeper tissue takes more time for backscatter than superficial tissue. The low coherence light is optically split using beam splitter, one beam (reference arm) is directed towards the movable mirror to reflect at a particular distance, and the other beam (sample arm) is reflected from the retinal tissues. The reflected beams are recombined, and the interference energy is converted as intensity in the image using a photosensor. Different intensities represent the different depths observed from the backscatter giving depth scan (typically called A-scan). The 3D OCT image is obtained by combining A-scan of two directions (x and y axis).

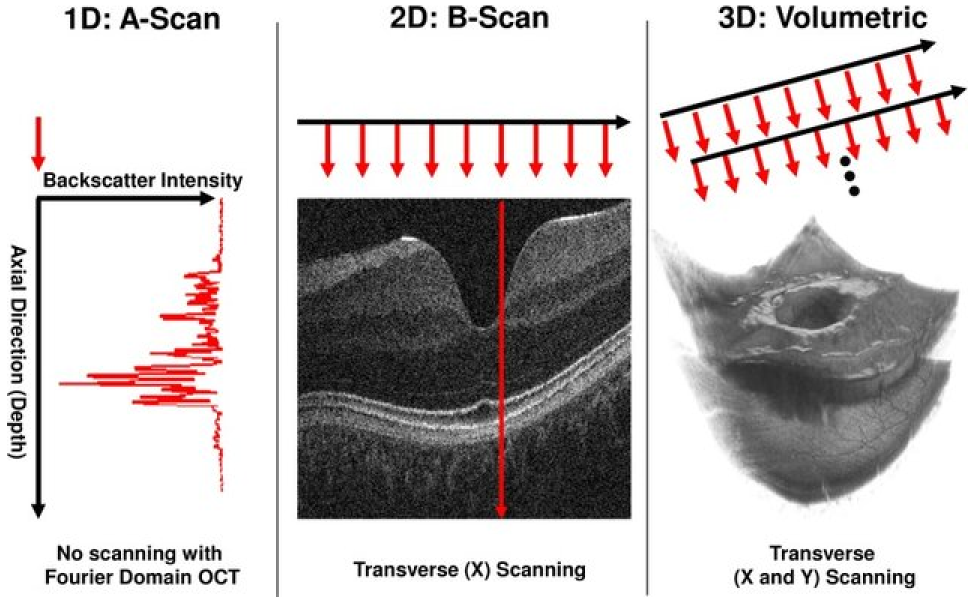


Figure 1.4: Illustration of OCT scans. Left: A depth profile (A-scan) of backscattered intensity. Middle: The beam is scanned in a transverse direction to obtain 2D imaging (B-scan). Right: Multiple B-scans are acquired to obtain 3D volume

Different methods have been developed for improving shorter A-scans interval. The initial imaging method was Time-domain OCT (TD-OCT), where the reference mirror is moved mechanically, limiting the acquisition time of A-scan. Later, Spectral-domain OCT (SD-OCT) was developed, which is based on spectrometer in the receiver. It uses the Fourier principle to estimate depth from the spectrum of reflected light on the retina. This method increases the speed of acquisition of A-scan and emerge in a single sweep of depth scan over retina linearly or circularly to obtain 2D slice or B-scan. Consequently, multiple A-scan and B-scan gives a 2D and 3D image of retina. The resolution of image depends on the number of A-scan and B-scan obtained. The OCT scanning and its coordinate system are shown in Figure. 1.4.

OCT images are widely used by an ophthalmologist for detailed imaging of retina to view distinctive retinal layers and help in diagnosis by assessing layer thickness. It is used for detection and treatment of glaucoma, age-related macular degeneration, and diabetic retinopathy. Recently, it has been used in other medical imaging to diagnose coronary artery disease and dermatology. It is also popular in industrial nondestructive testing to check the quality of the material.

1.3 Image Quality

To understand and diagnose any disease requires a good visualization of the anatomy. Optical properties of the eye is a significant advantage in imaging retina as it allows light to illuminate and capture an image in a non-invasive manner. However, image quality is degraded either due to physical properties of the light source or quality of optics and sensors used or low and uneven light condition.

1.3.1 Degradation in OCT Image

The principle of OCT image formation is based on coherence interference of reflected beams from a scattering tissue. The coherence plays both the strength and weakness of OCT. The spatial and temporal coherence of the backscattered from the tissue is measured using interferometry to obtain OCT image. Besides, the same coherence strategy gives speckle, an insidious form of noise that degrades the quality of OCT images. Hence, speckle noise occurs as a natural consequence from the scattering tissue by adding constructive and destructive interference, which are appeared as bright and dark dots in the image, as shown in Figure 1.5. Apart from OCT, speckle noise commonly arises in active radar, synthetic aperture radar (SAR), medical ultrasound imaging. In the OCT image, speckle noise reduces the contrast of the image and weaken the strength of retinal structures. It also complicates the computational systems like retinal layer segmentation, disease detection such as AMD, retinal fluid like cysts.

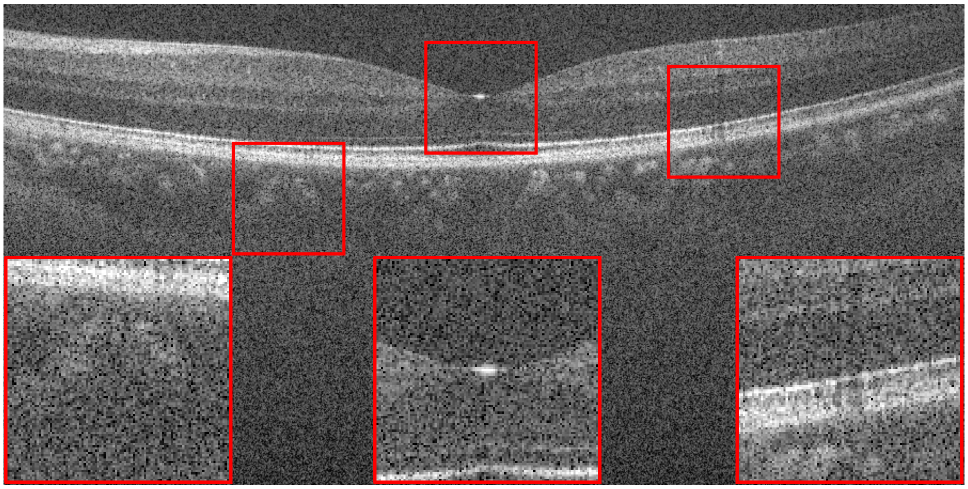


Figure 1.5: Illustration of OCT image. Inset image from left to right shows degradation of retinal structures in choroid, macula and blood vessel regions

1.3.2 Degradation in Smartphone Camera based Fundus Image

Standard fundus camera captures a retinal image by the illumination reflected from the retinal surface. It is capable of a high level of zoom due to the complex optics of a low power microscope. However, due to the curved surface of the retina, all regions cannot be illuminated uniformly; hence, most retinal images suffer from non-uniform illumination. The uneven illumination causes darker periphery in the image. This issue remains even in smartphone-based imaging. Although imaging retinal using smartphone camera is a cost-effective and offers scalability and portability, however quality of image is degraded due to uses relatively low-cost lens and sensor, LED flash. Figure 1.6 shows a sample retinal image captured by a standard fundus camera and a smartphone camera. The quality of image differs in colour, signal-to-noise, definition/detail, especially of small objects. Sometimes, images are also affected by flash/dust artefacts as shown in Figure 1.7

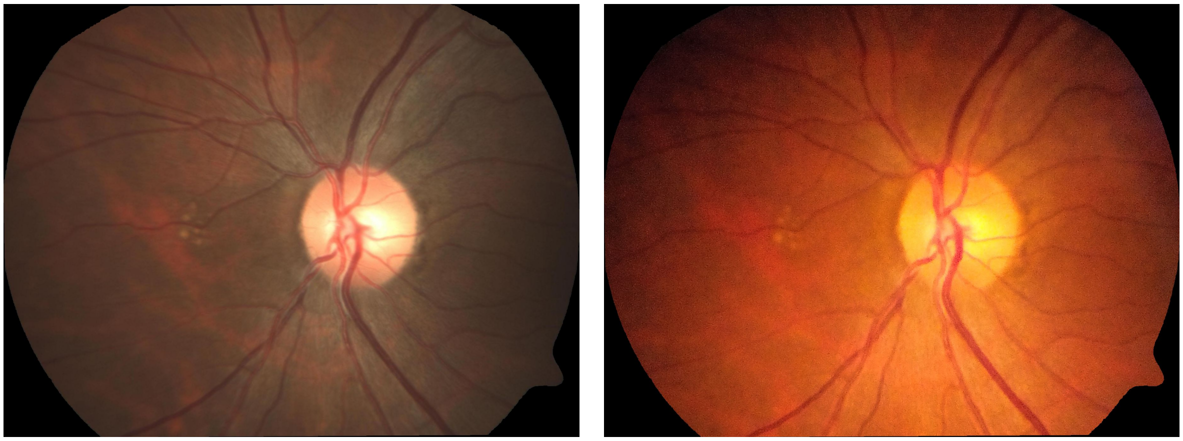


Figure 1.6: A sample retinal image captured by a standard fundus camera (left) and a smartphone camera (right)

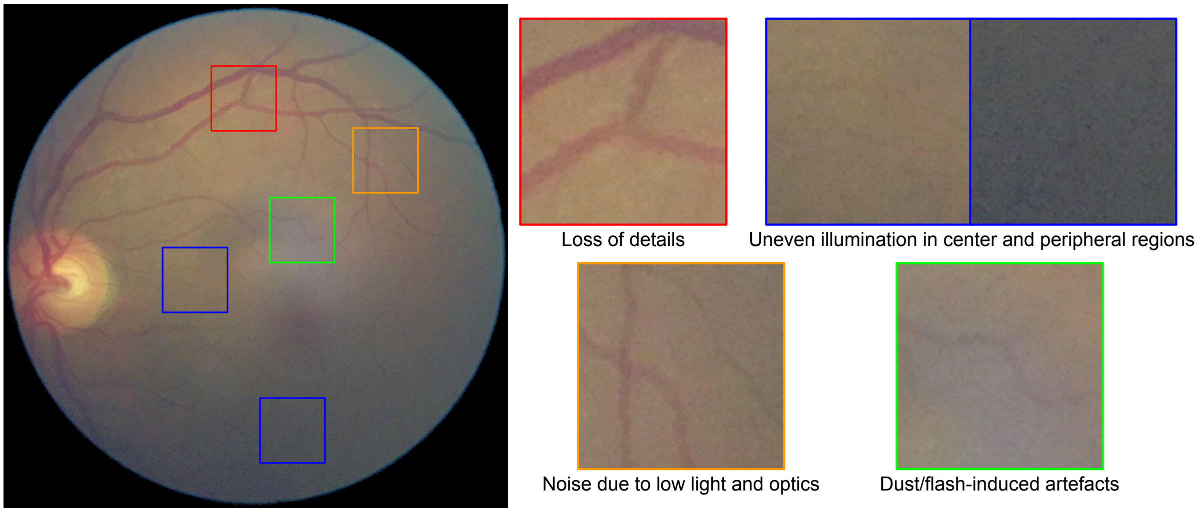

Loss of details
Uneven illumination in center and peripheral regions
Noise due to low light and optics
Dust/flash-induced artefacts

Figure 1.7: Challenges in fundus images using smartphone camera

1.3.3 Other Challenges

Since retina is externally illuminated by light source to capture retinal image, the size of the pupil matters as it complicates the imaging. If the illumination and imaging beams overlap, it results in corneal and lenticular reflections diminishing or eliminating image contrast. Hence a dilation (mydriasis) of eye is preferred before imaging. Also, motion artifacts such as saccades results in blurry or ungradable image, safety requirements limiting the amount of light that can be projected onto the retina and patient comfort limit the time per image or volume, especially in OCT.

1.4 Thesis Focus

In this thesis, we focus on image quality problems associated with two modalities of retinal imaging, namely, OCT and fundus imaging by a smartphone camera. The aim is to develop a retinal image quality improvement algorithm via learning based approaches, namely neural network techniques, while preserving the integrity of anatomical structures and with no artefacts introduced. This task is of interest in better visualisation and diagnosis by the ophthalmologist/optometrist and also in CAD systems.

OCT images are corrupted by speckle noise due to the underlying coherence-based strategy. In noisy conditions, ophthalmic experts will find it challenging to analyse the image, and it can become erroneous (especially in pathological case) with a slower throughput. Hence, speckle suppression/removal in OCT images is of interest as it plays a significant role in both manual and automatic detection of diseases, especially in early clinical diagnosis. In the first part of this thesis, we develop and validate the novel architecture for denoising OCT image using a small set of unpaired data. We show that the proposed method is robust to change in scanner and image quality and can handle both healthy and pathological cases.

Fundus imaging with a Smartphone camera is a cost-effective solution for the assessment of retina. However, imaging at high magnification and low light levels results in loss of detail, uneven illumination, noise especially in the peripheral region and flash artefacts. In this condition, reading images will require some adaptation by both manual and automatic diagnosis system. While the ophthalmic experts routinely see/read images in hospitals/clinics acquired by a standard fundus camera images, hence an adaptation is necessary. On the other hand, automatic systems are mostly developed for standard fundus camera which needs pre-processing to improve the quality. Also, imaging without mydriasis further suffers in quality of image, which is common in screening scenarios, where the reading can become much more erroneous and relatively slows down the screening process. Solving all these problems at once is very challenging and can be attempted by learning an appropriate mapping of images from smartphone camera images to standard camera images, which are presented in the next part of this thesis. We develop a robust method and a new framework for improving the quality of fundus images obtained by smartphone camera using both unsupervised and supervised techniques. We also show that the method is robust to images with/without pathologies as well as images acquired with/without mydriasis.

1.5 Contributions

The major contributions in developing an end-to-end retinal image quality improvement systems via learning based techniques, described in the thesis are:

- A novel autoencoder based architecture for denoising of Optical Coherence Tomography images
- An improved method for unpaired image-to-image translation to match the characteristics of fundus images from a smartphone camera to standard fundus camera images.
- A novel architecture to enhance the image quality along with denoising for a smartphone based fundus image in an end-to-end and supervised manner.

1.6 Outline of the Thesis

The purpose of the thesis is to improve the quality of the retinal image via learning, such that it can aid for better diagnosis by both ophthalmologist and computer-aided diagnosis (CAD) system. In this aspect, various deep learning approaches have been explored in this thesis are outlined as follows. Chapter 2 describes the traditional speckle removal methods in OCT and propose a new architecture for speckle suppression. Chapter 3 and 4 discuss challenges in quality of retinal images obtained by a smartphone camera and propose the image quality improvement methods in an unsupervised and supervised manner, respectively. Conclusions and future scope are presented in Chapter 5.

Chapter 2

Denoising of Retinal OCT Images

A Semi-supervised Approach

Optical coherence tomography (OCT) is a type of imaging modality which is commonly used by ophthalmologists to capture retina layers and thereby diagnosis of retinal diseases. This imaging technique is also gaining popularity in dermatology, cardiology, dentistry, and cancer research due to its ability of imaging in a non-invasive manner. OCT images are corrupted by speckle noise due to the underlying coherence-based strategy. Speckle noise affects the structural information and reduces contrast, thus causing difficulty in diagnosing images. Hence speckle suppression/removal in OCT images plays a significant role in both manual and automatic detection of diseases, especially in early clinical diagnosis.

2.1 Introduction

OCT imaging aids a cross-sectional 3D imaging of biological tissues. It is widely used for manual or automatic detection and diagnosis of many diseases such as diabetic macular edema (DME), age-related macular degeneration (AMD), glaucoma. It uses low-coherence light to capture micrometer-resolution. A major problem with OCT image is corruption with speckle noise. The noise is primarily due to the coherence-based strategy used in imaging where multiple scattering and phase deviations of a light beam are common [59]. It degrades the quality of images by reducing the contrast and affects the boundaries/edges of tissue structures. In addition to speckle being tissue dependent, its distribution also varies from scanner to scanner. Figure 2.1 shows the variation of speckle noise from 5 different OCT scanners (namely Bioptigen, Cirrus, Nidek, Spectralis, and Topcon) from macula center region. An efficient and effective speckle denoising method is of interest.

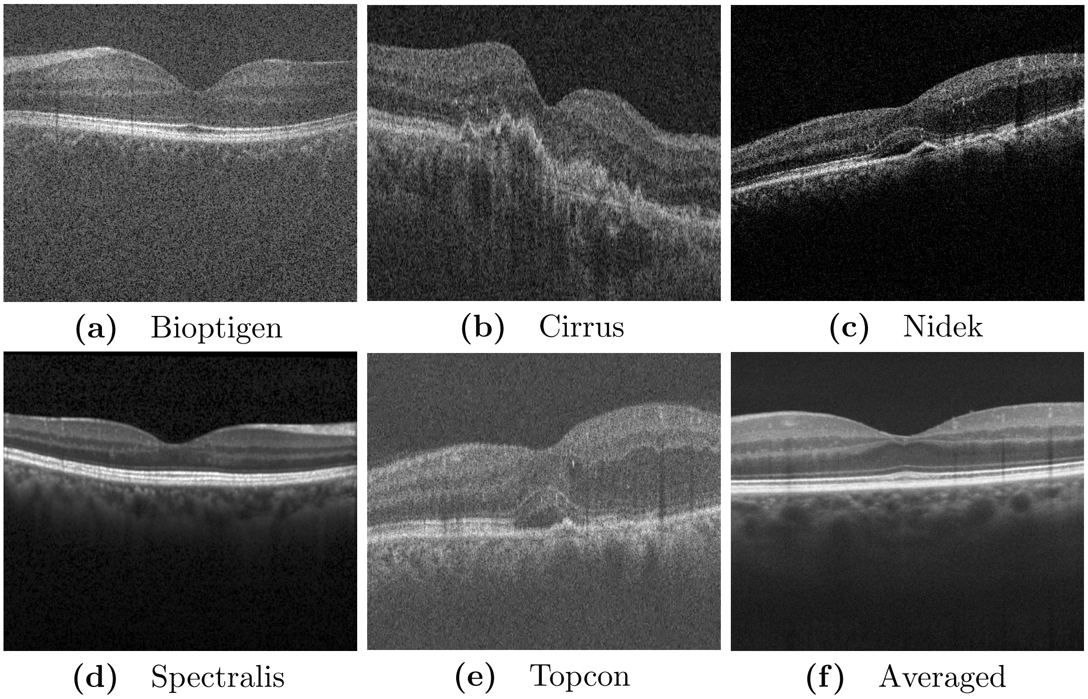

(a) Bioptigen (b) Cirrus (c) Nidek (d) Spectralis (e) Topcon (f) Averaged

Figure 2.1: OCT B-scan images from different scanner and an averaged OCT image

2.2 Related Work

There are many solutions have been proposed for speckle reduction. Early solutions were based on classic filtering with a median filter [56], Wiener filter [53, 26] and wavelets [6, 19]. These results exhibited blurring or over-smoothing and failed to preserve the sharpness of the edges in the image. A multi-frame averaging technique [58] was proposed, where the set of images are repeatedly acquired from an identical retinal position using a computerised alignment. These images are registered and averaged to create a less noisy image. This strategy increases the image acquisition time and cost. The slow acquisition is overcome in [42] by using wavelet decompositions. A smaller set of images are acquired, and its wavelet coefficients are weighted, averaged and reconstructed to obtain a denoised image. Anisotropic diffusion filtering techniques [71, 34, 2, 55] have also been explored for speckle removal. Most of these methods are sensitive to parameter initialisation with improper choice may destroy the edges in the images. The most recent variant of this approach is a diffusion potential based method [50]. This method is good for speckle removal, but with some loss of integrity of anatomical structures and edges.

Recently learning based approaches are gaining popularity. Fang et al. [11] proposed a multi-scale sparsity-based tomographic denoising, where pairs of high signal-to-noise ratio (SNR) and low-SNR slices/B-scan images are used for learning a dictionary. High-SNR images typically require slow acquisition time which implies a low throughput. This was circumvented in [10], by capturing multiple B-scans, registering and averaging them to obtain the approximate high-SNR images. A sample averaged high-SNR OCT slice is shown in Figure 2.1. This method requires a reference dataset which is obtained by custom scanning. In contrast, the dictionary learning approach in [25] is based on K-SVD using complex wavelets and is independent of high-SNR images. Here, learning is done from patches in the noisy image. Both learning and reconstruction process is lengthy.

Deep learning with Convolutional Neural Network (CNN) has also been explored for the reconstruction and denoising problem. Residual learning from paired noisy and reference images was demonstrated in [72] for natural images, is capable of denoising in the case of additive Gaussian noise with the unknown noise level. However, this method has not been extended to multiplicative noise such as speckle noise. In this chapter, we propose a CNN based solution for denoising OCT images that do not require any parallel data (i.e. a pair of noisy and reference/clean images). Instead, two autoencoders with a shared encoder are trained to reconstruct noisy and clean OCT images, respectively. The cross-model prediction gives denoised OCT images in an unsupervised way. The major strengths of proposed method are (i) a small set of unpaired noisy and clean images is used to reduce speckle noise effectively in end-to-end manner; (ii) preserves the anatomical structures with a good balance of speckle reduction; (iii) fast and robust to scanner variations and image quality; (iv) robust to both normal and pathological cases.

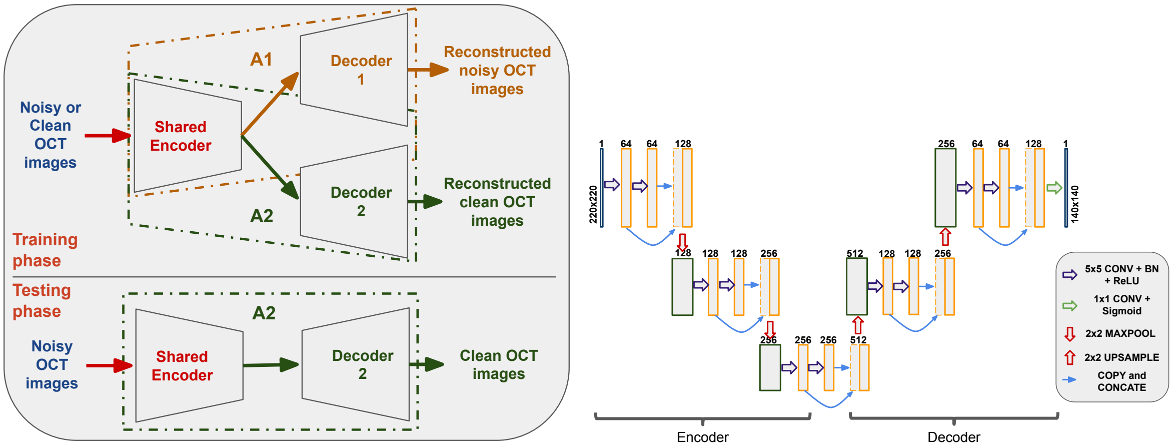
2.3 Proposed Method

The problem at hand is challenging because noise-free reference image is not readily available. Repeated scanning with the averaging or slow acquisition is not very practical as it increases imaging cost. This motivates us to explore a denoising strategy which is unsupervised and hence doesn't require any parallel data, i.e. the pair of clean/reference and noisy OCT images. A similar problem exists in machine translation where obtaining parallel corpora is resource intensive. [4, 35] propose to learn a model without any parallel data. This forms an inspiration to our proposed method which is based on autoencoders.

The pipeline of our proposed method is shown in Figure 2.2a consists of two autoencoders. Since the encoder layers are shared between the autoencoders, we call this architecture as a *Shared Encoder* (SE). In the training phase, SE is trained to reconstruct noisy and clean images with respective autoencoders. In the testing phase, given a noisy input image, a cross-model prediction is done using the decoder, which is used for clean image reconstruction.

2.3.1 Shared Encoder Architecture

The proposed SE consists of two simple autoencoders (A1, A2) consisting of a shared encoder and two decoders. One autoencoder (A1) learns to reproduce noisy OCT images while the other autoencoder (A2) learns to reproduce clean OCT images. The shared encoder of A1, A2, is trained alternatively with



(a) (b)

Figure 2.2: (a) The proposed Shared Encoder (SE) architecture (b) Schematic representation of the encoder and decoder layers in SE. Solid boxes represent multi-channel feature maps. Dashed boxes represent copied feature maps. Number of channels is denoted on the top of the box

noisy and clean images. This strategy is to help the decoder of A1 and A2 to learn to reconstruct noisy and clean OCT images, respectively. The shared encoder, on the other hand, learns the common features of the noisy and clean image using *backtranslation* technique, which is explained in section 2.2.1.

In the following, $CBR_{f,k}$ denotes the Convolution-BatchNorm-ReLU layer with f filters of kernel size $k \times k$; CC denotes the concatenation of the previous and current $CBR_{f,k}$ layer outputs; $CS_{f,k}$ denotes the Convolution-Sigmoid layer with f filters of kernel size $k \times k$; DS_k and US_k denote downsampling and upsampling by $k \times k$, respectively. The encoder-decoder architecture of the proposed SE network can be written as:

$$CBR_{64,5} \rightarrow CBR_{64,5} \rightarrow CC \rightarrow DS_2 \rightarrow CBR_{128,5} \rightarrow CBR_{128,5} \rightarrow CC \rightarrow DS_2 \rightarrow \\ CBR_{256,5} \rightarrow CBR_{256,5} \rightarrow CC \rightarrow US_2 \rightarrow CBR_{128,5} \rightarrow CBR_{128,5} \rightarrow CC \rightarrow US_2 \rightarrow \\ CBR_{64,5} \rightarrow CBR_{64,5} \rightarrow CC \rightarrow CS_{1,1}$$

The U-net [57] architecture is widely used for segmentation or restoration task. It is modified by removing the skip connections from encoding to the decoding end. The skip connections if retained will lead to the signal *and* noise to be transferred to the decoder which is inappropriate as the task at hand is denoising. However, skip connections are used between adjacent layers, which helps the network to learn better features [61]. The specific encoder-decoder architecture used is illustrated in Figure 2.2b. The encoding layer consists of the repeated two blocks of 5×5 unpadded convolutions (CONV), batch normalization (BN) [22] and rectified linear unit (ReLU). Between two blocks of CONV-BN-ReLU layer, a dropout layer [60] (with probability 0.2) is included. Dropout layer prevents over-fitting, BN layer enables faster and more stable training, ReLU layer helps to keeps the sparsity of the convolution kernel. The output of the two blocks CONV-BN-ReLU are concatenated and downsampled with a 2×2 maxpooling operation with stride 2. Decoder layer is similar to the encoder layer with one exception:

maxpooling is replaced by the upsampling layer which helps to reconstruct an output image. The final layer is a 1×1 convolution layer with a sigmoid activation function which gives the reconstructed output image. Finally, in our proposed SE architecture, two such decoder layers are combined with an encoder layer as shown in Figure 2.2a.

2.3.2 Network Training and Testing

In the training phase, the proposed SE network is first trained alternatively with noisy and clean OCT images using autoencoders A1 and A2, respectively. Since the weights of the shared layers get updated from both A1, A2, the gradients for weight update need not be similar which can lead to instability of the shared encoder. Hence, a second step is introduced in the training process, which is a *backtranslation* type of training [4] described next.

2.3.2.1 Backtranslation Training:

The *backtranslation* technique has 2 steps: (i) obtain decoded output from cross-model, forming pseudo-pair of data; (ii) train the network to reconstruct the input images using pseudo-pair of data. For example, a clean image I_c is first decoded with A1 to derive a pseudo-noisy image A1(I_c). The pair A1(I_c) and I_c is used to train A2, which learns to translate a pseudo-noisy image to a clean image. Likewise, a noisy image is decoded by A2 to derive a pseudo-clean image, which in turn is used to train A1. This aids learning common features and achieve a balanced/smoother weight update. The backtranslation technique is similar to cycle-consistency condition, except that our network has a single encoder.

The overall training procedure for SE consists of the following steps:

1. The A1 is first trained with a batch of noisy images.

2. The A2 is next trained with a batch of clean images.

3. *Backtranslation training of A1:* A2 predicts the batch of noisy images which in turn is used to train A1.

4. *Backtranslation training of A2:* Similar to step 3, A1 predicts the batch of clean images which in turn is used to train A2.

5. Repeat steps 1-4, until convergence.

The proposed SE network is trained as mentioned above in a patch-wise manner. The patch size of the input image is set to 220×220. Given that there are unpadded convolutions layers and two downsampling and upsampling layers, the reconstructed output is smaller in size, namely, 140×140. The reference patch of size 140×140 from input patch (after centring) is obtained to match reconstructed output size. A stochastic gradient descent (SGD) optimiser was used to optimise the proposed network to minimise a combination of per-pixel loss and the perceptual based loss function which is described in detail in section 2.3.

Finally in the testing phase, once the proposed SE model is trained, the decoder of the autoencoder A1 is removed. The noisy image with original size is first padded with edge value, then given to autoencoder A2 to produce a clean image of size same as given noisy image. This way, denoised OCT image will be obtained by using cross-model prediction, as shown in Figure 2.2.

2.3.3 Loss Function

The mean squared error (MSE), a reference-based metric and Peak Signal-to-Noise Ratio (PSNR) are popular error measures for denoising. In deep learning, MSE is widely used as a loss function for many applications. However, neither MSE nor PSNR correlates well with human perception of image quality. Structure similarity index (SSIM) [66] is a reference-based metric that has been developed for this purpose. The SSIM is measured at a fixed scale and may only be appropriate for a certain range of image scales. A more advanced form of SSIM is the multi-scale structure similarity index (MS-SSIM) [67]. In addition to choosing perceptually correlated metric, it is also of interest to preserve luminance and intensity. So we choose a combination of per-pixel loss and MS-SSIM to define the loss function.

Let I_y and I_x be the input and ground truth image of size $M \times N$, respectively and let $f(\theta)$ be the network mapping. The l_1 loss function is defined as:

$$L_{l_1}(\theta) = \frac{1}{NM} \sum_{j=1}^{N} \sum_{i=1}^{M} \left| f(I_{y_{i,j}}; \theta) - I_{x_{i,j}} \right| \tag{2.1}$$

A second component of the loss function is defined based on a multi-scale SSIM which is defined as:

$$\text{MS-SSIM}(I_x, I_y) = [L(I_x, I_y)^{\alpha K} \cdot \prod_{i=1}^{K} C_i(I_x, I_y)^{\beta_i} \cdot S_i(I_x, I_y)^{\gamma_i}] \tag{2.2}$$

where L, C and S denote luminance, contrast and structure of a patch in I_x and I_y. The contrast (C) and structure (S) components are measured at multiple-scale with weights β_i, γ_i and the luminance (L) component is weighted only at the coarsest scale with weight α. Number of scale is defined by K. The L, C and S are defined as follows:

$$L(I_x, I_y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}; \quad C(I_x, I_y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}; \quad S(I_x, I_y) = \frac{\sigma_{xy} + c_3}{\sigma_x + \sigma_y + c_3} \tag{2.3}$$

The variable μ_x, μ_y, σ_x, and σ_y denote mean and standard deviation of a local image patch of I_x and I_y. The variable σ_{xy} denotes covariance between a local image patch. The constants c_1, c_2, and c_3 are small constant values added for numerical stability.

In our work, all the weights $\alpha, \beta_i, \gamma_i$, for $i = 1, ..., K$ are set to 1 [73], and $K = 3$ scales are used. We can approximate the MS-SSIM loss as:

$$L_{\text{MS-SSIM}}(\theta) = 1 - \text{MS-SSIM}(I_x, f(I_y; \theta)) \tag{2.4}$$

Finally, the loss function for our autoencoders A1, A2 is defined as:

$$L(\theta) = \delta \cdot L_{\text{MS-SSIM}}(\theta) + (1 - \delta) \cdot L_{l_1}(\theta) \tag{2.5}$$

where δ denotes weight between MS-SSIM and l_1 loss. In our SE network, I_x and I_y will be the same since autoencoders A1, A2 with a common encoder are used to reconstruct (using $f(I_y;\theta)$) the noisy and clean OCT image, respectively.

2.4 Dataset and Experiments Details

2.4.1 Dataset and Parameters

Three datasets were used in our experiment for assessment of the proposed method. Dataset1 had 28 synthetic and 39 real OCT data (slices) [10] while Dataset2 had 17 real OCT data [11]. It should be noted that both synthetic and real data represent noisy OCT image. The two datasets also provide reference (with reduced noise or 'clean') images derived by averaging, details of which can be found in [10]. A third dataset considered was Chiu's dataset [7, 8] which has 220 real data (slices) from 10 volumes (11 B-scans per volume) of normal and Diabetic Macular Edema (DME) cases each. The data in the 3 datasets are captured with two different scanners (Bioptigen and Spectralis).

After removing images with the high level of blur, a training set was created with 38 synthetic images and 42 reference images from the two datasets plus 220 real images from the Chiu dataset. A test set was created exclusively from Dataset1 with 18 synthetic and 39 real images. The SE training was done patch-wise with a patch size of 220 × 220. The output image is of size 140×140 as our network has an unpadded convolution layer. Patches are selected from images with 30% overlapping resulting in a total of 2244 noisy patches and 894 reference patches. An SGD optimiser was used for training over 70 epochs with a batch size of 16 and an initial learning rate of 0.1; the momentum of 0.7 is set for both autoencoders. The learning rate is exponentially decayed by a factor of 0.00002 for autoencoder A1 and 0.00006 for A2. This differential in decay for the two autoencoders is to take care of the imbalance (3:1) in the number of noisy/reference patches. The weight δ in loss function is set to 0.85 [73]. The training of SE was done on NVIDIA GTX 1080 GPU, with 12GB of GPU RAM on a core i7 processor. The entire code was implemented in Keras library using python.

2.4.2 Experimental Setting

Denoising with the proposed model was done with different model variants using SE. These are described next:

2.4.2.1 Unsupervised Shared Encoder

This is the main scenario of our work, where training is done with unpaired clean and noisy data. The autoencoders A1 and A2 of SE are trained with patches of noisy and clean images respectively, in an unsupervised manner. All network parameter and dataset details for training are mentioned in the Dataset and Parameters section 2.4.1.

2.4.2.2 Stacked Shared Encoder

In order to study the effectiveness of stacking autoencoders, a second variant was considered. A stack of two SE's (with the same configuration) was created as shown in Figure 2.3. This is referred to as the $StackedSE$ (StSE) model. Training of SE's was done as before. First, stage 1 of StSE has trained alternatively with noisy and clean images. Then, stage 2 of StSE is trained using the noisy and clean outputs of stage 1. Image size is handled appropriately by padding border pixels. In the test phase, the denoised OCT image is obtained by cross-model prediction via stage 1 and stage 2 as shown in Figure 2.3.

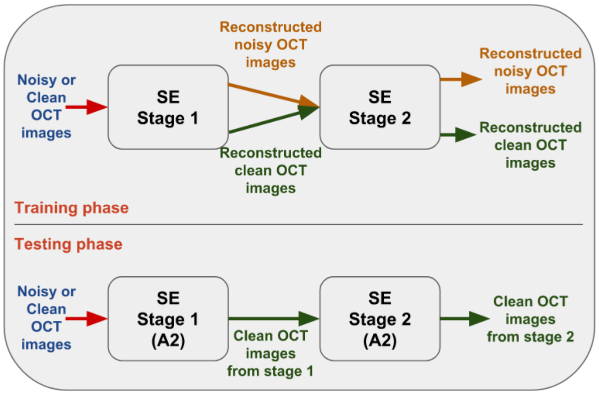

Noisy or Clean OCT images
SE Stage 1
Reconstructed noisy OCT images
Reconstructed clean OCT images
SE Stage 2
Reconstructed noisy OCT images
Reconstructed clean OCT images
Training phase
Testing phase
Noisy or Clean OCT images
SE Stage 1 (A2)
Clean OCT images from stage 1
SE Stage 2 (A2)
Clean OCT images from stage 2

Figure 2.3: The block diagram of Stacked Shared Encoder (StSE)

2.4.2.3 Semi-supervised SE

In order to study the effect of training with paired data, the trained SE was fine-tuned with paired data. In one version the paired data consisted of clean images and noisy version of the clean images. The noisy version was created by adding speckle noise of varying noise variance (0.01 to 0.1) to a clean image. We refer to this as Semi-supervised SE1 (SSE1). Since this network was fine-tuned, the learning rate was reduced to 0.01 and momentum was set to 0.9 with training over 30 epochs.

In a second version, a small set (10 OCT slices) of parallel noisy and clean/reference OCT images were used to do the final training (in a similar manner as for SSE1). These slices are excluded from the test set. This model is referred to as Semi-supervised SE2 (SSE2).

2.5 Performance Measures and Comparisons

All the SE variants were tested extensively. We first report qualitative results on the SBSDI [10] and Chiu's [7, 8] datasets. We next report results of quantitative evaluation on the test set which was described earlier. Both qualitative and quantitative results are compared with SBSDI [10] and Kafieh et al. [25] method. For convenience, we name the Kafieh et al. [25] method as CWDL (Complex Wavelet-based Dictionary Learning).

2.5.1 Qualitative Evaluation on SBSDI Dataset

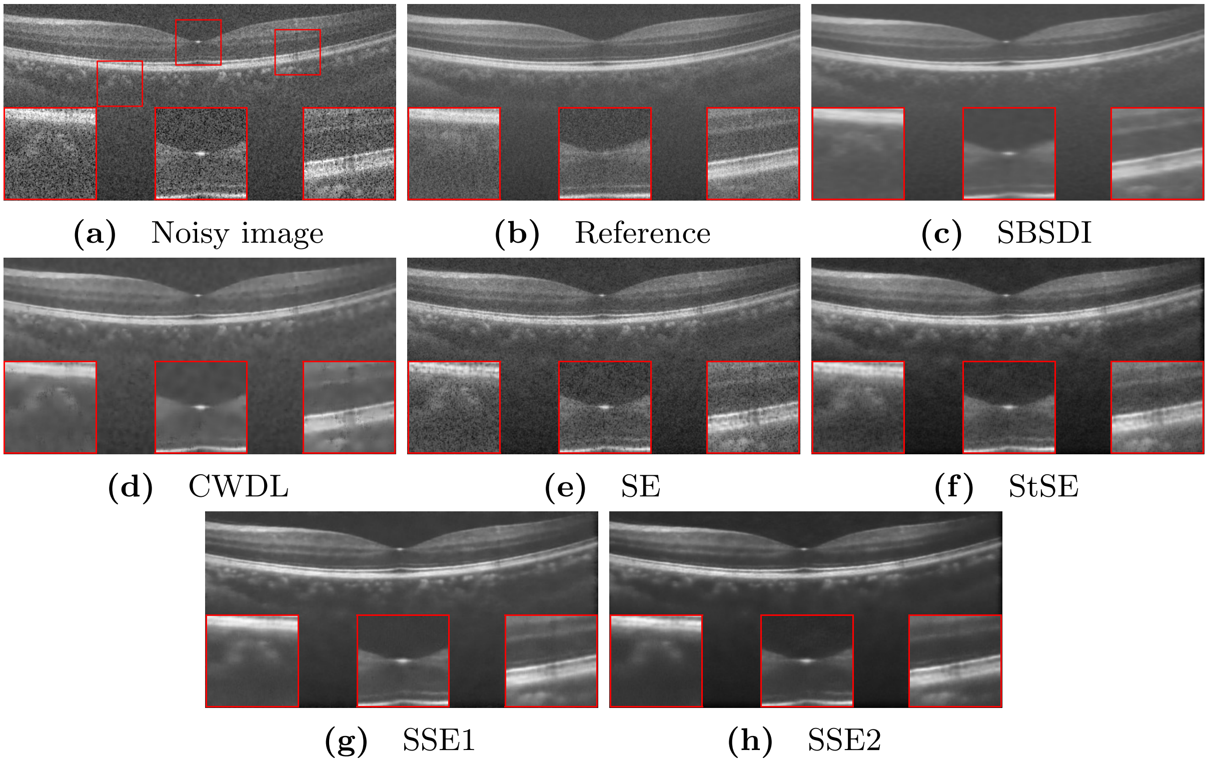


Figure 2.4: Illustration of denoising performance with different method for an example of OCT data from SBSDI [10]

The results of denoising a sample central slice/B-scan (from a Bioptigen scanner) from SBSDI [10] dataset is shown in Figure 2.4 along with results of methods reported in SBSDI and CWDL [25]. There are three insets at the bottom of each image which are zoomed versions of 3 key regions outlined in red colour, namely, the Choroid, macula and blood vessel. Figure 2.5a shows a noisy OCT example. The reference image in Figure 2.5b is obtained by registering and averaging B-scans as described in [10]. The results in Figure 2.4c shows that SBSDI method is effective in suppressing noise in the background but blurs out important structural details like blood vessels (right inset) and a layer (middle inset). The structural integrity is better in some parts of the result of CWDL. However, some artefacts can be seen in the right inset in Figure 2.4d and the layer (external limiting membrane) below the macula (middle inset) is completely blurred out. Figure 2.4e-h are results of the variants of the proposed SE. The Choroid region appears better in the results of all the variants as compared to SBSDI, with the SSE1 and SSE2 showing best clarity and integrity. Our SE and StSE results appear noisier than SBSDI and CWDL. However, the SSE variants exhibit better noise suppression while preserving the important anatomical structures.

A comparison with the latest speckle removal method [50] along with those of SBSDI, CWDL and SSE1 was also done for a sample with pathology. The relevant results are shown for a sample slice from the SBSDI dataset, in Figure 2.5. The inset in each image shows the zoomed region of normal anatomy (right inset) and lesion region (left inset). Figure 2.5a shows a noisy OCT example from SBSDI dataset used in [50]. The reference image in Figure 2.5b is obtained by registering and averaging B-scans as described in [10]. Figure 2.5c shows SBSDI method has good denoising performance, but some thin layers are not clearly visible. It can be seen in the inset image, where a thin retinal layer (external limiting membrane) is not clearly visible, and edges of layers are not sharper. Figure 2.5d shows CWDL method has little effect on noise removal and a lot of structural information such as retinal layers and their edge information are lost.

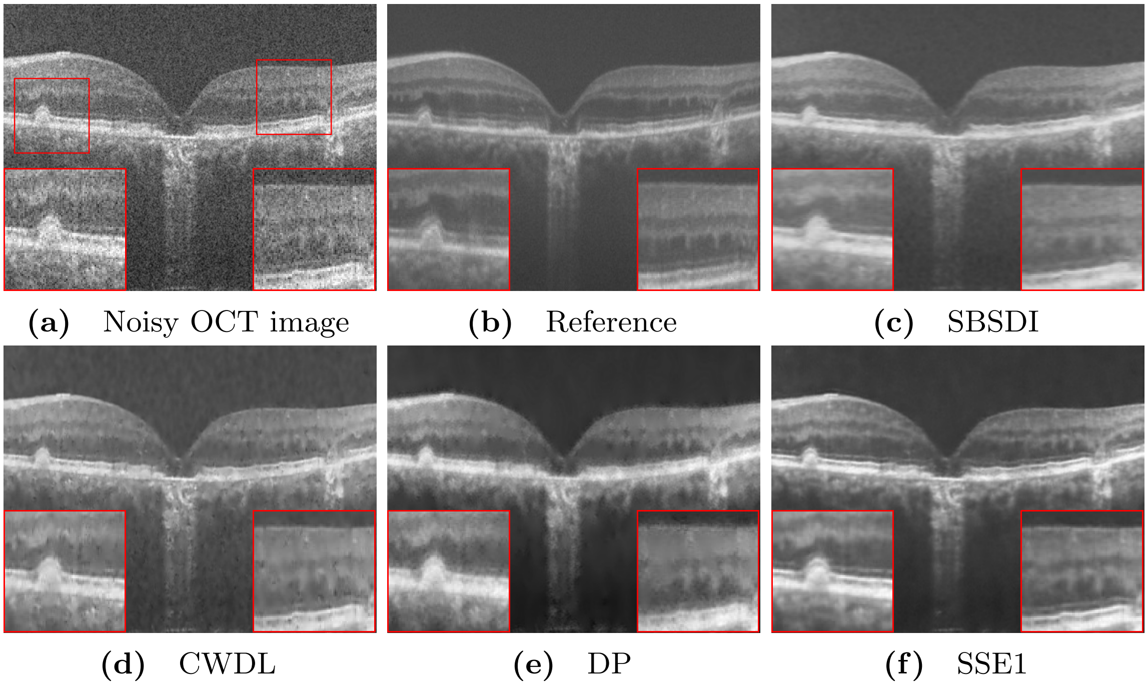

(a) Noisy OCT image
(b) Reference
(c) SBSDI
(d) CWDL
(e) DP
(f) SSE1

Figure 2.5: A OCT denoising performance comparison with DP [50] method

Figure 2.5e shows results of a recent method (taken from [50]) based on diffusion potential (for convenience, we name this as DP method). It can be seen that it reduces speckle well, but the retinal layer boundaries are blurred. All the images in Figure 2.5 are reduced in size to match the scale of the image in [50]. Figure 2.5f shows the result of our best method SSE1, which is seen to be effective at noise removal and maintain the structural information. Further, from the inset images, it can also be seen that the thin OCT layer (external limiting membrane) and lesion (drusen) region are sharper.

2.5.2 Qualitative Evaluation on Chiu's Dataset

We assessed denoising on Chiu's [7, 8] dataset which has images with and without pathologies. This dataset does not provide a reference as it was initially meant for benchmarking segmentation algorithms. The denoising results of a *normal* peripheral B-scan from Chiu's dataset is shown in Figure 2.6. In this figure, the left inset is a zoomed region with two thin blood vessels (vertical black lines), and the right inset shows a zoomed region with a thick blood vessel. The results in Figure 2.6b-g indicate the following: all methods reduce noise with SE retaining some noise in the result. On a closer look, it appears that denoising by existing methods tends to blur fine details (thin layers, blood vessels). This is not the case with the variants of the proposed SE. This tendency is particularity clear when the left and right inset images are compared across the methods.

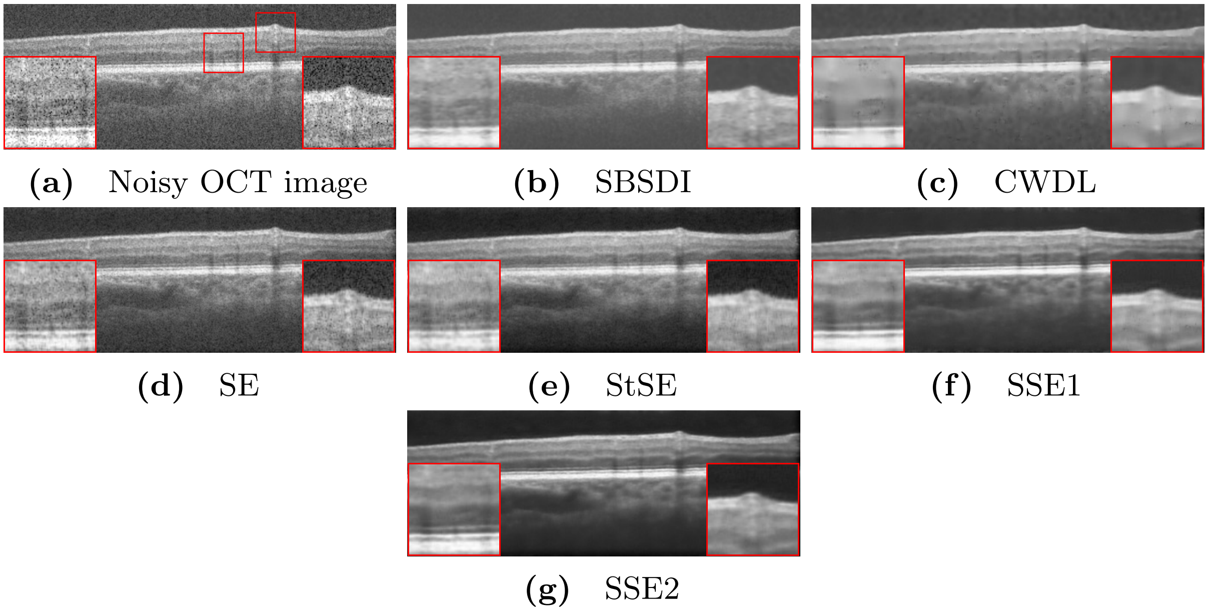

(a) Noisy OCT image
(b) SBSDI
(c) CWDL
(d) SE
(e) StSE
(f) SSE1
(g) SSE2

Figure 2.6: Illustration of denoising performance with different method on Chiu's normal OCT data [7]

SBSDI [10] method performs fairly good in noise removal in the background region, and little effect of noise removal in layer region and also some structural information is lost like the thin blood vessel, and the thin layer is not clearly visible in the left inset image. Figure 2.6c shows that CWDL [25] method, a lot of structural information such as retinal layers is lost. Figure 2.6d-g illustrates the results of the proposed SE methods. Our SE and StSE results are noisier, but it retains good structural information, and it is improved using SSE1 and SSE2 model which gives improved denoised results. In the left inset image, we can clearly see the thin blood vessel and retinal layers in SSE1, SSE2 than any

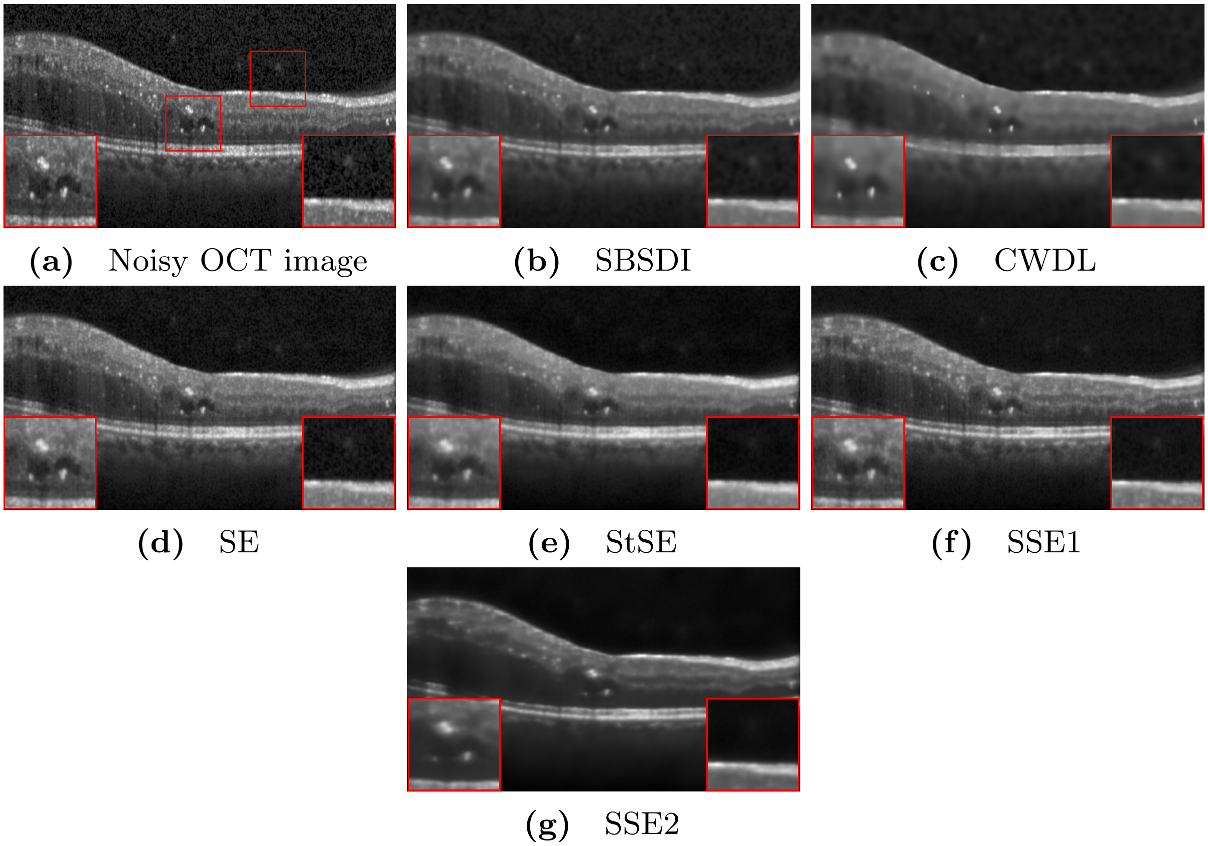


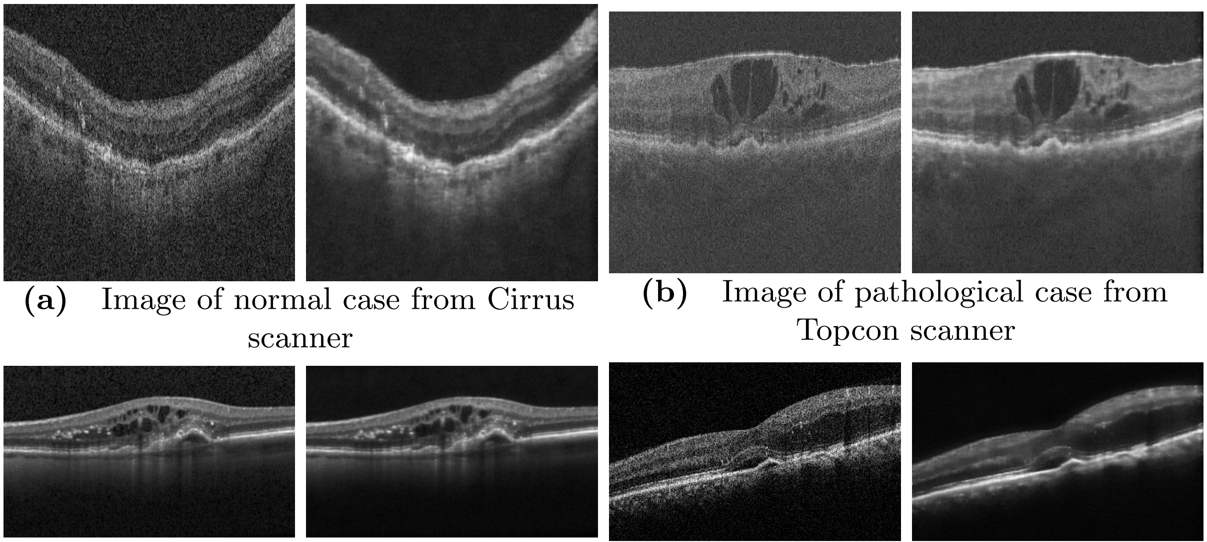
Figure 2.7: Illustration of denoising performance with different method on Chiu's DME OCT data [8]

other method. Also, there is a thin layer (external limiting membrane) which is very clearly perceivable in SSE1 and SSE2. We can see SSE1 and SSE2 performance of noise removal both in the background and upper retinal layers is better than other methods in the right inset image.

The denoising results for a B-scan with *pathology* is shown in Figure 2.7. Here, the left and right inset images show a zoomed view of regions with retinal fluid and lesion (bright spot), respectively. The results in Figure 2.7b-c are from existing methods. Both show a similar trend of reduction of noise at the cost of loss of important details, as in *normal* cases. The images in Figure 2.7d-g are results of the variants of the proposed SE. Of all the variants, SSE1 method is seen to be the best as there is a good balance of smoothing out the noise and retaining edges and details.

2.5.3 Denoising Performance with Scanner Variability

It is well known that OCT images suffer from speckle noise, but its distribution varies across scanners. Robustness of a denoising method to this variability is an indicator of its versatility. The denoising results from our best method SSE1 for sample OCT scans with and without pathologies from different scanners is shown in Figure 2.8. All the images shown are from *OPTIMA Cyst segmentation challenge,*



(a) Image of normal case from Cirrus scanner

(b) Image of pathological case from Topcon scanner

(c) Image of pathological case from Spectralis scanner

(d) Image of pathological case from Nidek scanner

Figure 2.8: Illustration of denoising results on OCT images from different scanner. Column 1 is noisy image and Column 2 is corresponding denoised image from our best method SSE1

MICCAI 2015[1] and they are different from the type of scanner used in training data (except image from Spectralis scanner). The proposed method exhibits stable performance for most of the scanners, showing its generalisation ability or adaptability. If a denoising solution is of interest for only one scanner type, it is also possible to further improve these results by tailoring the solution or training the SE with scanner-specific data.

2.5.4 Quantitative Evaluation

A quantitative assessment of the proposed denoising solution was done using PSNR, SSIM and the mean-to-standard-deviation ratio (MSR). The PSNR evaluates the quality of the reconstructed image, while the SSIM evaluates the luminance, contrast, and structural component of the images with respect to a reference image and the MSR evaluates the quality of noise removal using local statistics. The MSR is calculated at each location within a window using: $MSR_w = \frac{\mu}{\sigma}$, where μ and σ are mean and standard deviation of the intensity values inside the window respectively. The final MSR value is the average value of MSR_w for all the windows in the images. Higher MSR values indicate smoother the image.

The PSNR, MSR, and SSIM values were computed for 18 synthetic and 39 real images from the test set which was described earlier in section 2.4.1. The average values of these metrics are reported in Table 2.1 and 2.2, respectively for different denoising methods. The performance of SSE1 and SSE2 variants are on par with SBSDI. SSE2 is better than SBSDI and CWDL in terms of MSR. The PSNR

[1] `https://optima.meduniwien.ac.at/research/challenges/`

and MSR were developed for natural images and hence do not correlate well with human perception of image quality, unlike the SSIM. The SSIM values are low for all methods/models since OCT images have a very low percentage of structures and it is noteworthy that our method has similar SSIM values as SBSDI (best). Despite the quantitative results of our methods being inferior for synthetic data, it is not so for real data as it has the highest SSIM. In our experience, the medical experts appear to prefer noisy structures for reading/diagnosis over an over-smoothed one. This was corroborated by the results of a perceptual study conducted: two clinicians were asked to choose among the results of our method, SBSDI and CWDL on 20 images. Our results were preferred over others in 62.5% cases suggesting that our solution gives a good balance between denoising and preserving anatomical structures.

The PSNR value reported in latest speckle removal method [50] for the same set of synthetic dataset is 29.5 ± 2.9 dB and for real dataset is 26.2 ± 2.4 dB, which is $\sim$2 dB higher for synthetic and $\sim$3 dB higher for real dataset compare to our best method SSE1. However, our method performs better in preserving the anatomical structures.

	PSNR	MSR	SSIM
Noisy image	17.74 ± 0.45	2.52 ± 0.19	0.0867 ± 0.0197
SBSDI [10]	**28.31 ± 2.57**	24.79 ± 2.11	**0.6892 ± 0.0301**
CWDL [25]	27.50 ± 2.45	14.76 ± 0.99	0.6561 ± 0.0368
SE	21.81 ± 1.12	5.02 ± 0.28	0.3332 ± 0.0333
StSE	20.27 ± 1.22	6.89 ± 0.27	0.5126 ± 0.0328
SSE1	27.53 ± 2.22	23.38 ± 2.12	0.6775 ± 0.0322
SSE2	26.14 ± 1.66	**31.31 ± 2.90**	0.6847 ± 0.0277

Table 2.1: Quantitative comparison of performance on the synthetic dataset

	PSNR	MSR	SSIM
Noisy image	18.97 ± 0.25	2.54 ± 0.06	0.3946 ± 0.0110
SBSDI [10]	**24.28 ± 1.00**	31.03 ± 1.88	0.3147 ± 0.0257
CWDL [25]	24.07 ± 0.98	18.39 ± 0.92	0.3233 ± 0.0247
SE	20.48 ± 1.17	5.27 ± 0.09	**0.4521 ± 0.0170**
StSE	19.08 ± 0.89	7.49 ± 0.21	0.3604 ± 0.0193
SSE1	23.40 ± 1.19	28.57 ± 1.47	0.3139 ± 0.0245
SSE2	22.40 ± 1.11	**37.36 ± 1.96**	0.2923 ± 0.0243

Table 2.2: Quantitative comparison of performance on the real dataset

2.6 Conclusions

We presented a CNN-based *Shared Encoder* (SE) architecture for OCT denoising using an unsupervised learning. To the best of our knowledge, this is the first work which shows a lot of potential for OCT denoising using CNN. The entire SE is trained using a small set of unpaired noisy and clean images to reduce speckle noise effectively. A key strength of the proposed method is it preserves the integrity of anatomical structures with a good balance of smoothing out the speckle noise, a feature which is very essential in clinical diagnosis. Results demonstrate that the method is robust to change in scanner and image quality and is able to handle both normal and pathological cases. Our method can also be used as a preprocessing stage for the automatic segmentation/detection system, which has to be explored in the future.

Chapter 3

Matching the characteristics of Fundus and Smartphone camera images

An Unsupervised Approach

Fundus images are used for imaging retina by the ophthalmologists to diagnose retinal diseases, with diabetic retinopathy being a major example. Imaging of retina by a Smartphone camera offers a cost-effective solution for the assessment of retina. However, imaging at high magnification and low light levels results in loss of details, uneven illumination, noise especially in the peripheral region and flash-induced artefacts. It causes difficulty in both manual and automatic diagnosis. Hence, it calls for an image quality enhancement to use the smartphone-based retina imaging in the diagnostic centre.

3.1 Introduction

A fundus camera (FC) is a digital camera capable of high level of zoom due to the complex optics of a low power microscope at the front end. Thus, enabling high quality and high-resolution imaging of the fundus (or retina). It is therefore expensive and bulky. Recently, the smartphone camera (SC) has been explored for retinal imaging with a relatively low-cost lens attachment [54, 5]. This innovation has two significant advantages: much lower cost and a high degree of portability. However, even without a special lens, natural images of a scene captured by an SC (e.g. iPhone) and a standard DSLR camera (e.g. Canon) can differ as seen in Figure 3.1. In addition to a colour shift, there is a loss of definition/detail of small objects in the iPhone images. Imaging of the retina is even more challenging: calls for capturing a 45° field of view (FOV) of the retina (spanning 132.32 sq. mm [33]) with an SC with a special lens, under illumination of a LED-based flash. This limits the ability to capture fine details such as capillaries.

Challenges in SC images includes (i) noise due to low light conditions and CMOS sensors; (ii) uneven illumination, with typically darker periphery due to the curved retinal structure; (ii) dust/flash-induced artefacts; (iv) variable image quality depending on camera specification of the mobile device. Both (i) and (ii) are acute in non-mydriatic imaging conditions.

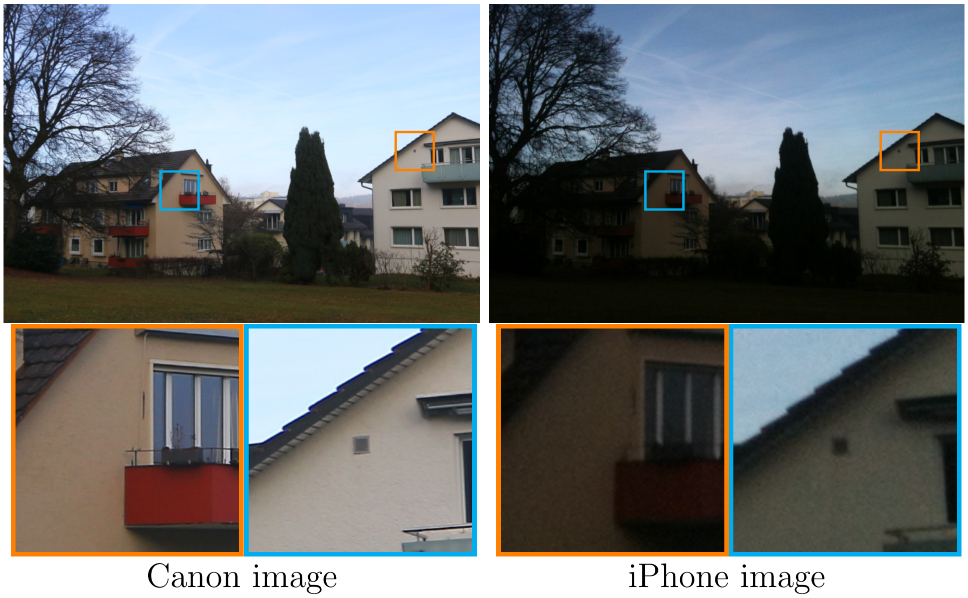


Figure 3.1: Comparison of images of a scene taken by a standard camera (left) and an SC (right)

Ophthalmic experts routinely see/read images in hospitals/clinics acquired by an FC. Hence, reading images acquired with an SC in screening scenarios will require some adaptation, without which screening can become erroneous with a slower throughput. Matching the standards/quality of the images from SC and FC is a solution. Standard image enhancement approaches proposed for FC images [24, 74] are inappropriate for this task, given the complex sources of problems in SC images. Kohler et al. [32] offer a solution to improve retinal image acquired with a custom-designed, low-cost camera with an adaptive and incremental frame averaging. Imperfect alignment of the frames blurs the image, and hence registration is done before averaging which increases the acquisition time. To our knowledge, the matching SC image standards to FC remains an open problem. Standardizing SC-sourced retinal images is essential for adaptability and reliable diagnosis.

In this chapter, we propose a mapping solution to transform the SC retinal images (henceforth just referred to as SC images) such that its characteristics are closer or similar to those of FC images. The mapping will aim to preserve the integrity of structural details and introduce no artefacts. Noise removal is not within the scope of this work.

3.2 Method

The SC image requires illumination correction, structure enhancement (such as vessels, optic disk (OD), lesions) and flash artefact suppression for better clinical and automatic diagnosis. Further, it is also desired to match its characteristics to that of an FC image to facilitate experts who are used to reading FC images. Solving all these problems at once is very challenging and can be attempted by learning an appropriate mapping from SC to FC image. The problem at hand is similar to image-to-image translation [23] which relies on paired image data. In the medical domain, acquisition of paired data is very challenging. Hence, the need is to learn image-to-image translation *without* paired data. Among the many solutions proposed for unsupervised image-to-image translation [29, 69, 37], the CycleGAN [75] has shown excellent results and hence, is taken as a source of inspiration for the proposed method.

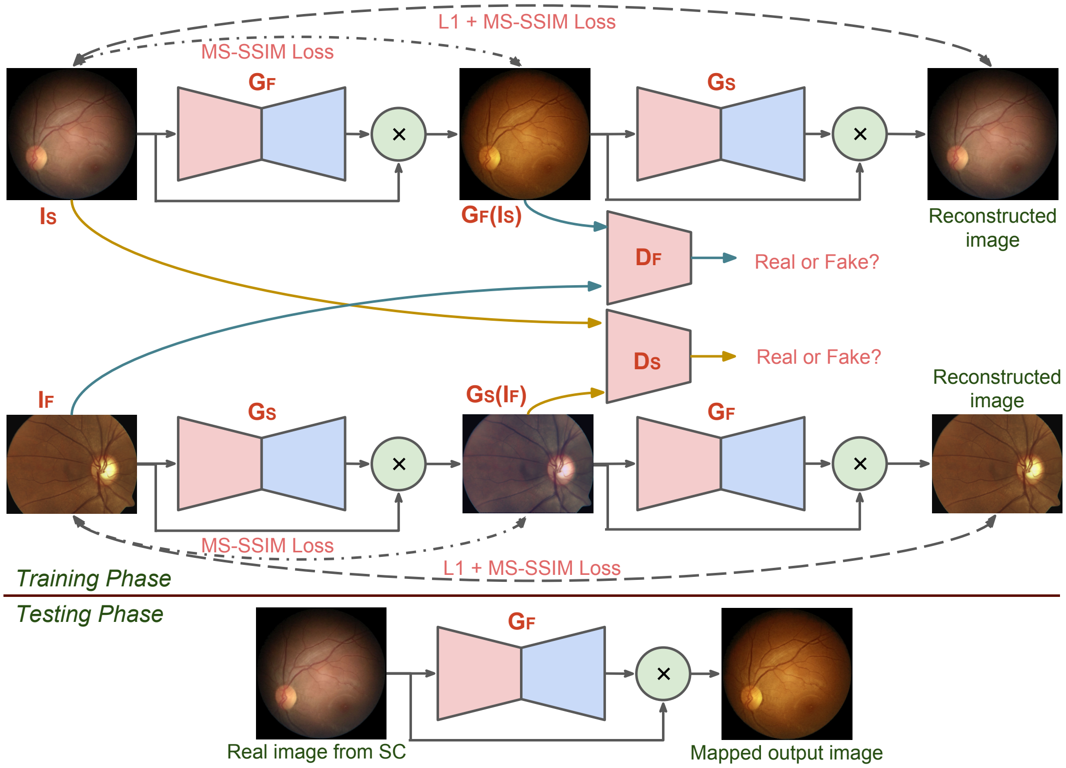


Figure 3.2: Schematic representation of the proposed architecture. The red and blue color trapezoid represent encoder and decoder layer, respectively. The green color circle represent residual connection

3.2.1 Proposed Architecture

Our aim is to learn mapping functions between SC and FC images (more compactly referred to as S and F respectively) in an unsupervised manner. We pose the problem as learning a gain matrix (K) for a given input image (I_i) to get a mapped output (I_o).

$$I_o(\mathbf{r}) = K(\mathbf{r}) \circ I_i(\mathbf{r}); \tag{3.1}$$

where $\mathbf{r}$ is a spatial vector.

The CycleGAN [75] learns to map an image from a source to the target domain with the two domains being quite different, for example, horse $\leftrightarrow$ zebra, winter $\leftrightarrow$ summer, etc. In our problem, the source and target domain are same (retina), and the aim is to only change the characteristics of an image without losing any structural details. Thus, the CycleGAN is modified by introducing a residual connection between the generator from input to the output end and network learns the required gain matrix (K). The proposed architecture is called as ResCycleGAN, is as shown in Figure 3.2. It consists of two generators G_F and G_S, which learn the mapping from S to F and F to S, respectively. Besides, two discriminators D_S and D_F learn to distinguish between real/fake S and F images, respectively. The ResCycleGAN is trained to minimise an objective function made of three terms: an adversarial loss [16], a cycle-consistency loss, and an identity loss. These are described next.

3.2.2 Loss functions

3.2.2.1 Adversarial Loss

The adversarial loss generally serves to match the distribution of the generated output with the reference image. Here, it is used match the characteristics of SC to FC domain. This loss is applied to both the generator G_F and G_S. A least-squares function [41] is used for adversarial loss for stable training and generating high-quality results. The adversarial loss for the generator G_F and its corresponding discriminator D_F is given as

$$\mathcal{L}_{GAN}(G_F, D_F) = D_F(G_F(I_S))^2 + (1 - D_F(I_F))^2 \tag{3.2}$$

where I_S and I_F denote *unpaired* SC and FC images. In the training phase, G_F tries to generate an image $G_F(I_S)$ close to real FC image, while D_F tries to distinguish between the generated image $G_F(I_S)$ and real sample I_F. G_F aims to minimize this loss against an adversary D_F that tries to maximize it, i.e. $\min_{G_F}\max_{D_F}\mathcal{L}_{GAN}(G_F, D_F)$. Similarly an adversarial loss for generator G_S and its discriminator D_S are also defined, i.e $\min_{G_S}\max_{D_S}\mathcal{L}_{GAN}(G_S, D_S)$.

3.2.2.2 Cycle-Consistency Loss

The adversarial loss is insufficient to guarantee that the learned mapping is to a target distribution as it can map to any random permutation of images in the target domain. Hence, Cycle-Consistency loss is used, which measure the reconstruction capability of the network. i.e. The reconstructed images from $G_S(G_F(I_S))$ and $G_F(G_S(I_F))$ are needs to be identical to their inputs I_S and I_F. The l_1 or l_2 norm is a popular choice for the loss function in a reconstruction problem, but they do not correlate well with the human perception, which is critical in our application as the end user can be a medical expert. The multi-scale, structure similarity index (MS-SSIM) [67] based loss addresses this issue while handling the variations in scale. Hence, we define the cycle-consistent loss function as a combination of l_1 norm and MS-SSIM and define it as follows

$$\begin{aligned}\mathcal{L}_{cycle}(G_F, G_S) = {} & \delta_1 \cdot \mathcal{L}_{\text{MS}}(G_S(G_F(I_S)), I_S) + (1 - \delta_1) \cdot \mathcal{L}_{l_1}(G_S(G_F(I_S)), I_S) \\ & + \delta_2 \cdot \mathcal{L}_{\text{MS}}(G_F(G_S(I_F)), I_F) + (1 - \delta_2) \cdot \mathcal{L}_{l_1}(G_F(G_S(I_F)), I_F)\end{aligned} \tag{3.3}$$

where $\mathcal{L}_{l_1}$ and $\mathcal{L}_{\text{MS}}$ are standard l_1 norm and MS-SSIM metric, respectively. The weights are set to $\delta_1 = \delta_2 = 0.85$ as per [73] and MS-SSIM is computed over three scales.

3.2.2.3 Identity Loss

This loss generally helps preserve colour composition between the input and generated images, whereas, in the application at hand, the colour palette is camera-dependent. The generator has to learn a mapping to either SC or FC fundus images while preserving the integrity of anatomical structures. Hence, a structure similarity function (or MS-SSIM) is suitable for identity loss. This is defined as

$$\mathcal{L}_{ss}(G_F, G_S) = \mathcal{L}_{\text{MS}}(G_F(I_S), I_S) + \mathcal{L}_{\text{MS}}(G_S(I_F), I_F) \tag{3.4}$$

MS-SSIM is once again computed over three scales.

3.2.2.4 Overall training Loss

The overall training loss for the network is defined as a combination of the three losses as

$$\begin{aligned} \mathcal{L}(G_F, G_S, D_F, D_S) = {} & \mathcal{L}_{GAN}(G_F, D_F) + \mathcal{L}_{GAN}(G_S, D_S) \\ & + \lambda_1 \cdot \mathcal{L}_{cycle}(G_F, G_S) + \lambda_2 \cdot \mathcal{L}_{ss}(G_F, G_S) \end{aligned} \tag{3.5}$$

where λ_1 and λ_2 are weights for the loss terms.

3.3 Implementation and Dataset Details

3.3.1 Network Architecture

The architecture of our ResCycleGAN is adopted from CycleGAN [75]. The encoding layer in the generator has 4 blocks of 4×4 convolution (CONV) of stride 2 followed by LeakyReLU activation and Instance Normalization [63]. The decoding layer has blocks of 4×4 CONV of stride $\frac{1}{2}$, followed by ReLU activation and Instance Normalization. Skip connections were used from encoding to decoding layer for blocks having the same size. The final layer combined the decoded feature map with a 4×4 CONV with ReLU. The input and the final CONV layer are multiplied to derive the generator output as shown in Figure 3.2. The final CONV layer learns the correction required for SC image to match to FC image. The discriminator network has layers similar to the encoding layer, followed by a 4×4 CONV with ReLU.

3.3.2 Training Details

The ResCycleGAN was trained to minimize the objective function $\mathcal{L}$ (Eq. 3.5) by alternatively updating $G_{F/S}$ with fixed $D_{F/S}$ and vice versa. The network was trained with patches of size 256×256

after normalisation to a range of [0,1]. The weights are set to $\lambda_1 = 10$ and $\lambda_2 = 1$. The optimisation was with an Adam solver [30] with an initial learning rate of 0.0002 and batch size of 1. The network was trained for 200000 iterations. The entire code was implemented in Keras library using python and executed on NVIDIA GTX 1080 GPU with 12GB RAM on a core i7 processor. In the testing phase, only the generator G_F is used. The SC image with the original size is given to the generator G_F to produce a mapped image (with characteristics similar to the FC images) is derived as shown in Figure 3.2.

3.3.3 Dataset Details

265 FC images acquired (with mydriasis) with a Zeiss FF450 Plus camera were obtained from the authors of a Diabetic Retinopathy study [54]. A total of 540 SC images, the majority without mydriasis, were obtained from the *Fundus on Phone* (a product of Remidio Innovative Solutions Pvt. Ltd.) at $45°$ FOV using iPhone 6. Both SC and FC images included pathological cases and were of varying quality. A 50% split was done to form the training and testing datasets for SC images. All FC images were used for training the network.

3.4 Performance Analysis

Both qualitative and quantitative evaluation of the proposed ResCycleGAN was done. A quantitative assessment was done using two metrics: Q_v score [31] and the Bhattacharyya distance D_b for comparing the characteristics (histograms) of mapped and FC image.

3.4.1 Qualitative Assessment

Sample original SC images (first column) and their mapped results (last column) are shown in Figure 3.3 along with magnified views of two sub-regions per image (middle two columns). The ResCycleGAN results (whole as well as sub-regions) in Row 1 indicate an improvement in contrast of structures such as OD and vessels as well as a reduction in bluish LED noise in the periphery. The horizontally oriented very thin vessels within OD and thin, dull vessels are distinguishable from the background in the magnified results. Similarly, the mapping is seen to improve the lesion (hard exudate in top and microaneurysm in the bottom sub-image) contrast in Row 2, which can be seen in the magnified image. Overall, the mapping is seen to change the colour profile and produce a balanced illumination and contrast.

3.4.2 Comparison with CycleGAN [75]

In order to assess the effectiveness of the modification done to a CycleGAN, two mappings were generated: one with CycleGAN (trained with the same setting as ResCycleGAN) and the other with proposed ResCycleGAN. Two sample results are shown in Figure 3.4. The images shown are cases of

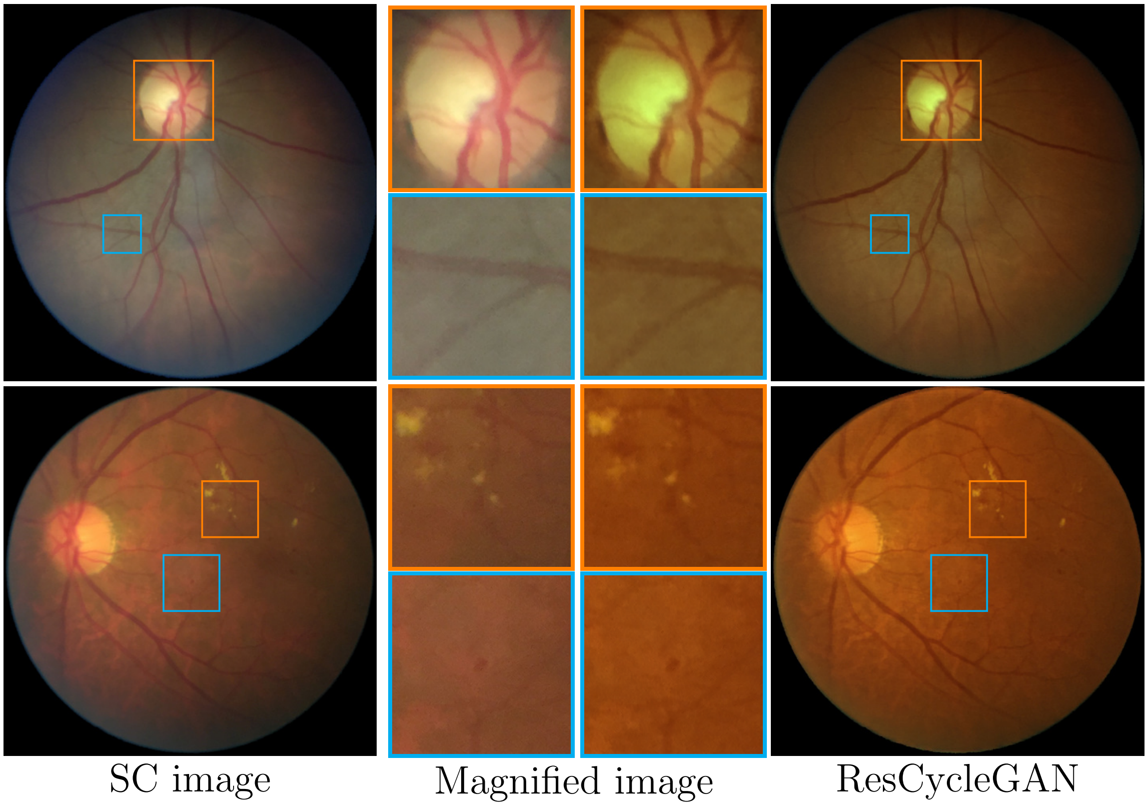


Figure 3.3: Sample results for ResCycleGAN for images without (top) and with pathologies (bottom)

imaging with/without (top/bottom) mydriasis. The tissue background in CycleGAN results look more synthetic (Row 1) with heavy smoothing of the background erasing vessel, vessel reflections; the OD is also saturated. In the second example in Row 2, the CycleGAN produces a completely uncommon palette with optic cup disappearing, which is unacceptable. The result of ResCycleGAN on the other hand has structural details with a balanced illumination and contrast. The CycleGAN was trained for 400000 iteration which is twice the number of iterations for the ResCycleGAN. The shorter training for the latter is due to the residual connection which helps in learning.

3.4.3 Quantitative Evaluation

A quantitative assessment is challenging when no reference image is available. To make a meaningful evaluation of the mapped results, we use a metric to assess the vessel quality (Q_v score [31]) and a metric to assess the similarity (D_b Bhattacharyya distance) between the mapped results (denoted as O) and FC images. Higher Q_v values indicate better quality in terms of noise and blur. This score was computed for 270 test images and is presented in Table 3.1. The similarity is assessed by computing D_b between colour (HSI space) histograms. Average histograms were computed over 270 SC images, their mapped outputs and 265 FC images. $D_b(FC, X)$; $X = SC$ or O, is computed for the average histogram pairs and reported separately for the chromatic (C: H and S) and achromatic (AC: I) components in Table 3.1.

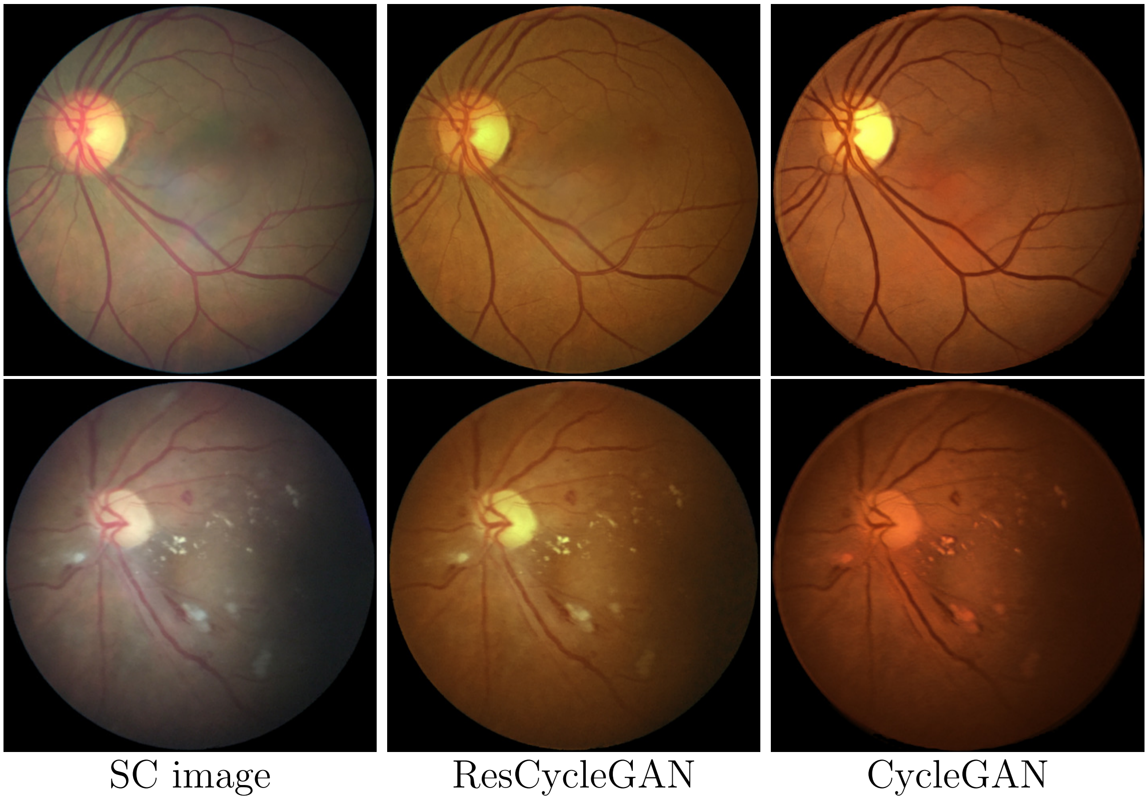


Figure 3.4: Comparison of ResCycleGAN with CycleGAN outputs for images without (Row 1) and with pathologies (Row 2)

The results indicate that ResCycleGAN outperforms CycleGAN in both Q_v (the difference is statistically significant as $p < 0.05$) and D_b values. This implies the mapping improves vessel contrast while attaining a good match with FC characteristics. Further, the match in characteristics is superior for both AC and C components.

3.4.4 Comparison against Standard Retinal Image Enhancement Method [74]

Finally, we present a comparison with a recently reported unsupervised enhancement method for retinal images [74]. Sample images (without mydriasis) along with the processed results are shown in Figure 3.5. Since [74] essentially stretches luminosity and contrast, it leads to a heightened contrast and luminosity (last column) in the results without a colour shift. However, an unwanted bluish peripheral artefact is seen in the results. In contrast, our results (middle column) exhibit an overall balanced improvement.

	Q_v score	D_b (C / AC)
SC images	0.0189 ± 0.0104	0.1656 / 0.0883
CycleGAN [75]	0.0263 ± 0.0143	0.0058 / 0.0288
ResCycleGAN	**0.0334 ± 0.0175**	**0.0014 / 0.0166**

Table 3.1: Quantitative comparison of performance using Q_v and D_b on SC images

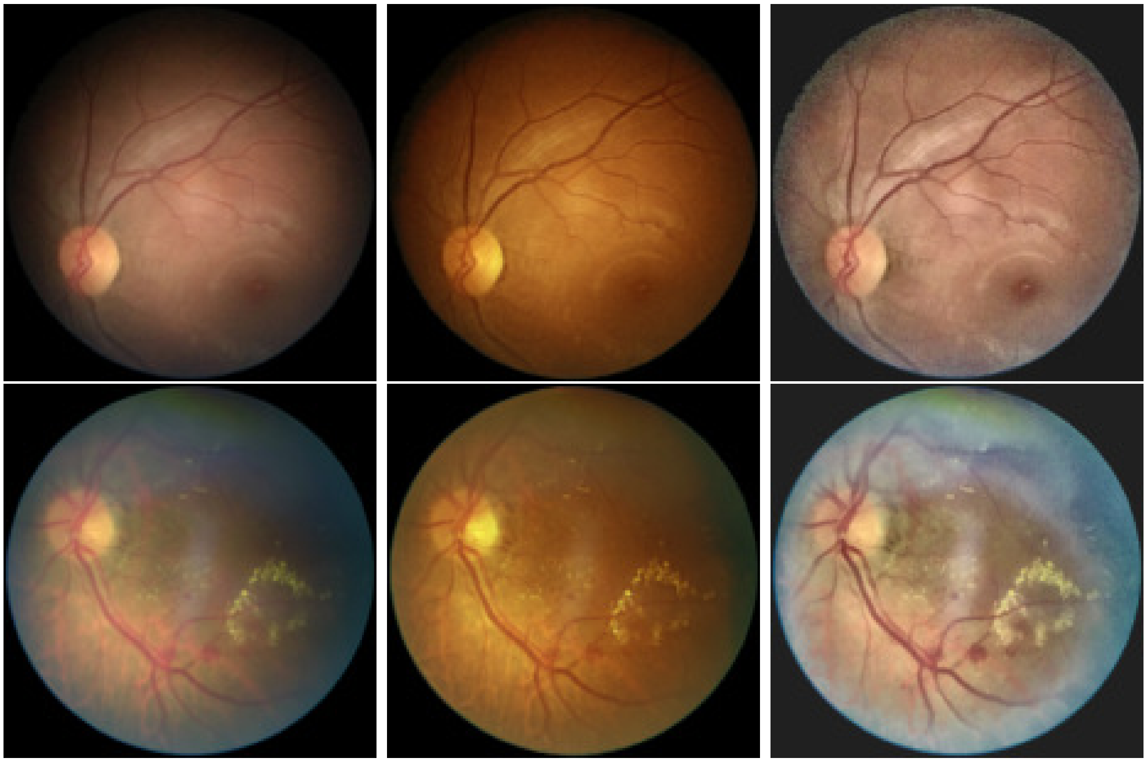


Figure 3.5: Comparison of standard retinal image enhancement with the proposed mapping. Left to right: SC image, results of our method and enhancement [74]

3.5 Conclusion

A ResCycleGAN solution was proposed to match the characteristics of SC images to mydriatic FC images successfully using an unsupervised learning. To the best of our knowledge, this is the first attempt to do such a mapping. The key strengths of our method are: it preserves the integrity of structures with a balanced illumination correction between the peripheral and centre region with no introduction of artefacts; the results are consistently good for images with/without pathologies as well as images acquired with/without mydriasis. Hence, our solution can aid ophthalmic experts; fast processing requiring 5.2 sec/image. One can also explore the method's use as a preprocessing stage for adapting CAD systems developed for FC images.

Chapter 4

Enhancement of Smartphone camera based Fundus Images

A Supervised Approach

In the previous chapter, an unpaired image-to-image mapping solution was proposed to enhance the quality of Smartphone-based fundus image using generative adversarial networks. The method maps the Smartphone Camera (SC) images to the standard Fundus Camera (FC) images such that the characteristics of SC images are closer/similar to those of FC images. However, noise removal was not handled. Noise in image cause difficulty for both manual and automatic diagnosis, especially in small lesion detection such as microaneurysms. It is the earliest indication of Diabetic Retinopathy (DR). Hence a noise removal along with other corrections such as illumination correction, structural enhancement and flash artefacts suppression is of interest.

4.1 Introduction

The conventional fundus camera imaging remains the gold standard method for screening retinal disease. Recent innovation of fundus imaging using a smartphone camera is more practical alternatives for screening a large number of people in resource-constrained (retina experts and funds) setting [54]. Also, it enables teleophthalmology due to the portable feature of SC device. The challenges lie in the quality of image due to the low-cost sensor, optics and low-light levels resulting in loss of details, uneven illumination, noise especially in the peripheral region and flash-induced artefacts. The ResCycleGAN proposed in chapter 3 provide a solution to most of these problems in an unsupervised manner. However, noise removal was not part of the work due to design constraints. In presence of noise, diagnosis can become laborious as well as erroneous as it is difficult to identify small structure and lesion such as microaneurysm which is the earliest clinical evidence of DR. Solution can be developed to handle noise with a pair of SC and FC images, by learning an appropriate mapping from SC to FC image. Obtaining paired data is challenging, especially in the medical domain, however in our case, the pair of SC and FC images can be acquired by imaging the same subject from both SC and FC device, but at the cost of multiple imaging and alignment issues. In a fixed clinical setting or commercial product, the pair data can be helpful in designing a better algorithm to improve the quality image, such that SC device can be adopted in clinics. Since acquiring such data is one-time setup and retina imaging is non-invasive, so it is feasible to obtain a small population of pair of SC and FC images.

In this chapter, we propose a new architecture to map SC images to FC images using the pair data. The mapping will aim to preserve the integrity of structural details along with effective noise and flash artefact suppression, balanced illumination and introduce no artefacts. We also show the model adaptability for cross-validation set from different SC device, which is not part of the training set and its robustness for the images with and without pathologies.

4.2 Method

The SC image needs illumination correction, denoising, structural enhancement (such as vessels, optic disk (OD), lesions) and flash artefact suppression for better visualization by the ophthalmic experts and effective automated diagnostic system. The problem at hand can be solved by learning an appropriate mapping from SC to FC image. The Convolutional Neural Network (CNN) is popular among learning-based methods for classification, detection and image-to-image translation problems. The problem at hand is similar to image-to-image translation, but it heavily relies on paired image data. For better enhancement, particularly in fixed clinical settings or commercial product, it is feasible and affordable to acquire a pair data such that appropriate image-to-image translation can be used to map SC to FC image. U-net [57] is simple and popular among many image-to-image translation methods, which forms an inspiration for the proposed method.

4.2.1 Proposed architecture

Our aim is to learn a mapping function from SC to FC images in a supervised manner. The U-net is widely used for image segmentation or restoration task, where it learns to map an image from one domain to other/same domain. The problem at hand is complex, as it involves multiple tasks such as denoising, illumination correction, artefact suppression, and enhancement of structure. Hence we pose the problem as learning a noise (N) and gain matrix (K) separately, for a given input image (I_i) to get a mapped output (I_o) as shown

$$I_o(\mathbf{r}) = K(\mathbf{r}) \circ (I_i(\mathbf{r}) - N(\mathbf{r})); \tag{4.1}$$

where **r** is a spatial vector. The noise matrix (N) will learn the correction needed for noise removal, whereas gain matrix (K) will learn the other corrections such as illumination correction, artefact suppression, and enhancement of structure. Learning separately to denoise and other correction helps in simplifying the task and thereby aid in better and stable results.

The pipeline of proposed method is shown in Figure 4.1. The method enhances the quality of SC images by learning the mapping function in an end-to-end, supervised manner. Hence we call our architecture as $SupEnh$. The architecture consists of an encoder (E) and two decoders (D_1, D_2). The D_1

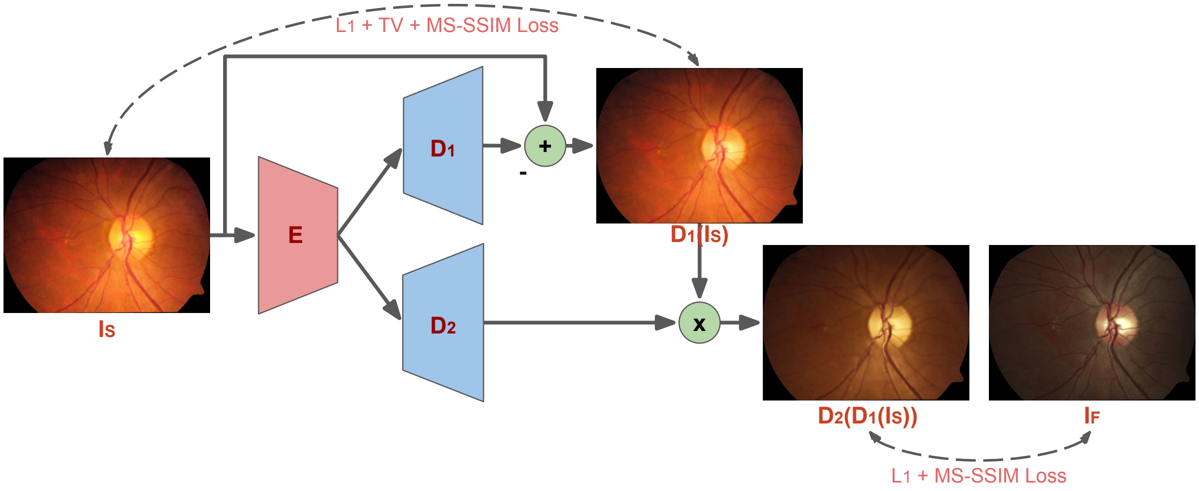


Figure 4.1: Schematic of the proposed architecture

and D_2 learns the noise matrix N and gain matrix K, respectively, whereas encoder E learns common features (such as structures) for two decoders. In our problem, the source and target domains are same (retina); hence, a residual connection is employed to learn only the required change in characteristics of an image without losing any structural details. This connection is used from the input of encoder to the output of decoders, as shown in Figure 4.1. The noise matrix N learned by decoder D_1 is subtracted from the input to obtain a denoised image, followed by it is multiplied with the gain matrix K to obtain mapped output image.

4.2.2 Loss function

The proposed SupEnh method is designed to learn noise and gain matrices independently using decoder D_1 and D_2. Hence we train the network to minimise two objective functions at output of D_1 and D_2. The FC images are of a high quality due to superior optics, sensor, and flashlight are used. Hence it serves as reference images for training decoder D_2. The loss function at D_2 measure the mapped SC image ($D_2(D_1(I_S))$) with the corresponding FC image (I_F). The l_1 or l_2 norm is a popular choice for the loss function, but they do not correlate well with the human perception. It is addressed using a multi-scale structure similarity index (MS-SSIM) [67] based loss, which handles the variations in scale as well. In addition to MS-SSIM, it is also of interest to map the luminance and contrast of SC to FC images. Hence, we define the loss function as a combination of l_1 norm and MS-SSIM loss as

$$\mathcal{L}_{D_2}(I_S, I_F) = \delta \cdot \mathcal{L}_{l_1}(D_2(D_1(I_S)), I_F) + (1-\delta) \cdot \mathcal{L}_{\text{MS-SSIM}}(D_2(D_1(I_S)), I_F) \tag{4.2}$$

where I_S and I_F denote *paired* SC and FC images, respectively and $\mathcal{L}_{l_1}$ and $\mathcal{L}_{\text{MS-SSIM}}$ are standard l_1 norm and MS-SSIM metric, respectively. The weight is set to δ = 0.85 as per [73] and MS-SSIM is computed over three scales.

For training D_1, obtaining only a noise-free reference pair is not feasible. Hence we define a loss function independent of FC images at D_1 and is trained to suppress noise in SC image in an unsupervised manner. Gaussian and median filters are simple techniques for denoising, and it is commonly used, but it smoothes the edges. Total variation loss overcomes this issue by preserving the edges while denoising. In addition to this, a l_1 norm is used to reconstruct the SC images, and an MS-SSIM metric is used to preserve the structures. The combined loss function at the output of decoder D_1 is defined as

$$\mathcal{L}_{D_1}(I_S) = \mathcal{L}_{l_1}(D_1(I_S), I_S) + \lambda_1 \cdot \mathcal{L}_{TV}(D_1(I_S)) + \lambda_2 \cdot \mathcal{L}_{\text{MS-SSIM}}(D_1(I_S), I_S) \tag{4.3}$$

where $\mathcal{L}_{TV}$ is standard total variation loss function. The weights are empirically set to λ_1 = 0.1 and λ_2 = 0.01 and MS-SSIM is once again computed over three scales.

4.3 Implementation and Dataset Details

4.3.1 Network Architecture

The encoder and decoder layers of our SupEnh is adopted from U-net [57]. Our architecture consists an encoder and two decoders and the skip connections are employed for both the decoders from encoding end. Also, skip connections are used between adjacent convolution layers, which helps the network to learn better features [61]. The encoding layer consists of the repeated two blocks of 5×5 unpadded convolutions (CONV), batch normalization (BN) [22] and Exponential Linear Unit (ELU) [9]. Between two blocks of CONV-BN-ELU layers, a dropout layer [60] is included, with rate 0.2. Dropout layer prevents over-fitting, BN layer enables faster and more stable training, ELU layer helps in faster learning by pushing the mean activation towards zero. The output of the two blocks CONV-BN-ELU are added and downsampled with a 2×2 maxpooling operation with stride 2. Decoder layers are similar to the encoder layer with one exception: maxpooling is replaced by the upsampling layer which helps to reconstruct an output image. The final layer is a 1×1 convolution layer with a tanh activation in decoder D_1 and sigmoid activation in decoder D_2 which gives the noise and gain matrices, respectively. Finally, the input, encoder E and decoders D_1 and D_2 are connected as shown in Figure 4.1.

4.3.2 Training Details

The SupEnh was trained to minimise two objective functions $\mathcal{L}_{D_1}$ and $\mathcal{L}_{D_2}$ simultaneously. The network was trained with patches of size 256×256 after normalisation to a range of [0,1]. A stochastic gradient descent optimisation was used with an initial learning rate of 0.1 and a momentum of 0.85. The learning rate is exponentially decayed by a factor of 0.0001, and network was trained for 50 epochs with a batch size of 8. The entire code was implemented in Keras library and executed on NVIDIA GTX 1080 GPU with 12GB RAM. In the testing phase, the SC image with the original size is given to the SupEnh network to produce a noise suppressed and enhanced image from D_1 and D_2 end, respectively.

4.3.3 Dataset Description

The SupEnh model is trained using a dataset obtained from the authors of a Diabetic Retinopathy study [54]. 208 subjects (out of 301) are acquired (with mydriasis) containing pairs of FC and SC images (each subject having at least two field of view (FOV) for each eye). After removing ungradable and non-overlapping images, 628 pair of images are obtained. The FC images acquired with a Zeiss FF450 Plus camera whereas SC images were obtained from the *Fundus on Phone (FOP)* (a product of Remidio Innovative Solutions Pvt. Ltd.) at $45°$ FOV using Micromax smartphone. Both SC and FC images included pathological cases and were of varying quality. Another test dataset consists of only SC images, majority without mydriasis, were obtained from FOP using iPhone, which are used for cross-validation of the model. The training and testing split of datasets are tabulated in Tabel 4.1.

Datasets	Training	Test
FC images	420	208
SC images (Micromax)	420	208
SC images (iPhone)	-	270

Table 4.1: Dataset description

4.3.4 Data Preparation

The pair of FC and SC images obtained by two different devices having a different resolution, protocols and also there are subject's eye movements while imaging, hence images need to be aligned. As the SC images are noisy and overlap between SC and FC images are less in some case; thus, a simple automated alignment algorithm will not work. Hence, we developed a semi-automated tool using MATLAB to align the images using Coherent Point Drift (CPD) [46] algorithm. First, we convert SC and FC images into vessel tree using multiscale line detector [47]. Then, overlapping regions are marked (bounding box) manually in both SC and FC images. These marked regions are used to aligning the images using the CPD algorithm.

4.4 Results

Both qualitative and quantitative evaluation of the proposed method was compared with U-net and its variants. A quantitative assessment was done using full reference and no reference metrics. We use Mean Squared Error (MSE), peak signal to noise ratio (PSNR), SSIM, and MS-SSIM metrics for full reference metric whereas for no reference metrics Q_v score [31] was used. We also assess the performance of enhancement of image in Computer-aided diagnosis (CAD) setting, using an existing red lesion detection algorithm [49].

4.4.1 Qualitative analysis

The results of SupEnh mapped images (middle column) for sample SC images (first column), and their corresponding FC images (last column) are shown in Figure 4.2 along with two sub-regions at the bottom of each image, which are magnified versions of regions outlined in blue and red colour. The SupEnh results as a whole, as well as sub-regions in Row 1, indicate an improvement in contrast of structures such as vessels and OD region as well as a reduction in noise and bluish flash artefacts. The horizontally oriented thin vessels with circular flash artefacts, noisy and dull vessels are distinguishable from the background in the magnified results. Similarly, the mapping is seen to improve the lesion (microaneurysm and hard exudate in the blue and red colour sub-image, respectively) contrast in Row 2. The algorithm can improve only the content present in SC images, whereas FC images show much more details due to the superior device. Overall, the mapping is seen to change the colour profile, reduces the noise and produce a balanced illumination and contrast within and across the image.

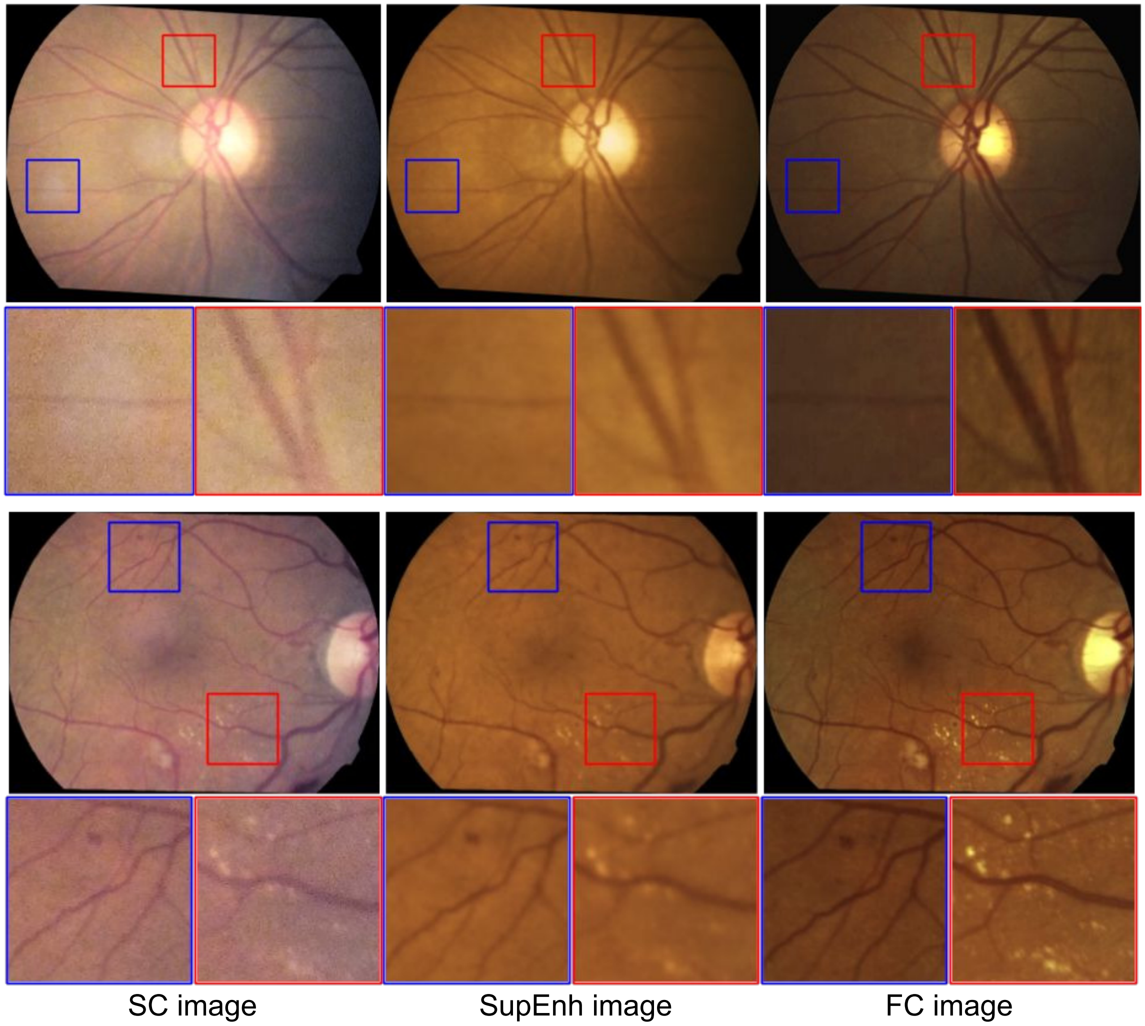


Figure 4.2: Sample results for SupEnh for images without (top) and with pathologies (bottom)

4.4.1.1 Comparison with U-net and ResCycleGAN

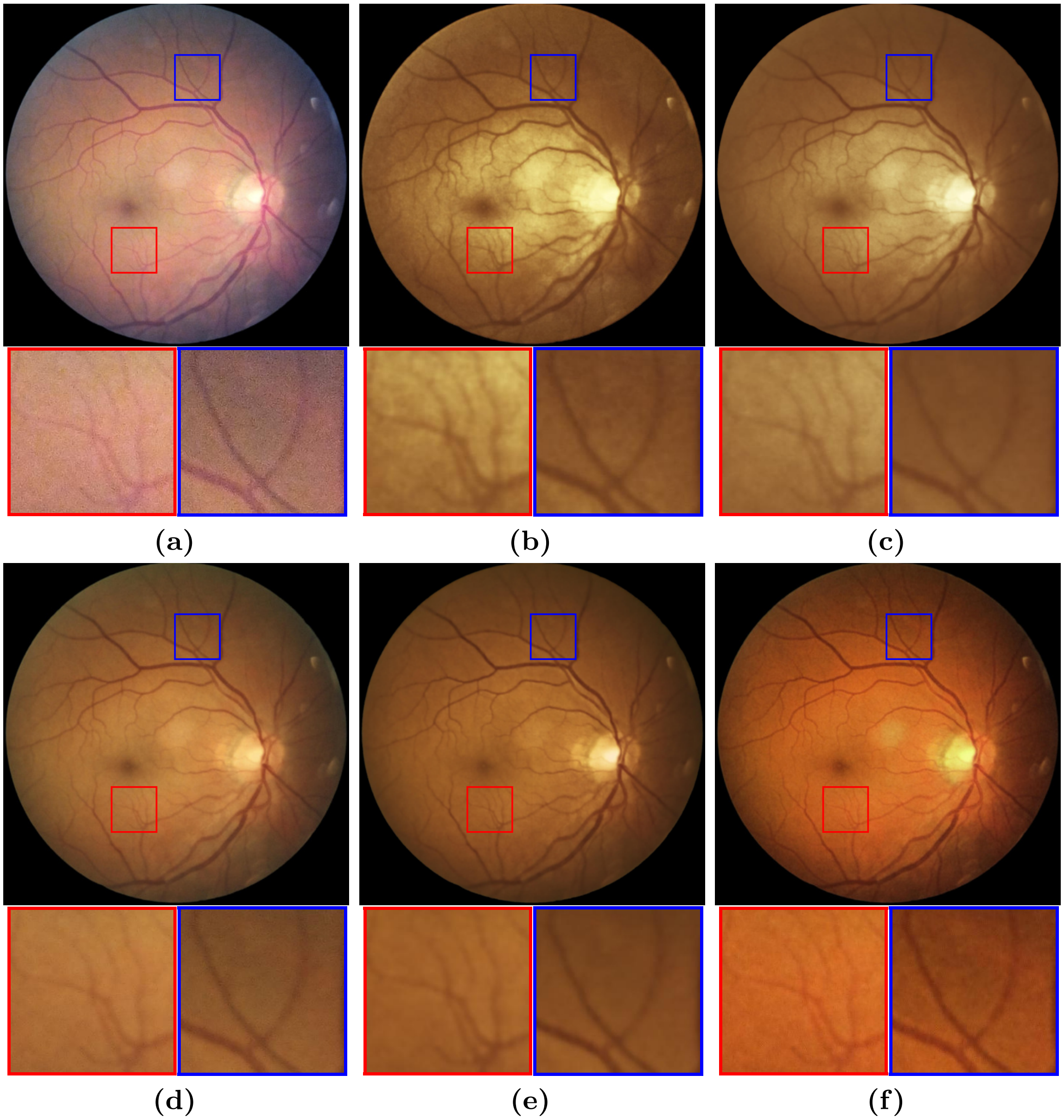


Figure 4.3: Comparison of mapped results for a sample image (a) with other methods: (b) U-net1, (c) U-net2, (d) R-U-net, (e) Proposed method, and (f) ResCycleGAN

To assess the effectiveness of the mapping, we generated four mappings with combination of loss function and residual connection, namely (i) U-net1: U-net with typical l_2 loss; (ii) U-net2: U-net with a combination of l_1 + MS-SSIM loss; (iii) Res-U-net: Residual connection from input to output end of U-net with l_1 + MS-SSIM loss and (iv) proposed SupEnh method. All the mappings were trained with the same setting as SupEnh. Also, ResCycleGAN proposed (retrained with Micromax images) in chapter 3 is used for comparison. Sample results of a *normal* SC image using Micromax is shown in

Figure 4.3 along with two magnified regions showing thin and thick vessels with a noisy background. Figures 4.3b-c show the result of U-net1 and U-net2 having saturated OD with bright background in middle and washed out effect, respectively. Both U-net1 and U-net2 performs good in noise removal, but both are having uneven illumination, which can be seen in two magnified regions. The results of Res-U-net and SupEnh are shown in Figure 4.3d-e, having a better mapping with a balanced illumination and noise suppression with good contrast. Results of SupEnh looks better in OD region is having clear boundaries as compared to Res-U-net result. The ResCycleGAN result as a whole, as well as sub-regions in Figure 4.3f possess a good definition of vessels, but having high contrast and noisy background.

To assess the noise removal in the mapped results, a line profile is shown in Figure 4.4 for a sub-region (outlined in blue colour) of image in Figure 4.3. The I channel of image is used for line profile, and it is normalized using mean value to show all plots in same range. The line position is indicated in the inset image covers two vessels; hence, the profile should have valley at the location of the vessels. The results indicate the ResCycleGAN and R-U-net having noisier than U-net1, Unet2 and SupEnh. The valley in ResCycleGAN shows the good definitions vessel compared to all other methods.

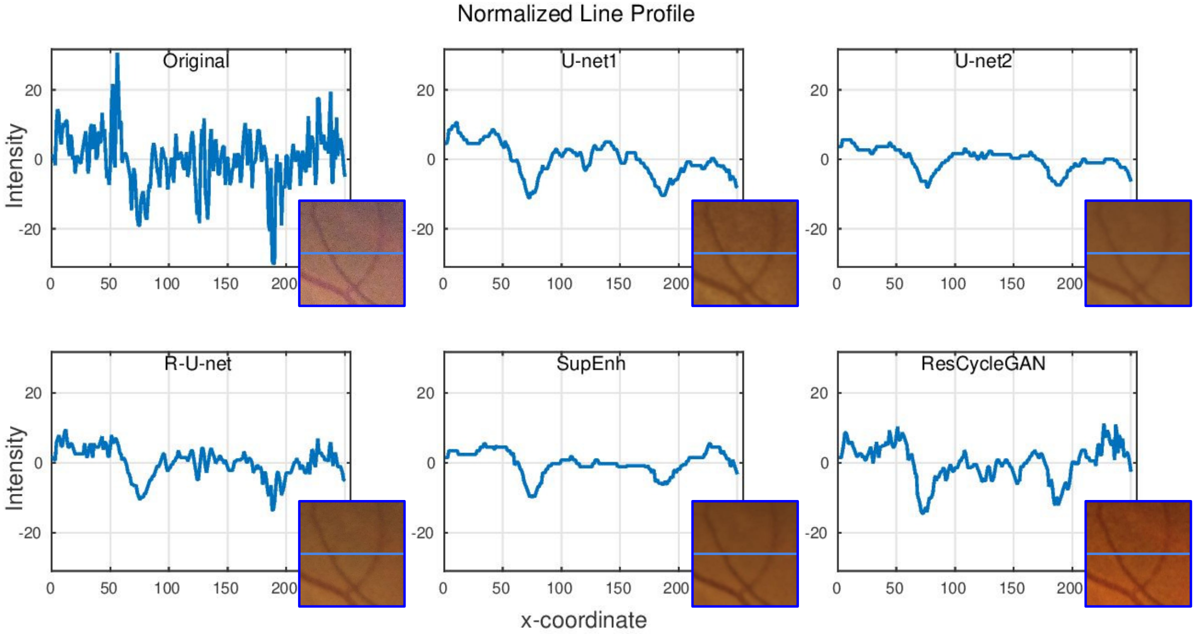


Figure 4.4: Line profile comparison of mapped results for a sub-region (shown in inset image)

The mapped results for a sample image with *pathologies* is shown in Figure 4.5 along with magnified regions containing hard exudates (bright spot) and microaneurysm (dark spot) outlined in red and blue colours, respectively. The results in Figure 4.5b-c of U-net1 and U-net2 shows a similar trend of uneven background, as in *normal* case and it can cause difficulties in analysis brighter structure (OD) and lesions (hard exudates are barely visible in the sub-region shown in red colour). The Figure 4.5d-f shows the results of R-U-net, SupEnh and ResCycleGAN having balanced illumination and bluish flash artefact suppression. However, the R-U-net produces the whitish background affecting vessels and lesions contrast (shown in the magnified region). The ResCycleGAN results posses good structural details with high contrast and noisy background similar to *normal* case. The SupEnh results (Figure 4.5e) is seen to be the best as there is a good balance of smoothing out the noise with better illumination and contrast around structures. The U-net1 and U-net2 were trained for 100 epochs which are twice the number of epochs for the SupEnh whereas R-U-net was trained for 50 epochs. The shorter training for the proposed and R-U-net method is due to the residual connection which helps in learning.

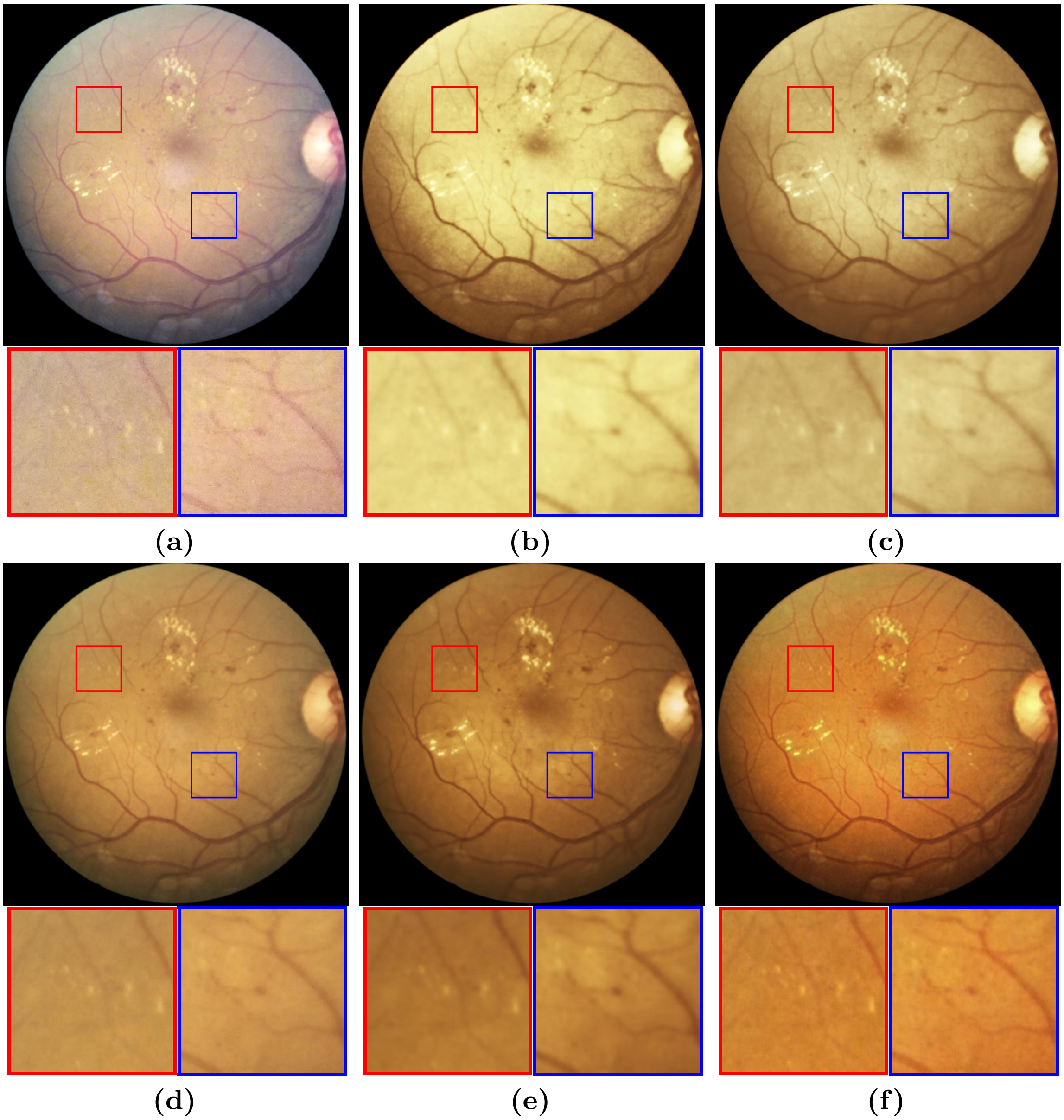


Figure 4.5: Comparison of mapped results for a pathological image (a) with other methods: (b) U-net1, (c) U-net2, (d) R-U-net, (e) Proposed method, and (f) ResCycleGAN

4.4.1.2 Cross-validation with iPhone images

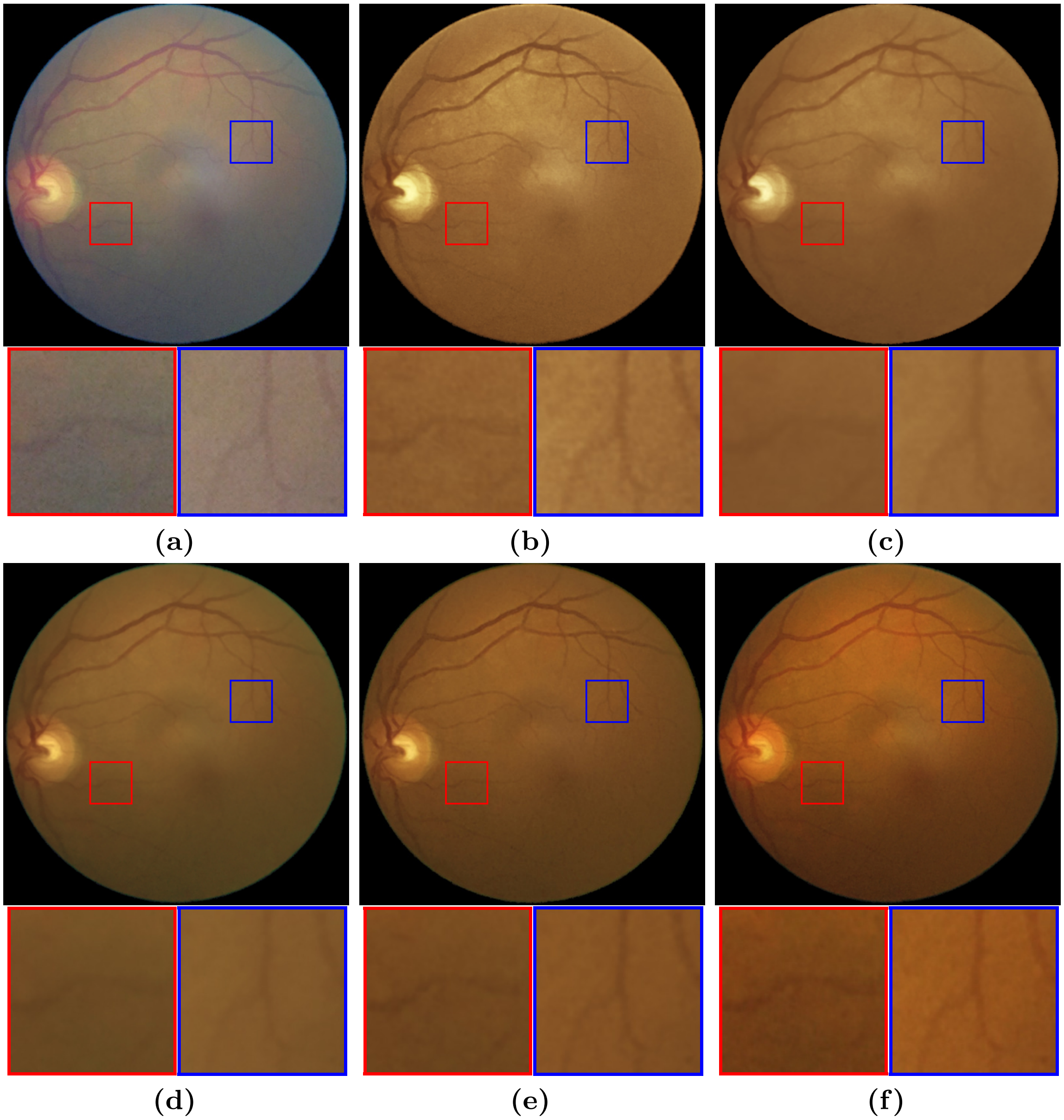


Figure 4.6: Comparison of mapped results for a sample image from iPhone (a) with other methods: (b) U-net1, (c) U-net2, (d) R-U-net, (e) Proposed method, and (f) ResCycleGAN

Smartphone images can have variable image quality depending on the specific brand/generation of device. Performance of the mapping for the different SC images infer about the robustness and adaptability of the model. To assess this, SC images captured using iPhone are used for cross-validation of the model. A sample image (without mydriasis) along with mapped results are shown in Figure 4.6 along with two noisy sub-regions. Results in Figure 4.6b-c shows a similar trend for U-net1 and U-net2 with noisy background and loss of details shown in respective magnified regions. The mapping results of R-U-net, the proposed method and ResCycleGAN are shown in Figure 4.6d-f exhibits balanced illumination and contrast with flash artefact suppression. However, the magnified regions indicate over-smoothing in R-U-net and noisy in ResCycleGAN. The SupEnh results showing overall balanced improvement along with good noisy removal (shown sub-regions) showing its generalisation ability. If the SC image quality improvement is of interest for a fixed clinical setting or commercial FOP product, the quality of image can be further improved by customising the solution or retraining the SupEnh with specific SC image and corresponding pair of FC image.

4.4.2 Quantitative analysis

Quantitative evaluations of the mapping are performed using three ways: full-reference based metric, no-reference based metric and performance of early stage DR detection. These are described next.

4.4.2.1 Evaluation using Full-Reference based Quality Metric

Full-Reference based quality metric needs a pristine image with no distortion to evaluate the performance of image quality. FC image is of high quality with better details, and it is aligned with SC image as described in section 4.3.4. Hence, it serves as the pristine/reference image. Evaluation of mapped solutions is done using standard full-reference metrics, namely MSE, PSNR, SSIM, and MS-SSIM. The MSE and PSNR evaluate the quality of the mapped image whereas the SSIM and MS-SSIM evaluate the perceptual image quality by measuring the luminance, contrast, and structural component of the image to that of a reference image. Also, MS-SSIM evaluates at multiple scales to handle variation in resolutions.

The MSE, PSNR, SSIM, and MS-SSIM values were computed for 208 images from the test set from Micromax, which was described in section 4.3.3. The mean values of these metrics are reported in Table 4.2 for the ResCycleGAN, SupEnh and variant of U-net described earlier. The performance of R-U-net, ResCycleGAN and SupEnh are better that U-nets with SupEnh method is best among them in all the metric.

	MSE	PSNR	SSIM	MS-SSIM
SC images	0.0329	11.42	0.5570	0.7493
U-net1	0.0243	14.30	0.6382	0.8345
U-net2	0.0207	14.70	0.6672	0.8636
R-U-net	0.0132	15.95	0.7091	0.8812
ResCycleGAN	0.0123	15.67	0.7108	0.8621
SupEnh	**0.0110**	**16.82**	**0.7347**	**0.8896**

Table 4.2: Full-reference based quantitative comparison of performance on the SC images obtained using Micromax

4.4.2.2 Evaluation using No-reference based Quality Metric

The reference image is not feasible to obtain (especially in medical imaging) all the time; hence, a no-reference image quality metric is commonly used to evaluate the quality of the image. Generally, a no-reference metric is defined based on the use of statistical features of the image. To achieve a proper no-reference evaluation of retinal image quality, we use a Q_v score [32] which assess the noise and blur around the vessels. The scores are calculated for R-U-net, ResCycleGAN and SupEnh methods which are better in both qualitative and full-reference quantitative assessment.

The average and standard deviation scores were computed for two test sets containing 208 images using Micromax and 270 images using iPhone, which was described in section 4.3.3 and score are presented in Table 4.3. The results of Q_v score in both the datasets indicates that SupEnh method outperforms both R-U-net and ResCycleGAN methods, implies the mapping has balanced noise removal and vessel contrast. Although ResCycleGAN has trained separately for Micromax and iPhone images, Q_v score are low compared to SupEnh method, due to noise present in the image. Also, ResCycleGAN performance better than R-U-net method for iPhone dataset due to better signal-to-noise-ratio in iPhone than Micromax images, which shows ResCycleGAN method can't handle noise in the image as stated in Chapter 3. Even though SupEnh and R-U-net methods are trained using Micromax images, SupEnh method scores are better than other methods in both similar and cross dataset, which indicates the versatility of the model.

	Q_v score (Micromax dataset)	Q_v score (iPhone dataset)
SC images	0.0273 ± 0.0195	0.0189 ± 0.0104
R-U-net	0.0306 ± 0.0195	0.0275 ± 0.0126
ResCycleGAN	0.0268 ± 0.0192	0.0334 ± 0.0175
SupEnh	**0.0338** ± **0.0193**	**0.0387** ± **0.0187**

Table 4.3: No-reference based quantitative comparison of performance on the SC images obtained using Micromax and iPhone

4.4.2.3 Performance of Early Stage Diabetic Retinopathy Detection

To assess the effectiveness of enhancement of the image in CAD setting, we compare the early stage Diabetic Retinopathy (DR) detection for the SC images before and after enhancement. As per Early Treatment Diabetic Retinopathy Study (ETDRS) [43] guideline, early stage DR (or mild non-proliferative DR (NPDR)) contains only microaneurysms (MA). MA is the first clinically visible changes of DR and appears as small red dots. Existing red lesion detection method proposed in [49] is used for evaluation of early stage DR detection. The method is based on the ensemble of CNN and hand-crafted features, followed by Random Forest (RF) classifier to get probability map for MA. We use the provided

CNN and RF models[1] trained on DIARETDB1 dataset [27] were used to evaluate the performance of early stage DR detection before and after enhancement using SupEnh.

A different private dataset obtained from the FOP product using the iPhone is used for our experiment, it contains the DR grading from two clinicians. Grading is done as per ETDRS guidelines containing five levels: no DR, mild, moderate, severe NPDR and proliferative DR along with some ungradable images. The dataset has 332 cases, each containing three fields/images per case. Since each case includes the three images, the probability map for each image is combined to find the DR grading. For evaluation, we obtain a total of 92 cases containing 33 mild and 59 no DR cases having common grading between two clinicians and eliminating other grades and ungradable case. The performance before and after enhancement using SupEnh is shown as Receiver operating characteristic (ROC) [20] curve in Figure 4.7. After enhancement using SupEnh method, the area under the curve (AUC) for mild DR detection is improved by nearly 5%. The SupEnh model was trained on Micromax images and cross tested on iPhone images here, which shows the potential of the proposed method. The results could be further improved by retraining the model with specific pair data.

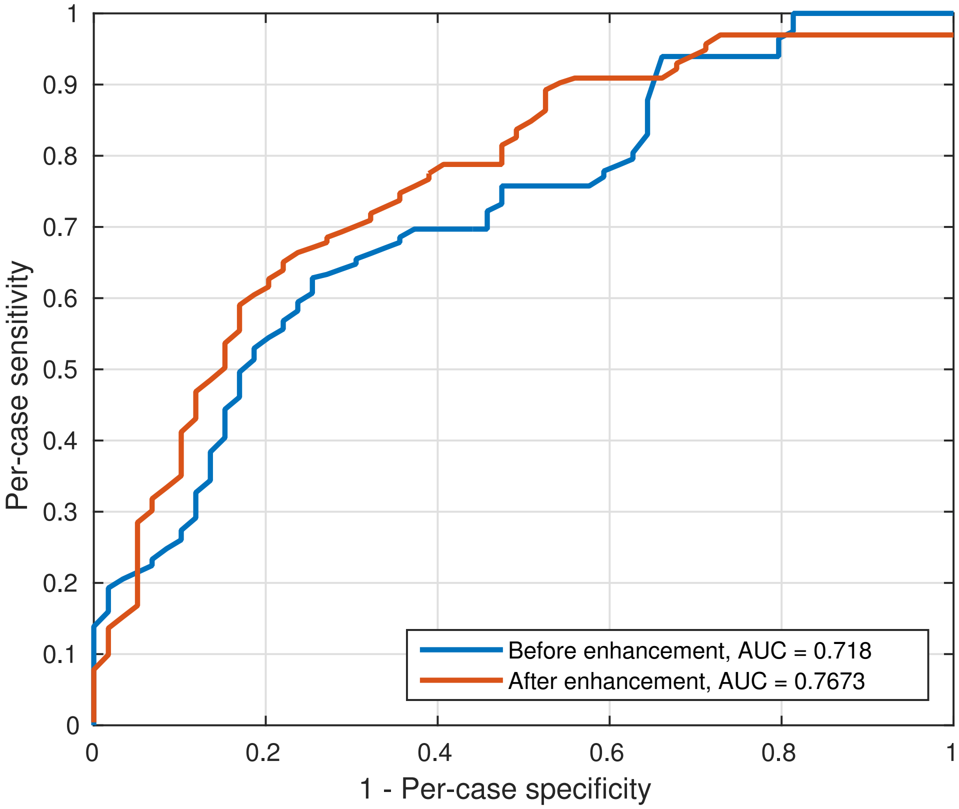


Figure 4.7: ROC curve for early stage DR detection for per case

[1] https://github.com/ignaciorlando/red-lesion-detection

4.5 Conclusions

A SupEnh solution was proposed to enhance the quality of smartphone-based fundus image using supervised learning. It learns to match the characteristics SC images to FC images along with denoising of the image in an end-to-end manner. The key strengths of our method are: it preserves the integrity of structures with excellent noise and flash artefact suppression, balanced illumination correction between the peripheral and centre region, consistent contrast within and across images with no introduction of artefacts. The results are invariably good for images with/without pathologies and change in image quality as well as images acquired with mydriasis. Cross-validation with iPhone images demonstrates that the method is robust and adaptable. Hence our solution can aid ophthalmic experts. The improvement in the result of early stage DR detection illustrates that it can be used for screening purpose. The method can be further explored to use as a preprocessing stage is comprehensive CAD systems in future.

Chapter 5

Conclusions

This thesis is focused on the problems associated with the quality of the retinal images and developing solutions for image quality improvement via learning based techniques for two modalities of retinal imaging, namely, OCT and fundus image using a smartphone camera. We have presented three deep learning approaches to enhance/improve the quality of retinal images by preserving the integrity of anatomical structures and no introduction of artefacts.

Firstly, we started the thesis by looking into the problem of speckle suppression in OCT images, which arises due to the physical properties used in OCT imaging. There are exhaustive works of literature available for speckle suppression in OCT with traditional approaches. Chapter 2 presented a new architecture based on autoencoder to denoise the OCT images using *unpaired* data. The solution uses two autoencoders with shared encoder strategy, hence it was named as Shared Encoder (SE). This method was the first work to our knowledge, which shows a lot of potential for OCT denoising using Convolutional Neural Network in an unsupervised manner. The entire SE is trained using a small set of unpaired noisy and clean images to reduce speckle noise effectively. We have been able to develop such end-to-end trainable module due to the advancement in deep learning, which means the solution doesn't require any other module. Many variants of SE are presented, while fine-tuning SE with Semi-supervised approach performing best among them. Shortage of clean data hinders this deep learning model from outperforming baseline methods. The quantitative results of the proposed methods are on par with the existing tradition methods. However, our method preserves the integrity of anatomical structures with a good balance of smoothing out the speckle noise, a feature which is essential in clinical diagnosis. Also, the results of a perceptual study with two clinicians show our results were preferred over other methods. A key strength is it handles the variability of scanner and image quality as well as pathological cases.

Next, in this thesis, we focused on the problem of image quality enhancement of smartphone camera-based fundus image (or SC image). The imaging of the retina with a smartphone is entirely new and emerging as the device is portable with a much lower cost than standard fundus camera (FC). But the quality of images are degraded due to low-cost sensors and optics, low-light levels resulting in loss of details, uneven illumination, introduce noise and flash artefacts. A solution is presented in Chapter 3 to solve all these problems by learning a mapping function from SC to FC images. The proposed ResCycleGAN method learns to match the characteristics of SC images to standard fundus camera images using an *unpaired* data. To our knowledge, no such attempt to improve the quality of images are made, and the solution provides stable results over a CycleGAN with reduced training time. The proposed solution preserves the integrity of structures with a balanced illumination correction between the peripheral and centre region. The method is fast and robust to images with pathologies as well as images acquired with/without mydriasis. Noise removal was not within the scope of this work. Chapter 4 presented a new architecture which handles noise removal using *paired* data. The proposed SupEnh solution enhances the quality of SC images by learning a mapping function from SC to FC in an end-to-end, supervised manner. It aims to match the characteristics SC images to FC images along with denoising of the image. The method preserves the structural details with good suppression of noise and flash artefact, balanced illumination correction, consistent contrast within and across images. It also handles images with/without pathologies as well as the change in image quality. Cross-validation with different images without explicit training demonstrates that the method is robust and adaptable. Hence, the solution can aid ophthalmic experts. The improvement in the result of early stage DR detection illustrates that it can be adapted to screening purpose.

5.1 Future Work

The following are the possible directions for future work

- **Automated OCT analysis system:** In chapter 2, the method suppress the speckle noise in OCT images with good integrity of anatomical structures. It can be used as a preprocessing stage in automated OCT anatomy analysis system such as segmentation of retinal layers. Our method can handle images with pathologies; hence, it can be used as a preprocessing stage even in disease detection and segmentation system.

- **Screening system:** Chapter 3 and 4 gives a solutions for enhancing the quality of smartphone-based fundus image. The method is robust to images with pathologies as well as the change in image quality and improvement of early stage DR detection after enhancement suggests that it can be used as a prepossessing stage in screening system for detection of DR, glaucoma, Age-related Macular Degeneration, Diabetic Macular Edema.

- **Data augmentation:** Chapter 3 has an auxiliary task of mapping FC to SC images which helps in the reconstruction of SC images for Cycle consistency loss. This mapping introduces the uneven illumination, flash noise and contrast variation. It can be used as data augmentation to improve the automated retinal analysis system to handle the practical variations occurs while capturing a retina image.

- **Adapting to smartphone-based fundus imaging:** The solution proposed in Chapter 4 can be adopted in a fixed clinical or commercial product for better visualization of retinal images. The solution can be further improved by tailoring to a specific pair of images having good alignment.

Related Publications

- *”Shared Encoder based Denoising of Optical Coherence Tomography Images”*
  Sukesh Adiga V, Jayanthi Sivaswamy
  Proceedings of 11^{th} Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP), 2018, Hyderabad, India

- *”Matching the Characteristics of Fundus and Smartphone Camera Images”*
  Sukesh Adiga V, Jayanthi Sivaswamy
  Proceedings of 16^{th} IEEE International Symposium on Biomedical Imaging (ISBI) 2019, Venice, Italy

- *”Enhancement of Smartphone Camera-based Fundus Images”*
  Sukesh Adiga V, Jayanthi Sivaswamy
  Under preparation

Bibliography

[1] M. D. Abràmoff, M. K. Garvin, and M. Sonka. Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, 3:169–208, 2010.

[2] S. Aja-Fernández and C. Alberola-López. On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering. *IEEE Transactions on Image Processing*, 15(9):2694–2701, 2006.

[3] L. Allen. Ocular fundus photography*: Suggestions for achieving consistently good pictures and instructions for stereoscopic photography. *American Journal of Ophthalmology*, 57(1):13–28, 1964.

[4] M. Artetxe, G. Labaka, E. Agirre, and K. Cho. Unsupervised neural machine translation. In *Proceedings of the Sixth International Conference on Learning Representations*, April 2018.

[5] A. Bastawrous, M. E. Giardini, N. M. Bolster, T. Peto, N. Shah, I. A. Livingstone, H. A. Weiss, S. Hu, H. Rono, H. Kuper, et al. Clinical validation of a smartphone-based adapter for optic disc imaging in kenya. *JAMA Ophthalmology*, 134(2):151–158, 2016.

[6] M. I. H. Bhuiyan, M. O. Ahmad, and M. Swamy. Wavelet-based despeckling of medical ultrasound images with the symmetric normal inverse gaussian prior. In *International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages I–721. IEEE, 2007.

[7] S. J. Chiu et al. Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation. *Optics Express*, 18(18):19413–19428, 2010.

[8] S. J. Chiu et al. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomedical Optics Express*, 6(4):1172–1194, 2015.

[9] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

[10] L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu. Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. *IEEE Transactions on Medical Imaging*, 32(11):2034–2049, 2013.

[11] L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu. Sparsity based denoising of spectral domain optical coherence tomography images. *Biomedical Optics Express*, 3(5):927–942, 2012.

[12] J. Flammer, K. Konieczka, R. M. Bruno, A. Virdis, A. J. Flammer, and S. Taddei. The eye and the heart. *European Heart Journal*, 34(17):1270–1278, 2013.

[13] C. Flick. Centenary of babbage's ophthalmoscope. *The Optician*, 113(2925):246–246, 1947.

[14] K. M. Galetta, P. A. Calabresi, E. M. Frohman, and L. J. Balcer. Optical coherence tomography (oct): imaging the visual pathway as a model for neurodegeneration. *Neurotherapeutics*, 8(1):117–132, 2011.

[15] O. Gerloff. Uber die photographie des augenhintergrundes. *Klin Monatsblätter Augenheilkunde*, 29:397–403, 1891.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[17] J. E. Grunwald, J. Alexander, G.-S. Ying, M. Maguire, E. Daniel, R. Whittock-Martin, C. Parker, K. McWilliams, J. C. Lo, A. Go, et al. Retinopathy and chronic kidney disease in the chronic renal insufficiency cohort (cric) study. *Archives of Ophthalmology*, 130(9):1136–1144, 2012.

[18] A. Gullstrand. Neue methoden der reflexlosen ophthalmoskopie. *Berichte Deutsche Ophthalmologische Gesellschaft*, 36(8):326, 1910.

[19] S. Gupta, R. Chauhan, and S. Saxena. Locally adaptive wavelet domain bayesian processor for denoising medical ultrasound images using speckle modelling based on rayleigh distribution. *IEE Proceedings-Vision, Image and Signal Processing*, 152(1):129–135, 2005.

[20] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[21] D. Huang, 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. Optical coherence tomography. *Science*, 254(5035):1178–1181, 1991.

[22] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, pages 448–456, 2015.

[23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134, 2017.

[24] G. D. Joshi and J. Sivaswamy. Colour retinal image enhancement based on domain knowledge. In *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, pages 591–598. IEEE, 2008.

[25] R. Kafieh, H. Rabbani, and I. Selesnick. Three dimensional data-driven multi scale atomic representation of optical coherence tomography. *IEEE Transactions on Medical Imaging*, 34(5):1042–1062, 2015.

[26] M. Karaman, M. A. Kutay, and G. Bozdagi. An adaptive speckle suppression filter for medical ultrasonic imaging. *IEEE Transactions on Medical Imaging*, 14(2):283–292, 1995.

[27] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä. The diaretdb1 diabetic retinopathy database and evaluation protocol. In *British Machine Vision Conference*, volume 1, pages 1–10, 2007.

[28] C. R. Keeler. 150 years since babbage's ophthalmoscope. *Archives of ophthalmology*, 115(11):1456–1457, 1997.

[29] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim. Learning to discover cross-domain relations with generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1857–1865. JMLR.org, 2017.

[30] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proceedings of International Conference on Learning Representations*, volume 5, 2015.

[31] T. Köhler, A. Budai, M. F. Kraus, J. Odstrčilik, G. Michelson, and J. Hornegger. Automatic no-reference quality assessment for retinal fundus images using vessel segmentation. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 95–100. IEEE, 2013.

[32] T. Köhler, J. Hornegger, M. Mayer, and G. Michelson. Quality-guided denoising for low-cost fundus imaging. In *Bildverarbeitung für die Medizin*, pages 292–297. Springer, 2012.

[33] H. Kolb et al. Facts and figures concerning the human retina–webvision: The organization of the retina and visual system. 1995.

[34] K. Krissian, C.-F. Westin, R. Kikinis, and K. G. Vosburgh. Oriented speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 16(5):1412–1424, 2007.

[35] G. Lample, A. Conneau, L. Denoyer, and M. Ranzato. Unsupervised machine translation using monolingual corpora only. In *International Conference on Learning Representations*, 2018.

[36] G. Liew and J. J. Wang. Retinal vascular signs: a window to the heart? *Revista Española de Cardiología (English Edition)*, 64(6):515–521, 2011.

[37] M.-Y. Liu, T. Breuel, and J. Kautz. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*, pages 700–708, 2017.

[38] A. London, I. Benhar, and M. Schwartz. The retina as a window to the brainfrom eye research to cns disorders. *Nature Reviews Neurology*, 9(1):44, 2013.

[39] R. K. Lord, V. A. Shah, A. N. San Filippo, and R. Krishna. Novel uses of smartphones in ophthalmology. *Ophthalmology*, 117(6):1274–1274, 2010.

[40] R. N. Maamari, J. D. Keenan, D. A. Fletcher, and T. P. Margolis. A mobile phone-based retinal camera for portable wide field imaging. *British Journal of Ophthalmology*, 98(4):438–441, 2014.

[41] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2794–2802, 2017.

[42] M. A. Mayer, A. Borsdorf, M. Wagner, J. Hornegger, C. Y. Mardin, and R. P. Tornow. Wavelet denoising of multiframe optical coherence tomography data. *Biomedical Optics Express*, 3(3):572–589, 2012.

[43] P. Mitchell and S. Foran. Guidelines for the management of diabetic retinopathy. 2008.

[44] M. Mohammadpour, Z. Heidari, M. Mirghorbani, and H. Hashemi. Smartphones, tele-ophthalmology, and vision 2020. *International Journal of Ophthalmology*, 10(12):1909, 2017.

[45] U. Mutlu, P. W. Bonnemaijer, M. A. Ikram, J. M. Colijn, L. G. Cremers, G. H. Buitendijk, J. R. Vingerling, W. J. Niessen, M. W. Vernooij, C. C. Klaver, et al. Retinal neurodegeneration and brain mri markers: the rotterdam study. *Neurobiology of Aging*, 60:183–191, 2017.

[46] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, 2010.

[47] U. T. Nguyen, A. Bhuiyan, L. A. Park, and K. Ramamohanarao. An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern Recognition*, 46(3):703–715, 2013.

[48] H. R. Novotny and D. L. Alvis. A method of photographing fluorescence in circulating blood in the human retina. *Circulation*, 24(1):82–86, 1961.

[49] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko. An ensemble deep learning based approach for red lesion detection in fundus images. *Computer Methods and Programs in Biomedicine*, 153:115–127, 2018.

[50] A. Paul, D. P. Mukherjee, and S. T. Acton. Speckle removal using diffusion potential for optical coherence tomography images. *IEEE Journal of Biomedical and Health Informatics*, 2018.

[51] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3):158, 2018.

[52] J. Purkinje. Erstes bandchen, beitrage zur kenntniss des sehens in subjectiver hinsicht. *Prague*, 1823.

[53] J. U. Quistgaard. Signal acquisition and processing in medical diagnostic ultrasound. *IEEE Signal Processing Magazine*, 14(1):67–74, 1997.

[54] R. Rajalakshmi, S. Arulmalar, M. Usha, V. Prathiba, K. S. Kareemuddin, R. M. Anjana, and V. Mohan. Validation of smartphone based retinal photography for diabetic retinopathy screening. *PloS One*, 10(9):e0138285, 2015.

[55] G. Ramos-Llordén, G. Vegas-Sánchez-Ferrero, M. Martin-Fernandez, C. Alberola-López, and S. Aja-Fernández. Anisotropic diffusion filter with memory based on speckle statistics for ultrasound images. *IEEE Transactions on Image Processing*, 24(1):345–358, 2015.

[56] E. Ritenour, T. Nelson, and U. Raff. Applications of the median filter to digital radiographic images. In *International Conference on Acoustics, Speech, and Signal Processing*, volume 9, pages 251–254. IEEE, 1984.

[57] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[58] B. Sander, M. Larsen, L. Thrane, J. Hougaard, and T. Jørgensen. Enhanced optical coherence tomography imaging by multiple scan averaging. *British Journal of Ophthalmology*, 89(2):207–212, 2005.

[59] J. M. Schmitt, S. Xiang, and K. M. Yung. Speckle in optical coherence tomography. *Journal of Biomedical Optics*, 4(1):95–106, 1999.

[60] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[61] R. K. Srivastava, K. Greff, and J. Schmidhuber. Highway networks. *arXiv preprint:1505.00387*, 2015.

[62] B. C. Toy, D. J. Myung, L. He, C. K. Pan, R. T. Chang, A. Polkinhorne, D. Merrell, D. Foster, and M. S. Blumenkranz. Smartphone-based dilated fundus photography and near visual acuity testing as inexpensive screening tools to detect referral warranted diabetic eye disease. *Retina*, 36(5):1000–1008, 2016.

[63] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6924–6932, 2017.

[64] A. Van Trigt. Trajecti ad rhenum. 1853. *Dissertatio Ophthalmologica Inauguralis de Speculo Oculi*.

[65] M. E. van Velthoven, D. J. Faber, F. D. Verbraak, T. G. van Leeuwen, and M. D. de Smet. Recent developments in optical coherence tomography for imaging the retina. *Progress in Retinal and Eye Research*, 26(1):57–77, 2007.

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

[67] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *37th Asilomar Conference on Signals, Systems & Computers*, volume 2, pages 1398–1402. IEEE, 2003.

[68] R. H. Webb and G. W. Hughes. Scanning laser ophthalmoscope. *IEEE Transactions on Biomedical Engineering*, (7):488–492, 1981.

[69] Z. Yi, H. Zhang, P. Tan, and M. Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2849–2857, 2017.

[70] W. Yip, P. G. Ong, B. W. Teo, C. Y.-l. Cheung, E. S. Tai, C.-Y. Cheng, E. Lamoureux, T. Y. Wong, and C. Sabanayagam. Retinal vascular imaging markers and incident chronic kidney disease: a prospective cohort study. *Scientific Reports*, 7(1):9374, 2017.

[71] Y. Yu and S. T. Acton. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 11(11):1260–1270, 2002.

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

[73] H. Zhao, O. Gallo, I. Frosio, and J. Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2017.

[74] M. Zhou, K. Jin, S. Wang, J. Ye, and D. Qian. Color retinal image enhancement based on luminosity and contrast adjustment. *IEEE Transactions on Biomedical Engineering*, 65(3):521–527, 2017.

[75] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, 2017.