Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Extraction of Retinal Blood Vessels from Fundus Images using Laplacian Pyramids and Anisotropic Diffusion

Abishek Vaithylingam

Supervisor: Prof. Rozenn Dahyot

A dissertation submitted to the University of Dublin,
in partial fulfilment of the requirements for the degree of
M.Sc. in Computer Science (Data Science)
Trinity College Dublin, University of Dublin
September, 2020

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar. I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Abishek Vaithylingam

September 7, 2020

Acknowledgments

I would like to convey my deepest gratitude and regards to my Dissertation supervisor, Prof. Rozenn Dahyot for providing me with invaluable knowledge, support and feedback during the course of this dissertation. Her lectures on Computer Vision and Deep Learning were truly fascinating and inspired me to choose this research problem. I am grateful to the School of Computer Science and Statistics, Trinity College Dublin for providing me with the resources and foundation I needed to conduct this study.

I would also like to acknowledge the help and support extended to me by my Undergraduate thesis guide, Aditya Pai, my brother, Kartik Vaithylingam, my sister, Varsha Balu and my friends, Nidhi Prakash and Saniya Rohida for proof-reading my work at ungodly hours. Their help has been instrumental in the completion of my thesis.

None of this would have been possible without the love and support given by my Mom, who has truly been a beacon of hope during this time. Lastly, I would like to thank my Dad, who would've been proud to watch me graduate from this prestigious university. Wherever you are, I owe it all to you and Mom.

Thank you.

Abstract

Retinal Fundus images are commonly studied by ophthalmologists and bio-medical researchers to diagnose or predict the onset of various systemic diseases. These Fundus images are examined for anomalies to make accurate diagnoses. Since the process of manually inspecting and annotating retinal blood vessels from Fundus images is time consuming and relies entirely on the skill and experience of the human experts, computer-aided approaches are progressively becoming better alternatives, thus creating a growing need for more efficient retinal vessel segmentation algorithms. These computer-aided approaches could be used for clinical studies, pre-screening the patient for signs of systemic diseases and for developing effective treatment plans. In this thesis, we propose a framework for extracting retinal blood vessels from Fundus images based on their multi-layer, multi-scale Laplacian Pyramid representation. The algorithm is implemented using two publicly available datasets, the High-Resolution Fundus dataset and the DRIVE dataset. We provide a robust pre-processing pipeline for enhancing retinal vessel features in Fundus images. We evaluate the performance of the technique by comparing the segmentation results with a manually annotated ground-truth image, based on performance metrics like specificity, sensitivity, accuracy, precision and recall. The framework is capable of detecting and extracting both, thin and coarse retinal vessels and is competitive with other state-of-the-art techniques in terms of computation time, quality of segmentation, accuracy, sensitivity and specificity. It is robust to parameter changes and is capable of handling noise in the image without blurring the edges of retinal vessels. Our proposed technique shows an average accuracy of 95.85%, sensitivity of 79.33%, and specificity of 97.54% on High-Resolution Fundus images of healthy patients.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Existing techniques for screening hypertensive conditions, cardiovascular diseases, glaucoma, pre-stroke conditions, ocular diseases and diabetic retinopathy are inadequate, arbitrary and fail to identify early onset of these diseases [4]. Modern, computer-assisted quantitative studies have enabled researchers to analyze the effects of these systemic diseases on retinal vasculature. Specialists are progressively moving towards adopting these computer-assisted approaches as they provide a means to screen for early indications of these diseases in a non-invasive way, thus avoiding expensive and possibly painful disease screening procedures. There exists a need for more efficient screening methods to detect retinal blood vessels accurately, identify anomalies in these retinal images and predict the risk of occurrence of these diseases. In order to analyze retinal blood vessels for anomalous occurrences efficiently, it is essential to first detect and classify retinal blood vessels accurately. Automated retinal blood vessel segmentation algorithms are capable of identifying and extracting both, asymmetrical and symmetrical patterns of retinal vasculature. However, these algorithms continue to face challenges with noisy images and detecting fine retinal vessels with high levels of accuracy. Occurrences of pathological patterns, background structures or lesions in these retinal images also hinders the accuracy of these algorithms. While using supervised learning based approaches does increase the accuracy of segmentation, they are still heavily reliant on using labelled training data, which is difficult to obtain. Supervised learning models also take a lot of training time. Thus, there remains a need for an efficient framework to extract retinal vessels from different types of retinal images accurately, while preserving all the image features that are significant for the analysis of retinal vessels, like torsion, colour, peripheral width, fine vessels, edges and topological features. This framework would also be required to handle high levels of noise and anomalous structures in the retinal images.

1.2 Overview

The field of Ophthalmology has seen significant development since the 19th century. The only blood vessels in the human body that can be studied in vivo using non-invasive techniques are the retinal blood vessels [5]. Retinal blood vessel caliber is studied extensively to draw correlations between the appearance of retinal vessels and the occurrence of various systemic conditions caused by anomalies in the blood vessels. Retinal cardiovascular caliber is affected by determinants with a physiological, modifiable nature like the patient's Body Mass Index (BMI), race, glucose levels, sex, Blood Pressure (BP), age as well as determinants with a pathological nature like the patient's smoking habits and occurrence of conditions like dyslipidaemia and atherosclerosis [6]. The most common method used to study retinal vasculature is through Fundus imaging. Abràmoff et al. defines Fundus imaging as the process of using reflected light to obtain a two-dimensional representation of the semi-transparent, three-dimensional retinal tissues on an imaging plane [7]. Retinal scans obtained by Fundus imaging are called retinal Fundus images.

A study by Mulvany et al. [8] shows that peripheral narrowing of retinal arteries is directly associated with cardiovascular hypertension conditions in the patient. The studies showed a reduction in the diameters of retinal arteries prior to the occurrence of cardiovascular hypertension in lab rats [9]. Studies on human subjects also showed that patients with a history of cardiovascular hypertension in their family showed peripheral narrowing of retinal arteries, leading to reduced retinal arteriolar diameters as compared to patients with no history of cardiovascular hypertension in their family [10]. Other studies have indicated that high amounts of glucose levels in the blood plasma could lead to a peripheral widening of retinal veins [11]. It has also been identified that a significant number of human subjects with a wider retinal vein diameter show increased risk of hyperglycaemia. While secondary retinal ischemia and hyperglycaemia cause an increase in blood flow, a change in the glucose metabolism causes a release of nitric oxide and inflammatory cytokine, thus leading to Coronary Artery Diseases (CAD) like failure of the endothelial layer in arteries [12]. Dumitrescu et al. [6] conducted a study that correlated changes in retinal vasculature to the onset of systemic diseases like diabetes mellitus, cardiac hypertension, brain diseases, cardiac heart diseases, kidney diseases and various ocular diseases. Diabetic Retinopathy, which has become an increasingly common cause of blindness due to high blood glucose levels and cardiac hypertension, can be detected at the early onset stages through screening of the retinal blood vessels.

Traditional techniques used by ophthalmologists to study retinal blood vessels are hindered by noise and low levels of discrimination between the contrasts of retinal vessels and background structures. The manual process of annotating retinal vessels from Fundus images is time consuming and not

always accurate. Furthermore, the quality of annotation and diagnosis depends entirely on the skill and experience of the specialist. Modern advancements in medical imaging technology and image processing techniques have aided ophthalmologists and researchers greatly in the analysis of retinal vessels. The usage of computer-assisted techniques for studying retinal vessels is found to be as authentic and accurate as that of traditional techniques like direct ophthalmoscopy and Fundus imaging, while taking a fraction of the time for the processing and analysis. Besides this, computer-aided techniques have better capability to detect tiny retinal vessels that cannot be identified easily by the human eye. We broadly classify techniques for extracting retinal blood vessels as rule-based image processing techniques and supervised learning-based techniques. While rule-based image processing techniques are capable of extracting retinal blood vessels from unlabelled retinal Fundus images, they are limited in functionality compared to supervised learning-based techniques due to anomalies, pathological patterns and high levels of noise in the retinal scans. Supervised learning-based approaches, on the other hand, require a labelled dataset of images manually annotated by human experts. This labelled dataset is then used to train a classifier model to classify the image pixels as retinal blood vessel pixels and non-retinal blood vessel pixels. To make an accurate diagnosis, it is essential that the technique has high levels of accuracy, sensitivity and specificity. Supervised techniques require large amounts of training data and hours of training time, but attain high levels of accuracy, sensitivity and specificity. Rule-based techniques attain competitive accuracy and specificity figures with low computation time, but their ability to generalize and attain high sensitivity figures is limited.

Our study aims to improve the performance and quality of segmentation in the technique proposed by Dachsel et al. [13] by introducing a robust pre-processing pipeline and adopting the technique of Anisotropic Diffusion for smoothing and removing noise from the image. In the proposed technique, we first apply the technique of gamma correction [14] and grey-scaling to increase the contrast ratio and extract the green channel from the image. The histogram of the image is then equalized using CLAHE [15], so that pixel intensity values in the image are stretched out and equalized between 0 and 255. In the next step, we further enhance the image contrast with the use of an edge-aware local contrasting technique. This is done to increase the discrimination between the contrasts of the retinal vessels and the background. We use the technique of anisotropic diffusion for smoothing and removing noise in the image. After pre-processing the image, we compute the Gaussian layers of the image iteratively with the help of a reduction operation. The Laplacian layers of the image are then generated iteratively by computing the pixel-wise difference of two adjacent layers in the Gaussian pyramid. We then binarize the layers of the Laplacian Pyramid using intensity thresholding and discard disconnected pixel objects and tiny connected structures that are not retinal vessel pixels. Finally, the

binarized image layers are resized to the resolution of the original image and fused to produce the resultant segmented image.

The chapters in this thesis are organized as follows: Chapter 2 provides a comprehensive review of various state-of-the-art techniques used for the segmentation and analysis of shapes, keenly focusing on the segmentation of retinal blood vessels. Chapter 3 explains the workflow of our proposed technique for the extraction of retinal blood vessels from Fundus images. Chapter 4 contains the experimental results and observations from our proposed technique. Chapter 5 provides a comprehensive quantitative and qualitative evaluation of our proposed technique along with its performance measures based on various metrics. Finally, Chapter 6 contains the overall conclusions drawn from our study and ideas for future research works. The complete workflow of our proposed technique is shown in Figure 3.13.

1.3 Research Problem and Objectives

Computer-aided analysis of retinal blood vessels is progressively taking over traditional techniques of analysing retinal blood vessels. Ophthalmologists and medical researchers are constantly searching for more efficient algorithms to segment retinal vessels. In order to make accurate diagnoses, it is essential to first extract retinal blood vessels from Fundus images accurately. The biggest challenge faced by retinal vessel segmentation algorithms is the presence of noise, low contrast and anomalous structures in the background, which significantly hinders the performance of the segmentation.

Given a high-resolution Red Green Blue (RGB) Fundus image *I* in a large pixel domain $\Omega \rightarrow [0,1]$ with background pixels *B*, the objective of this study is to simplify the representation of the image *I* to a binary segmentation map matrix *S(i,j)* containing labels 1 (indicating retinal blood vessel pixels) and 0 (indicating non-retinal vessel pixels), such that -

$$S(i,j) = \begin{cases} 1, & \text{if } (i,j) \in \Omega \cap B \\ 0, & \text{if } (i,j) \in B \end{cases} \tag{1.1}$$

The resultant segmentation map would have to be adequately representative of the retinal blood vessels. The orientation and position of the extracted retinal blood vessels with respect to the image boundaries should remain unchanged. The algorithm should also be able to preserve the connectedness of the retinal vessels and handle image noise without losing significant features and topological information.

1.4 Data

This study makes use of two publicly available datasets, the DRIVE dataset and the High-Resolution Fundus dataset. The DRIVE dataset was compiled by a clinic in the Netherlands during an eye study program. It is a compilation of 40 low-resolution retinal images, divided into 20 training images and 20 testing images. Each image has a resolution of 565px × 584px. Each retinal image in the DRIVE dataset comes with a corresponding field-of-view mask and ground truth image, which is manually annotated by human experts. The High-Resolution Fundus dataset [16] contains 30 high resolution retinal Fundus images. The dataset is divided into 3 categories containing 10 images each of retinal scans of healthy patients, patients with glaucoma and diabetic retinopathy. Each image has a resolution of 3504px × 2336px. Each retinal image in the High-Resolution Fundus dataset comes with a corresponding field-of-view mask and ground truth image, which is manually annotated by human experts.

Chapter 2

Literature Survey

This section provides a comprehensive review of state-of-the-art techniques used for the segmentation and analysis of retinal blood vessels. Section 2.1 covers different rule-based image processing techniques and Section 2.2 covers different supervised learning-based architectures adopted for the segmentation of retinal blood vessels from Fundus images.

2.1 Rule-based Image Processing Techniques

To overcome the limitations of traditional techniques of retinopathy, there have been several studies [17, 18] that follow quantitative, computer-assisted approaches for classifying retinal vasculature and understanding the correlation between systemic diseases and retinal vascular caliber. These computer-assisted approaches include classical image processing techniques for segmentation of images and modern deep learning based segmentation techniques. For the purpose of our study, we look into various segmentation and skeletonization algorithms used in the analysis of bio-medical images. The analysis of objects or shapes in images is greatly hindered by image noise. Traditional techniques of pre-smoothing the images makes it a challenge to extend the technique to 3-dimensional objects and shapes in the images. Narrow regions in the shapes may have insufficient pixels to perform high-order smoothing operations on the images. It is also unclear how the smoothing operation affects the shape skeletons. To overcome these limitations, Shah et al. [19] proposed a technique for shape analysis in images based on significant numbers assigned to unique points on the shape skeleton, also called the “grey-skeleton”. This technique used a pruning based approach to extract a noise-free skeleton from the grey-skeleton. A segmentation function based on the grey-skeletons using fast algorithms was adopted for pruning and generating the grey skeletons. Lam et al. [20] proposed a skeletonization algorithm that used the technique of thinning to detect and remove pixels at

the boundaries of objects in an iterative manner, based on a set of constraints. The thinning technique is used to reduce the amount of information to be processed to the minimum necessary information required to perform the task of object or pixel classification. The technique has been used in studies dating back to the 1950s when Dinnen et al. [21] conducted a study on pattern recognition where it was observed that using a high threshold for an averaging operation over a square window helped in thinning of the input image. This technique has also been used widely in various other applications like analysing chromosomes [22], analysis of coronary arteries [23], analysis of white blood cells (WBCs) [24], analysis of X-ray images [25], etc. The thinning algorithm uses sophisticated conditions for stopping the process of deleting object boundary pixels and for maintaining the connectivity and topology of the skeleton. While this algorithm is efficient with reference to the number of computations performed and retains the topology and connectedness of the object pixels, it appears to face difficulties in retaining some of its geometric properties and detailed features. The specific geometric properties and features it should retain remains application dependent.

Arcelli et al. [26] proposed a skeletonization technique using distance transforms to compute the distance map of objects in the image. They adopt a process of labelling to assign values to object pixels that are equivalent to the distance of the object pixels from the complement of the object. The technique performs peak and ridge detection for segmentation of the object in the image by computing its Euclidean Distance Map (EDM) and has reduced computational cost. The algorithm computes the EDM of the object in the image and extracts two different types of pixels from the distance map. The first type of pixel points are strong ridge pixel points and peak ridge pixel points, which are detected and extracted in parallel by making use of local operators. The second type of pixel points are weak ridge pixel points, which are extracted sequentially by tracking paths along the steepest gradient [26]. The experimental results showed that a single raster scan is sufficient to detect strong ridge pixel points and peak ridge pixel points, while the weak ridge pixel points are detected in the same scan as sequential elements that form paths. These paths begin from strong ridge pixel points and are traced towards either the strong ridge pixel points or the peak ridge pixel points. The extracted set of pixel points from this operation is centred and connected within the structure of the object, typically like the skeleton system is to the human body. To retain the width information of the extracted set of points, the technique employs a second raster scan. The pixel points on the obtained set of points are labelled based on the Euclidean distance of the pixel points from the complement of the object [26]. Thus, the technique preserves its width information and the resultant skeleton obtained from this technique can be used for analysis of the object's shape which is distinguished by non-uniform width.

Another technique proposed by Ogniewicz et al. [27] used the object boundary to compute the Voronoi diagram (VD) [28] to perform skeleton extraction of objects in images. The technique adopted in the

paper replaces the distinct distance map used in these thinning algorithms with the Voronoi diagram of a set of distinct boundary points to overcome various shortcomings of traditional raster-based algorithms used for thinning. Every edge in the Voronoi diagram is attributed with a residual function which is used to measure its significance [27]. The resulting Voronoi diagram generated by the algorithm is well-defined in a continuous space and has proven to be predominantly consistent with reference to the image noise, subject to various image transformations like scaling, resizing, translation, rotation, etc. The approach demands complex stages of post-processing for pruning the branches. It also demands that the boundary points should have a high sampling rate to generate the most optimal Voronoi skeleton.

Both the techniques proposed by Ogniewicz et al. [27] and Arcelli et al. [26] require that the extracted skeleton must be pruned if the boundaries of the object are found to be noisy, in order to avoid generating spurious branches. To address this limitation, Shen et al. [29] proposed a technique to prune these extracted skeletons based on a novel measure of significance, called Bending Potential Ratio (BPR), where the boundary segment context corresponding to each branch is used to determine whether the branch of the extracted skeleton should be pruned. Using the context information of the boundary segment, it is possible to make a better evaluation of the boundary segment's contribution to the resultant skeleton. This evaluation is generally dependent on the specific location of the boundary segment in the unabridged contour. The resultant skeleton is found to be connected and placed medially. Bai et al. [30] proposed an algorithm for skeleton pruning based on contour partitioning without re-positioning boundary pixel points and, by extension, without re-positioning the remaining skeleton pixel points. The algorithm uses the technique of Discreet Curve Evolution (DCE) [31] to perform contour partitioning and to choose the most optimal contour partition for pruning the skeleton [30]. The experimental results of this technique showed that the resultant pruned skeletons are stable and visually in accordance with human perception. It is found to be effective even in the presence of significant variations in the shape and noise, while retaining the topology and features of the original skeletons. It has been effective in removing inaccurate and spurious branches completely without shortening the main branches from the skeleton, which is a commonly encountered problem while using traditional pruning methods. This objective is also achieved without displacing any of the skeleton points, i.e. all the skeleton points on the pruned skeleton are centres of the maximal disks [32]. Krinidis et al. [33] presented a 2D deformable model based on Physics for extracting skeletons from objects without producing spurious branches. The model is found to retain the topology and significant features of the original shape and does not use any existing skeleton pruning techniques.

Further research studies have attempted to overcome the shortcomings of the technique of using distance transforms proposed by Arcelli et al. [32] and Kimmel et al. [34] with the use of smooth medial

functions. Siddiqui et. al [35] proposed an algorithm to extract object skeletons, which utilized smooth medial functions based on Newton's law. The approach reviews the Hamiltonian formulation [36] and measures the outward flux of the field of vectors underlying the Hamiltonian model. The outward flux measurements are used in combination with a thinning process for preserving the homotopy of the object skeleton. This technique is found to be robust while computing 3D and 2D skeletons as they have relatively moderate computational intricacies. In order to describe global properties of the object skeleton better than the method of using distance transforms, Grigorishin et al. [37] proposed a skeletonisation approach that is based on the theory of Electrostatic Fields which guarantees skeletisation of objects with optimal connectivity, detection of corners, thinning and multi-scale properties under a unified framework. Gorelick et al. [38] proposed a skeletisation approach to overcome the same limitation of distance transforms using the Poisson Equation. The approach extracts boundary conditions from the silhouette contours, which is used to solve Poisson's equation to compute a function. The function is used to extract different shape properties such as the rough skeleton, the object structure, the concave and convex segments of the object boundaries and the aspect ratio of different regions. The function performs better than the technique of distance transforms as it takes several boundary points into account, thus describing the global features and information of the shape better.

However, the techniques discussed above require that the images used for skeleton extraction are binary images and pre-segmented. Lindeberg et al. [39] presented a mechanism for skeleton extraction in scale-space from grey-scale images [40] by automating the process of selecting scale levels while detecting one dimensional edges and ridges. The original image is convolved with a Gaussian kernel to produce images with coarse resolutions for scale-space analysis. It defines a scale-space edge which permits the scale levels to differ along the edges of the shape. The fine scales are used to detect sharp edges in the object to reduce noise due to smoothing of the scale-space while the diffuse edges are used to detect coarse scales [39]. Tari et al. [41] further proposed a robust algorithm for skeleton extraction from gray-scale images with noise smoothing based on heat diffusion. The technique performs analysis of the level-set curves and shape analysis using an edge-strength function that is computed using the equation for linear heat diffusion. The most significant feature of this technique is that it computes the edge strength without computing the shape outline first from grey-scale images. The implementation of the algorithm is found to be fast and simplistic as the heat diffusion equation is linear in nature.

The drawback of using the linear heat diffusion equation is that it becomes a challenging task to deduce precise locations of skeleton features like edges and ridges at coarse scales. In addition to this, it is vital for the blurring to be spatially immutable. Other approaches of skeleton extraction from grey-scale images have adopted the use of gradient vector diffusion [42, 43]. Chenyang et al. [43]

introduced an external force for active contours called Gradient Vector Flow (GVF) which was used to relocate the active contours into their boundary concavities. GVF is computed from binary or grey-level images as a distribution of their gradient vectors and the correlated active contours are formulated by making use of a force-balance equation [44] instead of a traditional variation formulation. The GVF is also found to have a large range of capture. Zeyun et al. [42] proposed a segmentation free approach for skeletonisation of objects in grey-scale images that is based on anisotropic vector diffusion. The approach computes a skeleton strength map from the diffused vector field, which is used to measure the likelihood that a given pixel is a skeleton pixel. This skeleton strength map is used to trace back to the skeleton structure, thus imitating the role of the edge detection while computing the edge strength map of the image. The Gradient Vector Flow based approach proposed by Chenyang et al. [43] is found to be sensitive to noise. This can be overcome by performing smoothing operations on the images or by using a heuristic approach [42] based on heat diffusion. This method, however, can only aid in increasing tolerance to noise and can lead to the loss of significant features after performing smoothing operations on the image.

Direkoglu et al. [45] presented an algorithm for skeleton extraction that computes a medial function based on the anisotropic heat diffusion equation, also known as the Perona-Malik diffusion [46, 47] equation. Anisotropic diffusion is a space-invariant and non-linear process that can be used on both grey-scale and binary images. A Skeleton Strength Map (SSM) is obtained by making use of the mean curvature measure and a binary, thin skeleton is obtained by hysteresis thresholding and non-maxima suppression [48]. A medial function is defined by diffusing the image in the direction normal to the boundaries of the features and a diffusion coefficient is employed to regulate the relative contribution of tangential diffusion with reference to normal diffusion in the heat diffusion equation. Thus, the diffusion coefficient is a unique parameter used to regulate the medial function. The algorithm denoises the image without losing significant information and high quality features in the image like edges, level-sets and contours. The medial function is found to be better than that of other linear diffusion techniques like Poisson equation, distance transforms and Gaussian filtering, particularly in noisy images. Since the medial function also takes tangential diffusion into consideration in the diffusion equation, the algorithm preserves the centeredness of the skeleton. The algorithm performs image diffusion without the use of any object shape extraction or segmentation techniques. It does not produce spurious branches and hence, does not require any skeleton pruning. The algorithm also does not need any boundary initialization or computation of edge maps like other skeleton extraction techniques. Consequently, the technique of anisotropic diffusion is an iterative process that is terminated when the desired degree of smoothing is achieved. Hence, it is not possible to determine the number of iterations required for the diffusion process accurately.

Lam et al. [49] proposed a novel multi-concavity framework to handle bright lesions while segmenting retinal blood vessels from Fundus images. The technique makes use of a concavity measure that is locally normalised to handle non-uniformly distributed noise caused due to the variation of spherical intensities in Fundus images. This technique shows promising results particularly on Fundus images of unhealthy patients. However, this technique has a high computation time and fails to detect fine scale retinal vessels. Rezaee et al. [50] proposed a method to achieve higher segmentation capabilities by making use of a skeletonization and thresholding technique derived from fuzzy entropy. They adopt the use of a Wiener's filter [51] to de-blur the image. The discrimination between retinal vessels and other tissue structures is increased by defining an optimal threshold based on fuzzy entropy [52]. Finally, adaptive filtering is used to extract the retinal vessels. Bahadar et al. [53] proposed a technique for automating the technique of retinal vessel detection with a hessian-based morphological approach. The technique uses morphological filters and adaptive histogram equalization to discard low frequency noise and enhance the image contrast. The eigen values and hessian matrix of the image is computed to extract vessels at different image scales. Finally, the retinal vessels are segmented using otsu-thresholding. Adelson et al. [54] proposed a more efficient method for analyzing different image features at different scales using Gaussian Pyramids. Gaussian layers are generated by using a reduction operator to down-sample the images in an iterative manner. Each layer of the Gaussian pyramid efficiently represents the original image at different scales, thus making it possible to analyze and extract significant image features from each layer. High level feature information can be analyzed and extracted from high-resolution layers in the bottom of the Gaussian Pyramid while low level feature information can be analyzed and extracted from low-resolution layers at the top of the Gaussian Pyramid.

Dachsel et al. [13] presented a robust algorithm for the segmentation of retinal blood vessels, keenly focusing on adapting to the problems of high-resolution Fundus images. The algorithm utilizes the multi-scale description of the Laplacian Pyramid to achieve state-of-the-art computation speeds, while also detecting finely branched and coarse retinal blood vessels. In order to compare and transfer features from one layer to another, the algorithm performs down-sampling and up-sampling operations using the technique of bilinear interpolation [13], which also has smoothing properties. This improves the overall quality of the extracted retinal blood vessels. The green channel of the original image is used to extract the Laplacian layers of the image and various thresholding, re-scaling and fusion operations are performed on it. A Gaussian pyramid is computed by recursively applying the Reduce Operator for smoothing the image [54]. The pixel-wise difference of two adjoining Gaussian layers is computed to generate the Laplacian Pyramid [55] of the image. The large connected structures or vessels are retained and the small connected structures or noise is removed to generate an optimal

segmentation map. The algorithm is found to be robust, on par with state of the art methods, does not require any adjustment of parameters and requires minimal computation resources and time.

2.2 Supervised Learning: Convolutional Neural Networks

Automated frameworks for the segmentation of retinal blood vessels have found their use in diagnostic medicine and disease screening. Since these frameworks are used specifically for medical diagnostic and screening purposes, they are expected to have high generalization capability and accuracy. The segmentation of retinal blood vessels is a particularly challenging task due to the low discrimination between the contrast of the retinal blood vessels and the background caused by inadequate illumination in the medical imaging device. Another challenge faced while designing a model to segment retinal blood vessels is the complicated asymmetrical and symmetrical patterns in the retinal blood vessels. Thus, it is essential to address these challenges while designing an automated framework to segment retinal blood vessels to make it robust and generalized.

In the 1980s, Japanese scientist Kunihiko et al. [56] first introduced the concept of a basic neural network in the “Neocognitron” for performing simple image recognition. Consequently, convolutional neural networks (CNN) or ConvNets were established for the first time by a Computer Science researcher, Yann LeCun in the 1980s, who built on the concept of Neocognitrons to design the early LeNet CNN architecture for handwriting digit recognition. These convolutional neural networks were then used in various niche markets for recognition of alphanumeric characters. However, they were highly limited in the aspect of scalability as they required very high computing resources and training data to perform efficiently on high-resolution images. Thus, they were primarily utilized only for applications that required low resolution images.

In September 2012, Alex Krizhevsky and a team of scientists from the University of Toronto designed a convolutional neural network called AlexNet that was trained on millions of labelled images from the ImageNet dataset for the “ImageNet Large Scale Visual Recognition Challenge” with an accuracy of 0.85 and the runner-up scoring an accuracy of 0.75 [57]. With vast computing resources at their disposal, the AlexNet architecture was capable of performing highly complex computer vision tasks like recognition, classification and segmentation that were previously thought to be unfeasible. The convolution neural network consisted of 8 convolution layers, max-pooling layers, 650,000 neurons, 60 million parameters and dropout layers to prevent over-fitting of the model. The model was found to be computationally expensive, as its performance largely relied on the depth of the model. Thus, highly efficient graphics processing units (GPUs) were used for training the model in order to meet its computational requirements. Since then, convolutional neural networks have advanced tremendously,

which led to the development of several other CNN architectures like ZFNet (2013) [58], Inception/-GoogleNet (2014) [59], VGGNet (2014) [60] and ResNet (2015) [61].

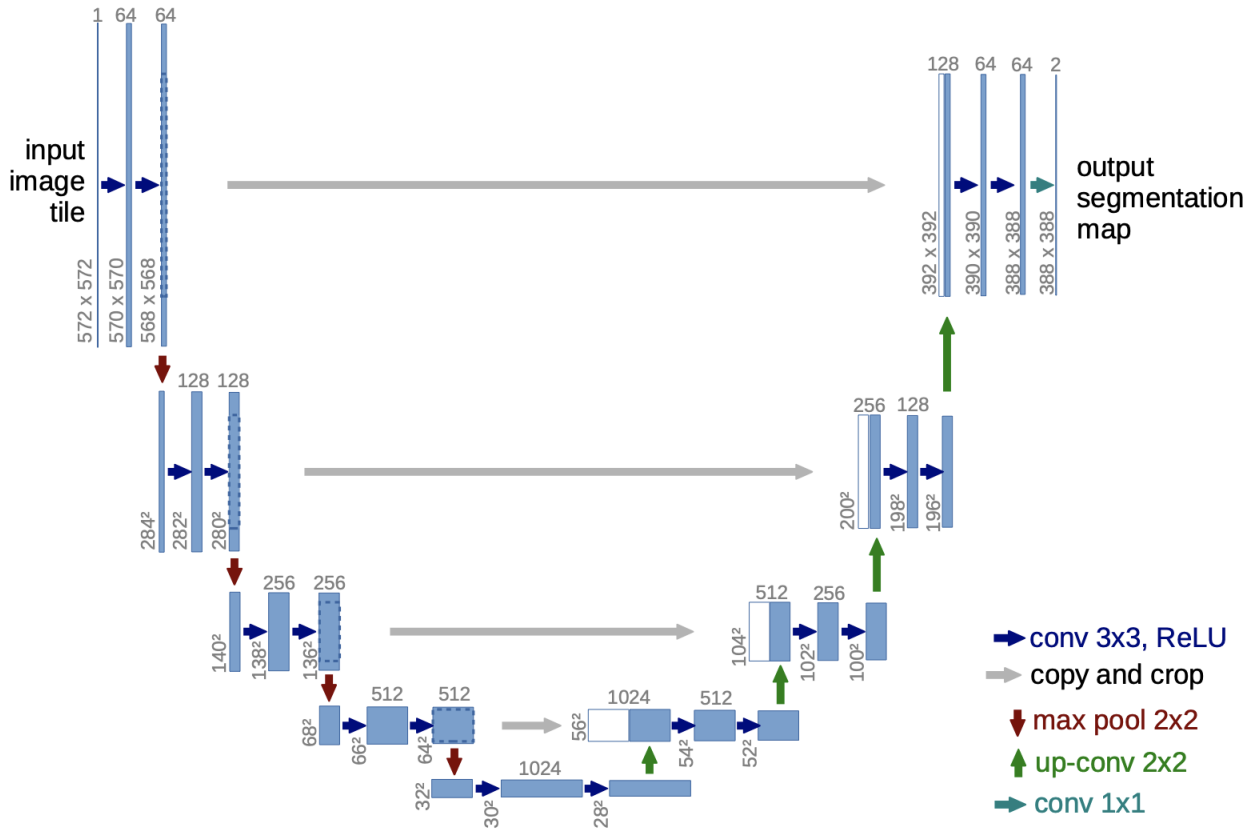

Figure 2.1: Architecture of UNet model proposed by Ronneberger et al. [1]

Convolutional neural networks have gained popularity in the field of bio-medical imaging with tasks like object detection, segmentation, classification, etc. Ronneberger et al. [1] designed a deep convolutional neural network, UNet, for the segmentation of bio-medical images which was built on the architecture of the popular "fully convolutional network" [62]. The architecture of UNet is extended to perform more efficiently with a smaller dataset of training images and shows a higher accuracy of segmentation. The regular contracting layers in the architecture are supplemented with successive layers in which up-sampling operators replace the pooling operators, thus increasing the output image resolution. High-resolution image information and features are merged with outputs from the up-sampling operators, thus allowing the consequent convolution layers to learn to generate more accurate outputs based on these high resolution features and information.

Separating objects of the same class that touch each other is a common challenge faced during biomedical applications like segmentation of cells. The architecture makes use of a weighted loss, such that the background labels that separate the region between cells in contact with each other are

assigned a greater weight and have a higher contribution to the loss function. This technique shows more promising results than that of the results obtained by the artificial neural network architecture proposed by Cirean et al. [63].

Jiang et al. [2] developed a deep, fully convolutional neural network (FCN) framework to automate the segmentation of retinal blood vessels. The framework uses a novel "Random Crop and Fill" (RCF) method of data augmentation that is used in combination with various standard data augmentation techniques to improve the results of the model. The framework shows improved performance with the use of a module for processing multi-scale inputs, is capable of diversifying into a multi-path fully connected network and obtaining the resultant segmented image by fusing multiple outputs. It also improves performance with the use of various image pre-processing techniques like image gray-scaling, gamma correction, Contrast Limited Adaptive Histogram Equalization (CLAHE) [64] and normalization.

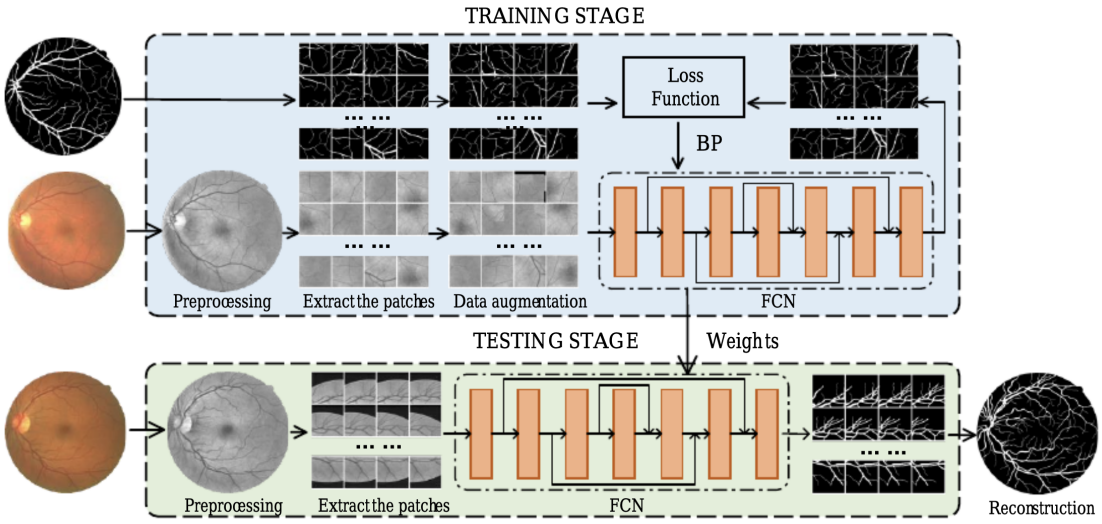

Figure 2.2: FCN architecture proposed by Jiang et al. [2]

The resultant segmented image is obtained by fusing the output patches and reconstructing them to produce a single output image. The model results surpass those of the state-of-the-art methods with an average accuracy of 0.9773, 0.9777 and 0.9706 respectively on each data set (DRIVE, STARE and CHASE_DB1 datasets). The fully convolutional neural network model also shows better robustness and generalization capability due to a deeper architecture, better data pre-processing methods, extraction of input patches, data augmentation techniques and appropriate model training strategies. The deep network architecture makes use of more parameters, which significantly increases the capability of the model and makes it possible to extend it to other diagnostic applications that involve the segmentation of symmetric and asymmetric patterns from medical images.

Fu et al. [3] approached the retinal vessel segmentation problem by formulating it as a task for boundary detection using a CNN architecture. Four multi-level CNNs make use of side-output layers to generate multi-scale hierarchical representations of the input Fundus images. Each CNN layer contains one side-output layer and multiple Rectified Linear Unit (ReLU) and convolutional layers. The technique also uses a Conditional Random Field (CRF) [65] model to efficiently represent the wide range of interactions between retinal vessel pixels and background pixels. The four CNN models and CRF model are then consolidated into one deep neural network, identified as the DeepVessel architecture [3]. Side-output layers adopt smaller receptive field sizes, thus aiding in the detection of fine-scaled retinal vessels while large scale details are represented more accurately by the deeper layers. Finally, the technique of otsu thresholding is adopted to binarize and segment retinal vessels from outputs of the DeepVessel network. Figure 2.3 shows the architecture of the DeepVessel model adopted by Fu. et. al [3].

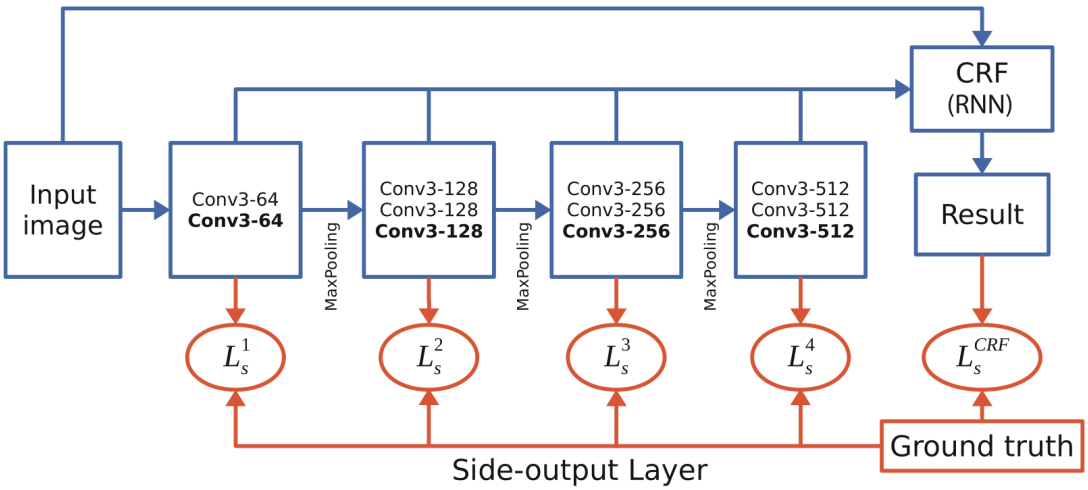

Figure 2.3: DeepVessel architecture proposed by Fu et al. [3]

The model achieves an accuracy of 0.9523 and a sensitivity of 0.7603 on the DRIVE dataset. It requires a longer training time and higher computational resources due to its multi-layered, multi-scale architecture. By fusing the results of the deep-side outputs, the DeepVessel architecture also handles green spots in Fundus images of patients with Diabetic Retinopathy better than other state-of-the-art techniques.

Chapter 3

Proposed Methodology

The implementation of our proposed methodology is based on the technique of computing Laplacian pyramids by Dachsel et al. [13]. The Laplacian Pyramid is generated by iteratively computing the pixel-wise difference of two adjacent layers of the Gaussian Pyramid. The Gaussian pyramid is computed by performing an iterative reduction operation on the original image, thus successively smoothing the image. We use the technique of Anisotropic Diffusion based on the Perona-Malik heat diffusion equation [47] for smoothing and removal of noise. Various pre-processing techniques like gamma correction, adaptive histogram equalization and edge-aware local contrasting have also been adopted to enhance the image features and to increase the discrimination between retinal blood vessel structures and the background without losing significant edge information. The subsequent equations used for this implementation are adopted from [13] and [46].

The primary objective of this study is to represent the RGB image $I_{rgb} : \Omega \rightarrow [0,1]$ where $\Omega = \{1,...N\} \times \{1,...M\}$ in a simplified form that contains the segmented retinal blood vessels, indicated by a binary segmentation map matrix, $S(i,j)$ of labels (0 or 1) by segmenting retinal blood vessel structures from the background of the image B, such that-

$$S(i,j) = \begin{cases} 1, & \text{if } (i,j) \in \Omega \cap B \\ 0, & \text{if } (i,j) \in B \end{cases} \tag{3.1}$$

where label 1 indicates pixels which are retinal blood vessel pixels and label 0 indicates background pixels. Here, $S(i,j)$ is the segmentation map of retinal blood vessels. The secondary objective of this study is to ensure that the resultant segmentation map is adequately representative of the required regions on the image. Additionally, it is also important to ensure that the resultant image retains features and topological information of the original image.

3.1 Data pre-processing

The first step in our methodology involves pre-processing the data to convert it to a form that contains an optimal representation of the image information that is significant for our task of segmenting retinal blood vessels. This is primarily done to reduce the complexity of the image, enhance image features that are of importance to our segmentation task, increase the performance of the model and discard irrelevant image information that degrades or does not improve the performance of the model. There are four stages of pre-processing used in our approach- gamma correction, grey-scaling, Contrast Limited Adaptive Histogram Equalization (CLAHE) and edge-aware local contrasting. The image is first pre-processed with gamma correction to encode and decode luminance values in the image. It is a non-linear function which is highly responsive to the corresponding differences between low-contrast regions than high contrast regions in the image, effectively enhancing regions that are of significance and diminishing regions that are of no significance to our application. The gamma correction function reduces pixel intensity values in the lower grey-scale areas and increases pixel intensity values in the higher grey-scale areas, thus increasing the contrast ratio of the image [14]. The original image and the image after gamma correction can be seen in Figure 3.1.

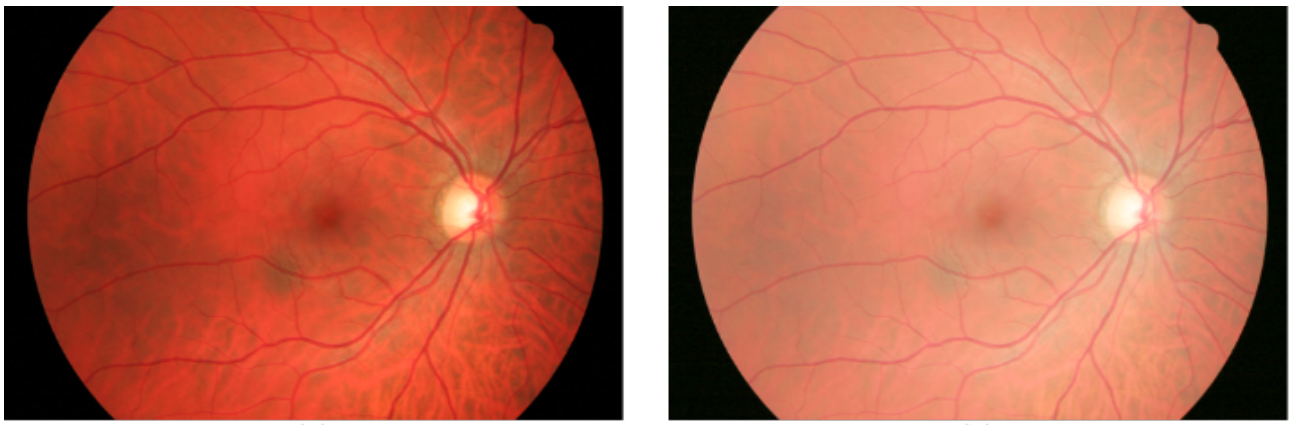

Figure 3.1: (a) Input retinal Fundus image (b) Visualization of the image after gamma correction

The resultant RGB image after gamma correction is converted to a mono-channel grey-scale image. Figure 3.2 shows the three monochrome channels (Red, Green and Blue channel) extracted from the original image. Figure 3.2 shows that the Red channel has the least degree of discrimination between the background and the retinal vessel structures, while the Blue channel is relatively more representative of the vessel structures. The Green channel is found to have the highest degree of discrimination between the retinal blood vessel structures and the background.

In comparison with the Green monochrome channel, the Red and Blue channels are found to have very low contrast and high levels of noise. The formula given in Equation 3.2 was used to convert the RGB image, I_{rgb} into a monochrome Green channel image-

$$I_{Grey} = 0.299r + 0.587g + 0.114b \tag{3.2}$$

where *r* represents the Red channel, *g* represents the Green channel and *b* represents the Blue channel. As per the equation, the contribution of the Green channel to the grey-scale image is the highest, with the Green channel contributing 58.7%, the Red channel contributing 29.9% and the Blue channel contributing 11.4% to the grey-scale image.

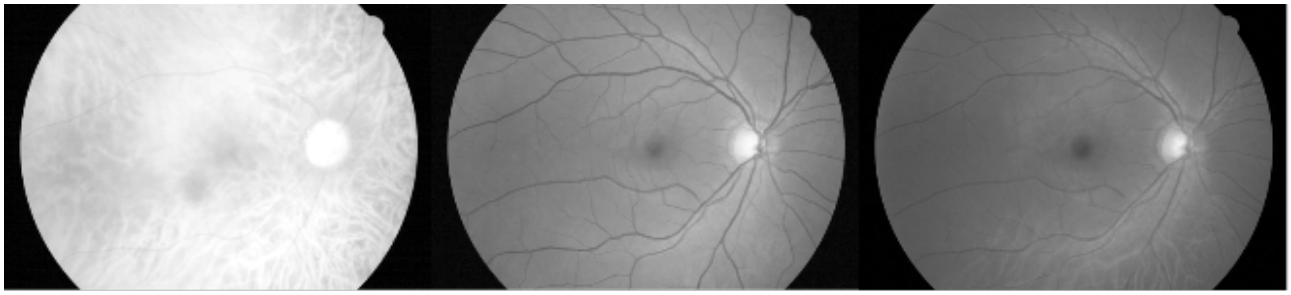

Figure 3.2: (a) Red Channel (b) Green Channel (c) Blue Channel

The next pre-processing step adopts the use of a type of histogram equalization, called Contrast Limited Adaptive Histogram Equalization (CLAHE) [15]. This technique aids in improving the contrast of the image by computing multiple histograms relative to discrete regions in the image and using these histograms to stretch out and equalize the intensity values across the image. Thus, it enhances edge information in every region of the image and improves local contrast. CLAHE is more effective than traditional Adaptive Histogram Equalization (AHE) in this application because AHE tends to over-amplify noise in regions with same or similar contrasts, which is solved with the CLAHE technique by clipping regions of the histogram that exceed a clip limit and redistributing it proportionally across all the histogram bins.

Figure 3.3 shows a comparison between the green channel image and the image after performing CLAHE operation along with their histograms. From the two histograms in Figure 3.3, it can be clearly seen that the CLAHE operation has stretched out and equalized the intensity values across the image.

While the Green channel image histogram shows a concentration of pixel intensity values between 100 and 200, the histogram of the image after the CLAHE operation shows that the pixel intensity values

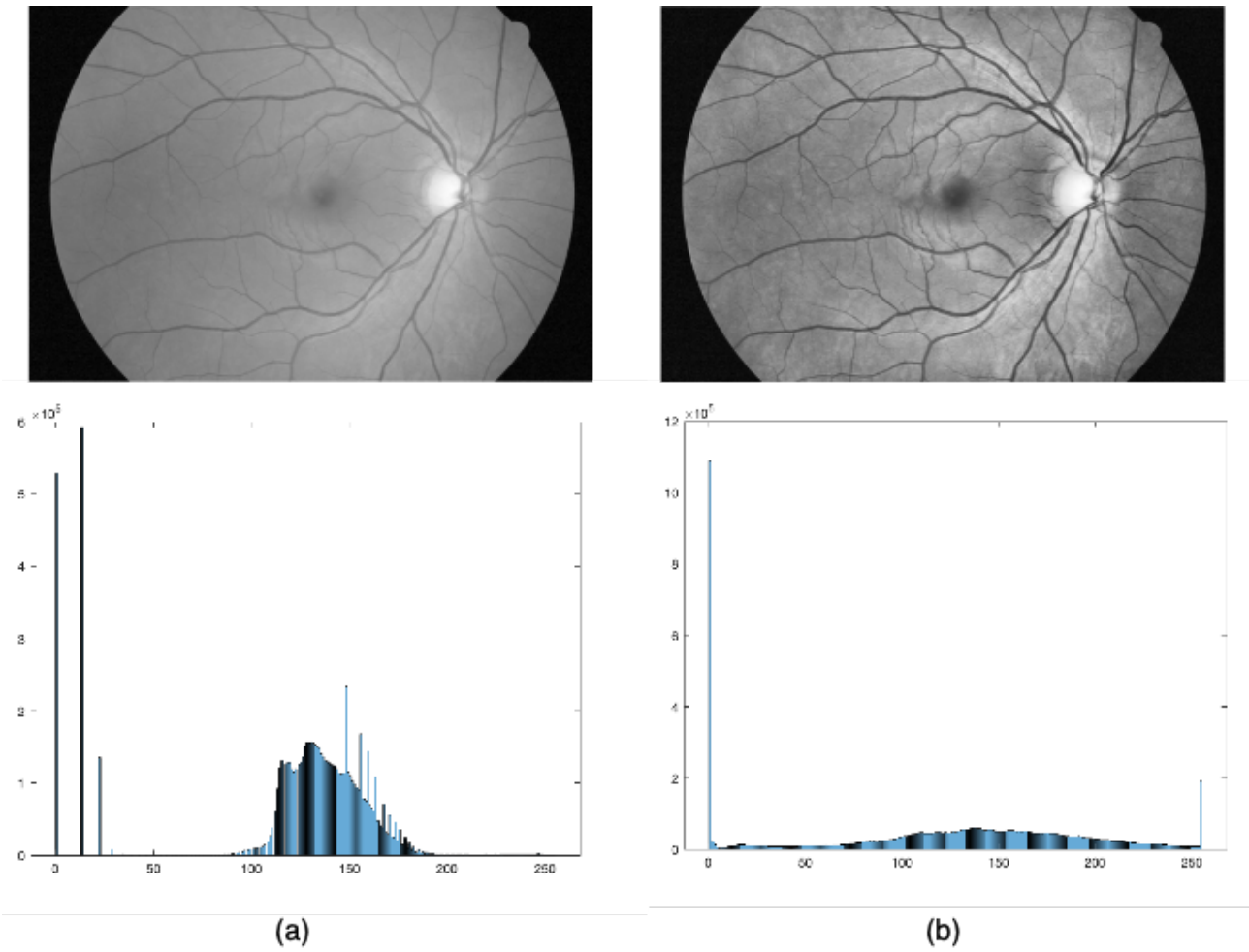

Figure 3.3: (a) Green Channel Histogram (b) Histogram after CLAHE

have been equalized and spread out across the range from 0 to 255, thus enhancing the contrast between the background and the foreground in the image.

In the next pre-processing step, we use an edge-aware local contrasting technique to smooth details in the image and differentiate between small-scale and large-scale edges. This is done by defining an edge threshold on the pixel intensity values, thereby enhancing the contrast in the image without developing halos in high contrast regions or degrading important edge information. We also make use of an intensity threshold to regulate the intensity of contrasting performed on the image.

The technique of edge-aware local contrasting was adopted to further increase the discrimination between contrast levels of retinal blood vessel structure pixels and the background pixels. It is also the most computationally heavy step in our proposed algorithm as it takes up 55.17% of the total computation time. Retinal vessel edges have now been enhanced and appear to be more well-defined, thus making even the finer retinal structures more visible. This increases the sensitivity and accuracy of our segmentation task.

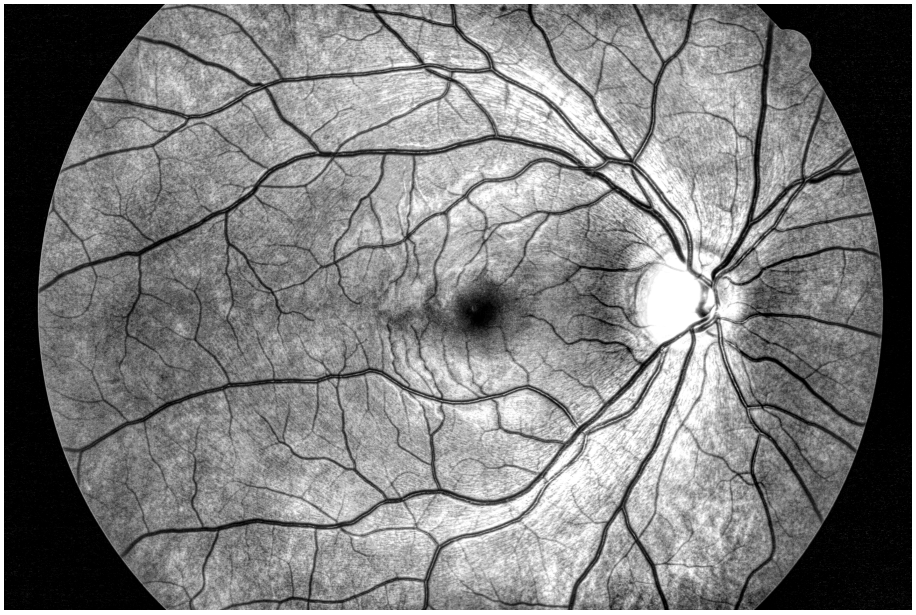

Figure 3.4: Visualization of the image after edge-aware local contrasting

3.2 Anisotropic Diffusion

The resultant image of the contrasting operation was found to be noisy, which would significantly hinder the performance of our skeletonization algorithm. Traditional techniques of smoothing and noise removal resulted in further degradation of the performance of our model as significant edge information was lost. Thus, we use an improvised version of the two-dimensional Gaussian filter, called Anisotropic diffusion. Anisotropic diffusion is a space variant, non-linear transformation performed on the original image for smoothing and removing noise from the image without degrading significant features and information in the image like edges, contours, ridges, etc [47]. This is done by generating a sequence of blurred images through an iterative process of diffusion. The sequence of images are convolved with a filter that is dependent on local pixel information of the original image to generate the resultant diffused image.

The rate of diffusion is regulated by a diffusion coefficient, *c* which is defined as a function of the gradient of the image, ∇. The implementation of anisotropic diffusion is derived from the Perona-Malik heat diffusion equation by defining the diffusion coefficient, c as an edge-seeking function to diffuse independent regions in the image without diffusing strong edges in the images, such as those on the retinal blood vessel structures [46]. By convention, ∇ represents the image gradient, Δ represents the image Laplacian, *div()* represents the divergence operator, *K* represents the edge sensitivity function used to regulate the sensitivity of the diffusion process to edges in the image, and *c (x,y,t)* represents

the diffusion coefficient used to regulate the rate of diffusion. Let the family of grey-scale images be represented by $I(.,t) : \Omega \rightarrow R$, such that $\Omega \subset \mathbb{R}^2$ represents a sub-set of the plane, we define the anisotropic diffusion process as follows-

$$\frac{dI}{dt} = div\ (\ c(x,y,t) \cdot \nabla I\ ) = \nabla c \cdot \nabla I + c(x,y,t) \cdot \Delta I \tag{3.3}$$

where the coefficient of diffusion, *c* is defined as an exponential function-

$$c = \frac{e^{-(\ ||\ \nabla I\ ||\ /\ K)^2}}{(\ ||\nabla I\ ||)} \tag{3.4}$$

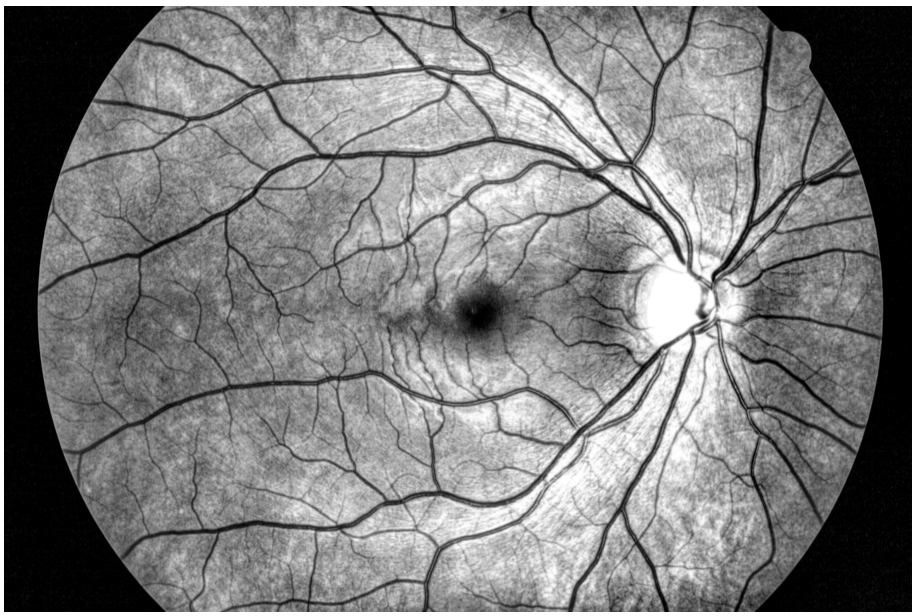

Figure 3.5: Visualization of the image after anisotropic diffusion

We carry out anisotropic diffusion through an estimation of the reducible Perona-Malik diffusion equation. The gradient at every pixel point is computed by taking the difference between that pixel and its neighbouring pixels. The image is then diffused by applying Equation 3.3 and 3.4 to the previously diffused image in an iterative manner. This iterative diffusion process is carried out till the desired level of noise is removed. The technique of anisotropic diffusion also functions identical to that of Gaussian blurring when the diffusion coefficient remains a constant as the equation for anisotropic reduces to the linear heat equation. While this technique removes noise, it is unsuitable for our purpose as it results in the blurring of edges. Thus, we define the diffusion coefficient as an edge seeking function to diffuse all the pixels surrounding each pixel while retaining significant curvatures and edges along the shape of the object.

3.3 Gaussian Pyramid

Since the retinal blood vessel structures are of varying width and proportions, we make use of a pyramid representation of the image, called Gaussian Pyramid to inspect and analyze the image at different scales [54]. Fine retinal blood vessel structures or high level information require a high resolution for efficient detection and analysis, whereas coarse retinal blood vessel structures or low level information can be efficiently analysed even at low resolutions. Thus, we adopt a multi-scale Gaussian Pyramid representation of the image presented by Dachsel et al. [13], which comprises of *k* layers of images with sequentially shrinking resolutions. We first make use of a separable, two-dimensional convolutional filter kernel, h_α to perform a convolution operation on the diffused image, such that $h_\alpha \in R^{5\times5}$. The convolution operation aids not only in sharpening and denoising the image, but it also performs edge detection. Equation 3.5 describes the convolution operation between the grey-scale diffused image *I* and two dimensional filter kernel, h_α -

$$(h_\alpha \cdot I)(i,j) = \sum_{a=-2}^{2} \sum_{b=-2}^{2} h_\alpha(3+a,3+b) \cdot I(i-a,j-b) \tag{3.5}$$

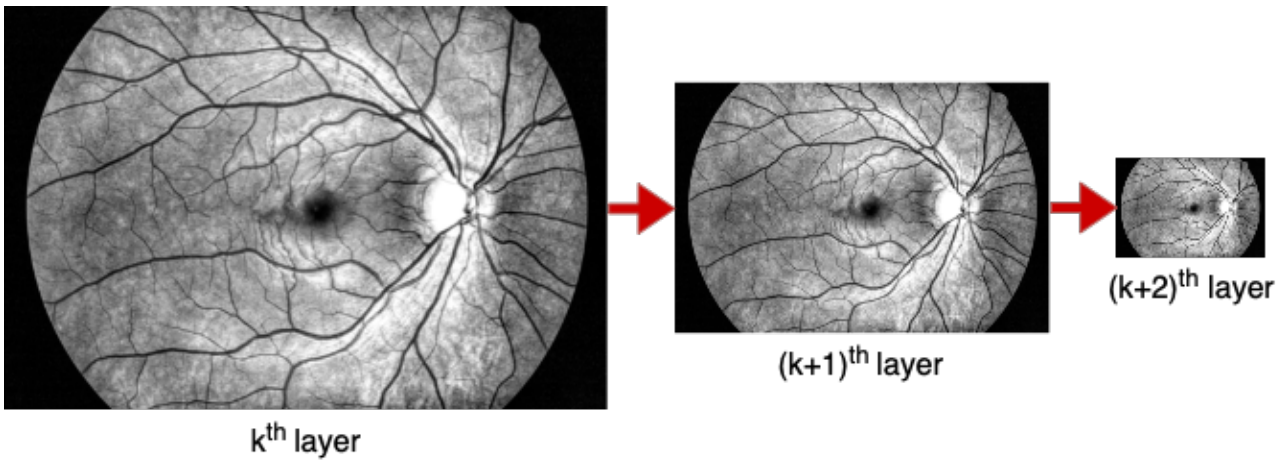

Figure 3.6: Gaussian Pyramid

The dimensions of the filter kernel are chosen experimentally based on the accuracy of the segmentation operation, i.e. the kernel size that results in the highest accuracy is chosen. We use the multi-scale representation of *I* to express the Gaussian Pyramid as a sequence of images of shrinking resolutions, denoted by $(G_0, G_1 ..., G_k)$.

$$\hat{R}(G_k)(i,j) = \sum_{a=-2}^{2} \sum_{b=-2}^{2} h_\alpha(3+a,3+b) \cdot G_k(i-a,j-b) \tag{3.6}$$

$$G_{k+1} = \hat{R} \cdot (G_k) \tag{3.7}$$

The first layer of the Gaussian Pyramid, G_0 has the highest resolution of the original image, while the last layer of the Gaussian pyramid, G_k has the lowest resolution of the original image. We make use of a reduction operator, $\hat{R}$ to down-sample the image from scale G_k to G_{k+1} as shown in Equation 3.6. The grey-scale image *I* is used as the first layer of the Gaussian pyramid and the remaining layers are generated recursively by using the reduction operator defined in Equation 3.7.

3.4 Laplacian Pyramid

The next step in our proposed technique involves computing the Laplacian Pyramid from the Gaussian Pyramid generated in the previous step. The Laplacian Pyramid is used to extract information related to all the scales of the image. The first four layers, (G_0, G_1, G_2, G_3) of the Gaussian pyramid are used as the foundation for computing the Laplacian pyramid. The pixel-wise difference of two consecutive Gaussian layers from the Gaussian Pyramid is computed to construct the three Laplacian layers, (L_1, L_2, L_3) of the Laplacian Pyramid using Equation 3.8.

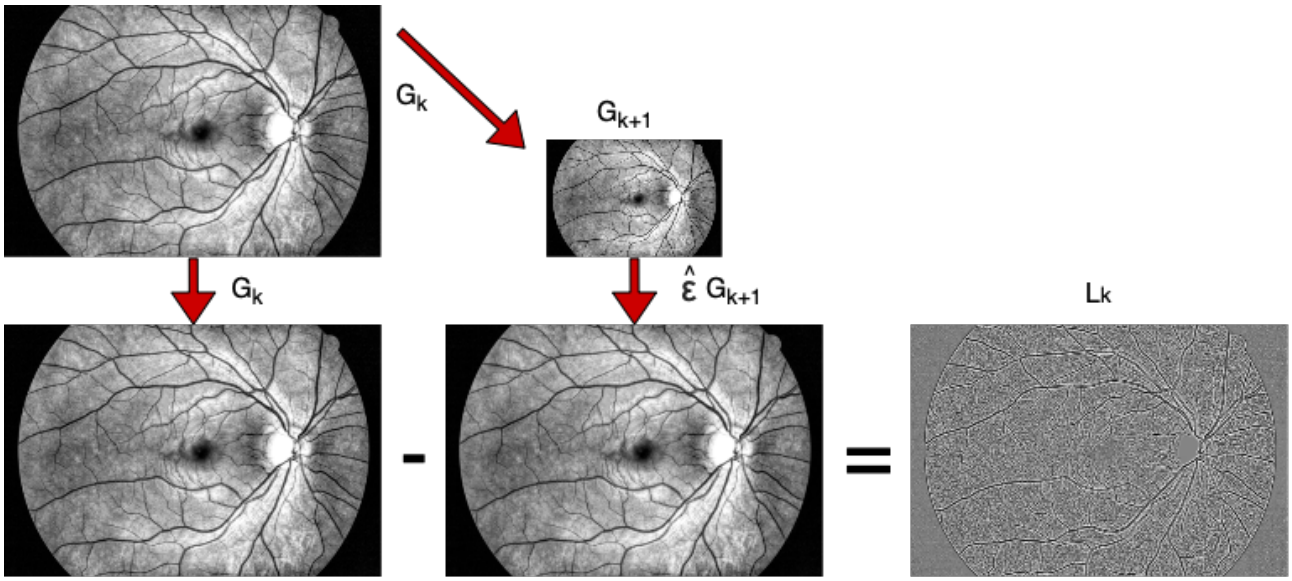

Figure 3.7: Computation of Laplacian layers

$$L_k = G_k - \hat{\varepsilon} \cdot (G_{k+1}), \qquad k = 0, 1, 2...r-1. \tag{3.8}$$

where $\hat{\varepsilon}$ is the expansion operator used to expand the resolution of layer G_{k+1} to G_k by performing

an interpolation operation on the missing pixels. The Interpolation operation is bi-linear in nature and acts as a convolution filter performing an expansion operation, defined by Equation 3.9 given below -

$$\hat{\varepsilon} \cdot (G_k)(i,j) = 4 \times \sum_{a=-2}^{2} \sum_{b=-2}^{2} h_\alpha(3+a,3+b) \cdot G_{k+1}\left(\frac{i+a}{2}, \frac{j+b}{2}\right) \tag{3.9}$$

The first Laplacian layer, L_0 and layers after L_3 are discarded as they do not contain information of consequence for our task of segmenting retinal blood vessel structures from background pixels. The last stage of this step involves using histogram equalization to redistribute pixel intensity values equally between 0 to 255 with 512 discreet quantization steps. We do not construct the complete Laplacian pyramid as we only focus on the pertinent range of scales that are essential for our purpose, thus reducing the computation cost of constructing the Laplacian Pyramid. The resultant Laplacian Pyramid is shown in Figure 3.8.

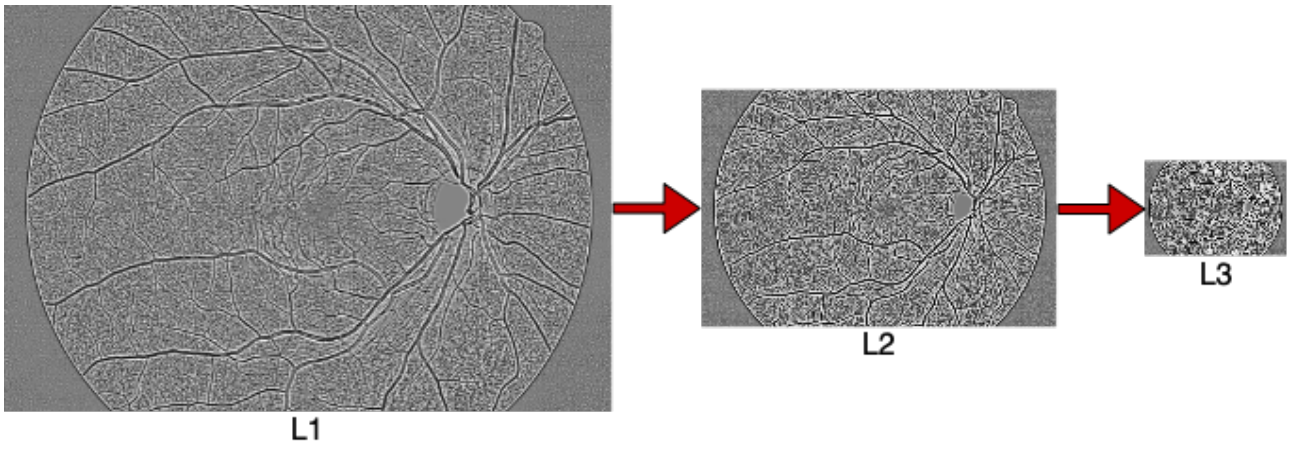

Figure 3.8: Laplacian Pyramid

The Laplacian layers resemble the result of using an efficient Band-pass filter (BPF) on the original input image due to the pixel-wise difference between two adjacent Gaussian layers with different levels of smoothness. From Figure 3.8, we can see that the Laplacian layers represent the original input image in a way that is beneficial for our task of retinal vessel segmentation. On close observation, it can be seen that the thick, coarse vessel structures are retained in the resultant image while the small, insignificant structures become diminished. Thus, this aids us in segmenting and analysing retinal blood vessel structures on various scales, from fine retinal blood vessel structures to coarse retinal blood vessel structures.

3.5 Binarization and discarding disconnected pixels

After generating the Laplacian pyramid for the retinal scan image, it is now adequately processed for the segmentation task. The final stage of the segmentation task involves binarization of the image to distinguish the retinal blood vessel pixels from the background pixels. Binarization techniques like adaptive thresholding, global thresholding and local thresholding are popularly used in segmentation tasks. Background pixels are separated from potential retinal blood vessel pixels by making use of a global threshold of intensity z^*, such that $\Omega_k^* = (i,j) \in \Omega_k : L_k(i,j) \leq z^*$ represents the set of potential retinal blood vessel pixel positions.

The biggest challenge with binarization of images for segmentation tasks is the estimation of an appropriate threshold to separate the foreground or object from the background [66]. Retinal blood vessels account for approximately 10% of the total pixels in the image. Background pixels are lighter and have low contrast levels compared to blood vessel structures, which are darker and have higher contrast than the background. Thus, these blood vessel structures can be detected at low intensity values compared to the background pixels. The pixel intensity values are sorted in ascending order and an intensity threshold, β is used to filter out pixels whose intensity values exceed the intensity threshold. The largest intensity value from the sorted list of pixel intensity values is considered to be the global intensity threshold, z^*. We represent z^* as follows in Equation 3.10 -

$$z^* = max\{z_1, z_2, z_3..., z_{512}\} \tag{3.10}$$

where $z_1, z_2, z_3..., z_{512}$ are the intensity values of all the pixels that fall under the local intensity threshold, β sorted in ascending order. For equation 3.10 to hold good, it must satisfy the condition mentioned in Equation 3.11 -

$$\sum_{l=1}^{512} \sum_{i=1}^{N/2^k} \sum_{j=1}^{M/2^k} \delta_{z_1, L_k(i,j)} \leq \frac{\beta \cdot N \cdot M}{2^k} \tag{3.11}$$

where *N* and *M* represent the number of pixel rows and columns respectively and the Kronecker Delta, $\delta_{z_1, L_k(i,j)}$ is a bi-variable function that is equal to 1 if both the variables are equal and 0 if the variables are unequal. Another noteworthy observation is that the use of a local intensity threshold instead of a global intensity threshold yields lower segmentation performance and higher computation times. This is due to the fact that local thresholding works best on uniform neighbors and does not work well in case of asymmetrical patterns of the retinal blood vessels.

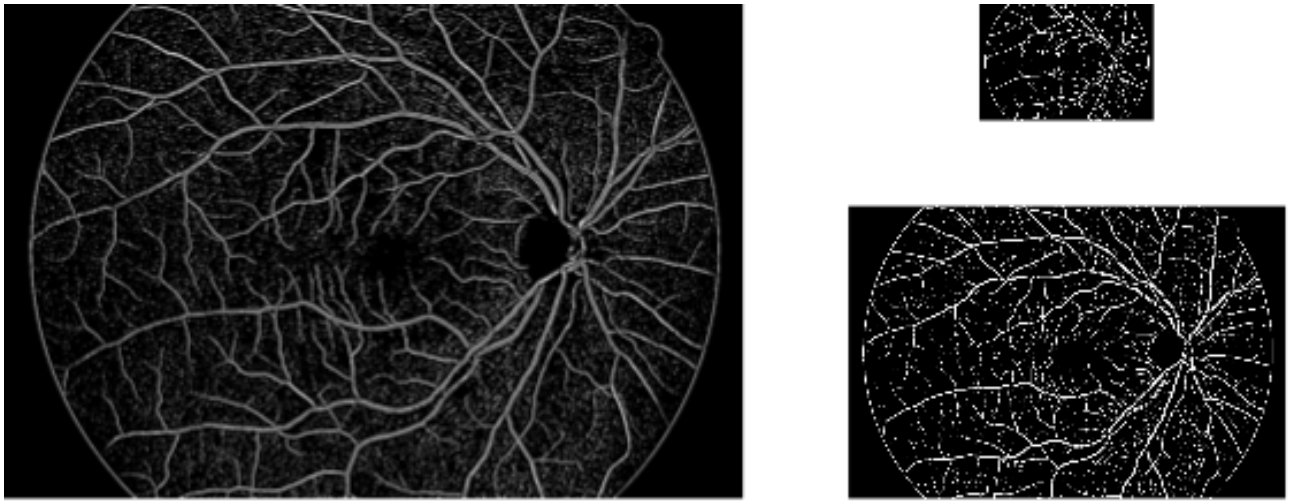

Figure 3.9: Binarization of Laplacian layers using intensity thresholding

Once the retinal blood vessels are segmented from the background of the image, we label all the pixels in δ^* that have adjoining pixels in either the diagonal, horizontal or vertical direction as connected pixels. Pixels that share either a corner or an edge are considered as connected components. The size of the objects formed by the connected pixels, *d* is computed and a threshold size, γ is used to discard all the connected pixel objects whose size falls under the given threshold. As a result, all the connected pixels that are part of the background and not of the retinal vessels are discarded. The segmentation map is refined using Equation 3.12 and the resultant image after discarding the disconnected components is shown in Figure 3.10.

$$\Gamma_k = \left\{ (i,j) \in \Omega_k^* : d \leq \frac{\gamma}{\sqrt{N \cdot M \cdot 2^k}} \right\} \tag{3.12}$$

where γ is the threshold size, Ω_k^* is the domain of connected pixels and *d* is the size of the connected object. Finally, we define the general binary skeleton map for the k^{th} layer with a set exclusion operation as follows -

$$S_k(i,j) = \begin{cases} 1, & \text{for } (i,j) \in \Omega_k^* \setminus \Gamma_k \\ 0, & \text{else} \end{cases} \tag{3.13}$$

The resultant image layers are fully segmented, noise-free, and do not contain spurious branches. The technique of using a threshold to discard disconnected pixels and small connected objects also achieves the task of skeleton pruning. The technique is also computationally more efficient than other state-of-the-art skeleton pruning techniques.

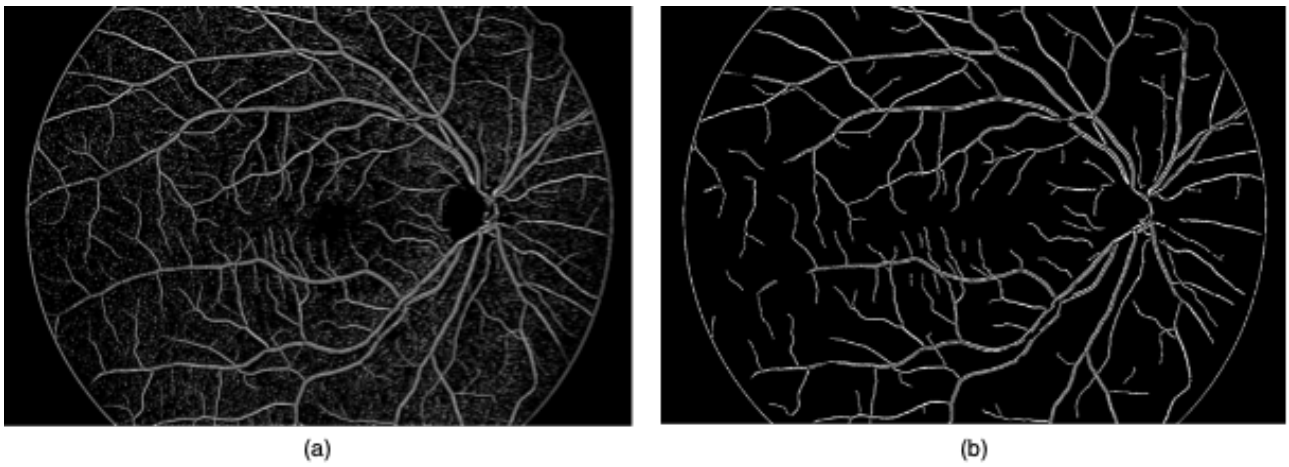

Figure 3.10: (a) Segmented image after binarization (b) Segmented layer S_k

3.6 Re-scaling and fusion of Layers

The last step of our proposed technique involves re-scaling the layers of the pyramid to the resolution of the original image. This re-scaling process is done by using the expansion operator defined in Equation 3.9. This is done in order to bring all the segmented layers of various scales to the same resolution. The re-scaled images can be seen in Figure 3.11.

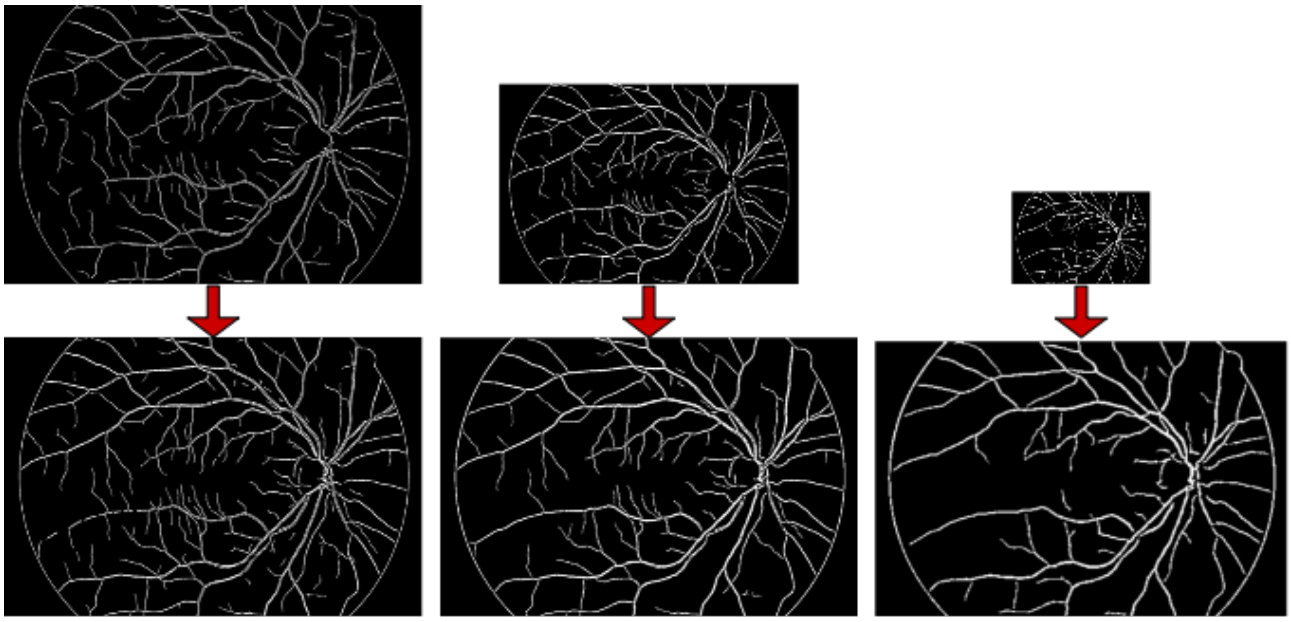

Figure 3.11: Re-scaled, segmented layers

Finally, we complete the task by fusing the re-scaled layers of the pyramid to generate a single image consisting of the extracted blood vessels from the input Fundus image. The fusion operation was computed by using a logical conjugation operation as shown in Equation 3.14. The resultant segmentation

map is effectively representative of the retinal blood vessels in the original Fundus image and retains all of its significant features and topology. Figure 3.12 shows the resultant output segmented image using our proposed technique. Figure 3.13 shows the workflow pipeline of our proposed technique.

$$S(i,j) = \begin{cases} 1, & \text{for } \hat{\varepsilon} \cdot S_1(i,j) = 1 \vee \hat{\varepsilon^2} \cdot S_2(i,j) = 1 \vee \hat{\varepsilon^3} \cdot S_3(i,j) = 1 \\ 0, & \text{else} \end{cases} \tag{3.14}$$

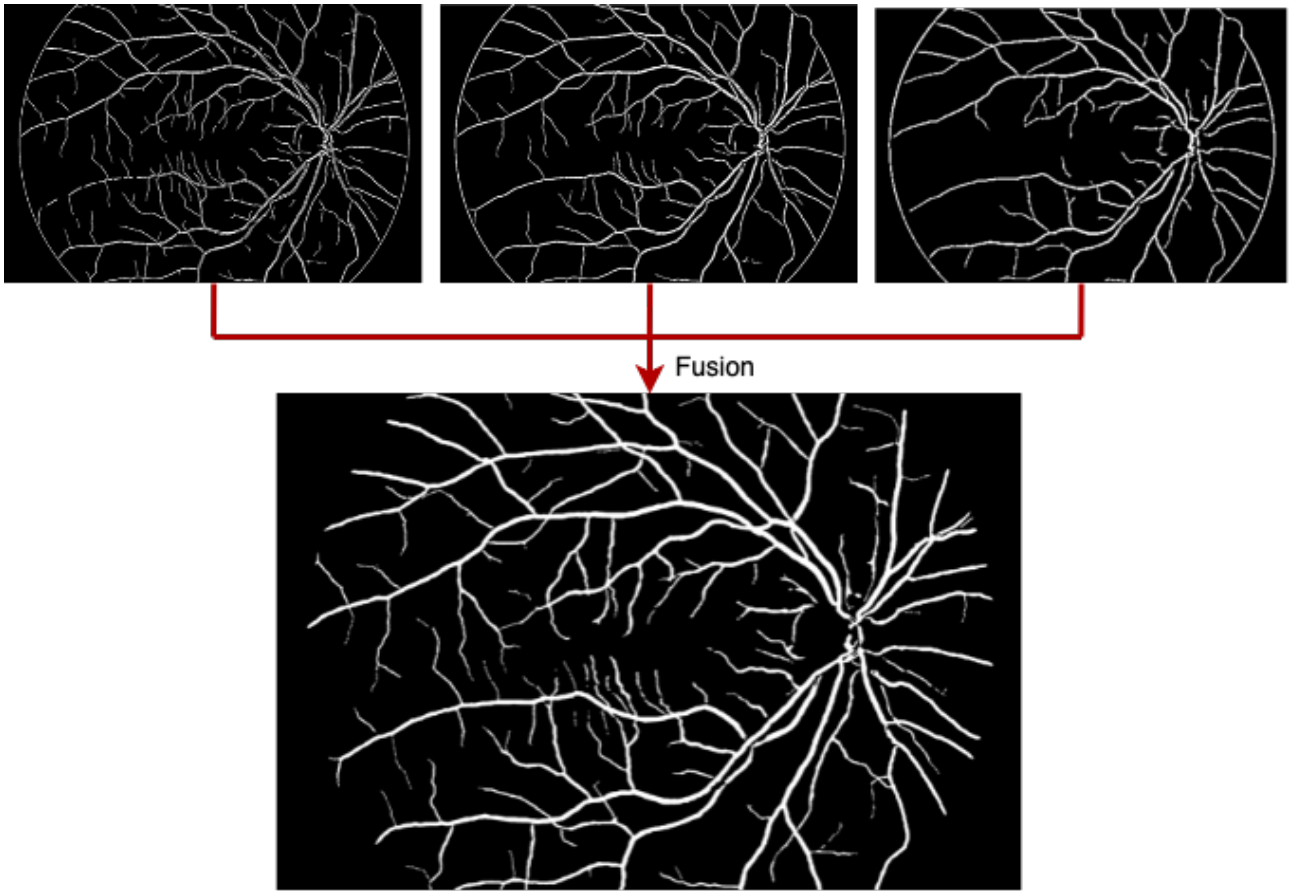

Figure 3.12: Resultant image after fusing the segmented layers (Resized for better visualization)

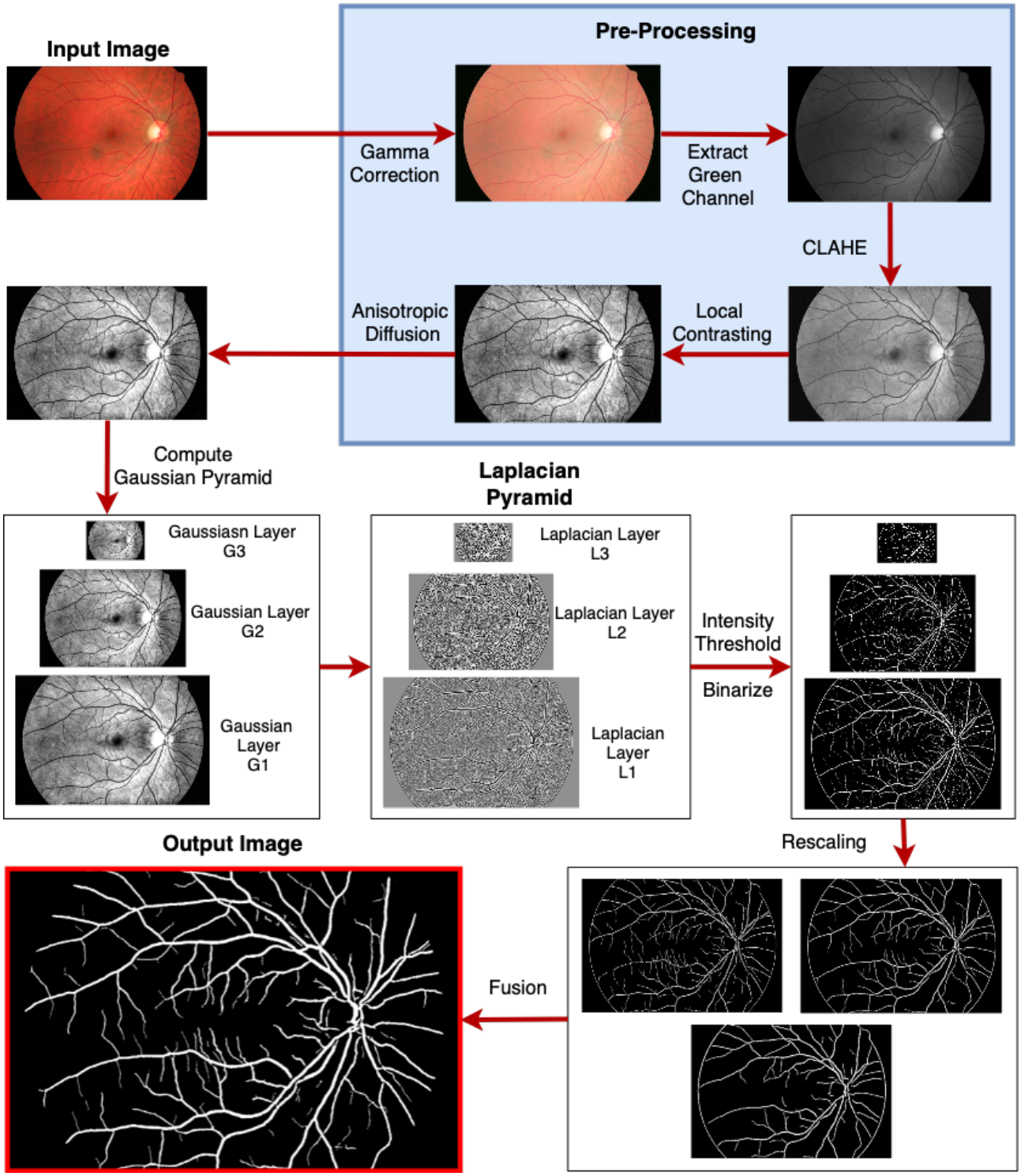

Figure 3.13: Workflow of our proposed retinal vessel extraction technique

Chapter 4

Experiments

This section contains the quantitative and qualitative results of our algorithm, along with some sample output results of the computations made in the algorithm. Initial experiments with our technique were first tested on low-resolution images from the DRIVE dataset. This was primarily done to study different pre-processing and segmentation techniques and their effects on the quality of segmentation, the accuracy, specificity and sensitivity of our model. The algorithm was then used to extract retinal blood vessels from three categories of retinal scans of patients (Healthy, Glaucoma, Diabetic Retinopathy). The experiments were conducted on 5 randomly selected images from each category of patient in the High-Resolution Retinal Fundus dataset and average performance figures were computed to provide a generalized understanding of the effect of various pre-processing steps in the technique. Images from the DRIVE dataset were resized from a resolution of 565 $\times$ 586 to 2048 $\times$ 2048 to generate accurate representations of the images with the Gaussian and Laplacian Pyramids.

4.1 Experimental Setup

The programming logic for this study was written using Matlab due to the vast range of image processing features and tool-kits offered by the platform. Some aspects of the implementation were adapted from Dachsel et al. [13]. The experiments were conducted on a 2.3 GHz quad-core Macintosh system with mac-OS Catalina 10.5.2, an i5 processor and an 8 gb RAM. The algorithm computes in 10.28 seconds on the DRIVE dataset and 14.54 seconds on the High-Resolution Fundus dataset, and the same hardware setup was used throughout the experiments. The application of the algorithm can be extended to include a GPU-based architecture to support further complexities in the algorithm as well as faster computation times.

4.2 Experimental Results

Figure 4.1 shows the comparison of the results of the algorithm on sample images from the DRIVE dataset with two manually annotated ground truth images. The results of the algorithm were computed for 10 randomly chosen images and the average sensitivity, specificity and accuracy were 74.63%, 97.57% and 95.65% respectively.

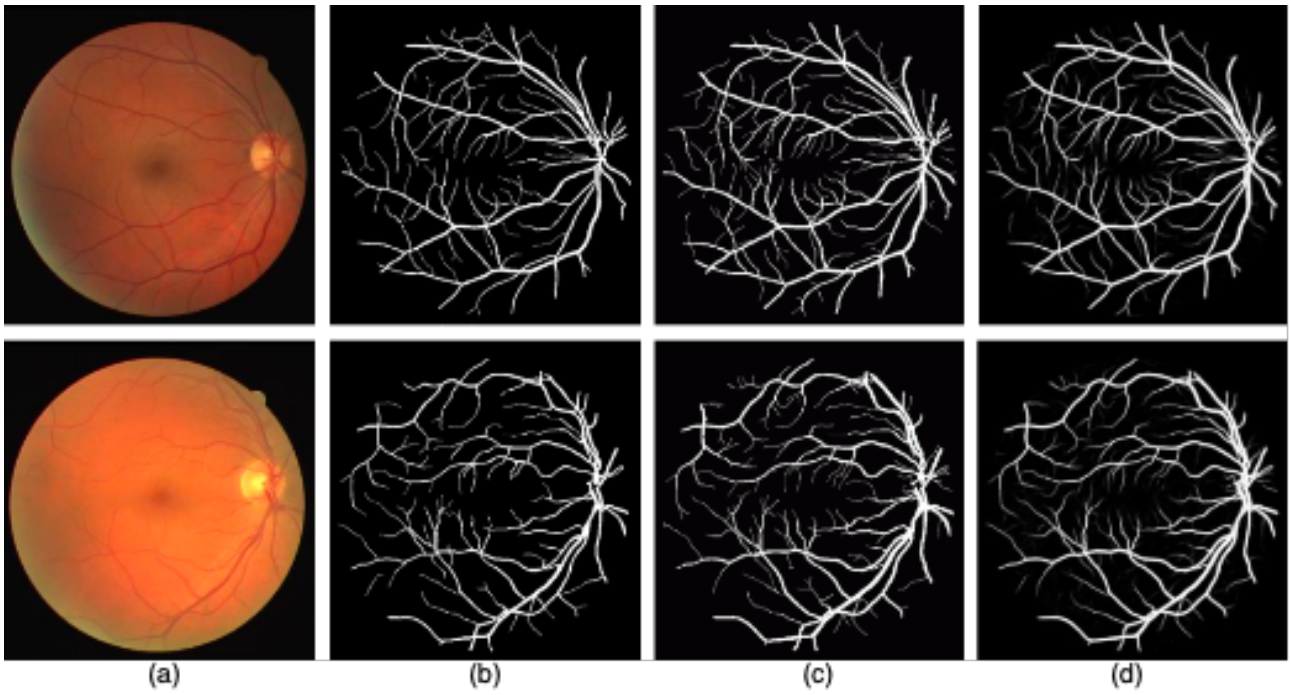

Figure 4.1: (a) Original image (b) Ground Truth (c) 2^{nd} manual annotation (c) Resultant image

The next phase of our experiment involved implementing the algorithm with High-Resolution Retinal Fundus images with a resolution of 3504 × 2336. The input dimensions of the image were preserved as they were adequate for extracting relevant level-set information from the Gaussian and Laplacian Pyramids. The results were computed separately for 5 images each from all three categories, and were averaged for a better understanding of the variations in classification patterns between images from healthy, glaucoma and diabetic retinopathy categories. Figures 4.2 and 4.3 show a comparison between the original images, ground truth images and resultant segmented images of healthy patients and patients with Glaucoma.

The performance of the algorithm on scans of healthy patients and patients with Glaucoma are almost the same. The algorithm efficiently extracts the retinal blood vessels with an accuracy of 95.85%, sensitivity of 79.33% and specificity of 97.54%. The results of the same are shown in Table 5.2. The model shows significant improvements with increase in resolution and quality of imaging. However,

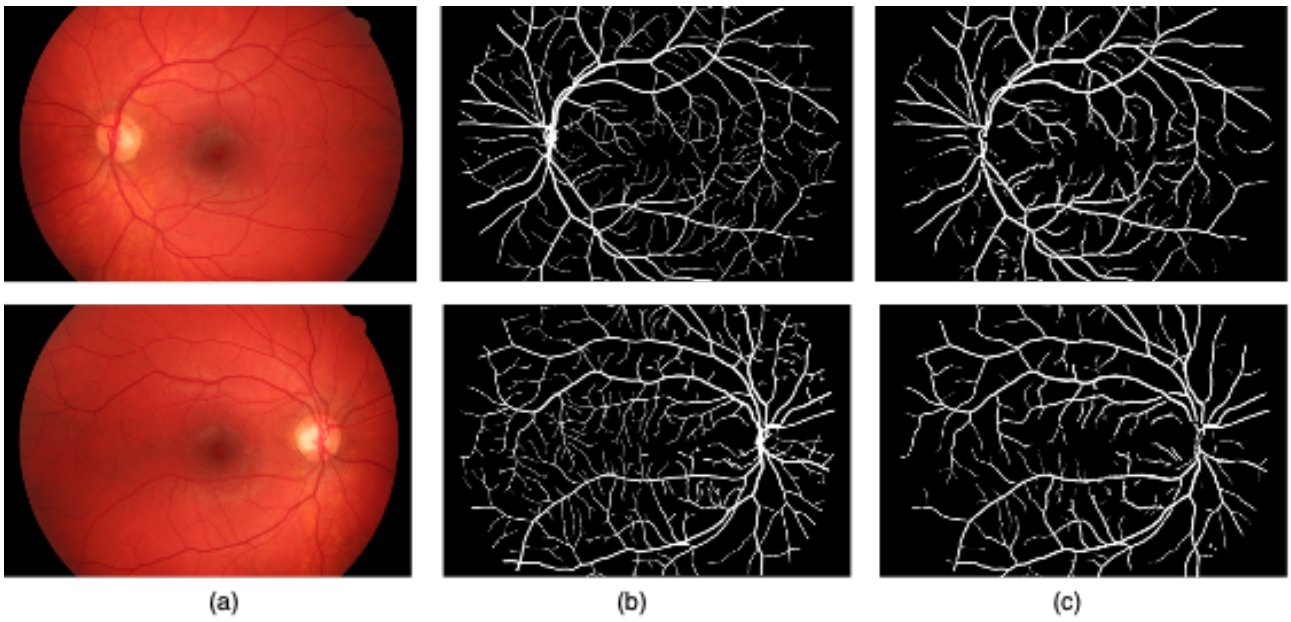

Figure 4.2: Healthy Fundus scans: (a) Original image (b) Ground truth (c) Resultant image

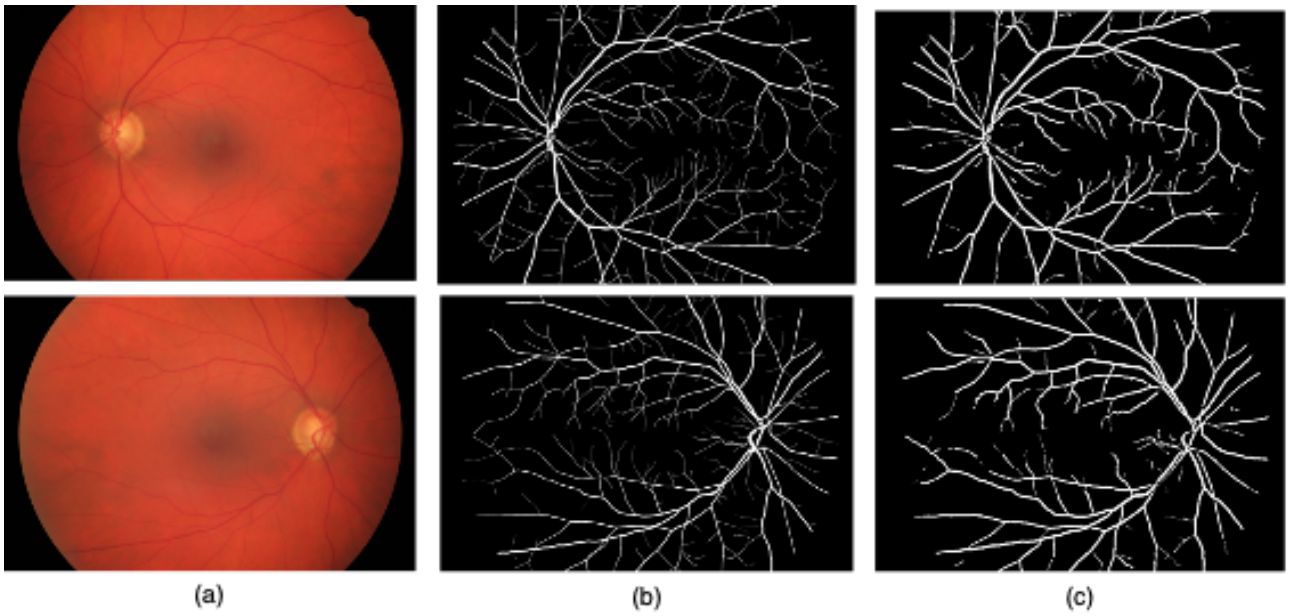

Figure 4.3: Glaucoma Fundus scans: (a) Original image (b) Ground truth (c) Resultant image

the model tends to show a reduction in performance with images of patients with Diabetic Retinopathy. Figures 4.4 shows a comparison between the original images, ground truth images and resultant images of patients with Diabetic Retinopathy due to green spots in the background. These green spots are enhanced while extracting the green channel of the image, thus having a higher contrast than the background pixels. From the figure, it can be seen clearly that the green spots on the original image are wrongly being classified as retinal blood vessel pixels, thus causing an increase in the number of False Positives and the FPR.

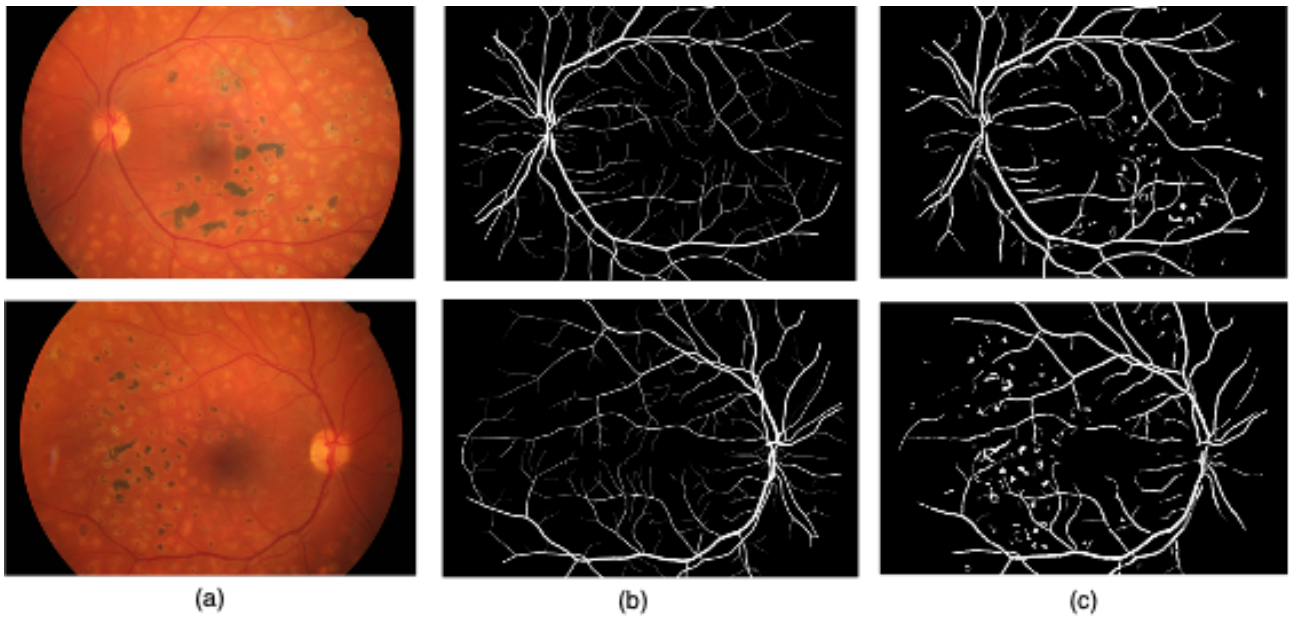

Figure 4.4: DR Fundus scans: (a) Original image (b) Ground truth (c) Resultant image

We found that the technique of edge-aware local contrasting resulted in a significant improvement in the precision, recall, sensitivity, specificity and accuracy of the algorithm. However, it should be noted that sub-optimal tuning of the parameters leads to a reduction of these performance figures. A low value of edge threshold and intensity leads to a reduction in the precision of the algorithm. A high value of edge threshold, *ET* and intensity, *I* leads to a highly saturated image, loss in significant image features, high noise levels and an inaccurate representation of the original image. Thus, these parameters were tuned and maintained as constants throughout our experiments.

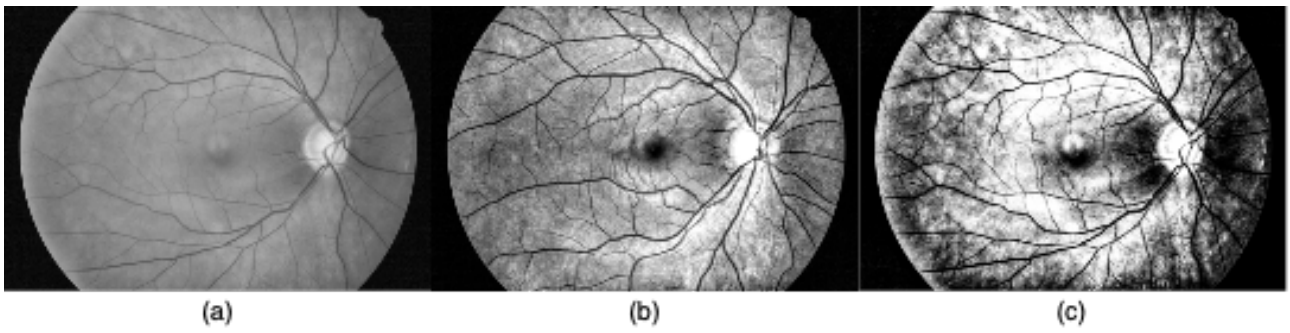

Figure 4.5: (a) ET = 0.1, I = 0.1 (b) ET = 0.6, I = 0.4 (c) ET = 0.9, I = 0.9

The parameters were chosen experimentally and the most optimal values of edge threshold and intensity were found to be 0.60 and 0.40 respectively. Figure 4.5 shows the resultant image with different parameter values for edge threshold and intensity. We can see clearly from the images that a low value of *ET* = 0.1 and *I* = 0.1 results in negligible visible changes to the contrast levels in the image.

The resulting image with ET = 0.9 and I = 0.9 shows high levels of saturation with significant loss of important features in the image. It has high levels of noise, the algorithm fails to detect tiny retinal blood vessels accurately, and the connectivity of retinal vessels is broken due to the formation of shadow regions in the image. Figure 4.5 (b) shows the resultant of edge-aware local contrasting with optimally tuned parameters, i.e. ET = 0.60 and I = 0.40. Additionally, we observed that increasing the edge threshold by 0.05 leads to an increase in the specificity and reduction in the sensitivity of the classification. Similarly, reducing the edge threshold by 0.05 leads to an increase in the sensitivity and reduction in the specificity of the classification, indicating an inverse relationship between the sensitivity and specificity. The same condition was observed with precision and recall, indicating an inverse relationship between them. Thus, the parameter values were carefully chosen to balance the trade-off between these performance measures. This step of pre-processing was the most computationally intensive step in our proposed technique, taking up almost 55% of the computation time.

One of the biggest challenges in the extraction of retinal images is image noise. This challenge is amplified when we perform contrasting operations on the image as they generate more noise. While there are plenty of image processing and supervised techniques for noise removal, our application required that the noise removal technique must not affect significant features and edge information in the retinal images. Thus, we adopted the technique of Anisotropic diffusion covered in Section 3.2. The algorithm for anisotropic diffusion, which was adopted from the Perona-Malik diffusion equation [47] has two parameters that define the level of diffusion or smoothing. The diffusion coefficient, c is used to control the rate of diffusion and the Diffusion Time, T_s is used to control the number of iterations required for the diffusion process. These parameters had to be tuned manually to achieve the desired level of smoothing while observing the effects of the diffusion on the performance of the algorithm. The effect of diffusion was found to be negligent when the coefficient of diffusion parameter was assigned a low value, thus not contributing to the objective of noise removal. Assigning a high value to the diffusion coefficient leads to over-smoothing and blurring of the image due to which significant edge information is lost, thus reducing the performance of the model. It is also observed that an increase in the Diffusion Time, T_s leads to excessive blurring of the image. The optimal parameter values were found to be c = 60 and T_s = 10. The resultant diffused preserves its significant edges and is noise-free at these values of c and T_s. The process of anisotropic diffusion contributes to about 20% of the computation time. Another noteworthy observation is that the overall computation time of the algorithm increases with an increase in the Diffusion Time, T_s.

Chapter 5

Evaluation and Analysis

5.1 Evaluation Metrics and Results

The primary outcome of this segmentation task is the classification of all image pixels as retinal blood vessel pixels or non-vessel/background pixels. Thus, we evaluate the algorithm based on the binary classification of all the pixels in every retinal scan image. Retinal blood vessel pixels are classified as 1 and other pixels are classified as 0 in the resultant segmentation map. The classification task could lead to four possible outcomes, i.e. two correct classifications (True Positives and True Negatives) and two mis-classifications (False Positives and False Negatives). We do a pixel-wise comparison of the resultant image with the corresponding ground truth image, which is manually annotated by a human expert and compute the number of correctly and incorrectly classified pixels. For the purpose of this study, we are considering that a pixel classification is a True Positive if a retinal blood vessel pixel is correctly classified as a retinal blood vessel pixel. A pixel classification is considered a True Negative (TN) if a background pixel (non-retinal blood vessel pixel) is correctly classified as a background pixel. A pixel classification is considered to be a False Positive (FP) if a background pixel is wrongly classified as a retinal blood vessel pixel and a False Negative (FN) if a retinal blood vessel pixel is wrongly classified as a background pixel. Using the above measures, we have chosen to use different performance metrics to evaluate the performance of our algorithm based on some globally accepted metrics used in other state-of-the-art methods as well as some additional metrics to get a complete understanding of the algorithm performance.

We first evaluate the correct classifications using measures of True Positive Rate and True Negative Rate. The True Positive Rate (TPR) indicates the fraction of retinal blood vessel pixels that are correctly classified as retinal blood vessel pixels, i.e. when a retinal vessel pixel occurs in both, the ground truth image as well as the original image. The True Negative Rate (TNR) indicates the fraction of

background or non-vessel pixels that are correctly classified as background pixels, i.e. when the background pixel occurs in both, the ground truth image as well as the original image.

Next, we measure mis-classifications using measures of False Positive Rates and False Negative Rates. The False Positive Rate (FPR) indicates the fraction of background or non-vessel pixels that are incorrectly classified as retinal blood vessel pixels. The False Negative Rate (FNR) indicates the fraction of retinal blood vessel pixels that are incorrectly classified as background or non-vessel pixels.

The average TPR, TNR, FPR and FNR for 10 high-resolution Fundus images of healthy eyes are calculated using Equations 5.1, 5.2, 5.3 and 5.4 as follows -

$$\text{TPR} = \frac{TP}{TP + TN + FP + FN} = 0.0764 \tag{5.1}$$

$$\text{TNR} = \frac{TN}{TP + TN + FP + FN} = 0.8811 \tag{5.2}$$

$$\text{FPR} = \frac{FP}{TP + TN + FP + FN} = 0.0221 \tag{5.3}$$

$$\text{FNR} = \frac{FN}{TP + TN + FP + FN} = 0.0202 \tag{5.4}$$

This indicates that 7.64% of all the pixels are correctly classified as retinal blood vessel pixels, 88.11% of all the pixels are correctly classified as background or non-vessel pixels, 2.21% of all pixels are wrongly classified as retinal blood vessel pixels and 2.03% of all the pixels are wrongly classified as background or non-vessel pixels.

We also use sensitivity, specificity and accuracy as a measure of evaluation. Sensitivity is the ratio of correctly classified retinal blood vessel pixels to the total number of retinal blood vessel pixels in the image, used to indicate the ability of our algorithm to detect retinal blood vessel pixels. Specificity is the ratio of correctly classified background or non-vessel pixels to the total number of background or non-vessel pixels, used to indicate the ability of our algorithm to detect background or non-vessel pixels. Accuracy, which is a global measure of the performance of the algorithm, indicates the level of conformity of the resultant skeleton image to the gold standard ground-truth image. The quality and validity of diagnosis increases with an increase in sensitivity, specificity and accuracy. These measures are calculated as follows -

$$\text{Sensitivity} = \frac{TP}{TP + FN} = 0.7833 \tag{5.5}$$

$$\text{Specificity} = \frac{TN}{TN + FP} = 0.9754 \tag{5.6}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = 0.9585 \tag{5.7}$$

Additionally, we also provide two other metrics- Precision and Recall for evaluating our algorithm. Precision indicates the fraction of the retinal vessel predictions (1s) that are found to be correct predictions. It is used to indicate the validity of the classification. Recall indicates the fraction of the retinal blood vessel pixels (1s) that were correctly classified as retinal blood vessel pixels. It describes the ability of the algorithm to identify retinal blood vessel pixels. We calculate the Precision and Recall as shown in Equations 5.8 and 5.9 respectively-

$$\text{Precision} = \frac{TP}{TP + FP} = 0.7754 \tag{5.8}$$

$$\text{Recall} = \frac{TP}{TP + FN} = 0.7933 \tag{5.9}$$

An interesting observation was that the use of contrasting techniques and thresholding adaptive histogram equalization technique like CLAHE returned very good accuracy figures. However, they tend to have lower sensitivity due to their inability to extract very thin retinal blood vessels. This also results in lower True Positive Rates. The higher number of false positives can be attributed to non-uniform illumination of the retinal scans and background noise [67]. This problem is further enhanced in images from the DRIVE dataset, which contains low quality, low resolution images. We solve the problem of noise in the image by using the technique of anisotropic diffusion.

5.2 Comparison with state-of-the-art methods

Our proposed technique of extracting retinal blood vessels shows a promising specificity of 97.54 and accuracy of 95.85% on the High Resolution Fundus images of healthy patients, which is on par with other state-of-the-art techniques. The algorithm also shows a sensitivity of 78.33%, which is significantly better than other state-of-the-art methods. Our method shows an improvement in sensitivity of over 9% compared to the method proposed by Dachsel et al. [13]. The algorithm shows lower

performance figures on the DRIVE dataset due to the noise levels, non-uniform illumination and low resolutions of images in the DRIVE dataset. However, the performance of our algorithm on images from the DRIVE dataset is on par with the performance of other state-of-the-art algorithms on the same.

We also observe that the average computing time of the algorithm while processing images of healthy patients from the High-Resolution Fundus dataset was 14.54 seconds. This is significantly higher than the computation time of the technique proposed by Dachsel. et al. [13], which computes in 1.18 seconds. Our proposed algorithm achieves considerably better performance at the cost of increased computation time. However, the computation time and performance of our proposed technique does better than the other state-of-the-art image processing techniques. Table 5.2 shows a comparison of the performance of our approach with other techniques on images from the DRIVE dataset. Note that the performance figures are an average of performance figures of 10 randomly selected images from the DRIVE dataset. The performance varies based on the nature of illumination on the scan, noise levels, blurred edges, the presence of other anomalous structures, shadows, etc.

Method	Approach Used	Sensitivity	Specificity	Accuracy	Computation Time
Our approach	Image processing	0.7463	0.9757	0.9565	10.2828 sec
Budai et al. [68]	Image processing	0.6440	0.9870	0.9572	26.69 sec
Dachsel et al. [13]	Image processing	0.6534	0.9860	0.9535	1.81 sec
Frangi et al. [69]	Image processing	0.6600	0.9850	0.9570	39.29 sec
Odstrcilik et al. [70]	Image processing	0.7060	0.9693	0.9340	-
Bahadar et al. [53]	Image Processing	0.7460	0.9800	0.9620	-
Razaee et al. [50]	Image Processing	-	-	0.9463	-
Lam et al. [49]	Image Processing	-	-	0.9567	-
Jiang et al. [2]	Deep Learning	0.8325	0.9838	0.9706	-
Fu et al. [3]	Deep Learning	0.7603	-	0.9523	-
Ronneberger et al. [1]	Deep Learning	0.7937	0.9820	0.9531	-
2^{nd} Observer	Manual Annotation	0.7761	0.9725	0.9472	-

Table 5.1: Comparison of our technique with state-of-the-art techniques on the DRIVE dataset

From the above table, we can see that our algorithm achieves a higher sensitivity than the other image processing techniques used for segmenting retinal blood vessels from the DRIVE dataset as well as a higher accuracy than the baseline provided by the second human expert observer (benchmark provided by the DRIVE dataset). The algorithm also achieves an average accuracy and specificity on par with other state-of-the-art techniques, including deep learning based techniques. It also computes faster than most state-of-the-art image processing techniques and all the deep learning based techniques, which require a lot of training time and data. Table 5.2 shows a comparison of the performance of our approach with other techniques on images from the High-Resolution Retinal Fundus

dataset. Note that the performance figures are an average of performance figures of 5 randomly selected images from each category of patients (Healthy, Glaucoma, Diabetic Retinopathy) in the High-Resolution Retinal Fundus dataset.

Method	Category	Sensitivity	Specificity	Accuracy
Our approach	Healthy	0.7933	0.9754	0.9585
Odstrcilik et al. [70]	Healthy	0.7463	0.9757	0.9565
Budai et al. [68]	Healthy	0.6620	0.9920	0.9610
Dachsel et al. [13]	Healthy	0.7000	0.9880	0.9560
Frangi et al. [69]	Healthy	0.6210	0.9890	0.9550
Our approach	Diabetic Retinopathy	0.7503	0.9614	0.9545
Odstrcilik et al. [70]	Diabetic Retinopathy	0.7910	0.9640	0.9490
Budai et al. [68]	Diabetic Retinopathy	0.6870	0.9860	0.9650
Dachsel et al. [13]	Diabetic Retinopathy	0.7100	0.9790	9560
Frangi et al. [69]	Diabetic Retinopathy	0.6540	0.9840	0.9610
Our approach	Glaucoma	0.7989	0.9611	0.9515
Odstrcilik et al. [70]	Glaucoma	0.7460	0.9610	0.9440
Budai et al. [68]	Glaucoma	0.6580	0.9770	0.9550
Dachsel et al. [13]	Glaucoma	0.6740	0.9780	0.9520
Frangi et al. [69]	Glaucoma	0.5900	0.9720	0.9460

Table 5.2: Comparison of our technique with state-of-the-art techniques on the Fundus dataset

It would be noteworthy to mention that using supervised learning approaches like Convolutional Neural Networks for the segmentation of retinal blood vessels yields relatively better sensitivity figures than those of state-of-the-art traditional image processing approaches. This is due to the fact that Deep Learning models are better at identifying thin retinal blood vessels, thus producing a better quality of segmentation. The disadvantage of a deep learning based approach, however, is that the models rely heavily on correctly labelled training data. Our approach also does not require any explicit data preparation steps that are usually necessary while using supervised learning based approaches. The computation time of our algorithm would further reduce with the use of a GPU-based architecture. It is also important to remember while evaluating the algorithm that the standard of comparison with the benchmark is based on the accuracy of segmentation achieved by manual annotation by human experts. Note that this benchmark set by Odstrcilik et al. [70] is measured to have an accuracy of about 95%, considering the fact that an individual's expert interpretation could have a margin of error due to inaccuracies like disconnected annotations and missing thin retinal vessels. Compared to the benchmark, our technique achieves reasonable and competitive performance figures.

The technique is also found to be highly robust while extracting retinal blood vessels from different categories of retinal scans in the Fundus dataset (Healthy, Glaucomatous and Diabetic Retinopathy) as well as low resolution DRIVE images. Retinal images of patients with Diabetic Retinopathy show addition anomalous structures in the background, making it a challenge for a segmentation algorithm to classify these structures correctly as background pixels. The proposed technique shows similar performance results on retinal scans of healthy patients and patients with Glaucoma. Since the work of Dachsel et al. [13] served as the basis of our study, we compare the performance of our technique with theirs, in addition to comparisons with the benchmark standard [70] and other state-of-the-art techniques. We see a significant improvement of about 9% in the sensitivity of our algorithm's predictions compared to that of Dachsel et al. [13], while the sensitivity and accuracy of both techniques are competitive with other state-of-the-art techniques. However, the algorithm shows reduced performance figures on images of patients with Diabetic Retinopathy. This reduction in performance, particularly the sensitivity, is caused due to green spots or structures in the background. This caused an increase in the number of False Positives and, by extension, the False Positive Rate and rate of mis-classification. The parameters used in our algorithm were unchanged throughout our experiments on images from both the datasets, making our algorithm robust and scalable.

Chapter 6

Conclusion and Future Works

In this thesis, we have provided a comprehensive review of different image segmentation and skeletonization techniques used in bio-medical imaging, keenly focusing on the extraction of retinal blood vessels from retinal images. We propose an image processing framework for efficiently extracting retinal blood vessels from retinal Fundus images. We provide a robust data pre-processing pipeline that adopts the use of various image pre-processing techniques like gamma correction, grey-scaling, edge-aware local contrasting and Contrast Limited Adaptive Histogram Equalization to enhance significant features in the image and increase the discrimination between contrasts of retinal blood vessel pixels and the background pixels. We also make the use of a multi-layered Laplacian Pyramid, derived from the Gaussian layers of the original image to extract and analyse retinal vessels in different scales, i.e. from coarse scales to thin scales. Disconnected vessel pixels are then discarded to form coherent, well-connected retinal vessels. These layers are then resized and fused to produce a resultant image with the extracted retinal blood vessels. The parameters used in our pre-processing steps were kept constant throughout the course of our experiments on both the DRIVE and High-Resolution Fundus datasets, thus making our framework highly generalized, robust and scalable. We use evaluation metrics like TPR, FPR, TNR, FNR, Sensitivity, Specificity and Accuracy to evaluate the performance of our technique and compare it with the performance of other state-of-the-art techniques. Additionally, we also compute the Precision and Recall measures of our classification that can be used for comparison in future research.

The experimental results in our study indicate that the algorithm is competitive with other state-of-the-art algorithms in in terms of accuracy, specificity and computation time. It achieves better sensitivity than other state-of-the-art traditional image processing algorithms on both, the High-Resolution Fundus and DRIVE datasets. The capabilities of this technique could be improved by adopting a GPU-based architecture to decrease its computation time and increase its complexity. However, on

comparison with ground truth images, it is observed that the technique fails to extract tiny retinal blood vessels. It is also observed that retinal images of patients with Diabetic Retinopathy show green structures in the background that are wrongly detected as retinal blood vessels (False Positives). This could potentially lead to wrong diagnoses. Addressing this issue is a complex task, as using a simple threshold value to discard small regions (green spots) in the image would also lead to the loss of tiny, connected blood vessels as their size could similar to that of the tiny vessels. In out future work, we aim to make use of a convolutional neural network to treat this segmentation task as a multi-class classification instead of a binary classification to detect green spots along with the retinal vessel and background pixels. Supervised Learning architectures like Convolutional Neural Networks would also produce results with higher sensitivities. The resultant images of the Binarization task shown in Figure 3.9 could be fed as inputs to a Deep Learning model to extract tiny retinal vessels, as this representation of the image contains the most relevant information on all scales. We also aim to combine this technique with other image processing techniques to produce more optimal segmentations. Additionally, the use of a Deep Learning architecture can be adopted to detect anomalies in Fundus images to further extend the practical scope of this framework to diagnostic applications. This would further enhance the scope and accuracy of automated retinal vessel extraction frameworks in the practical domain of medical diagnostics.

Bibliography

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, (Cham), pp. 234–241, Springer International Publishing, 2015.

[2] Y. Jiang, H. Zhang, N. Tan, and L. Chen, "Automatic retinal blood vessel segmentation based on fully convolutional neural networks," *Symmetry*, vol. 11, p. 1112, 2019.

[3] H. Fu, Y. Xu, S. Lin, D. W. Kee Wong, and J. Liu, "DeepVessel: Retinal Vessel Segmentation via Deep Learning and Conditional Random Field," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, (Cham), pp. 132–139, Springer International Publishing, 2016.

[4] D. A. Salz and A. J. Witkin, "Imaging in diabetic retinopathy.," *Middle East African journal of ophthalmology*, vol. 22, no. 2, pp. 145–150, 2015.

[5] N. Patton, T. Aslam, T. Macgillivray, A. Pattie, I. Deary, and B. Dhillon, "Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: A rationale based on homology between cerebral and retinal microvasculatures," *Journal of anatomy*, vol. 206, pp. 319–48, 2005.

[6] A. G. Dumitrescu, L. Voinea, I. A. Badarau, V. A. Paun, M. Schowe, and R. Ciuluvica, "Update on retinal vascular caliber.," *Romanian journal of ophthalmology*, vol. 61, no. 3, pp. 171–180, 2017.

[7] M. Abràmoff and C. N. Kay, "Chapter 6: Fundus Imaging - Image Processing," in *Retina (Fifth Edition)*, pp. 151–176, London: W.B. Saunders, fifth ed., 2013.

[8] M. J. Mulvany, "Are vascular abnormalities a primary cause or secondary consequence of hypertension?," *Hypertension (Dallas, Tex. : 1979)*, vol. 18, no. 3 Suppl, pp. I52–7, 1991.

[9] J. L. le Noble, G. J. Tangelder, D. W. Slaaf, H. van Essen, R. S. Reneman, and H. A. Struyker-Boudier, "A functional morphometric study of the cremaster muscle microcirculation in young

spontaneously hypertensive rats.," *Journal of hypertension*, vol. 8, no. 8, pp. 741–748, 1990.

[10] Y. Muraoka, A. Tsujikawa, K. Kumagai, M. Akiba, K. Ogino, T. Murakami, Y. Akagi-Kurashige, K. Miyamoto, and N. Yoshimura, "Age- and hypertension-dependent changes in retinal vessel diameter and wall thickness: an optical coherence tomography study.," *American journal of ophthalmology*, vol. 156, no. 4, pp. 706–714, 2013.

[11] M. Reimann, W. Vilser, M. Gruber, S. R. Bornstein, and T. Ziemssen, "Insulin is a key determinant of elevated retinal arteriolar flicker response in insulin-resistant individuals," *Diabetologia*, vol. 58, no. 9, pp. 2154–2160, 2015.

[12] T. A. Ciulla, A. G. Amador, and B. Zinman, "Diabetic retinopathy and diabetic macular edema: pathophysiology, screening, and novel therapies.," *Diabetes care*, vol. 26, no. 9, pp. 2653–2664, 2003.

[13] R. Dachsel, A. Jöster, and M. Breuß, "Real-Time Retinal Vessel Segmentation on High-Resolution Fundus Images Using Laplacian Pyramids," in *Image and Video Technology*, (Cham), pp. 337–350, Springer International Publishing, 2019.

[14] X. Guan, S. Jian, P. Hongda, Z. Zhiguo, and G. Haibin, "An image enhancement method based on gamma correction," in *Second International Symposium on Computational Intelligence and Design*, vol. 1, pp. 60–63, 2009.

[15] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[16] A. Budai, J. Odstrcilík, R. Kolár, J. Hornegger, J. Jan, T. Kubena, and G. Michelson, "A public database for the evaluation of fundus image segmentation algorithms," *Investigative Ophthalmology Visual Science*, vol. 52, pp. 1345–1345, 2011.

[17] J. Gnanaselvi and M. K. G, *A Comprehensive Study of Retinal Vessel Classification Methods in Fundus Images for Detection of Hypertensive Retinopathy and Cardiovascular Diseases*, pp. 1239–1249. 2019.

[18] A. Imran, J. Li, Y. Pei, J. Yang, and Q. Wang, "Comparative analysis of vessel segmentation techniques in retinal images," *IEEE Access*, vol. 7, pp. 114862–114887, 2019.

[19] J. Shah, "Gray skeletons and segmentation of shapes," *Computer Vision and Image Understanding*, vol. 99, pp. 96–109, 2005.

[20] L. Lam, S. . Lee, and C. Y. Suen, "Thinning methodologies-a comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 869–885, 1992.

[21] G. P. Dinneen, "Programming pattern recognition," in *Managing Requirements Knowledge, International Workshop on*, vol. 1, (Los Alamitos, CA, USA), p. 94, IEEE Computer Society, 1955.

[22] C. J. Hilitch, "Linear skeletons from square cupboards," in *Machine Intelligence 4* (B. Meltzer and D. Michie, eds.), p. 403, Edinburgh University Press, 1969.

[23] T. Nguyen and J. Sklansky, "A fast skeleton-finder for coronary arteries," in *Proc. Int. Conf. Pattern Rec.*, p. 481–483, 1986.

[24] A. R. Dill, M. D. Levine, and P. B. Noble, "Multiple resolution skeletons.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 9, no. 4, pp. 495–504, 1987.

[25] K. Preston, M. J. B. Duff, S. Levialdi, P. E. Norgren, and J. Toriwaki, "Basics of cellular logic with some applications in medical image processing," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 826–856, 1979.

[26] C. Arcelli and G. Sanniti di Baja, "Ridge points in Euclidean distance maps," *Pattern Recognition Letters*, vol. 13, no. 4, pp. 237–243, 1992.

[27] R. L. Ogniewicz and O. Kübler, "Hierarchic Voronoi skeletons," *Pattern Recognition*, vol. 28, no. 3, pp. 343–359, 1995.

[28] J. M. Kang, *Voronoi Diagram*, pp. 1232–1235. Boston, MA: Springer US, 2008.

[29] W. Shen, X. Bai, R. Hu, H. Wang, and L. Jan Latecki, "Skeleton growing and pruning with bending potential ratio," *Pattern Recogn.*, vol. 44, no. 2, p. 196–209, 2011.

[30] X. Bai, L. J. Latecki, and W. Liu, "Skeleton pruning by contour partitioning with discrete curve evolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 449–462, 2007.

[31] B. P. Chacko and B. P. Anto, "Discrete curve evolution based skeleton pruning for character recognition," in *2009 Seventh International Conference on Advances in Pattern Recognition*, pp. 402–405, 2009.

[32] C. Arcelli and G. S. di Baja, "Euclidean skeleton via centre-of-maximal-disc extraction," *Image and Vision Computing*, vol. 11, no. 3, pp. 163–173, 1993.

[33] S. Krinidis and V. Chatzis, "A skeleton family generator via physics-based deformable models.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 18, no. 1, pp. 1–11, 2009.

[34] R. Kimmel, D. Shaked, N. Kiryati, and A. M. Bruckstein, "Skeletonization via distance maps and level sets," *Computer Vision and Image Understanding*, vol. 62, pp. 382–391, 1995.

[35] K. Siddiqi, S. Bouix, A. Tannenbaum, and S. W. Zucker, "Hamilton-Jacobi Skeletons," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 215–231, 2002.

[36] J. W. Leech, *The Hamiltonian Formulation*, pp. 42–54. Dordrecht: Springer Netherlands, 1965.

[37] T. Grigorishin, G. Abdel-Hamid, and Y. H. Yang, "Skeletonisation: An electrostatic field-based approach," *Pattern Analysis and Applications*, vol. 1, no. 3, pp. 163–177, 1998.

[38] L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt, "Shape representation and classification using the poisson equation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1991–2005, 2006.

[39] T. Lindeberg, "Edge Detection and Ridge Detection with Automatic Scale Selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 117–156, 1998.

[40] A. P. Witkin, "Scale-Space Filtering," in *Readings in Computer Vision* (M. A. Fischler and O. Firschein, eds.), pp. 329–332, San Francisco (CA): Morgan Kaufmann, 1987.

[41] Z. G. Tari, J. Shah, and H. Pien, "Extraction of Shape Skeletons from Grayscale Images," *Computer Vision and Image Understanding*, vol. 66, no. 2, pp. 133–146, 1997.

[42] Zeyun Yu and Chandrajit Bajaj, "A segmentation-free approach for skeletonization of gray-scale images via anisotropic vector diffusion," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1, pp. I–I, 2004.

[43] Chenyang Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, 1998.

[44] "A geometric model for active contours in image processing," *Numerische Mathematik*, vol. 66, no. 1, pp. 1–31, 1993.

[45] C. Direkoglu, R. Dahyot, and M. Manzke, "On using anisotropic diffusion for skeleton extraction," *International Journal of Computer Vision*, vol. 100, pp. 170–189, 2012.

[46] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[47] P. Perona, T. Shiota, and J. Malik, "Anisotropic diffusion," in *Geometry-Driven Diffusion in Computer Vision*, 1994.

[48] C. Direkoglu, R. Dahyot, and M. Manzke, "Skeleton extraction via anisotropic heat flow," in *British Machine Vision Conference, BMVC 2010 - Proceedings*, pp. 1–11, 2010.

[49] B. Lam, Y. Gao, and A. W.-C. Liew, "General retinal vessel segmentation using regularization-based multiconcavity modeling," *Medical Imaging, IEEE Transactions on*, vol. 29, pp. 1369 – 1381, 2010.

[50] K. Rezaee, J. Haddadnia, and A. Tashk, "Optimized clinical segmentation of retinal blood vessels by using combination of adaptive filtering, fuzzy entropy and skeletonization," *Appl. Soft Comput.*, vol. 52, pp. 937–951, 2017.

[51] W. S. Gan, *Wiener Filter and Kalman Filter*, pp. 31–36. Singapore: Springer Singapore, 2020.

[52] S. Al-Sharhan, F. Karray, W. Gueaieb, and O. Basir, "Fuzzy entropy: A brief survey," in *IEEE International Conference on Fuzzy Systems.*, vol. 3, (Melbourne, Victoria), pp. 1135 – 1139, 2001.

[53] K. Bahadar, A. A Khaliq, and M. Shahid, "A Morphological Hessian Based Approach for Retinal Blood Vessels Segmentation and Denoising Using Region Based Otsu Thresholding," *PLoS ONE*, vol. 11, no. 7, p. e0158996, 2016.

[54] E. Adelson, C. Anderson, J. Bergen, P. Burt, and J. Ogden, "Pyramid methods in image processing," *RCA Eng.*, vol. 29, 1983.

[55] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Communications*, vol. 31, pp. 532–540, 1983.

[56] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[57] P. Sidike, M. Z. Alom, T. Taha, and V. Asari, "The history began from alexnet: A comprehensive survey on deep learning approaches," 2018.

[58] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), (Cham), pp. 818–833, Springer International Publishing, 2014.

[59] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015.

[60] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

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

[62] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.

[63] D. Cirean, A. Giusti, L. M. Gambardella, and Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," *Proceedings of Neural Information Processing Systems*, vol. 25, 2012.

[64] A. W. Setiawan, T. R. Mengko, O. S. Santoso, and A. B. Suksmono, "Color retinal image enhancement using clahe," in *International Conference on ICT for Smart Society*, pp. 1–3, 2013.

[65] B. Yu and Z. Fan, "A comprehensive review of conditional random fields: variants, hybrids and applications," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4289–4333, 2020.

[66] J. Rogowska, "Chapter 5 - Overview and Fundamentals of Medical Image Segmentation," in *Handbook of Medical Image Processing and Analysis (Second Edition)* (I. N. BANKMAN, ed.), pp. 73–90, Burlington: Academic Press, second edition ed., 2009.

[67] S. K. Saha, D. Xiao, and Y. Kanagasingam, "A Novel Method for Correcting Non-uniform/Poor Illumination of Color Fundus Photographs.," *Journal of digital imaging*, vol. 31, no. 4, pp. 553–561, 2018.

[68] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *International journal of biomedical imaging*, vol. 2013, p. 154860, 2013.

[69] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing And Computer-Assisted Intervention - MICCAI'98*, vol. 1496, pp. 130–137, Wells, WM, Springer Verlag, 1998.

[70] J. Odstrcilik, R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, T. Kubena, P. Cernosek, O. Svoboda, and E. Angelopoulou, "Retinal vessel segmentation by improved matched filtering:

evaluation on a new high-resolution fundus image database," *IET Image Processing*, vol. 7, no. 4, pp. 373–383, 2013.

Appendix A

List of Abbreviations

BP - Blood Pressure

BMI - Body Mass Index CAD - Coronary Artery Disease

CLAHE - Contrast Limited Adaptive Histogram Equalisation

RGB - Red Green Blue

DRIVE - Digital Retinal Images for Vessel Extraction

WBC - White Blood Cell

EDM - Euclidean Distance Map

VD - Voronoi Diagram

BPR - Bending Potential Ratio

DCE - Discreet Curve Evolution

2D - Two Dimensional

3D - Three Dimensional

GVF - Gradient Vector Flow

SSM - Skeleton Strength Map

CNN - Convolutional Neural Networks

GPU - Graphics Processing Unit

FCN - Fully Convolutional Neural Network

RCF - Random Crop and Fill

STARE - Structured Analysis of the Retina CRF - Conditional Random Field

AHE - Adaptive Histogram Equalization

RAM - Random Access Memory

OS - Operating System

TP - True Positive

FP - False Positive

TN - True Negative

FN - False Negative

TPR - True Positive Rate

FPR - False Positive Rate

TNR - True Negative Rate

FNR - False Negative Rate

ET - Edge Threshold

Appendix B

Note

The scripts and programs used to implement this framework can be found at this GitHub Repository. The DRIVE dataset can be found here and the High-Resolution Retinal Fundus dataset can be found here.