Semantic Segmentation in Human Retinal Images

By

Qurrat-ul-ain

00000171378

Supervisor

Dr. Muhammad Moazam Fraz

Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree

of Masters of Science in Computer Science

In

School of Electrical Engineering and Computer Science,
National University of Sciences and Technology (NUST),
Islamabad, Pakistan.

Approval

Certified that the contents of thesis document titled "**Semantic Segmentation in Human Retinal Images**" submitted by Mr. **Qurrat-ul-ain** have been found satisfactory for the requirement of degree.

Advisor: Dr. Muhammad Moazam Fraz

Signature: _____

Date: _____

Committee Member1: Dr. Anis ur Rahman

Signature:

_____

Date: _____

Committee Member2: Dr. Omar Arif

Signature: _____

Date: _____

Committee Member3: Dr. Imran Mahmood

Signature: _____

Date: _____

Thesis Acceptance Certificate

Certified that final copy of MS/MPhil thesis written by Ms. Qurrat-ul-ain, (Registration No 171378), of School of Electrical Engineering and Computer Sciences (School/College/Institute) has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor: Dr. Muhammad Moazam Fraz

Date: _____

Signature (HOD): _____

Date: _____

Signature (Dean/Principal): _____

Date: _____

Certificate of Originality

I hereby declare that the research paper titled "***Semantic Segmentation in Human Retinal Images***" my own work and to the best of my knowledge. It contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or any other education institute, except where due acknowledgment, is made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic is acknowledged. I also verified the originality of contents through plagiarism software.

Author Name: Qurrat-ul-ain

Signature: _____

Acknowledgment

I pay my gratitude to Allah almighty for blessing me a lot and without His guidance I couldn't complete this task. I am thankful to Dr. Muhammad Moazam Fraz for guiding and encouraging me to complete my thesis. His timely and efficient contributions helped me to shape the thesis into its final form and I express my gratefulness for his sincere supervision all the way.

I am also thankful to Department of Computing, and the teachers for providing me with an academic base, which enabled me to complete this thesis.

I'd like to thank my parents who have provided me every comfort of life and their utmost affection and support made me reach where I am today.

Table of Contents

List of Tables

List of Figures

Abstract

Three leading causes of blindness in the world are diabetic retinopathy, glaucoma and macular degeneration. Vision loss due to these diseases can easily be avoided by early diagnosis and treatment. Many systemic and ocular diseases manifest themselves in retina, which is transparent in appearance making the in-vivo examination possible. An automated mass screening system based on retinal images can reduce the risk of blindness and vision loss. In this regard, the fundamental step is accurate segmentation of retinal pathologies (i.e., exudates, hemorrhages, and cotton-wool spots) and retinal landmarks (i.e. retinal blood vessels and optic disc). In this thesis an end-to-end framework based on encoder decoder fully convolutional neural network is proposed for simultaneous segmentation of retinal pathologies and landmarks. The proposed network is inspired from U-Net, with introduction of Batch Normalization layers and a down-sampling block to helps in segmentation of small retinal lesions. For the purpose of evaluation, a new public retinal image dataset is created from Messidor dataset, where the pixel-level multi-class annotations are done by the trained observers and verified by ophthalmologist. The proposed technique is evaluated on DRIVE, HRF, eOphtha and Messidor datasets, and it achieved remarkable performance in retinal anatomical structure segmentation using standard evaluation metrics.

Keywords: *semantic segmentation, computer vision, deep learning*

Chapter 1

Introduction

This thesis is about semantic segmentation applied on retinal images. This chapter provides general knowledge of the research to help in understanding this thesis, also introduce the problem statement along with its solution.

1.1 Semantic Segmentation

Semantic Segmentation is the process of partitioning an image into meaningful subparts each of which belongs to a particular class. In this process each pixel of the image is assigned an output class in order to get a pixel-wise dense classification [1]. Image segmentation is a high level task of computer vision which leads towards scene understanding. Importance of segmentation can be judged by the fact that number of applications that are based on knowledge about imagery is increasing rapidly [2]. In many application segmentation is to find desired object in image, that is to partitioned image into two parts object and background. Human visual system can perform segmentation naturally. We detect lines, edges, change in color and gradient and different shapes and take decision about object type based on this information. Semantic segmentation uses the same information to identify required objects that humans use to detect an object i.e. shape, colors etc. Algorithms are designed that use certain patterns and properties of interested object to perform segmentation. Semantic segmentation has achieved enormous performance in scene understanding which affirms the notion of its application in biomedical imaging. Tough a handsome amount of work has been done on segmenting common objects e.g. cat, table and person etc., but these types of techniques are not directly applicable to biomedical domain due to its specific requirements and constraints. A summary of state-of-the-art semantic segmentation Convolutional Neural Networks (CNN) along with their pros and cons is provided in Chapter 2.

1.2 Background

This section describes details about eye anatomy and working of retina and retina imaging.

1.2.1 Eye Anatomy

Humans visual sensory organ are eyes, which are responsible for sending visual stimulus to various visual centers of brain through optic nerve. Brain interprets this signal giving meaning through which we see the world. Eye is primarily spherical in shape and hollow organ from inside. Visible parts of eye include sclera, cornea, iris and pupil [3]. Sclera is the outer most white layer of eye whose function is to provide protection to inner parts of an eye. Front transparent part is cornea which allows light to pass through and covers the iris, pupil and internal chamber. Colored part of eye is iris, opening in iris called pupil. Light after passing through cornea and internal chamber become partially focused, this rays further passes through lens, pupil and anterior chamber with focuses the image on retina. Optic nerve then transmits the electrical signal to brain. Anatomy of eye is depicted in Figure 1.

1.2.2 Retina Anatomy

Retina is interior light sensitive layer of tissue in eye. It enables the incoming light to convert into neural signal, which is suitable for the visual cortex of brain for further processing. Function of retina is same as of film in camera. The retinal can be divided into following layers

1. Internal limiting membrane
2. Nerve fiber layer
3. Ganglion cell layer
4. Inner plexiform layer
5. Inner nuclear layer
6. Outer plexiform layer
7. Outer nuclear layer
8. External limiting membrane
9. Rods and cones cells layer
10. Pigment epithelium

Pigmented oval shaped area near retina center is called macula, which is accountable for color vision and high resolution vision. It is further divided into perifovea, parafovea and fovea. Central part of macula is fovea and it is highly packed with rods and cones cells. Average thickness of retina is 250-µm near optic nerve, this thickness is reached to 400-µm in macula and then reduces to 150-µm on fovea [4].

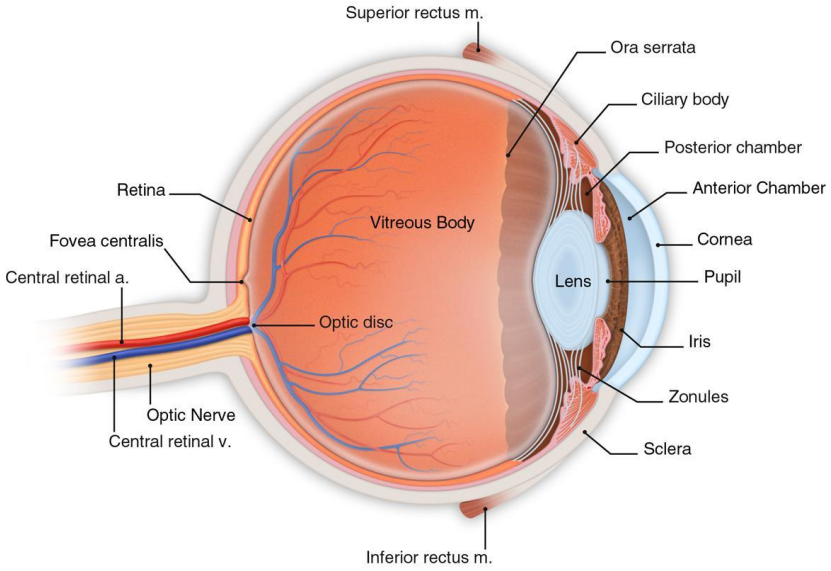

Figure 1: Anatomy of an eye

1.2.3 Retinal Imaging

All the ocular structure involved in image formation on retina are optically transparent because its function is to see the world. Therefore, retina is visible from outside which makes it possible to capture images of retina. Retinal imaging is of great interest for researchers and many techniques are developed to accomplish this task [3].

Fundus is 2-D representation of retina a 3-D structure as shown in Figure 2. The process of capturing fundus images refers as fundus imaging. Many techniques are developed for fundus imaging including fundus photography, color fundus photography, stereo fundus photography, hyper-spectral imaging, scanning laser ophthalmoscopy (SLO), adaptive optics SLO and fluorescence angiography. Visible ocular structure in retinal fundus comprises of optic disc, vessels (both arteries and veins), fovea, macula and pathologies/lesions (if present). Another famous technique for retinal imaging is Optical Coherence Tomography (OCT) in which depth of a

structure is estimated by change in refractive index. In this technique fundus images are used.

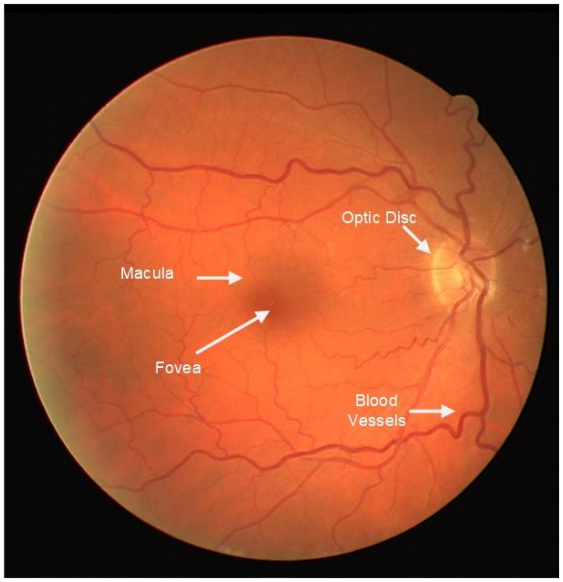

Figure 2: Retinal fundus

1.2.4 Pathogenesis

In general, the study of disease is called pathology; retinal pathology is study of structural abnormalities produced by different diseases in retina. There are many types of retinal pathologies or lesions but in this section only relevant pathologies morphology is discussed.

1.2.4.1 Hemorrhages

Retinal hemorrhages are formed due to bleeding of retinal capillaries at vessels end, which results in red blot or a deformed structure. This type of hemorrhages mainly caused due to hypertension, diabetes and retinal vein occlusion [5]. Hemorrhages can be seen in Figure 3 (d).

1.2.4.2 Hard Exudates

Hard exudates sometimes called exudates are tiny yellow lesions, which mostly present underlying the blood vessels. They are made up of lipids and lipoproteins, which are mostly caused by break down of blood-retina barrier. These lipids and proteins are residual of capillaries damage caused by diabetes, retinal vein occlusion and

neuroretinitis. Presence of hard exudates is chronic sign of macular edema [5] Figure 3 (c) show hard exudates.

1.2.4.3 Cotton Wool Spots

Cotton Wool Spots (CWS) also known as soft exudates, have patch shaped white fluffy appearance. They are caused by accumulation of nerve plasma within nerve fiber layer, this fluid discharge as a result of nerve fiber damage mostly due to hypertension and diabetes [5]. Morphology of CWS can be seen in Figure 3 (b).

1.2.4.4 Microaneurysms

Microaneurysms are small red dots having sac like appearance in superficial retinal layer, they are localized capillaries dilation caused by pericyte loss. They are considered as the earliest sign of diabetic retinopathy [5]. Microaneurysms are depicted in Figure 3 (f).

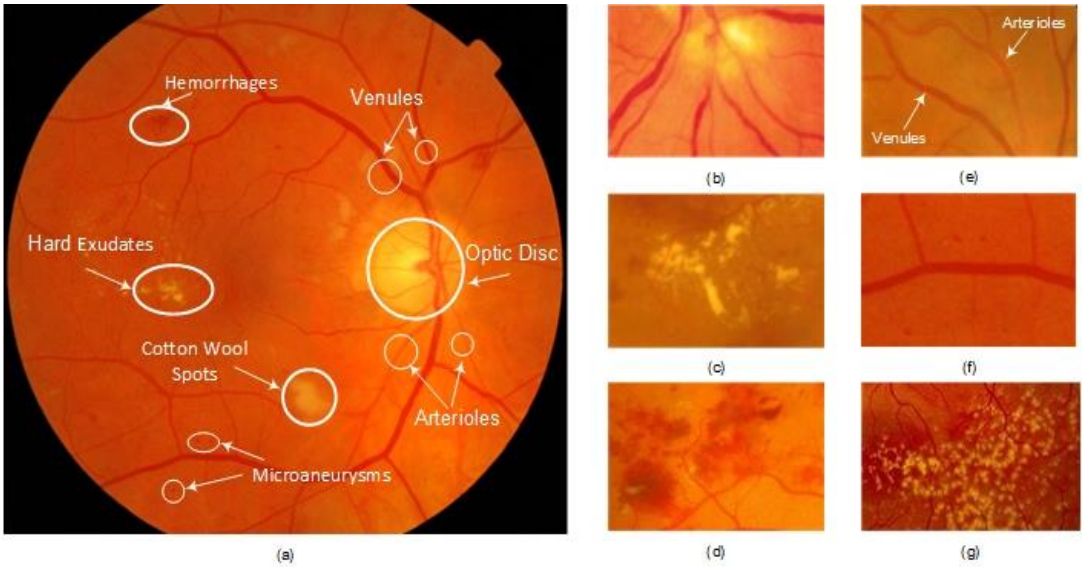

Figure 3: retinal pathologies, (a) retinal anatomical structure and retinal lesions (b) cotton wool spots (c) hard exudates (d) hemorrhages (e) retinal arteries and veins (f) microaneurosyms (g) drusens

1.3 Motivation

Diseases can be generally divided into two sub categories according to the area they effect and where symptoms appear. First type of disease which effects the whole body and not a particular organ is called Systemic diseases, for example, blood pressure and flu. Second type which effect a particular organ or tissue and reflects symptoms only

in effected organ. Human retina demonstrate changes for ophthalmic (also known as eye disorders) as well as for several systemic diseases. In some cases, patients complain of having eye pain and abnormalities later reported to be suffering with other systemic diseases. However, the information gathered through ocular examination is helpful in diagnosis of actual disorder. Almost every disorder effects retina in a specific way.

1.3.1 Systemic Disease and Ocular Manifestation

Significant number of diseases that have impact on retina and can be studied via eye imaging. Some of the popular diseases are mentioned in the following section.

1.3.1.1 Diabetic Retinopathy (DR)

According to World Health Organization (WHO) diabetes is defined as a condition of having elevated blood sugar level, 7.0 mmol/lit in fasting. There are different reasons for the onset of this disorder but its reason is not fully understandable up till now, but obesity, smoking, lifestyle, anemia and genetics bestow elevated risk of developing diabetes. Anti-hyperglycemic drugs, changes in diet and insulin injections are all types of treatments. Diabetes/Hyper-glycaemia disfigure nerve cells, in addition to small and large vessels, hence have its devastating effects on eyes, heart, brain and kidneys. Such complications of diabetes is known as Diabetic Retinopathy (DR).

The Diabetic Retinopathy (DR) contributes to 4.8% of the 37 million cases of blindness reported throughout the world [6]. Annual screening and early diagnosis can prevent vision loss and blindness in diabetes patients [7]. High blood glucose level lead to vasculature damage, more precisely vessels walls become more permeable, which in turn results in fluid leakage from vessels causing retinal exudates, microaneurysms and diabetic macular edema (DME). Another condition called ischemia, caused by growth of new hyper-permeable vessels which lead to retinal detachment.

1.3.1.2 Glaucoma

The glaucoma is the second dominant reason of blindness worldwide, it contributes to 12.3% of the 37 million cases of blindness throughout the world [6]. In a study it is estimated that by 2020 the approximate number of people suffering from glaucoma will reach 79.6 million [8]. Glaucoma is an eye condition, which damages the optic nerve in eye, leads to diminished vision acuity even blindness. It is characterized by increased intraocular pressure, due to collection of fluid in front part of eye which put

pressure on optic nerve. Primarily it is neuropathy, not a retinopathy. The peculiarity of glaucoma is cup-to-disc ratio, which is ratio of optic disc cup and neuroretinal border surface, this ratio is a significant factor in assessing the diagnosis and progression of disease. The cupping of optic nerve is easily visible in retinal fundus images.

1.3.1.3 Age-related Macular Degeneration

Age-related Macular Degeneration (AMD) is one the major causes of irreversible vision loss in people over the age of 50 years. With age, acellular debris deposits in retinal epithelium, known as **drusen**, which are pale yellowish in appearance and can be observed in retinal fundus images, can be seen in Figure 3 (g). Drusen can be present in macula and peripheral retina. AMD is an eye condition which damages the macula of eye, this part is needed for central and sharp vision, located near the center of retina, making it difficult for the affected person to see clearly in straight line of vision [9]. The advancement of disease varies patient to patient, in some cases it is reported that spread is faster which may lead to vision loss, whereas in some cases AMD progress slowly such that vision loss does not occur quickly. AMD is not characterized by complete vision loss or blindness. It is divided into two types:

- **Wet AMD** also called choroidal neovascularization (CNV) is associated with abnormal growth of highly permeable vessels beneath the macula. These vessels leak fluid into retina, which forms a scar in epithelium and then eventually leads to loss of central vision.
- **Dry AMD** can be distinguished from wet AMD by the presence of drusens in macula. Small number of drusens doesn't affect the vision, however as their number and size increased, they induced vision distortion.

1.3.1.4 Cardiovascular Diseases

Manifestation of cardiovascular diseases can be seen in many ways in retina, mostly these diseases effect the vasculature of retina. Atherosclerosis and hypertension provoke changes in diameters of veins and arteries, known as artery-to-vein (A/V) ratio. The veins become wide, whereas the arteries' diameter slenderized, hence A/V ratio decrease. This condition is also associated with increased risk of stroke [3]. Decrease in veins diameters induced vascular fluid leakage, which appear as a flame shaped retinal hemorrhages and retinal edema, severe retinal edema may result in

deposition of hard exudates near macular region. Another hallmark of accelerated hypertension is optic nerve swelling, which can be easily seen in retinal fundus [5].

Retinal ischemia can be enforced by hypertension, which cause dead retinal tissues to be appear as cotton wool spots and choroidal dead tissue to be visible as deep white spots of retina. Systemic diseases can induce arterial and venous occlusion (RAO, RVO), which is blockage in arteries and veins respectively, hindering the flow of blood in retina which in return deprived retinal tissue from oxygen. These condition is further divided into two sub-categories i.e. central and branch arterial occlusion (CRAO, BRAO) and central and branch venous occlusion (CRVO, BRVO) according to location of artery or vein which is affected [10].

1.3.2 Early Diagnosis from Retinal Fundus

By morphology eye is a lucid organ, making it possible to see through it or to capture the inner details of eye structures from fundus camera. As mentioned in previous section many systemic and ocular diseases manifest themselves in retina, which increase the significance of routine retinal screening. The cost associated with such regular screening programs is increasing day by day, as it involves trained ophthalmologist or eye care providers, these are expensive human experts and have limited amount of time per patient. Further humans are subjective; their decisions suffer from variability. Thus to increase the exactitude with which the patients are managed and saving the eye care provider time an automated system is required. Such system is economical for large scale screening of diseases such as diabetic retinopathy, hypertension and age-related macular degeneration.

Early detection of DR through mass screening programs can prevent vision loss with timely diagnosis and treatment. Glaucoma in its early stages have no symptoms other than retinal changes, in such a case routine eye examination can diagnose the condition in early stage, thus helping to cure before its get late. Similarly, timely diagnosis of cardiovascular diseases and macular edema through screening impede major health complication associated with these diseases. Such system is indispensable and cost efficient for patients.

1.4 Problem Statement

In related literature most of the works are oriented to the detection of each structure individually. According to best of my knowledge up till now no work has been performed on simultaneous segmentation of retinal pathologies and retinal landmarks. The objective is to obtain simultaneous detection of the main retinal structures and lesions in a fast and robust way.

1.5 Solution Statement

Segment retinal images to extract different anatomical structures and retinal lesions using deep neural network.

1.6 Application Areas

With the advancement in imaging technology data collected in form of retinal images is rapidly increasing. This information is not fully utilized by ophthalmologist, as they are expensive and have limited time. Giving rise the need of a comprehensive, mass screening system, this proposed system is economically extortionate apart from automated computer aided retinal images analysis.

One of the most important application is disease diagnosis through telemedicine, which is to remotely diagnose and provide treatment to patients. To achieve such functionality, mediocre quality fundus cameras have to be installed in places where patients are e.g. shopping malls, clinics and hospitals. Images captured by fundus camera will send to automated retinal image analysis for report generation, report will contain recommendations and diagnosis of a particular condition. Next the reports showing any alarming condition will send to doctor for further action. Automated retinal image segmentation will help to develop such system by segmenting pathologies (if present) and other retinal landmarks, whose size measurements are required for diagnosis of many diseases, for example macular degeneration required optic cup-to-disc ratio and artery-vein ration is required for hypertension retinopathy.

The proposed system is imperative for motherland, as in developing countries, the lifestyle and pollution accelerate the development of disease like DR, hypertension and glaucoma. This fact can be corroborated by the study carried out by British Journal of

ophthalmology on glaucoma, which concluded it as the second leading cause of blindness worldwide, affecting women and Asian. In 2020 total of 79.6 million people will suffer from glaucoma, form which 47% of open angle glaucoma (OAG) and 87% of angle closure glaucoma (ACG) will be represented by Asian [8]. However, in developed countries separate screening programs exists for glaucoma and age-related macular degeneration.

Eye care provider or other health care provider are humans, their decisions are subjective, it changes from individual to individual. Another application of computer-assisted disease detection system is to provide aid in grading the retinal images, which will help human to cope with inter and intra variability up to some extent.

1.7 Challenges

With advancement in technology and innovation, the computing power, machine learning methods and ways of communications are improving rapidly, providing opportunities for computer scientist and biomedical scientist and engineers to fulfill the needs of ever increasing clinical practices. Computer-aided systems for detection of disease from retinal images can help health care provider. To develop and test such system a considerable and diverse set of retinal images are required. To the very best of my knowledge there is no publically available dataset which provides gold standard for retinal pathologies and retinal anatomical structures. The publically available datasets either provides pixel level annotation for one object class e.g. vessels or exudates or doesn't provide pixel level annotation at all, DRIVE [11] and STARE [12] provides pixel level ground truth annotation for retinal vessels only whereas DRIONS-DB [13] and Messidor [14] don't provide pixel level annotations. Unavailability of retinal dataset hinders the research community for solving this problem. The development of public dataset will promote the research in automated retinal image analysis field. The biggest challenge in generation of automated retinal image analysis tool is availability of annotated retinal dataset, as it is laborious task and required training of annotator, supervision of domain experts and ophthalmologists.

Fine quality retinal image is fundamental for accurate and better results; however, usage of high quality camera will add significant factor in overall system cost. Retinal images are capture by fundus cameras of different resolution and quality, which induced significant variability in image quality, color and contrast, making it difficult for a single

technique to be used for retinal image analysis. A common normalization technique should be utilized which can handle the variability of input images.

Another challenge in simultaneous segmentation of retinal pathologies and retinal landmark is there is no end-to-end framework or architecture exists in literature to get the findings and observations to help in designing the proposed model.

1.8 Organization

This thesis is organized in four sections. Chapter 2 describes literature review until the time of this work on semantic segmentations techniques and segmentation techniques applied on retinal images. Chapter 3 presents our contribution on the field of segmentation. Proposed methodology is discussed in detail for segmenting retinal images. Chapter 4 will describe experimental details and performance measures for proposed method. Conclusion and future work are discussed in Chapter 5.

Chapter 2

Literature Review

In this chapter a general introduction of deep learning and its brief history is given, as well as why deep architectures are getting so much importance these days. It is necessary to grasp the main ideas of deep learning which in turn help in understanding the ideas and results interpretation presented in Chapter 3 and Chapter 4. This chapter also contains deep learning architectures used for semantic segmentation and will end with a brief review of research done so far for retinal fundus segmentation.

2.1 Deep Learning

Deep learning is subfield of machine learning and inspired by the working of brain. It enables the computational models to learn from experience and grasp environment in hierarchy of concepts [15]. Dividing the complex concepts into hierarchy making it possible for the artificial agent to comprehend complicated concepts which otherwise was not possible. Machine learning algorithms are linear whereas deep learning algorithms are hierarchical in level of abstraction and complexity. Deep learning models are composed of multiple processing layers which are known as artificial neural network (ANN) [16].

2.1.1 Neural Networks

Neural Network (NN) or artificial neural network (ANN) is a series of advance machine learning algorithms that attempts to identified the relationship between input patterns and output. The goal of NN is to discover high level representation from low level features. Neural networks imitate the working of brain, which is to divide the complicated concepts into sub-parts then gather the information on higher level to generate the general concept about the object. Different types of neural networks include convolutional neural network (CNN), feed forward neural network and recurrent neural network (RNN) etc.

2.1.1.1 Architecture

As the name suggests artificial neural network is inspired by the function of human brain. It has the same level of abstraction and hierarchy as human brain. For example, human visual cortex, the part of brain responsible of vision shows level of abstraction in the process of identification of the signal as an object. At low level features like edges, colors and gradient changes are identified, these low level features then further passed to next level for relatively high level features detection for example geometry detection. On final level complex visual shapes are identified. All of this process is carried out in visual by neuron cells, which are building blocks of brain. ANN works in similar manner, as depicted in Figure 4 at high level of abstraction man, horse can be seen, built upon lower level of abstraction like object parts (e.g. horse neck, legs etc.) which in turn built upon even more lower level of abstraction like color and edges. In ANN perceptron performed the functionality of neurons. Perceptron will be discussed in coming section.

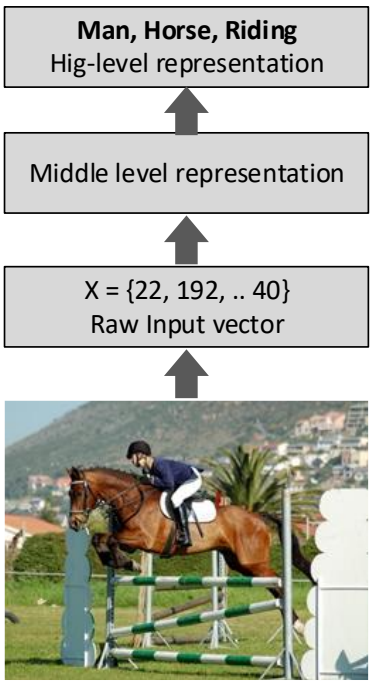

Figure 4 : Illustration of deep learning architecture

2.1.1.2 Perceptron

Neural networks are made up of perceptrons or neurons, which is nothing but a simple neuron that takes in input patterns (signals) in the form of real value vectors $X = \{x_1, x_2, \ldots, x_{n+1}\}$ with associated real values vectors of weights $W = \{w_1, w_2, \ldots, w_{n+1}\}$ and produce output *op* by the following function.

$$op = f\left(\sum_{i=1}^{n+1} w_i x_i + b\right) \qquad (1)$$

In above equation parameter b is **bias,** used to add offset to data and f(.) is non-linear activation function. Output of perceptron is called **activation**. Every input x_i is weighted by multiplication with respective factor w_i. Weighted inputs and bias are then summed up and passed through a non-linear activation function to get the final output of neuron [17]. Activation function used so that model can find nonlinear mappings. The basic structure of neuron is shown in Figure 5.

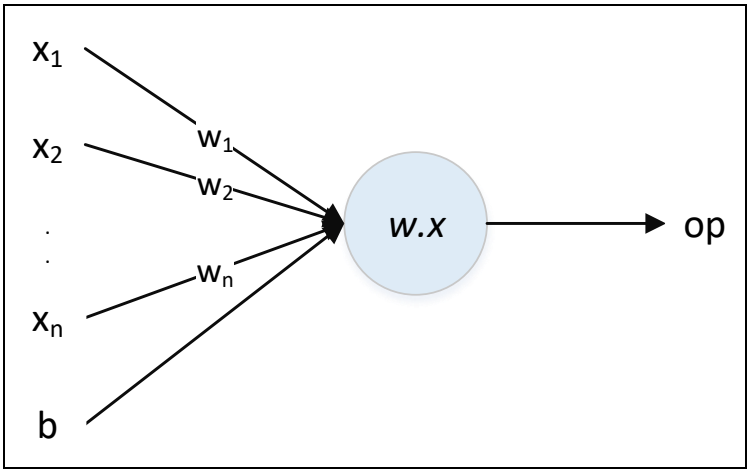

Figure 5 : Illustration of neuron

Choices for activation functions mostly used in deep learning are **sigmoid** function $f(x) = \frac{1}{1+e^{-x}}$ and **hyperbolic tangent** function $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. Recent research has proved a new and better non-liner function called **rectified linear** function (ReLU) $f(x) = \max(0, x)$ [18], it is not continuously differentiable like sigmoid and tanh. Output ranges of sigmoid is [0,1] and of tanh is [-1,1]. ReLU is piecewise linear and saturates at zero [16]. Plots of these functions are shown in Figure 6.

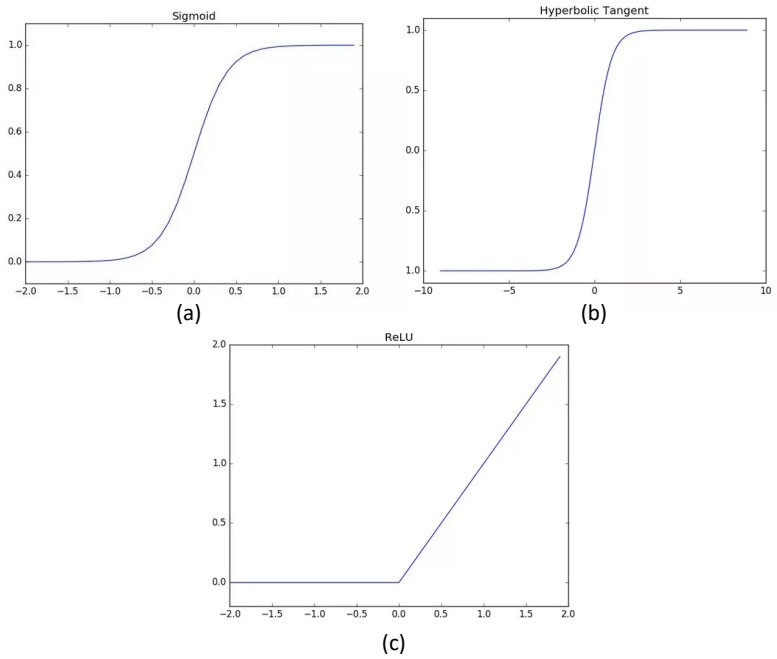

Figure 6: Non-linear activation functions. (a) sigmoid function (b) Hyperbolic tangent function (tanh) (c) Rectified linear function (ReLU)

2.1.1.3 Multilayer Perceptron

Basic computational unit of neural networks are neurons, single neurons connected with each other in a hierarchical fashion to made neural networks [17]. The output of one neuron is input of another neuron, as it can be seen in Figure 7 which show simplest model of neural network. This figure is an example of small multilayer perceptron or feed forward neural network, in which neurons are grouped into layers. Each neuron in preceding layer is connected to each and every other neuron of next layer without any loops or close connections [19]. The leftmost layer in Figure 7, L_1 is input layer, layer which receive input patterns in the form of vector, for every real value in input vector there must be a corresponding neuron in input layer. The rightmost layer known as output layer, a layer generates the final output of neural network, Layer L_3 in Figure 7. The middle layers are called hidden layers, because its value are not observed in training and testing phases [19] [17] shown by Layer L_2 in Figure 7.

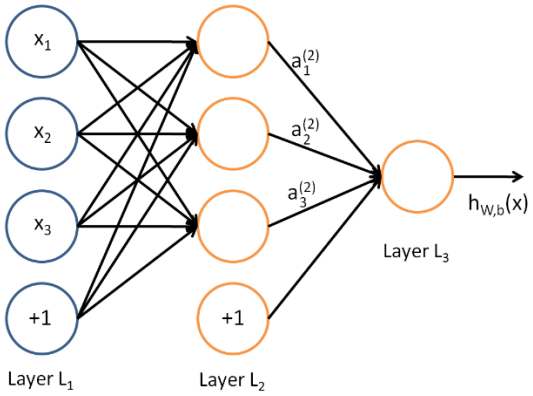

Figure 7: Feed forward neural network also known as multilayer perceptron

The output of neural network depicted in Figure 7 is mentioned to clarify the internal working of neural network. The first input layer is L_1 and parameters are given in matrix $W^{(1)}$ and $b^{(1)}$ is the bias value for the input layer. Similarly the parameters for layer L_2 are $W^{(2)}$ and $b^{(2)}$. Activations for hidden layers are computed as:

$$a_1^{(2)} = f(W_{11}^{(1)} x_1 + W_{12}^{(1)} x_2 + W_{13}^{(1)} x_3 + b_1^{(1)}) \quad (2)$$

$$a_2^{(2)} = f(W_{21}^{(1)} x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} x_3 + b_2^{(1)}) \quad (3)$$

$$a_3^{(2)} = f(W_{31}^{(1)} x_1 + W_{32}^{(1)} x_2 + W_{33}^{(1)} x_3 + b_3^{(1)}) \quad (4)$$

In above equations f(.) is non-linear activation function, $a_i^{(l)}$ denotes activation value for unit i in layer l. These activations are input for the hidden layer L_2. The output of the final layer L_3 is calculated as follows

$$output = a_1^{(3)} = \sigma(W_{11}^{(2)} a_1^{(2)} + W_{12}^{(2)} a_2^{(2)} + W_{13}^{(2)} a_3^{(2)} + b_1^{(2)}) \quad (5)$$

Whereas the σ is a function other than hidden layer activations. Computation of output from the mentioned functions is known as feed forward neural network, on the basis of the fact that inputs are being forwarded in the network [20]. A very basic structure of feed forward network is presented in this section, however number of hidden units in a layer and total number of layer in a network varies depending upon the problem at hand. Further it is not compulsory that the network can only have one output value, multiple output values are possible e.g. in case of multi-class classification problem.

2.1.1.4 Gradient Decent and Back Propagation

As the neural network goes deeper having more and more hidden layers and number of units tend to increase, an issue arises how to find the values of parameters (W, b) [19]. To train a neural network i.e. to find the values of parameters gradient decent and back propagation is used, although there are many methods to train the neural network but in this section the focus is on training of a neural network using back propagation and gradient decent. The objective of training is to find values of parameters weights and biases such that the error function E (W, b) is minimized. Error function also known as loss function is measure of disparity between the output generated by the deep neural network and the desired output. It is continuous function over the parameters of network.

Gradient decent belongs to class of optimization algorithms, it is a method to find the local minima of the function. In deep learning it is mostly used with back propagation to adjust the parameters of network proportional to the negative of gradient [21]. Back propagation required desired output value and generated output value and propagate the error value to every neuron of each layer, until each neuron has an error value proportional to its contribution in generated output value. The next step is updating the parameters value accordingly. Detailed discussion on the technique will go beyond the scope of this thesis.

2.1.1.5 Overfitting and Underfitting

The goal of machine learning is generalization i.e. to generalize model so that it can predict output values for unseen data, because no matter how much data is available during training, the chances to see the exactly same example in testing is very slim. Overfitting occurs when machine learning algorithm or model fits the training data too well, but not necessarily that it'll predict the test data accurately [22]. Another way to look into overfitting is by decomposing generalization error into **bias** and **variance**. Bias is the machine learning model's likelihood to learn the same wrong thing over and over again while variance is aptness to learn insignificant thing like noise irrespective of actual data trend. In overfitting the model displays low bias and high variance. To avoid overfitting cross validation is used. Let's consider an example for better understanding, where the data is generated by function $\sin(2\pi r)$ with some noise addition. We want to fit a polynomial of order M for this function. As shown in the

Figure 8 M=9 higher order polynomial fits training data perfectly, it covers each data point of training data meanwhile it performs oscillations along the normal curve making it a poor representative of actual function [19]. This polynomial performs excellent on training data but fails on unseen data, so if a model performs well on training data it doesn't grantee that it will perform good on test data.

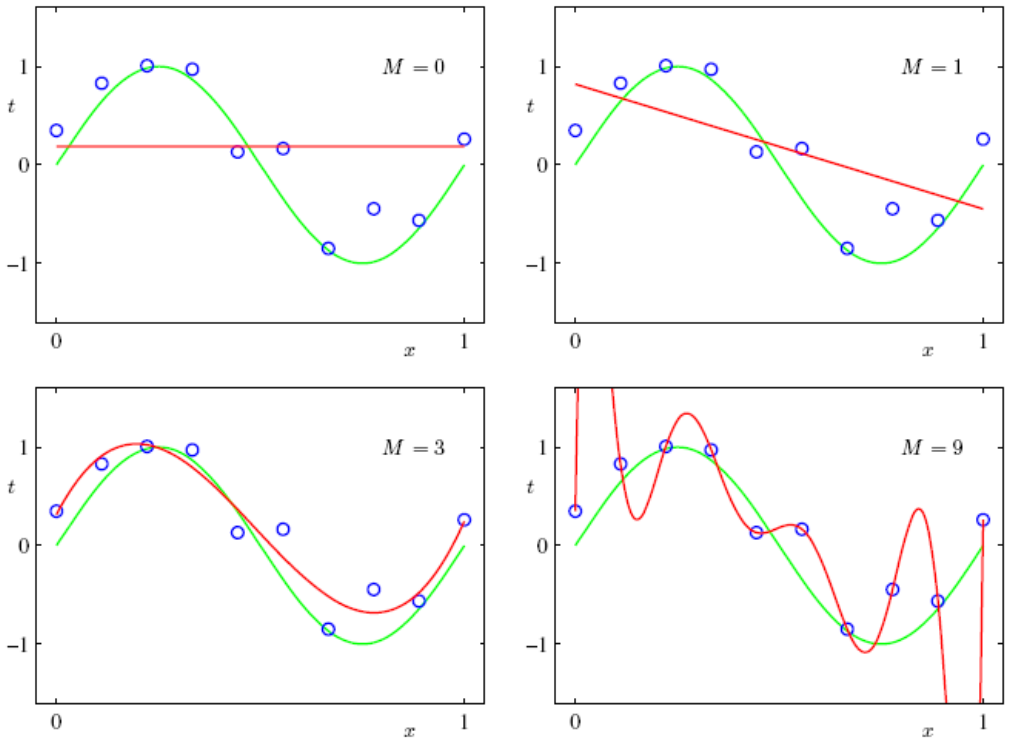

Figure 8: Polynomial plots with different order M are shown in red color and the actual function $\sin(2\pi r)$ is shown in green color.

Underfitting occurs when a machine learning algorithm or model failed to represent the trend of underlying data. In short underfit model does not fit the data very well on training and testing phases [22]. Precisely underfitting can be result if model display high bias and low variance. Let's continue with polynomial example, low order polynomial M=0 and M=1 are not capturing the underlying trend of data. These polynomial are too simple to express the function $\sin(2\pi r)$. As depicted in Figure 8 they miss most of the training data points [19]. Overfitting and Underfitting are among one of the most important factor that should be keep in mind while designing a machine learning algorithm, they lead to poor prediction if proper care is not followed. If more complex model is used it results in overfitting and if too simple model is used

that turns out a case of underfitting, as in the given example the correct polynomial that represents the function $\sin(2\pi r)$ is M=3.

2.1.1.6 Optimization Algorithms

Optimization algorithms are used to maximize or minimize the objective function [23]. These algorithms operate in cycles consisting of two phases propagation and weight update. The error values utilized by the back propagation process to update the weights values come from optimization algorithms. Some of the most used optimizer algorithms are discussed in this section.

2.1.1.6.1 Stochastic Gradient Decent

Stochastic gradient decent or SGD performs the parameters update of deep learning model for every training example. This technique is much faster than standard gradient decent but have high variance of parameters updates due to frequent updates of parameters, which in result of fluctuation of loss function. This fluctuation helps SGD to jump to a better local minima [23].

2.1.1.6.2 Momentum

The high oscillations of SGD make it difficult to reach convergence. To solve this problem with SGD a technique Momentum has been introduced [24], it helps in restraining fluctuations and accelerates SGD in relevant direction. It reduces the updates of the parameters whose gradient change direction and enlarge the updates for parameters whose gradient change is in the same direction [23]. It can be represented mathematically as below

$$V_t = \gamma V_{t-1} + \eta \nabla J(\theta) \qquad (6)$$

$$\theta = \theta - V_t \qquad (7)$$

In equations above θ is parameters of model, γ is momentum, η is learning rate, V_t and V_{t-1} are current and previous update vector respectively. The momentum γ is usually set to 0.9. All it does is adds a fraction (momentum) of the past step update vector to the current update vector.

2.1.1.6.3 Nesterov Momentum

Yurii Nesterov was the first person who observe a problem with momentum, that is at the lowest point of the curve due to high momentum the parameters are updated by a huge factor making it to miss the minima. Nesterov accelerated gradient (NAG) [25]

is the technique which solve this problem with momentum by adding a correction factor to the gradient. NAG evaluated the gradient after applying current velocity. In other words, it used the future parameter values instead of current parameters values in calculating the gradient [23].

2.1.1.6.4 AdaGrad

Adagrad is an adaptive method that takes into account the geometry of data [26]. It allows to set the learning rate based on the parameters, updates are immense for infrequent and small for frequent parameters. The learning rate is modified for every parameter at each step. The major advantages of Adagrad are the learning rate is not required to manually tuned and it is best suited for sparse data [23].

2.1.1.6.5 AdaDelta

AdaDelts is extension of Adagrad, which handles the monotonically decaying gradient issue faced by AdaGrad [27]. Instead of using all the previous squared gradient as in AdaGrad, it restricts the gradient square size to a fixed value.

2.1.1.6.6 Adam

Adaptive Moment Estimation (Adam) is another adaptive method that calculates learning rate for each parameter [28]. In addition to store the square gradient average like AdaDelta it also stores the exponentially decaying average of past gradients.

2.1.2 Convolutional Neural Network

Convolutional neural network or CNN also known as convolutional networks or ConvNets are similar to neural network which are discussed in the previous section, they are made up of neurons which have learnable parameters weights and biases. Each neuron perform a dot product of inputs and weights then followed by non-linearity. They also have a loss function and all other architecture tricks of a normal neural network. The change is, a CNN takes special assumption about the data i.e. used for data having grid like structure [15]. For CNN the inputs are mostly images, a grid of dimension height x width with certain depth (three for RGB and one for gray scale). CNN get its name from a mathematical operation **convolution**, which is the major part of these types of networks. CNN are neural networks are which use convolution in any one of their layer instead of general matrix multiplication [15]. In general, a CNN can be divided into three basic components

1. Convolutional layer

2. Pooling
3. Output layer

Where the pooling layer is optional, how to use pooling layer depends upon the architecture of CNN. Brief overview of each layer is discussed below.

2.1.2.1 Convolutional layer

As the name suggests, in this layer the input is transformed into output using a convolutional operation. The input is convolved with a kernel or filter to generate output. Network trainable weights matrix is called kernel or filter; hence the total number of weights depends on its size. Convolution operation is to multiply the input values with values of a kernel and sum them up.

Like in neural network every layer is connected to next layer but this type of connection is not fully-connected for entire grid of neurons instead the connection is made in a small region, for example in a region of 5×5 with corresponding 25 values. This is illustrated in Figure 9. The area of input image or grid under the filter is called **receptive field** it is like a little window of input layer. The filter is slide across the entire input image to generate the output, for input of 20×20 with filter of size 5×5 the output will be 16×16 with stride size of one. This reduction in size is because we can only slide the receptive field 15 time across the input before meeting the right end of input. **Stride** is the length of movement the sliding window takes, in the given example every time the window slides one pixel to right.

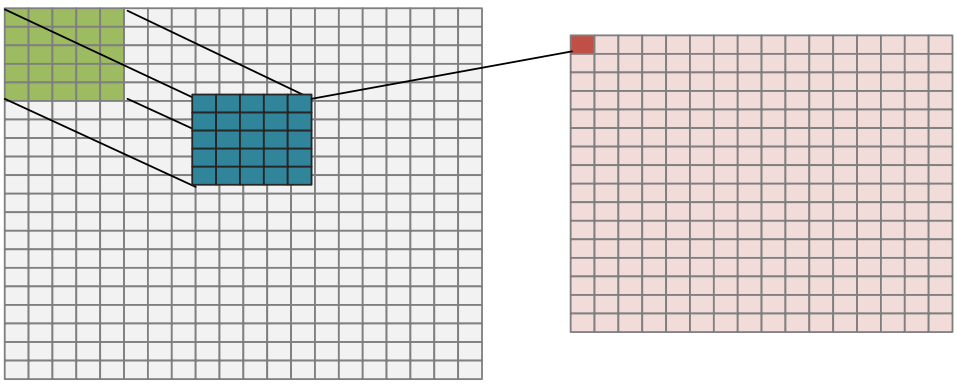

Figure 9 : Convolutional operation with a kernel. Receptive field is shown in green, kernel in blue and resulted output in red color.

The number of activation maps varies according to number of filters used for convolutions. If in the example given above, ten filters are used in input layer then ten activation maps would be generated, dimensions would be 16×16×10. In this convolutional layer, 10 filters of size 5×5 used therefore contain 5×5×10 = 250 weights.

2.1.2.2 Pooling

The pooling layer are not required part of CNN, its usage depends upon the input and architecture of model and targeted problem. In CNN pooling layers are used to simplified the information of convolutional layer. It is mostly followed immediately after the convolutional layer. It takes activation maps or features map from convolutional layer and compress the feature map information in smaller size [15]. The purpose of pooling layer is to reduce the spatial size of feature map and sometimes to reduce the number of trainable parameters, it is mostly used in CNN that works on large size of input images. Pooling operation is performed on each and every dimension so that the generated output will have same height and width, depth of feature map remains intact however height and width are reduced.

The most frequently used pooling type in deep learning is **max-pooling**, the pooling operation simply output the max value from the slided window. Size of the window is selected by the deep learning algorithm designer. For the example given in previous section 2.1.2.1, the feature map 16×16 if pass to a 2×2 size pooling layer the resultant feature map will have a size of 8x8. This is illustrated in the Figure 10 below.

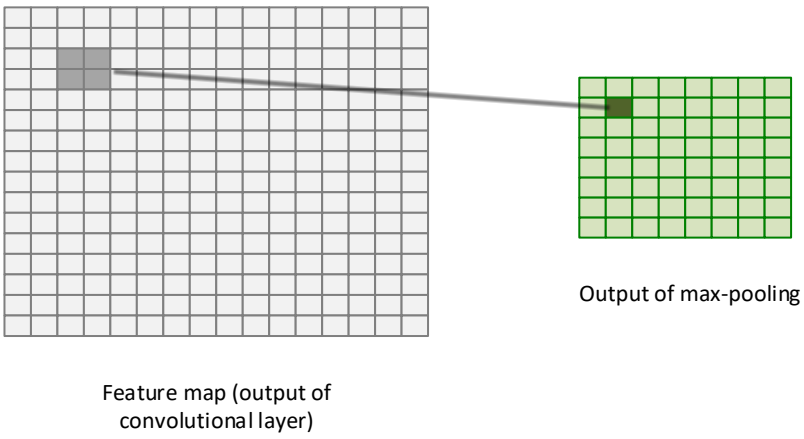

Figure 10: 2×2 max-pooling results on 16×16 input feature map and its result of 8×8 feature map

2.1.2.3 Output layer

Design of output layer depends upon the problem, what is required from a CNN. In general, the output layer of CNN doesn't differ from the normal neural network. For example, the CNN which classify the input image into certain number of output class the output layer would be a fully-connected layer having number of neuron equal to number of classes to distinguish between.

2.1.3 Semantic segmentation

Brief introduction of semantic segmentation is discussed in previous Section 1.1. in this section some of the famous deep learning based techniques and architecture are introduced, although they are not directly applied on retinal images but getting to know them will help to understand the proposed methodology. Almost every deep learning segmentation model is based on **encoder-decoder framework**. The encoder extract features from input image and reduced the spatial resolution of input image. Reduction in spatial resolution mostly perform by max-pooling layers. Second part is the decoder, which is responsible for recovering the spatial resolution and generating the dense classification by using the features learnt by encoder on pixel space [2]. The general framework of encoder decoder is shown in Figure 11.

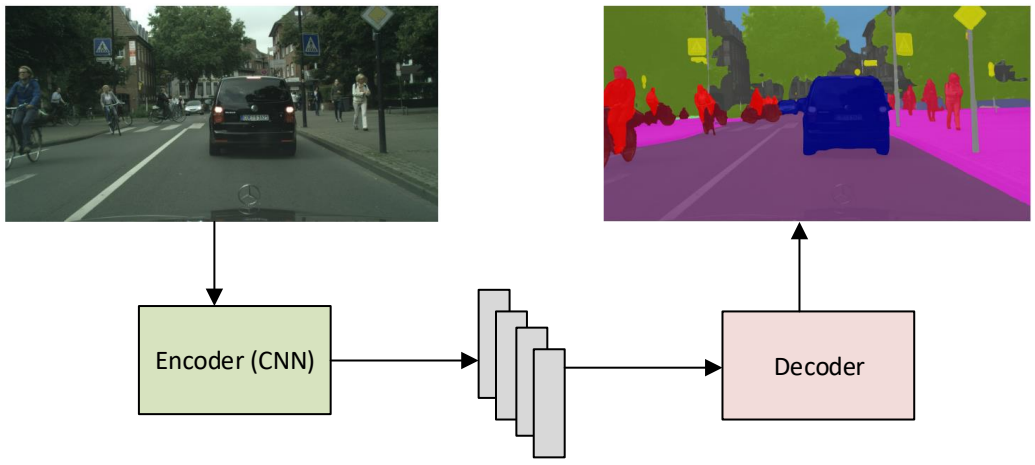

Figure 11 : Encoder decoder framework, the encoder extracts features which are utilized by decoder for pixel class label generation

2.1.3.1 Fully Convolutional Network

Fully convolutional neural network or FCN was first of the deep learning based model designed to performed segmentation of common objects. FCN has same architecture as of CNN but it has replaced the last fully-connected layers with convolutional layers. By incorporating convolutional layer instead of fully-connected layers makes the network to generate corresponding output segmentation map of arbitrary size input images [29]. One of the major flaw of this design is that it constructs coarse segmentation maps and object boundaries are not refined. To improve over the coarseness of spatial resolution and information loss due to up-sampling, skip connections are added between encoder and decoder pairs [29].

2.1.3.2 U-Net: Convolutional Networks for Biomedical Image Segmentation

U-Net [30] is the first deep learning based encoder decoder model. The encoder reduces the spatial dimensions gradually by applying pooling layers, decoder recovers the spatial resolution and object details. Skip connections are also added between encoder and decoder blocks to regain the lost information during the pooling operation, but still the segmentation maps generated by U-Net don't have crisp object boundary. U-Net was the first deep learning based architecture which was applied on biomedical images and heavily used data augmentation to cope up biomedical image limited data problem.

2.1.3.3 SegNet

FCN used skip connections and up-convolutions but still the segmentation map generated by FCN are coarse and delineated object boundaries, SegNet try to cover the flaws of FCN by adding max-pooling indices during up-sampling [31]. The remaining design is same as of FCN.

2.1.3.4 Multi-Scale Context Aggregation by Dilated Convolutions

Dilated convolutions or atrous convolutions is one of the among major architectures which has been presented in literature. Dilated convolutions are special type of convolutions, which is convolutions with defined gap. This technique replaces pooling operation by dilated convolutions. This change supports exponentially expansion in the receptive field size and keeping the resolution intact. The number of trainable parameters remains same and so as the original resolution of input [32]. It generates

the segmentation map of 1/8th size of input resolution which are then interpolated to get the final results. Dilated convolutions are not appropriate for extracting features of smaller objects due to gaps introduced to enlarge receptive field.

2.1.3.5 Fully Convolutional DenseNet

Fully convolutional DenseNet or FC DenseNet extends the a recently proposed CNN architecture "Densely Connected Convolutional Networks" or DenseNet. In the DenseNet architecture every layer is directly connected to all previous layers in a feed forward fashion [33]. Adding such connections have several advantages like vanishing gradient, better feature propagation and reduction in number of parameters. FC DenseNet utilize this network for encoder, for generation of segmentation maps decoder is introduced made up of transitions-up blocks. Transition-up blocks are made up of transposed convolutional layers that up-samples the previous feature maps [34].

2.1.3.6 Mask RCNN

Mask-RCNN is extension of Faster R-CNN, it uses the same intuition behind that faster RCNN performs so well for object detection than its extracted features are used for pixel-wise labeling too. it adds an additional branch for generation of object mask in parallel to object bounding box generation in faster R-CNN [35]. This special branch is just a fully connected neural network FCN which takes faster RCNN features as input and generates segmentation maps.

2.2 Semantic Segmentation Applied on Retinal images

For the last couple of decades, the retinal image analysis field has attracted a lot of research interest, where automated diagnosis of diabetic retinopathy (DR) has received substantial share of this interest. Another area which has received part of this interest is retinal landmark detection. Landmarks are retinal anatomical structures like optic disc, blood vessels, macula and fovea. In this research simultaneous segmentation of retinal landmarks and pathologies are addressed. Thus this section will start with brief review of available techniques in literature to segment retinal blood vessels, it is utilized as a prerequisite for DR detection especially new vessels detection methods. Further a review of optic disc and retinal pathologies (hemorrhages, exudates and cotton wool spots) segmentation procedures will be mentioned.

2.2.1 Vessels Segmentation

Segmentation of retinal blood vessels proves to have significant role in medical screening and early diagnosis of diseases, such as diabetes, hypertension, multiple sclerosis and choroidal neovascularization [36]. The primary traits of vessels that are used in segmentation techniques are their dark red color, contrast, huge gradient at vessel edges and distinction from background. According to literature vessels segmentation methods can be divided into four major groups based on morphology, matched filtering, vessel tracking and deep learning. Key concerns for vessels segmentation methods are detection of vague, fine and twisted vessels, which can further complicated by vessels crossing and branching. Pathologies mostly hemorrhages also impede the process accuracy and increase the false positive rate.

Deep learning (DL) based algorithms shows impressive advantages as compare to other methodologies, the state-of-the-art has been constructed for retinal vessels detection and segmentation. Melinscak *et al* [37] proposed a deep neural network to classify each pixel as vessels (foreground) or background in retinal fundus, but this work has two flaws, first it failed for local pathology region and second pixel-wise classification is time exhaustive for both testing and training phases. Whereas Fu *et al* [38] formularized the vessels segmentation into boundary detection, it utilized convolutional neural networks (CNN) and conditional random fields (CRF) to generate the probability maps for vessels. First a large receptive field CNN generates the probability map on pixel level, although which results in coarse vessels segmentation and then a fully-connected CFR is used to acquire binary vessels segmentation results.

Similar approach was used by Wang *et al.,* [39], which employed a Random Forest (RF) as classifier over the extracted features of CNN. The CNN has five alternating convolutional and subsampling layers which extracts features from input retinal fundus, after that an ensemble of three Random Forest is trained with features extracted by CNN. However the proposed system again suffers from two main issues as of Melinscak *et al* [37], the system takes plenty of time for pixel-wise classification in both training and testing phases, second it is failed to tackle the presence of pathologies. These issues are answered by the use of structured prediction, Liskowski and Krawiec [40] compared single pixel classification and structured prediction based vessels segmentation, which classify multiple pixels at a time instead of one. And

proved to be more computationally efficient. Fang *et al.,* [41] proposed somehow similar pipeline for retinal landmark i.e. vessels delineation, instead it get benefit from multiscale Hessian response on the probability maps for vessel segmentation. Firstly, a neural network output the label map then a probability map is constructing by synthesis strategy. Then eigenvalues and second derivatives are calculated to determine the Hessian matrix, the intuition behind this approach is that vessels have high Hessian response on probability maps.

Traditional methods for vessels segmentation used active contour with hybrid region [42], morphological component analysis (MCA) algorithms [43] and multiple-scale Hessian approach [44]. Level set method based on the local cluster value via bias correction which used a local cluster value to detect changes in local area [45]. Major benefit of using this technique is it can detect vessels in low contrast image. Another approach is a multi-scale tensor voting approach which is efficient for segmenting small retinal vessels, it utilize multiscale for tensor voting and line detection and adaptive thresholding is used to reconstruct small vessels [46].

2.2.2 Optic Disc Segmentation

Various methods are developed and evaluated on publically available dataset for optic disc (OD) segmentation. Srivastava, R., et al. [47] was the first attempt in literature that utilized deep neural networks (DNN) for delineation of OD. The approach works on colored fundus images for training and testing, it first detects OD in input images by applying circular transform then a region of interest (ROI) of 800×800 pixels is cropped around the detected OD center, this image then resized to 300×300 to keep the count of trainable parameters limited which in turn males the DNN lightweight. Contrast Limited Adaptive Histogram Equalization is applied on ROI to improve its contrast. Pixel intensity and distance features are extracted from ROI and a DNN is trained over these extracted features. Stacked Auto Encoders (SAE) are the main component of DNN, which is a type of Auto-Encoders a neural network which generates its input. The DNN is trained in two phases, first phase is unsupervised in which the SAE layers are trained in greedy fashion. Second phase is supervised, which utilized the weights learnt in first phase as initialization weights for the network, the whole network again trained in feed-forward backpropagation manner.

Maninis, K.-K., et al. [48] used VGG-16 [49] based fully-convolutional network (CNN) and transfer learning technique for OD delineation. This method achieved 0.97 Dice score which is more consistent to ground truths and approximate to human annotator. However, this approach has numerous disadvantages, the neural network used takes a lot of time for training and testing, the model has large number of network parameters and its file size is considerable so as GPU required memory.

A more recent and robust technique is presented in [50] which address the problems of previously discussed method. A pre-processing of retinal fundus is added to improve the accuracy of the deep learning network, Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to equalized the contrast and changing the color of image. The presented network is inspired from U-Net [30], based on encoder-decoder architecture. However, the network differs from original U-Net in terms of having less number of filters in each convolutional layer and every deeper layer the number of filter remains same. This change doesn't affect the recognition of OD but reduced the training time, make the network architecture more frivolous in terms of file size (memory to store model) and number of parameters required to train the network.

Recently a new architecture is proposed by Fu, H., et al. [51] which use M-net shaped architecture for joint OD and optic cup (OC) segmentation. This M-net is a U-shaped multi-scale convolutional network, compromising multi-scale input layers, multi-label loss function and side-output layers. It first localized OD detection mechanism is used to detect center of OD; the retinal fundus is transfer into polar coordinates based on the detected OD center. This transferred retinal image is passed to M-net which produced probability maps for OD, at the end the image again transforms back into Cartesian coordinates. It is an encoder-decoder architecture where the convolutional network is modified U-Net, with multiple input layers. Original image is rescaled and concatenated with respective feature maps during the contracting path of U-Net. Side-output layers generate local output maps for early layers and it serve the purpose of a classifier. To obtained multiple output classes (OD and OC) the multi-label loss function based on Dice coefficient is inducted.

Many traditional techniques are available to segment optic disc i.e. Circular Hough transform and grow cut algorithms [52], supervised gradient vector flow (SGVF) [53], local binary energy fitting [54].

2.2.3 Hemorrhages Segmentation

The first deep learning based work for hemorrhages segmentation was [55], which utilized deed forward back propagation neural network, followed by Larsen, M., et al. [56] doesn't describe the method in detailed for hemorrhages and micro-aneurysms detection. Delineation of red lesions is performed by other machine learning algorithms not necessarily CNN. Larsen, M., et al. [56] quantified the performance of four classifier for this task, MLP (Multi-Layer Perceptron), SVM (Support Vector Machine), RBF (Radial Basis Function) and MV (Majority Voting) and concluded that the finest outputs was obtained by RBF.

A recent approach is proposed by Tan, J.H., et al. [57], they devised a ten layered neural network (including input layer) for simultaneous segmentation of hemorrhages, micro-aneurysms, and exudates from retinal fundus images. The DNN is simple and lightweight, having three convolutional layers, each followed by max-pooling layer. Last two layers are fully connected layers followed by a Softmax function, weights for last layer is generated randomly by Gaussian distribution. Number of feature maps per layer is not more than 64, which reduces the trainable weights. This works produced poor delineation of hemorrhages boundary i.e. no precise localization and generated segments are of large size as compare to ground truth.

Other than deep learning based approaches for red lesions detection, traditional approaches are also available in literature mostly morphological based methods are commonly used for hemorrhages detection. The limited factors for such methods is vessels and hemorrhages having same color, most technique mark vessels as possible candidate then in next step vessels and hemorrhages are segregated, [58] use scale based method for vessels segmentation then apply gamma correction and global thresholding to delineate hemorrhages. [59] used a neural net to segregate vessels from hemorrhages, for detection of candidates pixels Moat operator is used. Similarly [60] also use CNN for separating vessels from hemorrhages, it utilize morphological operators and thresholding to segment blood vessels and hemorrhages.

2.2.4 Exudates Segmentation

In the literature, several methods are cited for automatic detection and segmentation of retinal pathologies, including hard exudates and soft exudates (cotton wool spots). In general, majority of the techniques start with retinal fundus preprocessing, next step

is candidate extraction for exudates and at the end a classification is employed over the candidate regions to get only exudates.

A combined detection algorithm is presented by Prentasic et al., [61], it detects retinal landmarks and its results are combined with exudates detection algorithm to improve the performance of the proposed system. Landmark detection is performed by using traditional image processing techniques to obtained probability maps of retinal landmarks. Exudates detection is performed by a ten layer CNN, having four convolutional and max-pooling layers with last two layers are fully connected. Softmax activation is added after the last layer to generate multi-class probability map. These probability maps are combined and final probability map is generated.

A novel technique based on the state-of-the-art residual network for exudates segmentation is presented in [62], however this approach also performs so well for classification of exudates i.e. presence of exudates in a patch which is sufficient for the clinicians to take appropriate action. Patch based learning is incorporated during training of CNN due to presence of small number of exudates and other bright spots in input images. This issue is well tackled by patch based learning, which reduce the false positive rate, leaving no need for post processing. The proposed network consists of nine ResNets blocks and a total of twenty convolutional layers. A ResNet block is made up of stacked of convolutional, batch normalization and ReLU layers. During the testing phase the entire image is pass through the trained network instead of patches by the reason of the network doesn't change the spatial resolution of input image, size of patches is large enough to represent the contextual information, hence the learnt weights of network are sufficient to applied on entire image, reducing the testing time.

To reduce the amount of data for training the neural network for exudates detection as the ground truth creation for retinal lesions are time consuming and expensive as it required trained observer and ophthalmologist to annotate the dataset. Otalora et al., [63] focused on active learning, which is an active area of research in machine learning. The main notion behind is the machine learning algorithms required less data samples if it choose data intelligently from training dataset. The CNN used in technique is based on LeNet, having total seven layers including patch input layer, two convolutional layers followed by max-pooling layers, Softmax function is added at the end of network for classification of a patch as being normal or have exudates present.

The deep neural network is label-efficient CNN having small number of layers to avoid computational cost and training time, most relevant patches selection for the training of network helps in early convergence to local optimum as compare to stochastic gradient decent.

Chapter 3

Proposed Methodology

To segment retinal images requires many stages regarding the dataset selection, network architecture and design and performance measures to evaluate the results of technique. This chapter discuss in detail the proposed methodology to solve the segmentation problem applied on retinal images and tools and technology used to implement the proposed method. Thereupon each stage is discussed in greater detail in subsequent sections which explained the reason for all choices made regarding to training network and selection of parameters.

An overview of the proposed network design is displayed in Figure 12. The design can be divided into following parts:

- Pre-processing to normalize input images
- Creation of patches
- Training of Network
- Generation of segmentation maps

The architecture can be divided into two main sub-parts i.e. encoder and decoder. Encoder is responsible for extracting features from input retinal fundus. Decoder generates pixel maps for each class from the features learnt by encoder on pixel space.

3.1 Pre-processing

Good quality retinal image is fundamental for accurate and better segmentation of retinal anatomical structure, which in turn used for diagnosis of many retinal diseases such as diabetic retinopathy, age-related macular degeneration, atherosclerosis and glaucoma etc. However, due to inadequate quality of retinal image approximately 12% of retinal images can't be analyzed clinically either manually or automatically [64]. Retinal images suffer from non-uniform illumination and inconsistency in contrast because of acquisition methods involved and anatomy of retinal fundus, insufficient pupil size, angle of fundus camera, movement of eye, opaque media and geometry of

sensor array. Therefore, the improvement of image quality can play an essential role in retinal images analysis. To use computer-aided system for diagnostic purpose such type of images need to be pre-process, so that output shouldn't affect the outcome of system.

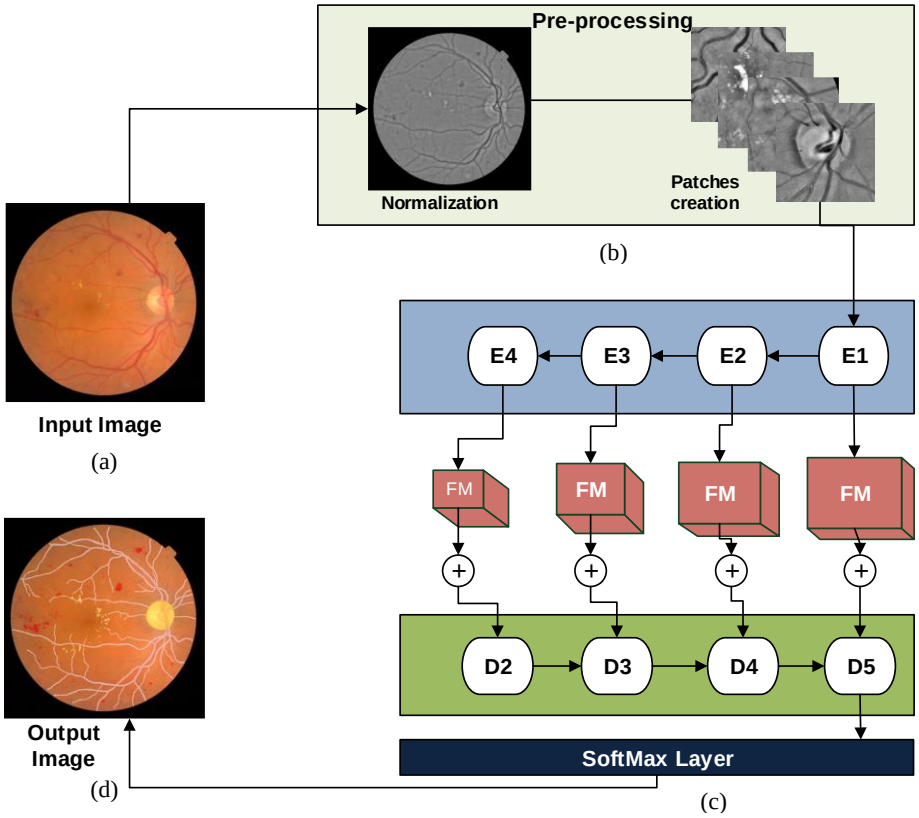

Figure 12: (a) Input Retinal fundus image (b) Pre-processing (c) Encoder Decoder based Fully Convolutional Neural Network (d) Segmented retina fundus overlaid on original image

In literature a lot of diverse algorithms are available for retinal images contrast and luminosity normalization but in this research I use pre-processing algorithm presented in [65] to obtain retinal images with enhanced contrast between retinal anatomical structure and normalized luminosity. This technique first estimates the background and later calculate the difference between the estimated background and green channel to generate the normalized image. To estimate background image, the original retinal image green channel is convolved with a large arithmetic mean kernel. This background image is denoted by I_{bg}. Kernel dimension is tuned in such a way so that the resultant blurred background doesn't contain any visible retinal anatomical structure, making the kernel size an insignificant parameter. The difference between the calculated background image (I_{bg}) and morphologically opened image (I_{open}) is

measure in order to obtain normalized image (I_{norm}), this can be mathematically represented as below

$$I_{norm}(x,y) = I_{open}(x,y) - I_{bg}(x,y) \quad (8)$$

There is significant intensity variation among images due to different illumination conditions in image acquisition process, which making it difficult to use same technique for all images. Thus making shade correction an indispensable part of proposed pipeline. A global linear transformation function is used to cater wide alternation in illumination condition. This transformation function alters each pixel value in such a way to utilize the whole gray levels range i.e. 0 to 255, which will rectify shade, enhance contrast and reduce luminosity variations.

$$I_{sc}(x,y) = \begin{cases} 0, & I_{norm}(x,y) < 0 \\ 1, & I_{norm}(x,y) > 1 \\ I_{adjust}(x,y), & \text{otherwise} \end{cases} \quad (9)$$

$$I_{adjust}(x,y) = I_{norm}(x,y) + 0.5 - I_{maxPxl}(x,y) \quad (10)$$

Where I_{sc} is the shade corrected homogenized image, I_{norm} is the normalized image, I_{adjust} is the adjusted image, and I_{maxPxl} is the gray level pixel value with the highest occurrence in the normalized image I_{norm}, pixels having intensity value I_{maxPxl} belongs to retinal background. The transformation will change pixel intensities of I_{maxPxl} pixel to 0.5, other background pixels having different illumination will normalize around this value. Transformation results are depicted in Figure 13.

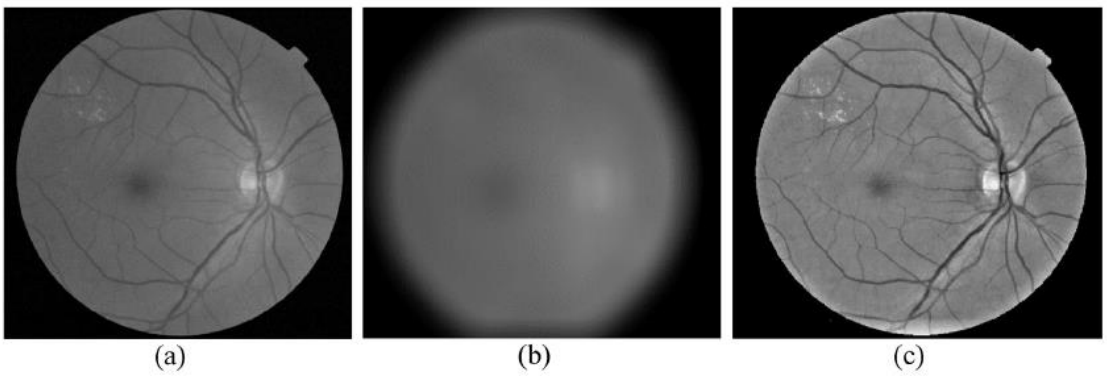

Figure 13 : Normalization of retinal image (a) Green channel input image (b) Background image (c) Normalized image

3.2 Patch-based Learning

Retinal fundus has several anatomical structures of different shape and geometry. A detail discussion on retinal anatomical structures is already provided in Chapter 1 . The size of optic disc and vessels remains persistent in almost all retinal images but the size of retinal lesion varies from image to image. In general, retinal lesion size is small in comparison to other retinal structures. Therefore, to segment optic disc and vessels in full retinal fundus is easy, whereas retinal lesions are difficult to segment due to capricious size.

In this research to solve this problem patch based learning is employed which enable the neural network to segment small structures quite efficiently. A patch is a small portion of complete image, it is also known as window. Patch-based learning is to divides the input image into small patches and use them as input of deep learning algorithm instead of whole image. For example, patch-based learning employed on a CNN to classify images then it will classify the patches individually and later final output classes has been determined from what patches have classified. This technique is becoming very popular and recent studies has shown to be useful and effective.

On the other hand, creation of patches also divides the pathologies and other anatomical retinal structures i.e. optic disc, which slim the chances for the network to get most from contextual information. To solve this problem related to patch-based learning and to get benefit from contextual information overlapping patches are practiced, as it suppresses patches with pathologies and optic disc near boundary to be misclassified. Figure 14 depicts the overlap patches creation, as it can be seen that if a pathology or retinal landmark divides in a patch that contextual information is somewhat restore in overlap patch.

Moreover, overlapped patches helps in accomplish data augmentation, as it applied translation on patches so that the object can be identified in any part of input image. Data augmentation making the network to train well even for smaller dataset size. Hence overlapping windows helps in lessening false positive rate. For generation of overlap patches a K×K size sliding window is slide over the original retinal image with stride size 'S'. The stride size 'S' is extend to the degree overlapping is required between patches e.g. K/2, K/4 etc. Only those patches are used for training the network which lie within field of view (FOV) as the region outside the FOV is of no practical use.

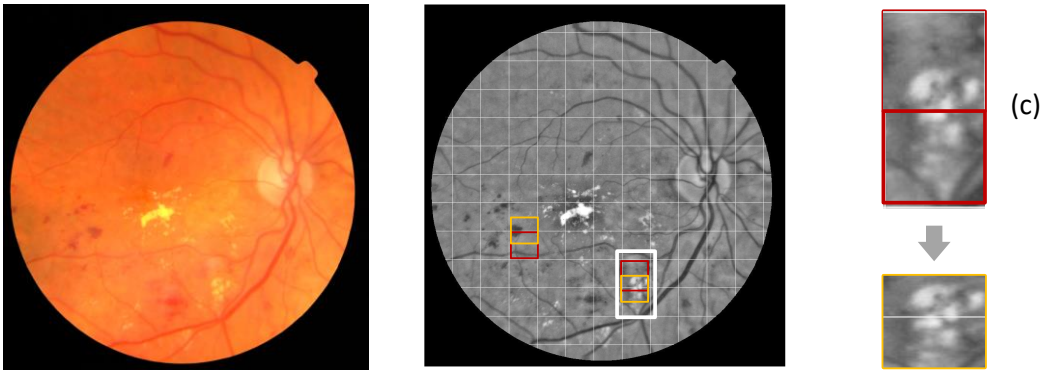

Figure 14 : Overlapped patches use contextual information and prevents small lesions to be misclassified (a) input image (b) overlap patches overlay with normalized image (c) two adjacent patches which divides the exudate (d) overlap patch show whole exudate

3.3 Network Details

The proposed network is based on encoder-decoder architecture. Fully convolutional networks are used for both encoder and decoder for semantic pixel-wise delineation of retinal anatomical structures. This work has taken the inspiration from U-Net [30] The network learns high level complex features and use these learnt features to generate pixel wise segmentation of retinal abnormalities and other anatomical structure by assigning particular object class label i.e. either 'exudates', 'hemorrhages', or 'vessels' to each pixel. Total number of classes in this work are six including one class for background. The classes are listed below.

1. Vessels
2. Optic disc
3. Hemorrhages
4. Exudates
5. Cotton wool spots
6. Background

3.3.1 Model Architecture

The Figure 15 shows the proposed network design. The proposed network is a modified version of U-Net [30]. The network can be divided into two sub parts i.e. encoder or down-sampling path and decoder or up-sampling path, both of them are

designed by stacking the layers of network over each other. The down-sampling path is composed of five blocks; a block is nothing else but just stacked of following layers

1. Convolutional Layer
2. Non-linear activation
3. Batch normalization (BN)
4. Pooling layer

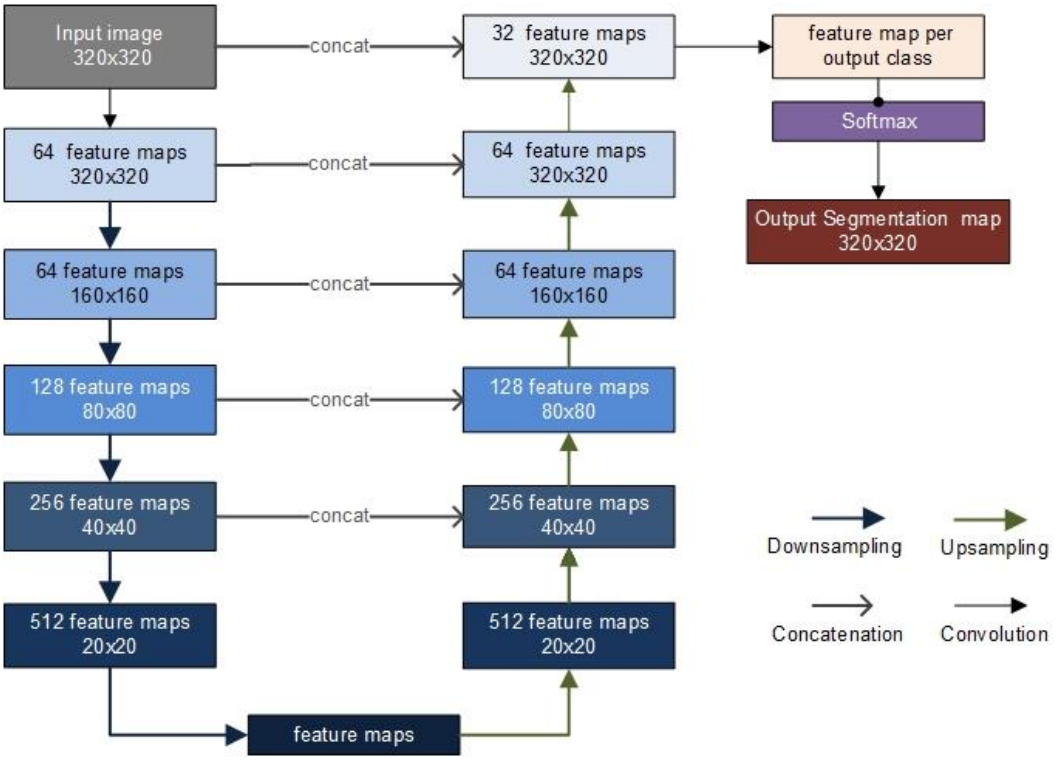

Figure 15 : Architecture of proposed U-Net variant

All convolutional layers followed by rectified linear unit (ReLU) and Batch Normalization (BN). The max pooling layers is placed at the end of each block that reduced the size by factor of two. All decoder block has same design, it is composed of up-sampling or un-pooling, concatenation, convolutions layers, ReLU and batch normalization layer. The order of layers' stack is as:

1. Un-pooling
2. Concatenation layer
3. Convolutional Layer
4. Non-linear activation

5. Batch Normalization

Detailed discussion of each layer is mentioned below. At the end of contracting path, a Softmax layer is used to generate final probability score of each pixel.

3.3.1.1 Convolutional Layer

The proposed network has two or three convolutional layer in each encoder block. All convolutional layer (composing encoder and decoder) in the network have 3×3 receptive field size, consequently the filter size is 3×3×d, where d is depth of input feature map and for first convolutional layer it's the depth of input image (three for RGB and one for gray level). Zero padding is added around the input feature maps to fabricate same size output feature maps. The proposed algorithm is implemented in Keras a neural network API (discuss in detail in section 3.5), in Keras this functionality is achieve by setting the value 'same' for padding parameters.

A single filter when convolution with input feature map produces one output feature map, whereas multiple filters are applied in a convolutional layer, generating multidimensional feature maps. For the first convolutional layer of encoder input is 320×320 patch of an image, zero paddings are added so that resultant feature map will have 320×320 size. The number of kernels per convolutional layer with their respective sizes are displayed in Table 1. For deeper layers of the encoding path of network, a larger number of kernels are applied because deeper layers' extract class specific and complicated features, which demands a larger number of trainable parameters.

3.3.1.2 Non-Linear Activation Layer

A non-linear activation function is preceded by all convolutional layers including encoder and decoder. The non-linear activation function used in this neural network is Rectified Linear Unit (ReLU) [18], given by following equation

$$f(x) = \max(0, x) \qquad (11)$$

This layer doesn't change the input size. It is the most used activation function in deep learning, it allows faster and effective training of neural network as compare to sigmoid and other similar non-linear activation functions.

3.3.1.3 Batch Normalization Layer

Batch Normalization (BN) layer subtracts batch mean and divide by standard deviation to normalized the output of previous layer feature map and enhance the network stability [66]. Convergence of network is fast if the inputs have zero mean and unit covariance and decorrelated properties [67]. Batch normalization is added in both encoder and decoder blocks, in contracting path BN is followed by ReLU layer to normalized the feature maps output and in expending path it is added after convolutional layer. The details of BN are listed in Table 1.

3.3.1.4 Max-Pooling Layer

In proposed network the last layer is pooling layer for encoder block. Max-pooling of size 2×2 is used, as it reduces the spatial size and trainable parameters hence the amount of computations.

Table 1 : Encoder network summary

	Layer Type	Filter size	Output shape	Padding	Stride	No. of params
Input Encoder Block	Input	320×320×3	-	-	-	0
	Convolution	3×3×64	320×320×64	[1 1 1 1]	[1 1]	1792
	ReLU	-	-	-	-	-
	BN	-	320×320×64	-	-	1280
	Convolution	3×3×64	320×320×64	[1 1 1 1]	[1 1]	36928
	ReLU	-	-	-	-	-
	BN	-	320×320×64	-	-	1280
	Max Pooling	2×2	160×160×64	[0 0 0 0]	[2 2]	0

3.3.1.5 Concatenation Layer

The concatenation layer in neural network takes input two or more input feature maps and returns a single feature map, the input feature maps should be of equal sizes. In proposed method the concatenation layer is part of decoder block, it joins the input feature map from previous layer and respective feature map of encoder. This layer is basically performed skip connection, to restore the lost information in down-sampling.

3.3.1.6 Up-sampling Layer

Up-sampling or un-pooling is technique to redeem the spatial resolution which is lost in contracting path of encoder-decoder architecture. It is opposite of max-pooling. It is first layer in decoder block followed by concatenation layer. This layer also halves the number of generated feature maps.

3.3.1.7 Softmax Final Activation Layer

The final activation in the network is differ from the activation used in feature extraction and in up-sampling. For segmentation problem the output classes are mutually exclusive and for the give problem i.e. retinal fundus segmentation there are multiple output classes so the best choice for final layer activation is Softmax. It is generally a logistic regression but used for multiple output classes [20]. Softmax is used at the output layer of network to obtained the probabilities of each class, it generates N number of output segmentation maps where N is number of output classes where each pixel location represents the probability of that pixel to be belongs to the respective class. Furthermore, all outputs add up to one.

Table 2 : Decoder Block Summary

	Layer Type	Filter size	Output shape	Padding	Stride	No. of params
Output decoder Block	Un-pooling	3×3×64	320×320×64	-	-	0
	Concatenation	-	322×322×67	[1 1 1 1]	[1 1]	0
	Convolution	3×3×32	320×320×32	[0 0 0 0]	[1 1]	19328
	ReLU	-	-	-	-	-
	BN	-	320×320×32	-	-	1280

3.3.2 Feature Maps

Number of feature maps and their sizes varies in correspondence to network layer. The Table 1 and Table 2 shows the feature maps sizes and numbers for first block of encoder and decoder respectively. As the network encoder move deeper and deeper the amount of feature maps grows and size reduced, whereas for decoder feature maps count decreases and size increases as the network goes to next layer. Features map at the deepest layer of encoder are the representation of input into extracted features.

3.4 Training Setup

The proposed network is trained on 140 images of retinal images taken from Messidor [14] dataset. Patch based learning is used so that small structure can be segment with high accuracy, size of the patch used for network is 320×320 pixels. For training overlapped patches are utilized whereas testing is carried out on non-overlapped patches. The weights of the network are randomly initialized. Network is trained for 200 epochs each having 290 iterations.

3.4.1 Data Preprocessing

For training of the network and production of segmentation maps, the normalized or pre-homogenized images (methodology discussed in section 3.1) is transform to have zero mean (it's a gray scale image so having a single channel). The mean value is estimated for the entire dataset of normalized images independently of the training and testing sets.

3.4.2 Parameters and Hyper-parameters

Hyper-parameters are neural network configurations, these are the parameters which need to be set up before training the network and their value doesn't come from data, for example stride size and padding to be added etc. [20]. All hyper-parameters are jot down in Table 3 with their respective values which are used throughout all experiments. The parameters are selected by looking at literature regarding semantic segmentation, for example the kernel size 3x3 and stride size [1 1] are used in pixel-wise labeling networks. Furthermore, most of the hyper-parameters are tuned by trial and error.

Table 3 : Hyper-parameters settings

Hyper-parameter	Setting	Hyper-parameter	Setting
Activation Function	ReLU	Optimizer	AdaDelta
Weights initialization	Random	Iterations	290
Kernel Size	3x3	No of epochs	200
Stride	[1 1]	Batch Size	15
Weights Regularizer	L2	Initial learning rate	-

The network has total 35 layers. Total number of parameters are 9,544,094 from which trainable parameters are 9,540,434 whereas 3,660 parameters are non-trainable.

3.4.3 Loss Function and Optimizer

The decisions of loss function and optimizers play important part in neural network performance. The loss function or optimization score function or objective function calculate the difference between input and generated output. The **Cross entropy loss function** is used for training the network. When using Keras categorical_crossentropy loss value is given to compile the model, because the output/targets are in categorical format. For example, total classes are six and output segmentation map should be a six dimensional vector having all zeros except for the pixel belongs to that particular class.

In literature most of the neural network use stochastic gradient decent technique to minimize the loss function for trainable parameters but in this research **AdaDelta** optimizer is used which is per dimension learning rate method for gradient decent with no manual setting for learning rate [27]. The parameters for AdaDelta was set as:

rho = 0.95 and zero decay value and epsilon = 1e-07.

3.4.4 DataSet Partitioning

The proposed methodology is judge on 140 images from Messidor dataset [14]. The details related to dataset will discuss in Chapter 4 . This dataset is divided into three sets, first set contains training images second set contains validation images and test images are in third set. All sets comprise of patches of input images, patch based learning is required to improve the performance of neural network and the full Messidor image are too large for the network as it required better specs hardware (in terms of larger RAM and GPU computational power). Further testing images selection is random. The split for training, validation and testing is 80, 20 and 40 respectively.

3.5 Tools and Technology

This section provide details related to tools and technology to implement the proposed approach in the previous section.

3.5.1 Software Level

Python is used as the basic language for implementation. Neural network implementation is achieved by using Keras and TensorFlow. Keras is application programming interface (API) written in python, designed to make deep learning algorithms, it executes over TensorFlow, Theano or Microsoft Cognitive Toolkit (CNTK). TensorFlow is a software library for machine learning so as Theano and CNTK. In this research Keras is run over the TensorFlow backbone to accomplish neural network.

Besides MALAB is used to perform pre-processing of input images, making them optimized for the neural network to work on. For evaluation process both Keras and MATLAB are used in parallel, per image evaluation is performed with Keras whereas complete test set evaluation measures are processed with MATLAB. Further self-written MATLAB script and 'evaluateSemanticSegmentation' function of Computer Vision System Toolbox of MATLAB is used, so that the results are compared and verified.

3.5.2 System Level

The proposed technique required two GPUs with minimum of 32GB RAM preferably 64GB RAM.

Chapter 4

Experiments and Results

In this chapter the last part of the thesis i.e. experiments and evaluation of proposed methods are discussed in detail. At the start of the chapter the datasets used for training and evaluate network performance, criteria or performance measures are described. At the end of chapter quantitative and qualitative outcome of the suggested approach are presented.

4.1 Materials

One of the major contribution of this research is the development of semantically segmented ground truths for publically available retinal dataset Messidor [14]. To the very best of my knowledge up till now there is no retinal dataset is available for retinal landmark (i.e. optic disc and vasculature) and pathologies (i.e. hemorrhages, cotton wool spots and hard exudates) segmentation. Messidor dataset provides the Diabetic Retinopathy (DR) grades annotation but the gold standard or ground truths doesn't contain semantic segmentation maps. The dataset comprises of pupil dilated and non-dilated retinal images, for generation of ground truth both dilated and non-dilated retinal images are used. The dataset is organized in to 12 bases each containing 100 retinal images, making the total of 1200 images in two resolution 2240x1488 and 1400x960. A total of 140 images are selected which having relatively good amount of pathologies. The number of images taken from each base in listed in Table 4.

Table 4 : Number of images selected from Messidor bases

Base Number	No. of images	Base Number	No. of images
Base 11	24	Base 21	14
Base 12	18	Base 22	12
Base 13	21	Base 24	15
Base 14	16	Base 33	15
Base 31	1	Base 34	4

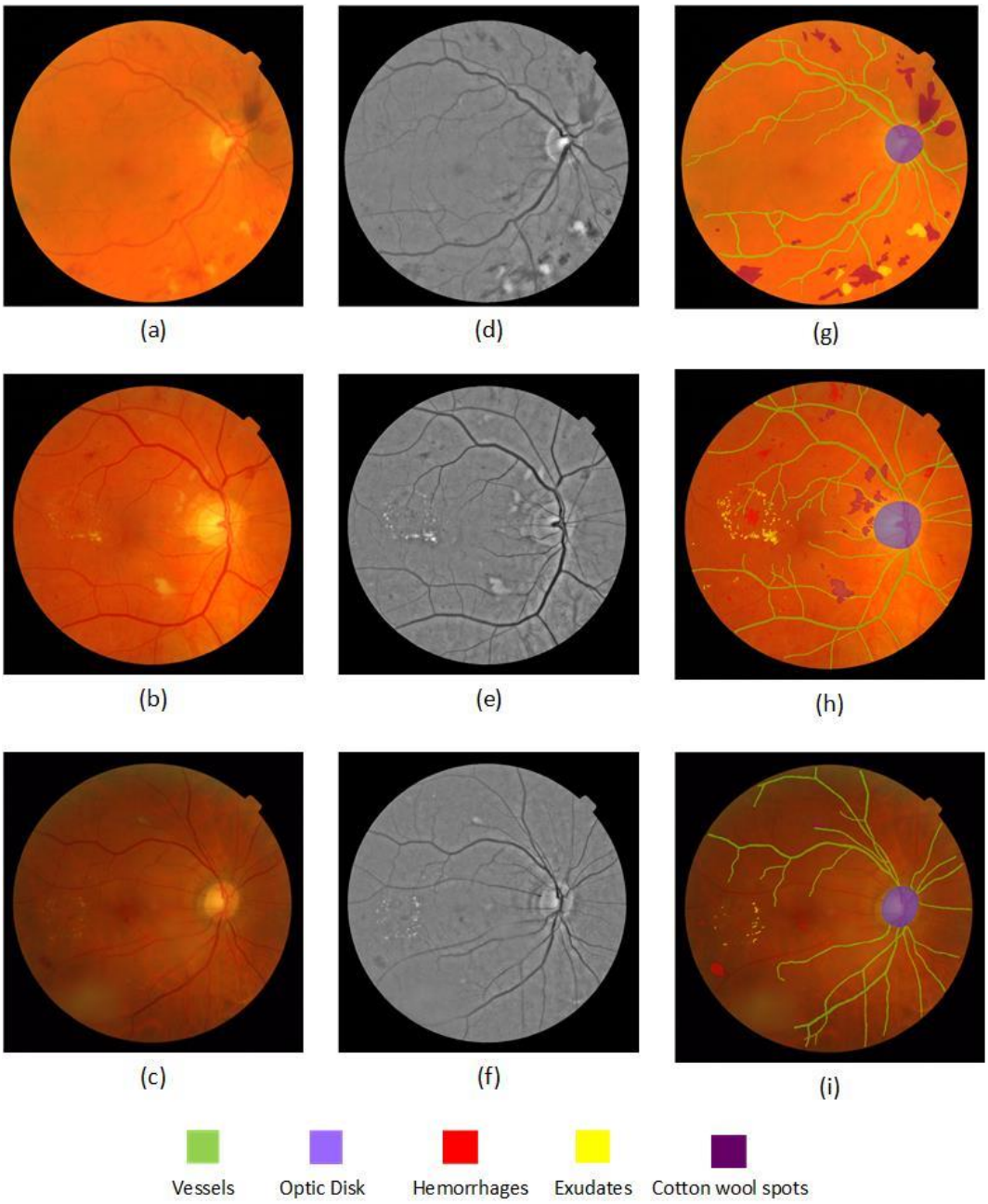

Figure 16: Messidor dataset Ground truth annotation, first column contains input images, second column is respective normalized images and third column represents annotated ground truths

Further to make the generated dataset more diverse, images from ChaseDB1 [68], Drive [11] and STARE [12] dataset are also added so that the network can learn weights according to them too. For Drive dataset 15,5 and 20 images are used for training, validation and testing respectively. Whereas 12, 5 and 5 for ChaseDB1 and 8, 4 and 8 images from Stare dataset are added in training, validation and test splits.

For increasing the visibility of exudates and hemorrhages during creation of annotations, the normalized images are used in parallel to input (RGB) images. Figure 16, (a) is the input image from Messidor dataset, (d) is respective normalized image, as it can be seen that pathologies are not clearly visible in original Messidor image but in normalized image hemorrhages and exudates are clearly visible, another major advantage of using normalization in ground truth creation is to deal with non-uniform illumination and inconsistency in contrast because of acquisition methods involved and anatomy of retinal fundus (detail is discussed in section 3.1). Figure 16 (c) is input Messidor image where (f) its normalized image the non-uniform illumination is handled in normalized image such that it will aid the observer marking annotations of dataset.

Another major fact which have given significant importance is overlapping segmentation classes. Several images have overlapping pathologies with retinal landmarks, i.e. hemorrhages, exudates and cotton wool spots overs vessels and sometime hemorrhages covers some part of optic disc, however in numerous images optic disc overlap with vessels as it is the originating point for vessels. Thus an order of precedence becomes a necessity, which is defined as: pathologies (hemorrhages, cotton wool spot and exudates all having same precedence and they don't overlap with each other), optic disc and vessels.

Retinal pathologies and anatomical structures are annotated in selected retinal fundus by trained people under the supervision of an ophthalmologist. The annotations are verified by proficient ophthalmologist and clinicians from Armed Forces Institute of Ophthalmology (AFIO), Islamabad, Pakistan. This semantically annotated dataset is created using MATLAB image labeler tool. This dataset is utilized to train and evaluate proposed technique. The dataset with image based ground truths is available and can be obtained on request.

4.2 Performance Measures

The evaluation criteria or performance measures to quantitatively evaluate the proposed semantic segmentation approach three measures are utilized. All of them have they own advantages, they are justified to be used for segmentation evaluation. Table 5 lists the evaluation measures used in this research.

4.2.1 Accuracy

Accuracy is ratio between correctly identified elements and all elements, its formula is given by

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

Where TP = true positive (positive elements, output as positive), TN = true negative (negative elements, output as negative), FP = false positive (negative elements, output as positive) and FN = false negative (positive elements, output as negative).

For semantic segmentation problem it is the number of pixels which are correctly assign output class over total number of pixels in input image, but only this measure is not sufficient for evaluation, as the output classes of retinal structures are small as compare to background so most of the segmentation map consists of background pixels, if the network assigned all pixels background class that would be a high accuracy network but not efficient as it failed to segment other retinal structures.

4.2.2 Sensitivity and Specificity

Sensitivity also known as recall and true positive rate is a measure of positives which are correctly identified as positives. Specificity is a measure of negatives which are correctly identified as negative. For example, percentage of diabetic people who are correctly identified as having diabetes is sensitivity whereas percentage of healthy people who are correctly identified as not having diabetes is specificity of the diagnostic system.

Table 5: Evaluation criteria for Retinal Image Analysis

Heading level	Description
Sensitivity (SN)	TP / TP+FN
Specificity (SP)	TN / TN+FP
Accuracy (Acc)	TP+TN / TP+TN+FP+FN

4.3 Evaluation Procedures

There are innumerable decisions which have to take for configuring and designing neural network. In this work mostly configurations are taken from literature, to empirically evaluate the performance of a particular setting and design decision a small experiment is carried out on real data. As we know deep learning algorithms required a lot of data (thousands of instances), hence they take a considerable time in training, in such scenario a robust test is well needed that able to estimate the performance on unseen data. To achieve fast evaluation of a model a 20 and 5 retinal images are for training set and testing set respectively. Moreover, final evaluation and training is performed on complete dataset. The dataset is divided into three sub-groups i.e. training set, validation set and testing set. Total of 80 images are used for training, whereas validation and test sets have 20 and 40 images respectively. This training, validation and test sets are partitioned explicitly outside the Keras environment.

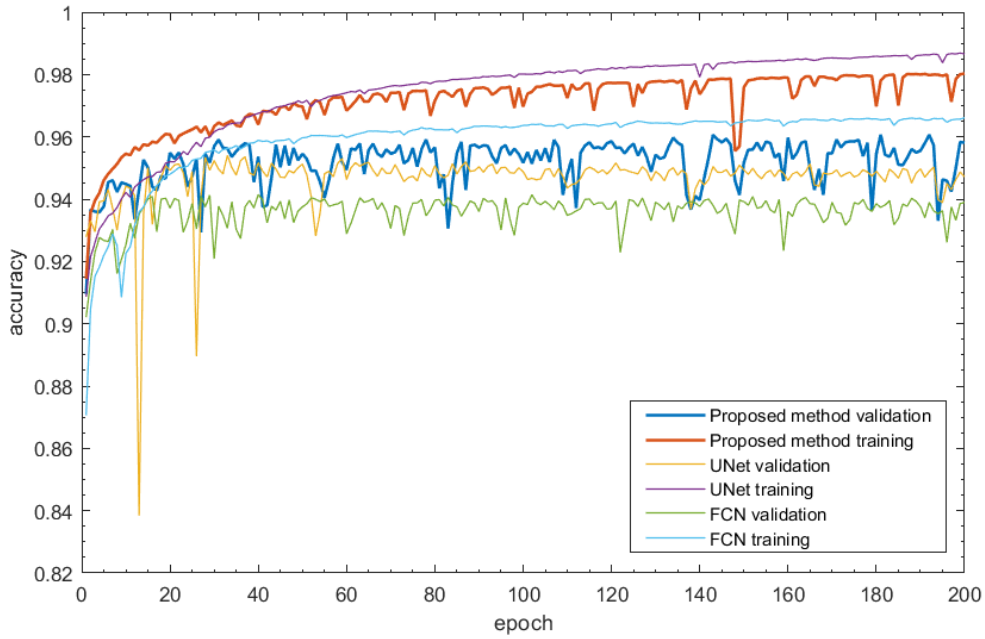

Figure 17 : Plots of accuracy, Proposed method, original U-Net and FCN on Messidor dataset

The model evaluation is performed during training process at the end of each epoch, the results of model accuracy and loss are plotted in Figure 17 and Figure 18 respectively to get the idea as the model learns. As it is depicted from Figure 18 the training accuracy quickly gets the value 95% at epoch 18 and then gradually increase

to 98% at epoch 200. Whereas the validation accuracy quickly takes up the value 95% on epoch 15 and fluctuate between 94%-95% up till epoch 200. The validation accuracy shots for certain epochs but returns to its range in next couple of epochs. The training loss or cost suddenly drops to value 0.2 in first four epochs then gradually decrease to value of 0.12 till 200th epoch. The validation loss shows lots of fluctuations same as validation accuracy, sometimes the loss value shots up then suddenly drops to its normal value, however the validation loss value doesn't show the same behavior as of training loss which keep on decreasing slowly.

To evaluate the performance of proposed technique, Fully Convolutional Network (FCN) and original U-Net neural network [30] is trained on the same Messidor dataset [14], so that a comparison can be made among models. In this section training and validation accuracies and losses plots are discussed, quantitative comparison of mentioned architecture will be discussed in next section.

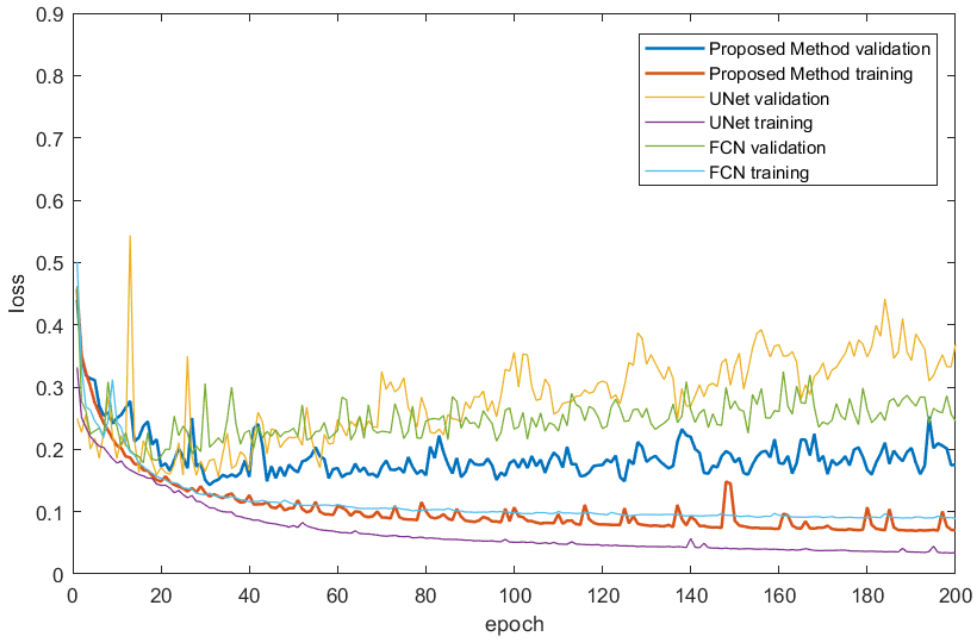

Figure 18: Plots of loss, Proposed method, original U-Net and FCN on Messidor dataset

Original U-Net training and validation accuracy and loss are also shown in Figure 17 and Figure 18 respectively. Training accuracy shows smooth increase up to 98%, this looks encouraging but by inspecting the corresponding validation accuracy, which starts declining after epoch 40, which is a sign of overfitting. Similar overfitting behavior is observed for original U-Net loss plots. Training loss graph depicts very

smooth decline which keep on decreasing on each epoch, however the validation loss shows fluctuations and the general trend is increasing for loss. Both accuracy and loss graphs shows the fact the original U-Net is overfitting for the same Messidor dataset on which the proposed network is trained.

Fully convolutional network FCN is also trained on same dataset. Figure 17 depicts the achieved accuracy for validation and training. The training accuracy shows the network learning is improving while the validation accuracy takes up the 94% value in 20 epochs, after that it shows stochastic fluctuation between the range of 92% to 94%, but the value is slightly decreasing. FCN loss is shown in Figure 18, validation loss is declining as the epochs increases whereas the training loss continues to drop smoothly. FCN is overfitting for 200 epochs, this result is easily deduced from loss graph while it cannot be confirmed only by looking at accuracy graph which don't show sharp decline in validation accuracy.

Both the networks (original U-Net and FCN) have shown overfitting for 200 epochs, it is clear from the increasing gap between training and validation plots for accuracy and loss. Whereas the proposed network doesn't suffer from overfitting problem due to steps which are taken into account in network design and training. Although the proposed network shows stochastic fluctuation in validation accuracy and loss due to weights adjustments in training process. The proposed network is trained by utilizing **L2 Regularization** (also called weight decay) [69] to keep the weights in a range by penalizing the large values of weights. Secondly a technique named **Early Stopping** is implemented to avoid overfitting, which is simple notion to stop network form training if the validation loss stops decreasing for certain number of epochs which is called patience factor.

4.4 Quantitative Results

To evaluate the performance of proposed technique, Fully Convolutional Network (FCN) and original U-Net neural network [30] pre-trained on VGG-16 weights is utilized, so that a comparison can be made among models. Fine tuning of U-Net is performed by training on the same Messidor dataset [14]. The acquired evaluation criteria (sensitivity, specificity and accuracy) are tabulated in Table 6 and Table 7 along with comparison of performance with FCN and originally proposed U-Net neural

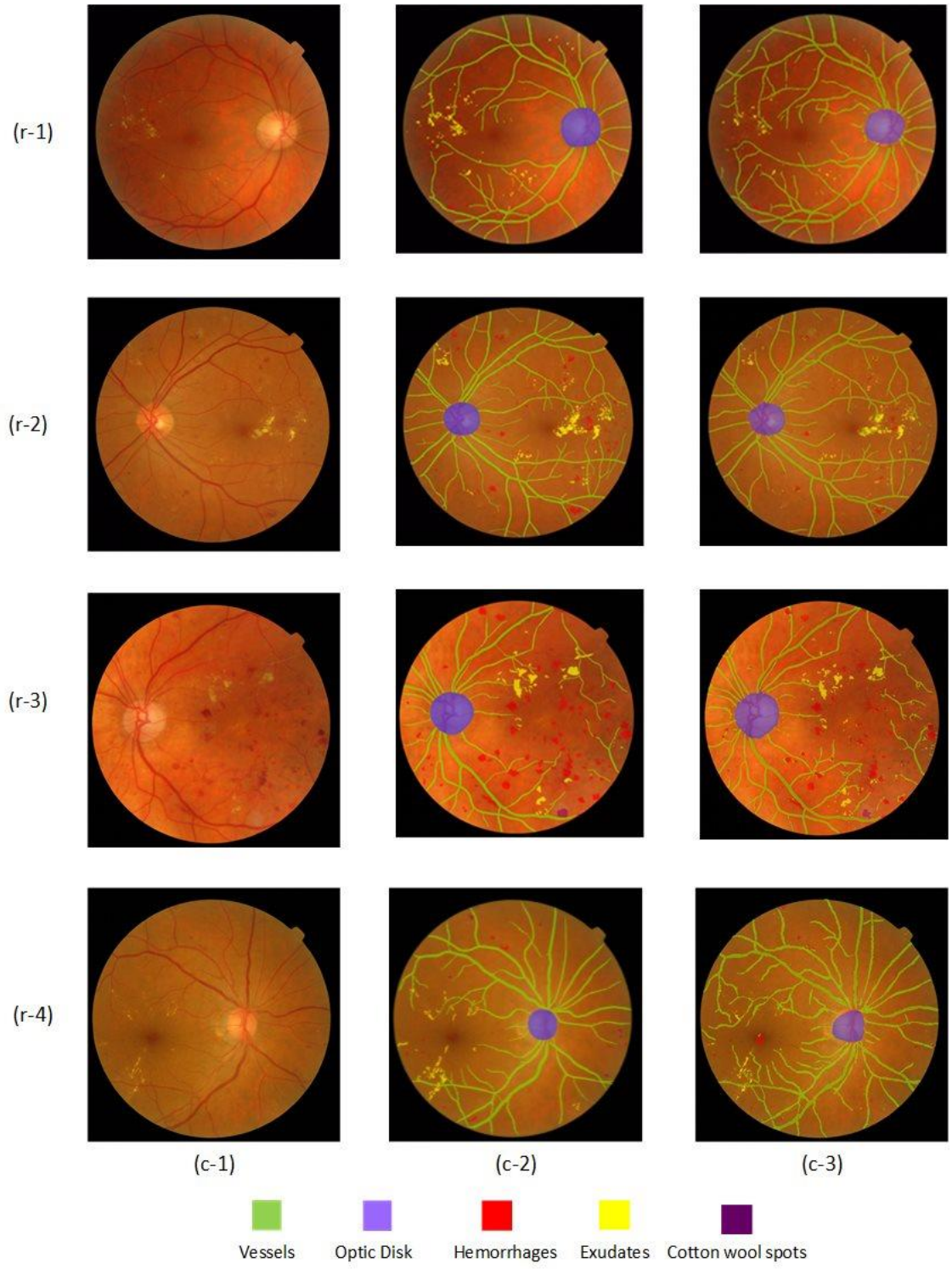

Figure 19 : Generated output segmentation map overlay input images on selected Messidor dataset, first column (c-1) is input images, second column (c-2) is ground truth and third column (c-3) is segmentation map generated by implemented methodology

network [30]. As it can be seen that proposed method has achieved better performance in segmentation of hemorrhages, exudates, cotton wool spot and vessels in

comparison to both other networks. The segmentation outcome and the ground truth overlaid on retinal fundus are depicted in Figure 19.

Table 6 : Optic disc and Vessels segmentation results on Messidor Dataset

Algorithm	**Optic Disc**			**Vessels**		
	Acc (%)	SP (%)	SN (%)	Acc (%)	SP (%)	SN (%)
FCN	98.26	99.97	58.34	96.28	98.21	55.62
U-Net (pre-trained on vgg-16)	97.82	99.75	55.96	96.75	99.83	59.80
Proposed Methodology	**98.69**	**99.43**	**90.06**	**95.38**	**95.80**	**88.06**

Table 7 : Retinal pathologies results on Messidor Dataset

Algorithm	**Exudates**			**Hemorrhages**			**Cotton wool spots**		
	Acc (%)	SP (%)	SN (%)	Acc (%)	SP (%)	SN (%)	Acc (%)	SP (%)	SN (%)
FCN	99.90	99.50	20.69	98.85	99.89	23.75	99.81	98.91	37.88
U-Net (pre-trained on vgg-16)	99.79	99.99	55.86	99.79	99.97	61.79	99.81	99.98	71.62
Proposed Methodology	**99.77**	**98.58**	**83.43**	**99.70**	**99.83**	**76.59**	**94.82**	**95.99**	**69.21**

As it is worth noting that sensitivity of proposed model is better for all output classes i.e. vessels, optic disc, hemorrhages, exudates and cotton wool spots. Sensitivity is eminence in evaluating the performance, as it gives information about the true positive rate. High sensitivity depicts the proposed architectural design is better in classification of pixels in comparison to original U-Net. The accuracy for vessels and all types of lesions are higher for original U-Net with low sensitivity values, this is due to the fact the original U-Net algorithm doesn't able to segment all pixels in to their respective

classes. It marks majority of pixels as background, which in return enhance the true negative rate hence increase accuracy. Similarly, the specificity of predicted output classes is escalated due to predicted high rate of true negatives.

4.5 Performance on Other Datasets

In this section, the evaluation measures and qualitative results of proposed approach on different retinal datasets are presented. Although no retinal dataset is known which provide segmentation of pathologies and retinal anatomical structures, thus only particular output label results are mentioned for which ground truth is available.

Table 8 : Performance summary of Proposed technique on other datasets

Dataset	**Object Class**	**Acc (%)**	**SP (%)**	**SN (%)**
DRIVE	Vessels	95.86	98.54	73.26
HRF	Vessels	96.19	97.07	78.79
eOphtha	Exudates	99.79	98.94	63.76

4.5.1 Drive

The Drive dataset [11] is established to study the retinal blood vessels segmentation. It consists of 40 images, collected from a diabetic retinopathy screening program. The Drive dataset only provides annotation for retinal blood vessels, marked by two domain experts, so proposed technique quantitative results are only calculated for vessels. Obtained performance measure 95.86, 98.54 and 73.26 for accuracy, specificity and sensitivity respectively. The achieved specificity value is less as compare to state-of-the-art because in drive ground truth very slender vessels are marked, this is not provided in annotated dataset used to trained the model. The segmentation results can be seen in Figure 20.

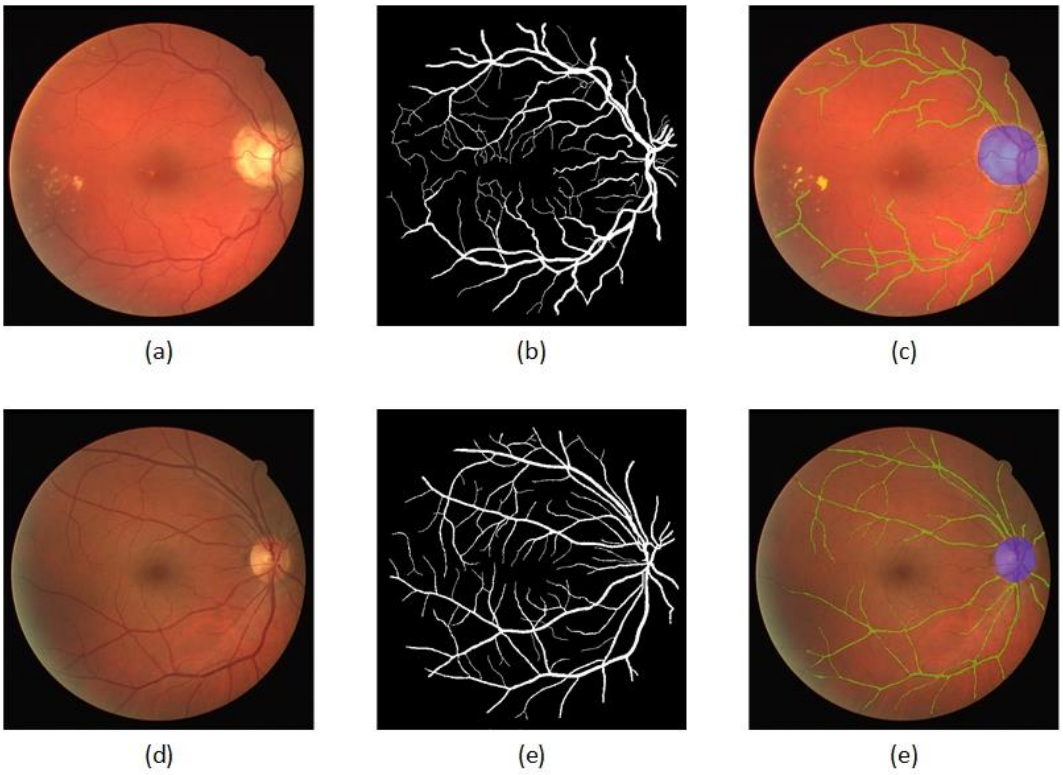

Figure 20: Proposed Method results on Drive results, first column is original images, second column contains vessels ground truth and third column represents overlaid segmentation map generated by proposed method

4.5.2 High-Resolution Fundus Image Database (HRF)

High-Resolution Fundus Image Database [70] is publically available retinal fundus dataset, contains total forty-five high resolution retinal images, from which 15 images are of healthy persons, 15 images of diabetic retinopathy and 15 images of glaucoma. Images resolution is 3504×2336 which is immense in comparison to Drive [11] and Stare [12] datasets, which is 565×584 and 700×605 respectively. Number of patches created per image is monumental due to high resolution, which results in better segmentation of small lesions and vessels, as this is depicted in Figure 21. The HRF provides gold standard only for vessels, so only vessels results are evaluated on this dataset. the proposed U-Net variant obtained 78.79, 97.07, 96.19 for sensitivity, specificity and accuracy respectively.

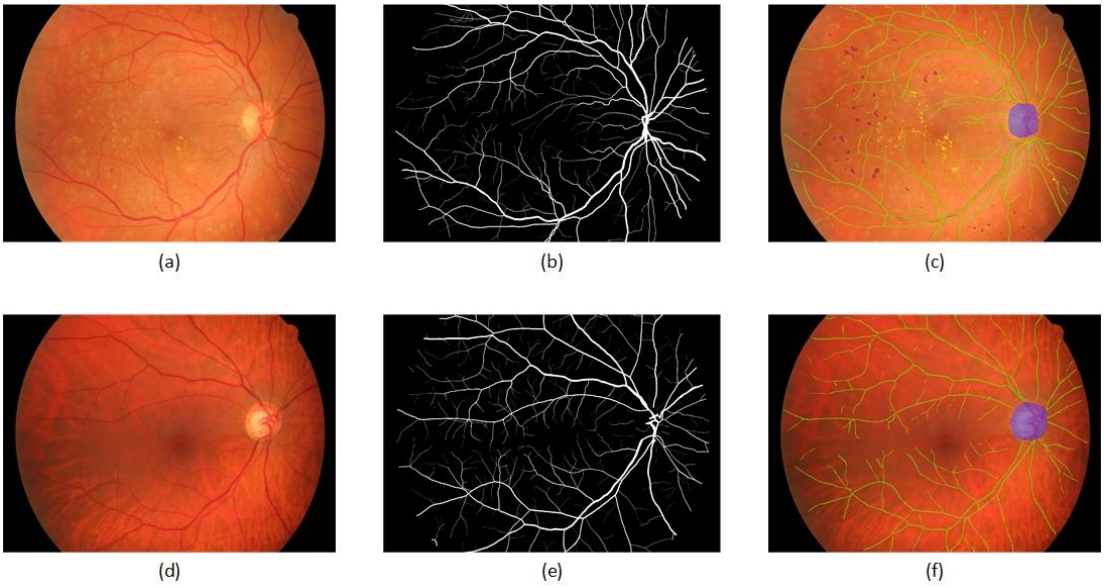

Figure 21: Proposed Method segmentation results on HRF datasets. First column is input image, second column is dataset ground truth and third column segmentation overlaid on input image

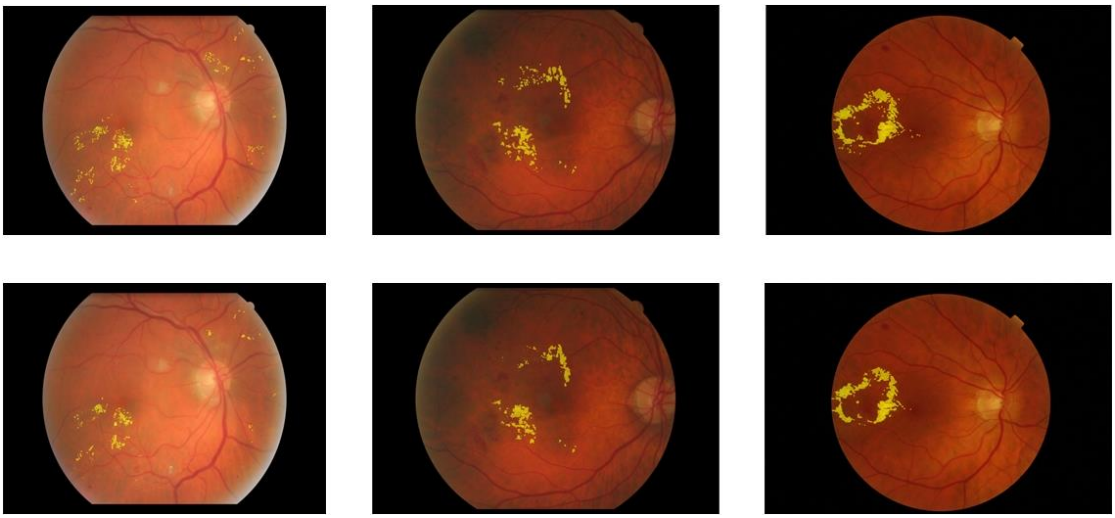

Figure 22: Proposed methodology results on e-Ophtha dataset, first row input image overlaid with ground truth, second row corresponding generated segmentation results

4.5.3 e-Ophtha

The e-Ophtha dataset is collection of color retinal images designed especially for DR detection in screening programs. It is divided into two sub-datasets, one for exudates named e-Ophtha-EX and second one is e-Ophtha-MA for microaneurysms detection. There are 47 images containing exudates in e-Ophtha-EX. The proposed technique achieved sensitivity of 63.76, whereas specificity and accuracy are 98.94, 99.79

respectively on this dataset for hard exudates. Segmentation results are depicted in Figure 22, for visual clarity only exudates are shown for proposed methodology.

4.6 Discussion

In this thesis the development of a deep learning based technique is presented, designed for automated detection of many systemic diseases e.g. diabetic retinopathy, hypertension, stroke and cardiovascular conditions. The proposed methodology attains excellent results for segmentation of retinal landmarks and lesions. For comparison study FCN and original U-Net pre-trained on VGG-16 weights is trained on same dataset, to evaluate performance of proposed U-Net variant. The proposed technique generates better results in comparison to FCN as well as original U-Net. However, the proposed U-Net variant fails to produce the expected results for particular inputs, all those scenarios are discussed in detail in this section.

The proposed methodology out performed original U-Net due to changes which are employed after detailed literature study and experimentations. Major change is the introduction of batch normalization layers after non-linearity (ReLU) to normalized its output thus helps the network to converge faster. Second alteration in model convolutional layers' design, high dimensional or deeper layer is removed and an initial layer is added in network. The intuition behind this change is retinal pathologies are small structures they are detected on second or third layer of network, as the network grow deeper the information about small structure starting to fade out. The proposed technique produced segmentation maps of same dimension as of input image in contrary to original U-Net.

One of the major requirements of automated diagnostic system is to detect pathologies precisely, retinal lesions provides more information about patient health. This fact is kept in view while designing the deep learning model and annotating dataset, retinal lesions have given preference over retinal landmark. The proposed network will detect pathologies in overlapping situation, the results are shown in first row of Figure 23, first row (a) exhibit cotton wool spot (CWS) is overlapping with optic disc, where the overlapped region is marked as CWS, likewise (c) column of first row shows cotton wool spot overlapped with vessels, whereas segmentation map produced correspondingly shown in (b) and (d). A similar preference is decided between optic

disc and vessels, where optic disc have given priority over vessels, such that the trained model can be utilized for glaucoma and other diseases that effect OD of eye.

For deep learning model to perform well on unseen data certain hyper-parameters are needed to be set carefully, number of iterations and epochs are among them. The network presented in this work is trained for 200 epochs, this number is selected after performing experiments. As the number of epochs increases the model's output dominated by plentiful classes i.e. optic disc and vessels, all weights it learnt is about these two classes. Whereas on small epochs the network trend is towards less complex objects i.e. retinal pathologies, weights are insufficient to represent complex objects like vessels. To avoid overfitting and underfitting the number of epochs just need to be right.

Another big challenge in design is imbalanced distribution of the output in dataset, cotton wool spots have least representation whereas hemorrhages and exudates are also underrepresentation. To cope with this issue many techniques are available in literature, in this work sampling technique is used to create balance distribution of training data. Oversampling is a sampling technique which is to duplicate the under representational samples to reach a balanced distribution, for this purpose rotational transformation is applied on patches containing cotton wool spots and hemorrhages. It results in reduction in overall error rate and improve performance for minority classes in training data.

Division of input retinal image into patches increase the performance for pathologies detection but reduces the accuracy for retinal landmark which required contextual information like optic disc, however vessels doesn't affect much from this division as they are consistent in terms of appearance and occurrence in whole retinal fundus. Although, this issue is overcome by overlapping patches but the overall OD accuracy reduced in comparison to whole input retinal fundus. Therefore, in some images the optic disc segmentation become deformed i.e. don't having oval shape and having a hole in oval boundary, or over-segmented i.e. other pixels are classified as optic disc. This issue is visible in row 4, column 3 of Figure 19. Similar is the case with macula which is taken as hemorrhages in some of the images, due to insufficient contextual information. In such cases the network classifies the pixels using only the relative pixel intensities. Note that macular region is marked as hemorrhage in image shown in row 4, column 3 of Figure 19.

Hemorrhages can be of any shape, they are blob like structure with no defined morphology. In some cases, the small hemorrhages are taken as vessels if it has a long and slender shape i.e. similar to vessels, increasing false positives. Besides, small part of hemorrhages is misclassified as vessels sometimes, as shown in Figure 23 (b) third row. Vessels and hemorrhages are similar in pixel intensity; the network mislead hemorrhages for vessels if they have vessel like appearance. In contrast vessels are not detected as hemorrhages due to training data composed of well representation of vessels.

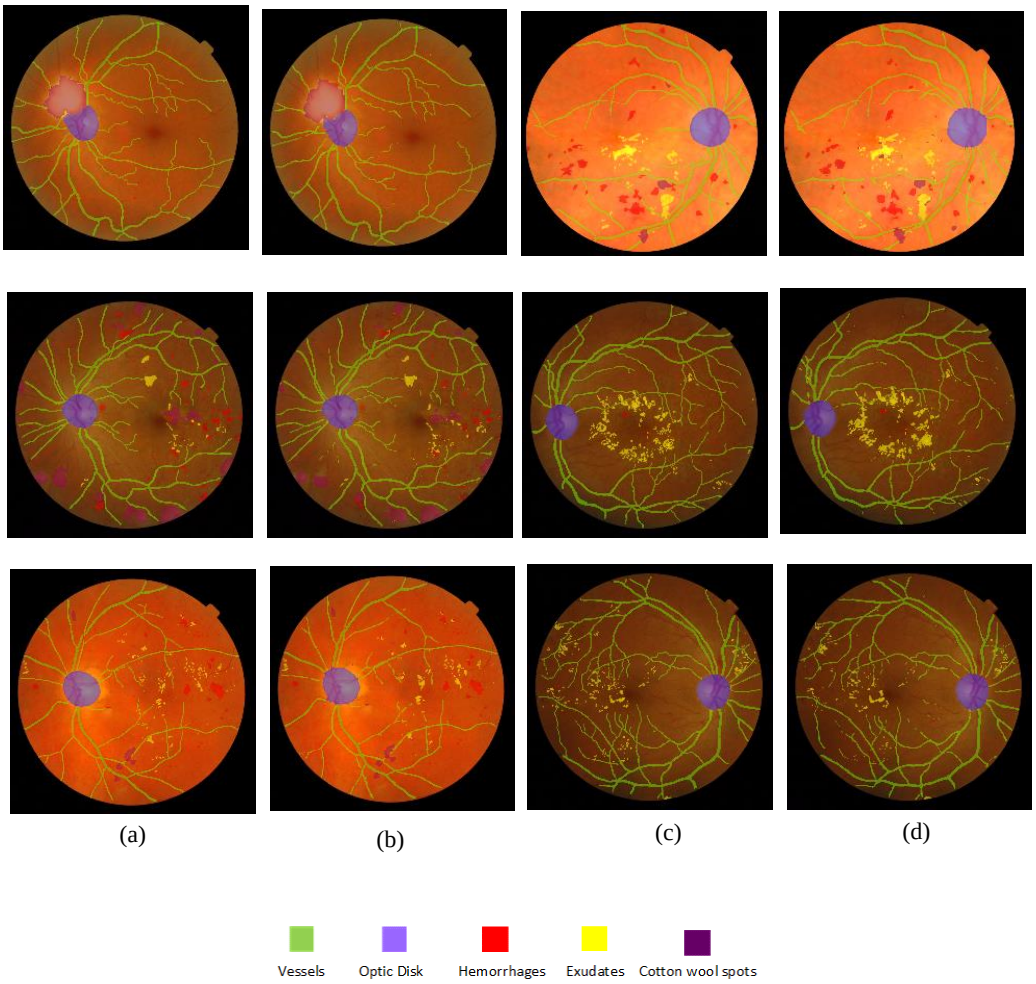

Figure 23 : Segmentation results on Messidor dataset, (a) and (c) are annotated ground truths, (b) and (d) are corresponding segmentation maps overlaid on input image

Chapter 5

Future Work and Conclusion

This work provides general introduction of machine learning and deep learning besides the anatomy of eye and its importance in early detection of systemic diseases. The screening process for systemic diseases can prevent vision loss and blindness. The automated detection system significantly reduce the manual grading time and workload thus helps in ensuring timely referrals. In this chapter major findings are summarized as well as future directions are indicated for further investigations.

5.1 Conclusion

This work proposed a novel deep learning based technique for simultaneous segmentation of retinal pathologies and significant retinal landmark. The proposed algorithm is enhanced version of famous deep learning architecture U-Net, which was originally presented to segments vessels in retinal images. This work extends the notion of U-Net for automated detection of vasculature, optic disc and pathologies. Deep learning algorithms have the potential to improve the efficiency of diagnosis of diabetic retinopathy, glaucoma and other cardiovascular diseases in screening programs through retinal images, thereby prevent visual loss and blindness.

5.2 Future Work

The results of proposed method on Messidor dataset is promising, as it is the same and only dataset used to train the network. However, the performance on other retinal fundus datasets are not up to state-of-the-art. To copped with this problem more diverse images should be added in dataset so that the network can learn the trainable weights according to them too. For this purpose, images should be diverse in terms of size, dpi, resolution, quality and up to some extend field of view (FOV) angle. Having a collection of diverse images in training dataset makes the model to be perform well on any unseen images, whether they vary in size or quality. Thus helps to use the very

same model with different types of fundus camera images. Currently the technique is evaluated on 140 images, which is miniature for deep learning algorithm, adding more images even the Messidor dataset in training dataset will improve the performance not only on annotated dataset but on all other datasets as well.

Other than adding more images another aspect that can be improved is the preprocessing. Retinal images are normalized before passing to the neural network, however it is observed the optic disc and cotton wool spots and even hard exudates have same appearance in normalized image for particular cases, which produce fallacious segmentation. Therefore, some more effective normalization that can distinguished between optic disc and large patches of cotton wool spot and exudates in some more effective manner.

Although this work study an extensive evaluation of hyper-parameters, further examination of hyper-parameters ranges and their influence on performance should be evaluated, like patch size, number of epochs and optimizer used. This algorithm has the potential to be used to annotate dataset that can be reviewed by ophthalmologists for further ground truth generation.

Figure 24 : Proposed Network Visualization

Bibliography

1. Thoma, M., *A survey of semantic segmentation.* arXiv preprint arXiv:1602.06541, 2016.
2. Garcia-Garcia, A., et al., *A review on deep learning techniques applied to semantic segmentation.* arXiv preprint arXiv:1704.06857, 2017.
3. Abràmoff, M.D., M.K. Garvin, and M. Sonka, *Retinal imaging and image analysis.* IEEE reviews in biomedical engineering, 2010. **3**: p. 169-208.
4. Dahl, A.A., *Retina Anatomy.*
5. Kanski, J.J. and B. Bowling, *Clinical ophthalmology: a systematic approach*. 2011: Elsevier Health Sciences.
6. Organization, W.H., *Prevention of blindness from diabetes mellitus: report of a WHO consultation in Geneva, Switzerland, 9-11 November 2005*. 2006: World Health Organization.
7. Klonoff, D.C. and D.M. Schwartz, *An economic analysis of interventions for diabetes.* Diabetes care, 2000. **23**(3): p. 390-404.
8. Quigley, H.A. and A.T. Broman, *The number of people with glaucoma worldwide in 2010 and 2020*. British journal of ophthalmology, 2006. **90**(3): p. 262-267.
9. Jager, R.D., W.F. Mieler, and J.W. Miller, *Age-related macular degeneration.* New England Journal of Medicine, 2008. **358**(24): p. 2606-2617.
10. Hamid, S., S.A. Mirza, and I. Shokh, *Branch retinal vein occlusion.* J Ayub Med Coll Abbottabad [Internet], 2008.
11. Staal, J., et al., *Ridge-based vessel segmentation in color images of the retina.* IEEE transactions on medical imaging, 2004. **23**(4): p. 501-509.
12. Hoover, A. and M. Goldbaum, *STARE public online database.*
13. Feijoo, J., et al., *DRIONS-DB: digital retinal images for optic nerve segmentation database*. 2014.
14. Decencière, E., et al., *Feedback on a publicly distributed image database: the Messidor database.* Image Analysis & Stereology, 2014. **33**(3): p. 231-234.
15. Goodfellow, I., et al., *Deep learning.* Vol. 1. 2016: MIT press Cambridge.
16. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436.
17. Krenker, A., J. Bester, and A. Kos, *Introduction to the artificial neural networks*, in *Artificial neural networks-methodological advances and biomedical applications*. 2011, InTech.
18. Hahnloser, R.H., et al., *Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit.* Nature, 2000. **405**(6789): p. 947.
19. Nasrabadi, N.M., *Pattern recognition and machine learning.* Journal of electronic imaging, 2007. **16**(4): p. 049901.
20. Karpathy, A., *Convolutional neural networks for visual recognition." Course notes of Standford CS class CS231n.*
21. Bengio, Y., *Practical recommendations for gradient-based training of deep architectures*, in *Neural networks: Tricks of the trade*. 2012, Springer. p. 437-478.
22. Domingos, P., *A few useful things to know about machine learning.* Communications of the ACM, 2012. **55**(10): p. 78-87.

23. Ruder, S., *An overview of gradient descent optimization algorithms*. arXiv preprint arXiv:1609.04747, 2016.
24. Qian, N., *On the momentum term in gradient descent learning algorithms*. Neural networks, 1999. **12**(1): p. 145-151.
25. Nesterov, Y. *A method of solving a convex programming problem with convergence rate O (1/k2)*. in *Soviet Mathematics Doklady*. 1983.
26. Duchi, J., E. Hazan, and Y. Singer, *Adaptive subgradient methods for online learning and stochastic optimization*. Journal of Machine Learning Research, 2011. **12**(Jul): p. 2121-2159.
27. Zeiler, M.D., *ADADELTA: an adaptive learning rate method*. arXiv preprint arXiv:1212.5701, 2012.
28. Kingma, D.P. and J. Ba, *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980, 2014.
29. Long, J., E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
30. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
31. Badrinarayanan, V., A. Kendall, and R. Cipolla, *Segnet: A deep convolutional encoder-decoder architecture for image segmentation*. IEEE transactions on pattern analysis and machine intelligence, 2017. **39**(12): p. 2481-2495.
32. Yu, F. and V. Koltun, *Multi-scale context aggregation by dilated convolutions*. arXiv preprint arXiv:1511.07122, 2015.
33. Huang, G., et al. *Densely connected convolutional networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
34. Jégou, S., et al. *The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation*. in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. 2017. IEEE.
35. He, K., et al. *Mask r-cnn*. in *Computer Vision (ICCV), 2017 IEEE International Conference on*. 2017. IEEE.
36. Fraz, M.M., et al., *Blood vessel segmentation methodologies in retinal images–a survey*. Computer methods and programs in biomedicine, 2012. **108**(1): p. 407-433.
37. Melinščak, M., P. Prentašić, and S. Lončarić. *Retinal vessel segmentation using deep neural networks*. in *VISAPP 2015 (10th International Conference on Computer Vision Theory and Applications)*. 2015.
38. Fu, H., et al. *Deepvessel: Retinal vessel segmentation via deep learning and conditional random field*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
39. Wang, S., et al., *Hierarchical retinal blood vessel segmentation based on feature and ensemble learning*. Neurocomputing, 2015. **149**: p. 708-717.
40. Liskowski, P. and K. Krawiec, *Segmenting retinal blood vessels with deep neural networks*. IEEE transactions on medical imaging, 2016. **35**(11): p. 2369-2380.
41. Fang, T., et al. *Retinal vessel landmark detection using deep learning and hessian matrix*. in *Image and Signal Processing (CISP), 2015 8th International Congress on*. 2015. IEEE.

42. Zhao, Y., et al., *Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images.* IEEE transactions on medical imaging, 2015. **34**(9): p. 1797-1807.
43. Imani, E., M. Javidi, and H.-R. Pourreza, *Improvement of retinal blood vessel detection using morphological component analysis.* Computer methods and programs in biomedicine, 2015. **118**(3): p. 263-279.
44. Annunziata, R., et al., *Leveraging multiscale hessian-based enhancement with a novel exudate inpainting technique for retinal vessel segmentation.* IEEE journal of biomedical and health informatics, 2016. **20**(4): p. 1129-1138.
45. Gongt, H., et al. *A level set method for retina image vessel segmentation based on the local cluster value via bias correction*. in *Image and Signal Processing (CISP), 2015 8th International Congress on*. 2015. IEEE.
46. Christodoulidis, A., et al., *A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images.* Computerized Medical Imaging and Graphics, 2016. **52**: p. 28-43.
47. Srivastava, R., et al. *Using deep learning for robustness to parapapillary atrophy in optic disc segmentation*. in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*. 2015. IEEE.
48. Maninis, K.-K., et al. *Deep retinal image understanding*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
49. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.
50. Sevastopolsky, A., *Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network.* Pattern Recognition and Image Analysis, 2017. **27**(3): p. 618-624.
51. Fu, H., et al., *Joint Optic Disc and Cup Segmentation Based on Multi-label Deep Network and Polar Transformation.* IEEE Transactions on Medical Imaging, 2018.
52. Abdullah, M., M.M. Fraz, and S.A. Barman, *Localization and segmentation of optic disc in retinal images using circular Hough transform and grow-cut algorithm.* PeerJ, 2016. **4**: p. e2003.
53. Hsiao, H.-K., et al., *A novel optic disc detection scheme on retinal images.* Expert Systems with Applications, 2012. **39**(12): p. 10600-10606.
54. Mittapalli, P.S. and G.B. Kande, *Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma.* Biomedical Signal Processing and Control, 2016. **24**: p. 34-46.
55. Gardner, G., et al., *Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool.* British journal of Ophthalmology, 1996. **80**(11): p. 940-944.
56. Larsen, M., et al., *Automated detection of fundus photographic red lesions in diabetic retinopathy.* Investigative ophthalmology & visual science, 2003. **44**(2): p. 761-766.
57. Tan, J.H., et al., *Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network.* Information Sciences, 2017. **420**: p. 66-76.
58. Mumtaz, R., et al., *Automatic detection of retinal hemorrhages by exploiting image processing techniques for screening retinal diseases in diabetic patients.* International Journal of Diabetes in Developing Countries, 2018. **38**(1): p. 80-87.

59. Sinthanayothin, C., et al., *Automated detection of diabetic retinopathy on digital fundus images.* Diabetic medicine, 2002. **19**(2): p. 105-112.
60. Matei, D. and R. Matei, *Detection of diabetic symptoms in retina images using analog algorithms.* World Academy of Science, Engineering and Technology, 2008. **45**: p. 408-411.
61. Prentašić, P. and S. Lončarić, *Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion.* Computer Methods and Programs in Biomedicine, 2016. **137**: p. 281-292.
62. Abbasi-Sureshjani, S., et al., *Boosted Exudate Segmentation in Retinal Images Using Residual Nets*, in *Fetal, Infant and Ophthalmic Medical Image Analysis*. 2017, Springer. p. 210-218.
63. Otálora, S., et al., *Training Deep Convolutional Neural Networks with Active Learning for Exudate Classification in Eye Fundus Images*, in *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. 2017, Springer. p. 146-154.
64. Rasta, S.H., et al., *A comparative study on preprocessing techniques in diabetic retinopathy retinal images: illumination correction and contrast enhancement.* Journal of Medical signals and sensors, 2015. **5**(1): p. 40.
65. Fraz, M.M., et al., *Multiscale segmentation of exudates in retinal images using contextual cues and ensemble classification.* Biomedical Signal Processing and Control, 2017. **35**: p. 50-62.
66. Ioffe, S. and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift.* arXiv preprint arXiv:1502.03167, 2015.
67. Orr, G.B. and K.-R. Müller, *Neural networks: tricks of the trade*. 2003: Springer.
68. research, K.U., *chase db1.* (January 2011).
69. Cortes, C., M. Mohri, and A. Rostamizadeh. *L 2 regularization for learning kernels*. in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. 2009. AUAI Press.
70. Budai, A. and J. Odstrcilık, *High Resolution Fundus Image Database*. 2013, September.