Project Report

Entitled

Diabetic Retinopathy Detection using Deep Learning

Submitted to the Department of Electronics Engineering in Partial Fulfilment for the Requirements for the Degree of

Bachelor of Technology

(Electronics and Communication)

: Presented & Submitted By :

Moksha Sood (U18EC145), Sindhu Guntha (U18EC149),

Dhruvkumar Patel (U18EC057)

B. TECH. IV(EC), 8^{th} Semester

: Guided By :

Dr. Jignesh N. Sarvaiya

Professor, DECE

&

Dr. Kishor P. Upla

Assistant Professor, DECE

(Year: 2021-22)

DEPARTMENT OF ELECTRONICS ENGINEERING

SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY

Surat-395007, Gujarat, INDIA.

Sardar Vallabhbhai National Institute Of Technology

Surat - 395 007, Gujarat, India

DEPARTMENT OF ELECTRONICS ENGINEERING

CERTIFICATE

This is to certify that the Project Report entitled "**Diabetic Retinopathy Detection using Deep Learning**" is presented & submitted by Moksha Sood (U18EC145), Sindhu Guntha (U18EC149), Dhruvkumar Patel (U18EC057) of B.Tech. IV, 8^{th} Semester in the partial fulfillment of the requirement for the award of B.Tech. Degree in Electronics & Communication Engineering for academic year 2021-22.

They have successfully and satisfactorily completed their **Project Exam** in all respects. We, certify that the work is comprehensive, complete and fit for evaluation.

Dr. Jignesh N. Sarvaiya
Professor & Project Guide

Dr. Kishor P. Upla
Assistant Professor &
Project Co-Guide

PROJECT EXAMINERS:

Name of Examiners	Signature with Date
1. Dr. Jignesh N. Sarvaiya	_____
2. Dr. Kishor P. Upla	_____
3. Dr. Suman Deb	_____
4. Dr. Raghavendra Pal	_____

Dr. P. N. Patel
Head, DECE, SVNIT

Seal of The Department
(May 2022)

Acknowledgement

We would like to express our profound gratitude and deep regards to our guide Dr. Jignesh N. Sarvaiya and our co-guide Dr. Kishor P. Upla for their guidance. We are heartily thankful for their invaluable suggestions and the clarity of concepts of the topic that helped us a lot for completing this work. We would also like to thank Dr. P N Patel, Head of the Electronics Engineering Department, SVNIT and all the faculties of ECED for their co-operation and suggestions. We are very much grateful to all our classmates for their support.

Moksha Sood
Sindu Guntha
Dhruvkumar Patel
Sardar Vallabhbhai National Institute of Technology
Surat

May 13, 2022

Abstract

PROJECT TITLE: Diabetic Retinopathy Detection using Deep Learning

Diabetic retinopathy is a disease caused due to Diabetes Mellitus in the diabetic patients. It is the major cause of blindness for several people. The longer the diabetic retinopathy affected the high chance that the person can go blind. 90% of cases reported that the cause of blindness due to the diabetic retinopathy occur due to the lack of proper treatment and monitoring the retinopathy before severe stage. Microaneurysms formation and macular edema in the retina is the initial sign of DR and diagnosis at the right time can reduce the risk of non proliferated diabetic retinopathy. The doctors can't find a cure to the severe stage of diabetic retinopathy but can be detected in early stage and prevent it. Hence automated computer diagnosis will help doctors in finding Diabetic Retinopathy at early stage with less cost and time. Therefore we need a simple, reliable and accurate model to automate this process.

In this project we used a modification of the U-Net architecture by introducing an attention block to achieve accurate blood vessel segmentation on 40 fundus images from DRIVE dataset including 7 abnormal pathology cases. The proper analysis of retinal vessels is required to get the precise result, which can be done by Retinal Segmentation which is the process of automatic detection of boundaries of blood vessels. We have also used a CNN classifier to identify the stage of Diabetic Retinopathy.

Keywords: Diabetic Retinopathy, CNN, Attention-UNET, Vessel Segmentation, DRIVE

Signature of the Students:

Student Name: Moksha Sood, Sindhu Guntha, Dhruvkumar Patel
Roll No. U18EC145, U18EC149, U18EC057
Guide Name: Dr. Jignesh N. Sarvaiya
Co-Guide Name: Dr. Kishor P. Upla
Date of Project Prelims: 13/05/2022
Timeslot: 10AM to 1PM

Table of Contents

List of Figures

List of Tables

List of Abbreviations

DR	Diabetic Retinopathy
PDR	Proliferative Diabetic Retinopathy
NPDR	Non-Proliferative Diabetic Retinopathy
DL	Deep Learning
LSTM	Long Short Term Memory
ML	Machine Learning
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
MS-CNN	Multi-Seiving Convolutional Neural Network
CBCE	Class Balanced Cross Entropy
FOV	Field of View
ICDRS	International Clinical Diabetic Retinopathy Scale
SDNN	the standard deviation of Neural Networks interval
DRIVE	Digital Retinal Images for Vessel Extraction
GAN	Generative Adversarial Network

Chapter 1
Introduction

1.1 Overview

The introduction chapter presents a brief discussion of the anatomy of the eye and Diabetic Retinopathy. It also describes image segmentaion theory and presents an introduction to optimization in deep learning. It covers various medical image segmentation architecture. It further summarizes the work completed till project preliminary and highlights our motivation and research objectives for this study.

1.2 Anatomy of the Eye

Eye has complex structure and functions. Each eye constantly adjusts the amount of light it lets in, focuses on objects near and far, and produces continuous images that are instantly transmitted to the brain . The structure of the eye is shown in Fig. 1.1

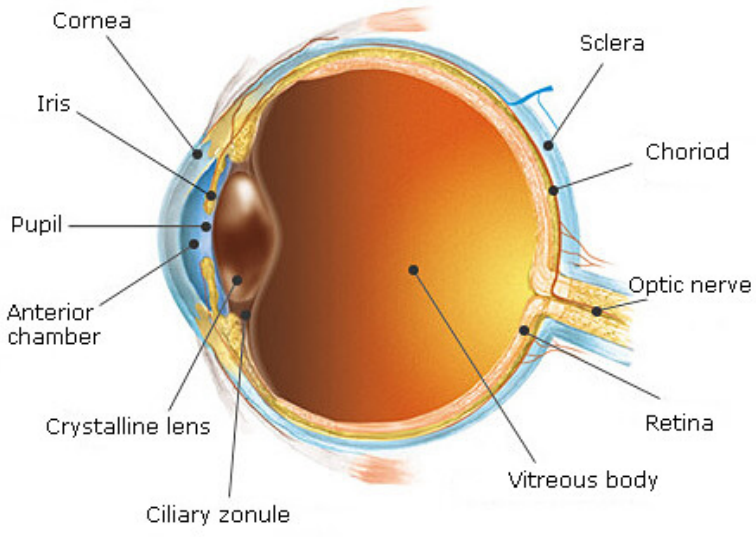

Figure 1.1: Structure of human eye [1]

The orbit is the bony cavity which contains the eyeball, muscles, nerves, and blood vessels, as well as the structures that produce and drain tears. The outer covering of the eyeball consists of a relatively tough, white layer called the sclera. The sclera is covered by a thin, transparent membrane called conjunctiva.

The iris surrounds the pupil and controls the amount of light that enters the eye. The size of the pupil is controlled by the action of the pupillary sphincter muscle and dilator muscle. Eye lens focuses light onto the retina. The retina contains the photoreceptors and the blood vessels that nourish them. The most sensitive part of the retina is macula, which has millions of tightly packed photoreceptors called cones.

Each photoreceptor is linked to a nerve fiber. The nerve fibers from the photoreceptors are bundled together to form the optic nerve. The photoreceptors in the retina convert the image into electrical signals, which are carried to the brain by the optic nerve. There are two main types of photoreceptors: cones and rods.

The anterior segment is filled with a fluid called the aqueous humor, which nourishes the internal structures. The anterior segment is divided into two anterior and posterior chambers. The anterior chamber extends from the cornea to the iris. The posterior chamber extends from the iris to the lens. Generally, the aqueous humor is produced in the posterior chamber, flows slowly through the pupil into the anterior chamber, and then drains out of the eyeball through outflow channels located where the iris meets the cornea.The posterior segment extends from the back surface of the lens to the retina and contains a jellylike fluid called the vitreous humor [14].

1.2.1 Diabetic Retinopathy

Diabetic Retinopathy (DR) is the most debilitating form of diabetes in which serious damage occurs to the retina and causes visual impairments. It harms the veins inside the retinal tissue, making them spill fluid and contort vision. Alongside maladies prompting visual impairment like, waterfalls and glaucoma, DR is one of the most continuous diseases.

There are five stages of DR including:
Stage 0: No apparent retinopathy
Stage 1: Mild Non-Proliferative Diabetic Retinopathy (NPDR)
Stage 2: Moderate NPDR
Stage 3: Severe NPDR
Stage 4: Proliferative diabetic retinopathy

Patients suffering from mild NPDR require regular screening, while appropriate laser treatment is needed for moderate/severe NPDR and PDR stage patients. Retinal blood vessels are damaged in NPDR stage leaking blood and fluid into the retinal

surface causing anomalies like Microaneurysms (MAs), Haemorrhages (HMs), Hard Exudates (EXs) and Cotton Wool Spots (CWs). Fig. 1.2. shows sample fundus images containing various types of lesions.

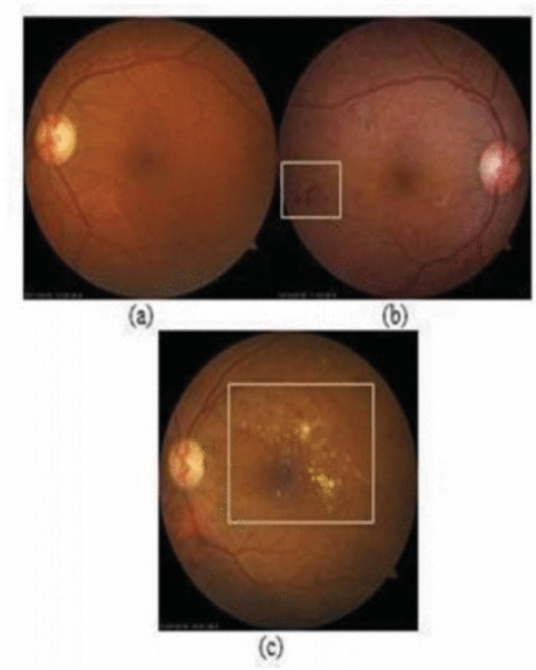

Figure 1.2: (a) A normal fundus image with no sign of DR, (b)Fundus image with red lesions, (c) Fundus image with exudates [2].

1.3 Image Segmentation

Segmentation is to subdivide an image into its constituent regions or objects. Segmentation attempts to partition the pixels of an image into groups that strongly correlate with the objects in an image. It is typically the first step in any automated computer vision application. It involves a simple level task like noise removal to common tasks like identifying objects, person, text etc., to more complicated tasks like image classifications, emotion detection, anomaly detection, segmentation etc. [15]

Basically, segmentation is identifying parts of the image and understanding what object they belong to. Segmentation lays the basis for performing object detection and classification. Image segmentation can be formulated as a classification problem of pixels with semantic labels or partitioning of individual objects. Within the segmentation process there are two levels of granularity and that are, I) Semantic Segmentation II) Instance Segmentation which are shown in Fig. 1.3.

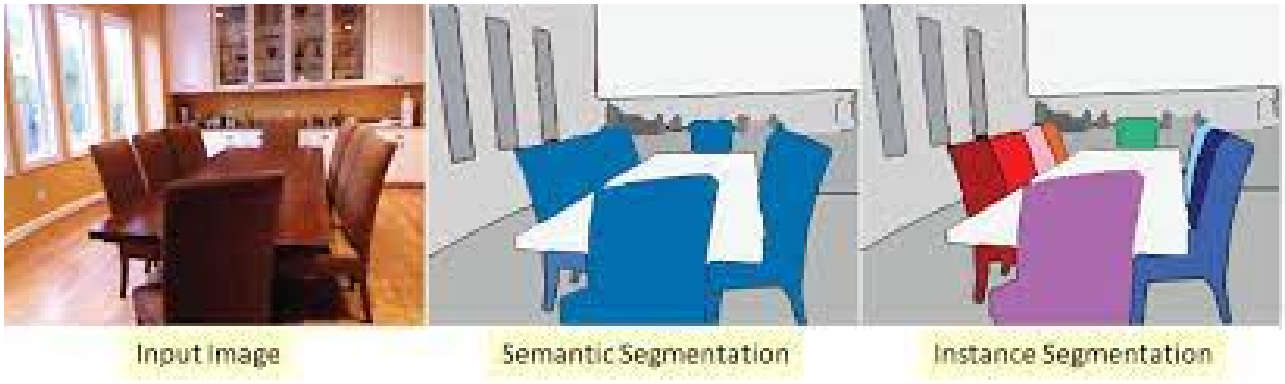

Figure 1.3: Image segmentation output of an RGB image [3]

1.3.1 Threshold Method

This is perhaps the most basic and yet powerful technique to identify the required objects in an image. Based on the intensity, the pixels in an image get divided by comparing the pixel's intensity with a threshold value [16]. The threshold method proves to be advantageous when the objects in the image in question are assumed to be having more intensity than the background (and unwanted components) of the image. At its simpler level, the threshold value T is considered to be a constant. But that approach may be futile considering the amount of noise (unwanted information) that the image contains. So, we can either keep it constant or change it dynamically based on the image properties and thus obtain better results. Based on that, thresholding is of the following types:

- **Simple Thresholding:** This technique replaces the pixels in an image with either black or white. If the intensity of a pixel (Ii,j) at position (i,j) is less than the threshold (T), then it is replaced with black and if it is more, then it is replaced with white. This is a binary approach to thresholding.

- **Otsu's Binarization:** In global thresholding, an arbitrary value for threshold value is used and it remains a constant. The major question here is, how can we define and determine the correctness of the selected threshold? A simpler but rather inept method is to trial and see the error. But, on the contrary, let us take an image whose histogram has two peaks (bimodal image), one for the background and one for the foreground. According to Otsu binarization, for that image, we can approximately take a value in the middle of those peaks as the threshold value. So, in simply put, it automatically calculates a threshold value from image histogram for a bimodal image. The disadvantage here, however, is for images that are not bimodal, the image histogram has multiple peaks, or one of the classes (peaks) present has high variance. However, Otsu's Binarization

is widely used in document scans, removing unwanted colors from a document, pattern recognition etc.

- **Adaptive Thresholding:** A global value as threshold value may not be good in all the conditions where an image has different background and foreground lighting conditions in different actionable areas. Need for an adaptive approach that can change the threshold for various components of the image. In this, the algorithm divides the image into various smaller portions and calculates the threshold for those portions of the image. Hence, we obtain different thresholds for different regions of the same image. This in turn gives us better results for images with varying illumination. The algorithm can automatically calculate the threshold value. The threshold value can be the mean of neighborhood area or it can be the weighted sum of neighborhood values where weights are a Gaussian window (a window function to define regions).

1.3.2 Region Based Segmentation

The region-based segmentation [16] methods involve the algorithm creating segments by dividing the image into various components having similar characteristics. These components, simply put, are nothing but a set of pixels. Region-based image segmentation techniques initially search for some seed points – either smaller parts or considerably bigger chunks in the input image. Next, certain approaches are employed, either to add more pixels to the seed points or further diminish or shrink the seed point to smaller segments and merge with other smaller seed points. Hence, there are two basic techniques based on this method.

- **Region Growing:** It's a bottom to up method where we begin with a smaller set of pixel and start accumulating or iteratively merging it based on certain predetermined similarity constraints. Region growth algorithm starts with choosing an arbitrary seed pixel in the image and compare it with its neighboring pixels. If there is a match or similarity in neighboring pixels, then they are added to the initial seed pixel, thus increasing the size of the region. When we reach the saturation and hereby, the growth of that region cannot proceed further, the algorithm now chooses another seed pixel, which necessarily does not belong to any region(s) that currently exists and start the process again. Region growing methods often achieve effective Segmentation that corresponds well to the observed edges.

- **Region Splitting and Merging:** The splitting and merging based segmentation methods use two basic techniques done together in conjunction – region splitting and region merging – for segmenting an image. Splitting involves iteratively dividing an image into regions having similar characteristics and merging employs

combining the adjacent regions that are somewhat similar to each other. A region split, unlike the region growth, considers the entire input image as the area of business interest. Then, it would try matching a known set of parameters or pre-defined similarity constraints and picks up all the pixel areas matching the criteria.

1.3.3 Watershed Based Methods

Watershed [17] is a ridge approach, also a region-based method, which follows the concept of topological interpretation. The slope and elevation of the said topography are distinctly quantified by the gray values of the respective pixels – called the gradient magnitude. The watershed transform decomposes an image into regions that are called "catchment basins". For each local minimum, a catchment basin comprises all pixels whose path of steepest descent of gray values terminates at this minimum. In a simple way of understanding, the algorithm considers the pixels as a "local topography" (elevation), often initializing itself from user-defined markers. Then, the algorithm defines something called "basins" which are the minima points and hence, basins are flooded from the markers until basins meet on watershed lines. The watersheds that are so formed here, they separate basins from each other. Hence the picture gets decomposed because the pixels are assigned to each such region or watershed.

1.4 Introduction to Optimization

Optimization is a big part of machine learning. Almost every machine learning algorithm has an optimization algorithm at its core. Gradient descent is an iterative machine learning optimization algorithm to reduce the cost function. This directly uses the derivative of the loss function and learning rate to reduce the loss and achieve the minima. The learning rate mentioned above is a flexible parameter which heavily influences the convergence of the algorithm. Larger learning rates make the algorithm take huge steps down the slope and it might jump across the minimum point thereby missing it. So, it is always good to stick to low learning rate such as 0.01. If the new value is not optimum one then this whole process is repeated. Generally, the maximum number of iterations in Gradient Descent is 1000.

This approach is also adopted in backpropagation in neural networks where the updated parameters are shared between different layers depending upon when the minimum loss is achieved. It allows us to obtain the weights in such a way that the error is reached to the global minima as shown in the Fig. 1.4.

It is easy to implement and interpret the results, but it has various issues. The weights

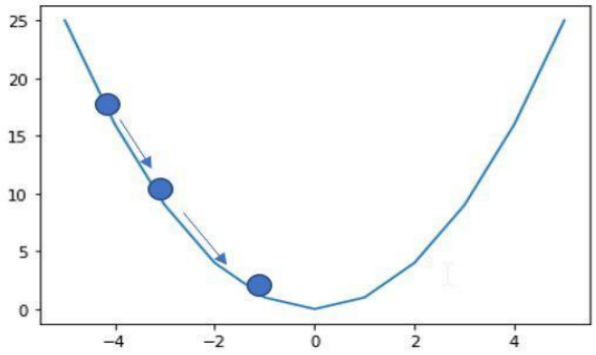

Figure 1.4: Gradient Descent Method [4]

are updated when the whole dataset gradient is calculated, which slows down the process. It also requires a large amount of memory to store this temporary data, making it a resource-hungry process. Though the idea behind this algorithm is well suited, it needs to be tweaked.

1.4.1 Types of Gradient Descent

There are three variants of gradient descent, which differ in how much use of data is done to compute the gradient of the objective function. Different types of Gradient descents are:

- Batch Gradient Descent or Vanilla Gradient Descent
- Stochastic Gradient Descent
- Mini batch Gradient Descent

In batch gradient, usage of the entire dataset to compute the gradient of the cost function for each iteration of the gradient descent is done and then update the weights [18]. Since the usage of the entire dataset to compute the gradient convergence is slow. If the dataset is huge and contains millions or billions of data points then it is memory as well as computationally intensive. Theoretical analysis of weights and convergence rates are easy to understand is the advantage whereas disadvantages are i) perform redundant computation for the same training example for large datasets ii) can be very slow and intractable as large datasets may not fit in the memory and iii) As one takes the entire dataset for computation, they can update the weights of the model for the new data.

In stochastic gradient descent, a single example is used to calculate the gradient and

update the weights with every iteration [19]. The first need is to shuffle the dataset so that one can get a completely randomized dataset. As the dataset is randomized and weights are updated for every single example, an update of the weights and the cost function will be noisy jumping all over the place. A random sample helps to arrive at a global minima and avoids getting stuck at local minima. While batch gradient descent converges to the minimum of the mountain of loss function the parameters are placed in, SGD's fluctuation, on the one hand, enables it to jump to new and potentially better local minima. On the other hand, this ultimately complicates convergence to the exact minimum, as SGD will keep overshooting. However, it has been shown that when we slowly decrease the learning rate, SGD shows the same convergence behaviour as batch gradient descent, almost certainly converging to a local or the global minimum for non-convex and convex optimization respectively.

Mini-batch gradient is a variation of gradient descent where the batch size consists more than one and less than the total dataset. Mini batch gradient descent is widely used and converges faster and is more stable. Batch size can vary depending on the dataset. As the batch with different samples is taken, it reduces the noise which is the variance of the weight updates and this helps to have a more stable and faster convergence.

Just using gradient descent, one cannot fulfil our thirst. Here Optimizer comes in. Optimizer's shape and mold our model into its most accurate possible form by updating the weights. The loss function guides the optimizer by telling it whether it is moving in the right direction to reach the bottom of the valley, the global minimum. Different types of optimizers are Momentum, RMSProp, Adaptive Moment Estimation (Adam),Adaptive Gradient Algorithm (Adagrad), Nesterov accelerated gradient (NAG), Nesterov-accelerated Adaptive Moment Estimation (Nadam).

1.5 Architectures for Medical Image Segmentation

The approach of using a "fully convolutional" network trained end-to-end, pixels-to-pixels for the task of image segmentation was introduced by Long et al. [5] in late 2014. The network architecture uses the VGG 16- layer net. The network is appended with a 1×1 convolution with channel dimension 2 to predict scores for lesion or liver at each of the coarse output locations, followed by a deconvolution layer to upsample the coarse outputs to pixel-dense outputs. The upsampling is performed in-network for end-to-end learning by backpropagation from the pixelwise loss. The FCN-8s DAG net was used as our initial network, which learned to combine coarse, high layer information with fine, low layer information. d the additional value of adding another lower level linking layer creating an FCN-4s DAG net. This was done by linking the Pool2 layer in

a similar way to the linking of the Pool3 and Pool4 layers in Fig. 1.5.

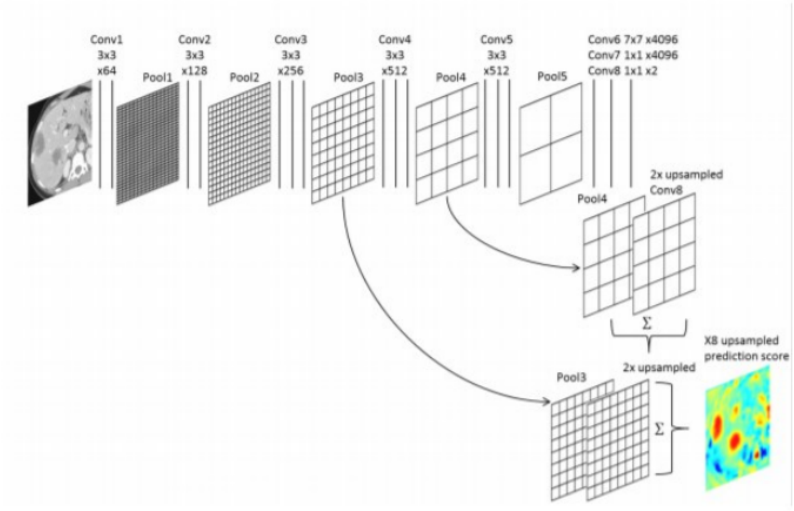

Figure 1.5: Architecture of FCN [5]

1.5.1 U-Net

U-net [6] was originally invented and first used for biomedical image segmentation. Its architecture can be broadly thought of as an encoder network followed by a decoder network. Unlike classification where the end result of the deep network is the only important thing, semantic segmentation not only requires discrimination at pixel level but also a mechanism to project the discriminative features learnt at different stages of the encoder onto the pixel space.

- The encoder is the first half in the architecture diagram (Fig. 1.6). It usually is a pre-trained classification network like VGG/ResNet where you apply convolution.

- blocks followed by a maxpool downsampling to encode the input image into feature representations at multiple different levels.

- The decoder is the second half of the architecture. The goal is to semantically project the discriminative features (lower resolution) learnt by the encoder onto the pixel space (higher resolution) to get a dense classification. The decoder consists of upsampling and concatenation followed by regular convolution operations.

- The main contribution of U-Net in this sense is that while upsampling in the network,the author is also concatenating the higher resolution feature maps from the encoder network with the upsampled features in order to better learn representations with following convolutions. Since upsampling is a sparse operation, we need a good prior from earlier stages to better represent the localization.

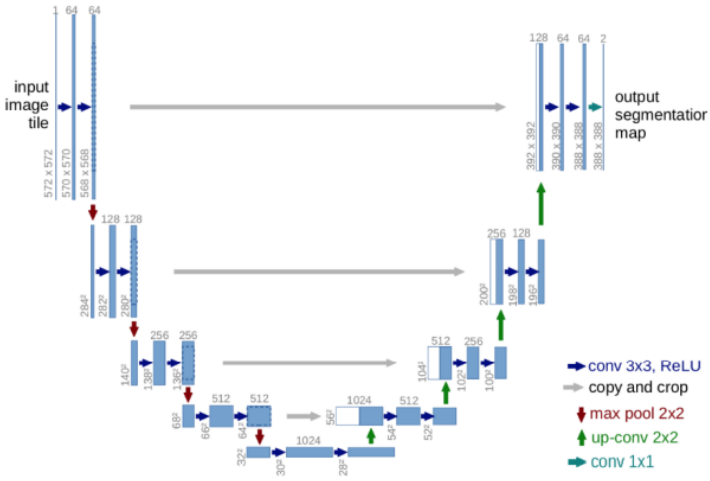

Figure 1.6: U-Net Architecture [6]

1.5.2 U-Net++

UNet++ [7] aims to improve segmentation accuracy by including Dense block and convolution layers between the encoder and decoder. Segmentation accuracy is critical for medical images because marginal segmentation errors would lead to unreliable results; thus, will be rejected for clinical settings. Algorithms designed for medical imaging must achieve high performance and accuracy despite having fewer data samples. Acquiring these sample images to train a model can be a resource-consuming process as requires high quality uncompressed and precisely annotated images vetted by professionals. The architecture of U-Net++ is shown in Fig. 1.7.

U-Net++ have 3 additions to the original U-Net:
I. Redesigned skip pathways (shown in green)
II. Dense skip connections (shown in blue)
III. Deep supervision (shown in red)

I. Redesigned skip pathways
In UNet++, the redesigned skip pathways (shown in green) have been added to bridge

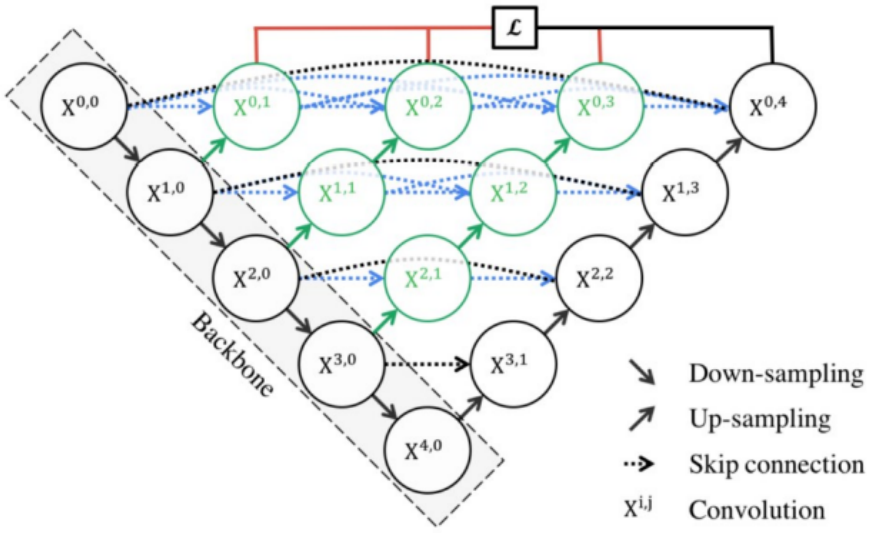

Figure 1.7: U-Net++ Architecture [7]

the semantic gap between the encoder and decoder subpaths. The purpose of these convolutions layers is aimed at reducing the semantic gap between the feature maps of the encoder and decoder subnetworks. As a result, it is possibly a more straightforward optimisation problem for the optimiser to solve. All convolutional layers on the skip pathway use kernels of size 3×3.

II. Dense Skip connections

In UNet++, Dense skip connections (shown in blue) has implemented skip pathways between the encoder and decoder. These Dense blocks are inspired by DenseNet with the purpose to improve segmentation accuracy and improves gradient flow. Dense skip connections ensure that all prior feature maps are accumulated and arrive at the current node because of the dense convolution block along each skip pathway. This generates full resolution feature maps at multiple semantic levels.

III. Deep Supervision

In UNet++, deep supervision (shown in red) is added, so that model can be pruned to adjust the model complexity, to balance between speed (inference time) and performance. For accurate mode, the output from all segmentation branch is averaged. For fast mode, the final segmentation map is selected from one of the segmentation branches. Zhou et al. conducted experiments to determine the best segmentation performance with different levels of pruning. The metrics used are Intersection over Union and inference time.

1.6 Motivation

Usually, patients suffering from diabetic retinopathy are unaware of their disease so its detection before the time is very important Diabetic retinopathy detection is very difficult because it is a time-consuming process that ultimately results in delayed treatment of the disease. With early diagnosis, it is estimated that 90% of patients can be cured of diabetic retinopathy. In manually detecting the DR, ophthalmologists need to be expert and don't need any technical assistance for it. Other limitations of the manual system are that these are more time consuming whereas proven to be inefficient when the large dataset is presented to them. The automated systems, on the other hand, can detect very small indications of DR in patients, as the patient's retina is visible to the doctors. They happen to eliminate the need for manual labor for detection.

1.7 Work Done till Project Prelim

We used DRIVE dataset for the training of the U-Net image segmentation model consisting of 20 images for training and 20 images for testing purposes. Such a small number of images are not sufficient for achieving optimal performance in the deep learning model. So, these images are resized to size 128 x 128 and they are divided into small patches of size 128 × 128 with the overlap of 64 pixels between adjacent patches. In this way, we obtained 4200 images which are sufficient for training of the model. The images are normalized and then, they are fed to the model for training. To test the model performance, the test image is divided into patches for segmenting the vessels and they are then fused together to obtain final output.

Input images, and the ground truth have been used to train the model with the following parameters:

- Number of epochs = 10
- Batch Size = 16
- Optimizer: Adam
- Learning Rate = 0.0001

This approach was implemented using TensorFlow library in Python. The model was trained on NVIDIA's K80 GPU on Google Colab. The results obtained after training are as follows:

- Dice coefficient = 0.9437

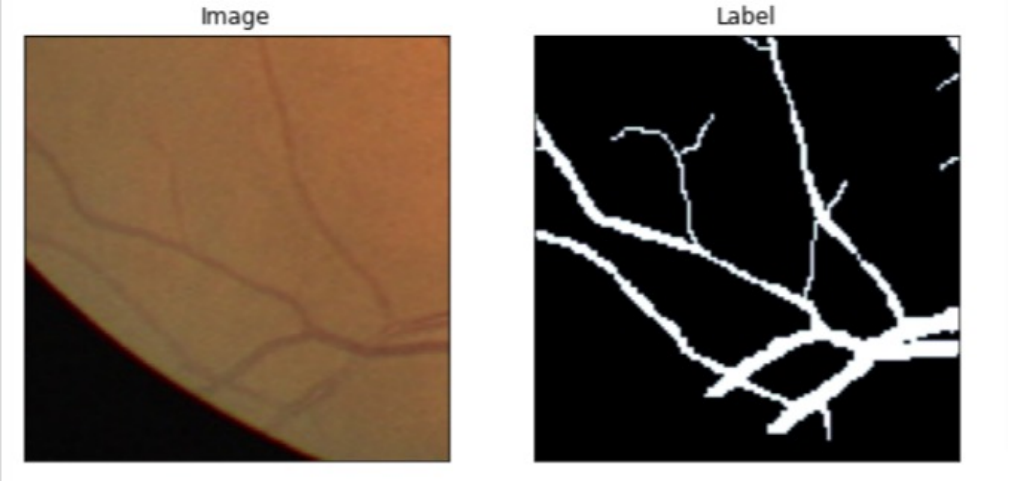

Figure 1.8: Training Result

- Validation Dice Coefficient = 0.9447

The input image and output label obtained from training data are shown in Fig 1.8.

In this report, we shall propose an architecture (Chapter 3) that will further improve the accuracy of retinal vessel segmentation. Further, we will present a CNN classifer to identify the stage of Diabetic Retinopathy.

1.8 Research Objective

The objectives of this report are to:

- To learn about the basic segmentation theory and the deep learning terminologies.
- To perform a literature survey on the research work that is being done in the field of diabetic retinopathy, and analyze the research gap.
- To learn about various deep learning techniques which are popularly used for retinal vessel segmentation.
- To build a CNN model for classification of Diabetic Retinopathy stage.
- To understand and implement codes in the Python programming language, learn about several libraries, and implement the deep learning architecture by utilizing the TensorFlow framework.
- To learn about various statistical parameters used to evaluate the performance of the model and to compare it with several state-of-the-art methods.

- To learn about the applicability of various deep learning algorithms on solving the same problem and comparing their performances.

1.9 Organization of the Report

Chapter 2 covers the literature survey of the state-of-the-art deep learning models for diabetic retinopathy detection published in various top-tier computer vision venues.

Chapter 3 elaborates the details of our proposed method which aims to overcome research gap found after literature survey.

Chapter 4 includes the simulation results of our proposed model and gives a comprehensive evaluation compared to other approaches.

Chapter 5 presents our conclusion of the work done on the project and discusses further work that can be done.

Chapter 2

Literature Survey and Related Works

2.1 Overview

In this chapter, we will discuss the information regarding the research work done using traditional methods to state-of-the-art approaches using deep learning in the field of diabetic retinopathy. Additionally, this chapter includes information regarding the different algorithms that can be used for segmentation and classification of the fundus images for Diabetic Retinopathy. A systematic survey on Google Scholar, PubMed, Scopus was performed by analyzing the research papers and journals from January 1990 to September 2021. An attempt is made to follow to a strategy based on Preferred reporting Items for Systematic Reviews ans Meta-analyses (PRISMA) reporting guidelines and Deep Learning -based study are included.

Diabetic Retinopathy screening program was initially introduced in 1980s and and early 1990s [20]. These programs were successful enough to reduce the incidence of DR related vision loss during that time. In Sweden, early diagnosis of Diabetic Retinopathy was initiated by mobile fundus photography research teams to decrease the incidence of Diabetic Retinopathy related vision loss by an average of 47% per year over 5 years follow up period in 1990s [21]. Recently, the International Council of Ophthalmology (ICO) provided some criteria for Diabetic Retinopathy screening that includes retinal examination with either (a) direct or indirect ophthalmoscopy or (b) mydriatic or non-mydriatic fundus imaging with grater than or equal to 30 degrees mono- or stereo photography.To measure the degree of DR severity, fundus image grading is essential and careful manual grading of these fundus images by an ophthalmologist can be labor-intensive, time consuming and subjective. An automated algorithm may play an important role in analyzing DR fundus images; although previously used algorithms generally relied on traditional approaches that were involved with manually selecting engineering images features for some classifiers methods such as random forest and support vector machine.

2.2 Literature Survey on Diabetic Retinopathy using Deep Learning and Other Techniques

In this section, the brief overview of various techniques used for Diabetic Retinopathy are discussed. From the pioneering steps of using ANNs for DR to current methods used for diagnosis. various methods of Image processing like Image segmentation plays an important role in detecting DR. From just a back propagation model to the use of Transfer Learning for Diabetic Retinopathy almost all the significant works are discussed and analysed.

2.2.1 The Pioneering Steps towards Diabetic Retinopathy using Artificial Neural Networks

In 1996, Gardener and team first introduced n artificial neural network (ANN) approach that was capable of automatically grading Diabetic Retinopathy with 88% sensitivity and 83% specificity relative to an opthalmologist [8].

Gardner and team in their paper mentions that 147 diabetic and 32 normal images were used which were captured from fundus camera, and were analysed using using back propagation model neural network. The network was trained to recognise the features in the retinal image. The network was trained to recognise the features in the retinal image. The effects of digital filtering techniques and different network variables were assessed. 200 diabetic and 101 normal images were then randomised and used to evaluate the network's performance for the detection of diabetic retinopathy against an ophthalmologist.

They employed a back propagation neural network for the analysis of fundi images of 147 patients with Diabetic retinopathy while 32 images were of normal subjects. The images of fundus were captured from the posterior pole. The images were scanned which provided the the image resolution of 700 x 700 pixels as shown in Fig. 2.1.

Followed by that the the input image were divided into the squares of 30 $\times$ 30 and 20 $\times$ 20 pixels, depending upon the detected feature. Eight bit data from the green plane of image were used to simulate red-free output images.

This study showed the world that it is possible to train a neural network to recognize features of Diabetic Retinopathy on the fundus images. The network can be used for identification for presence of vessels, exudates and haemorrhages with high predictive values. During that time haemorrhages were most difficult to recognise because of

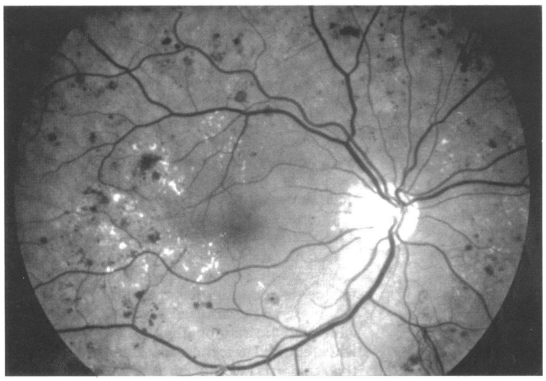

Figure 2.1: Input image example used by Gardner and team in 1996 [8]

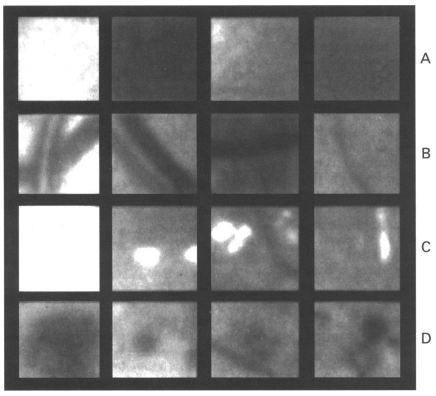

Figure 2.2: Example squares of A-normal background retina, B-normal vessels, C-exudates, and D-haemorrhages. [8]

similar pixel values to the blood vessels. Detection rates for the recognition of vessels, exudates, and haemorrhages were 91.7%, 93.1%, and 73.8% respectively. [8].

2.2.2 Diabetic Retinopathy and Image Processing

In earlier works, the automated systems struggled in classifying between exudates and blood vessels due to similar pixel values. In 2002, Thomas Walter and team proposed methods for detecting exudates present within the macular region with high sensitivity [22].

The contribution of image processing for diagnosing Diabetic retinopathy can be divided into following 3 groups:

1. Image enhancement
2. Mass Screening
3. Monitoring the Disease

Image Enhancement

Images taken at standard screening are noisy and not appropriately contrasted. In addition to that, there is presence of non-uniform illumination. So, the techniques for improving the contrast and sharpness are needed

- as an assistance for human interpretation
- as in initial steps towards automated analysis of the fundus images.

Standard contrast methods allow the enhancement of certain features like microaneurysms in [22]. Image restoration techniques for fundus images of very poor quality have been also applied [23].

Mass Screening

The mechanisms of diabetic retinopathy are not completely understood or clear but still its progress can be inhibited by early diagnosis. However the vision alters also in later stages of the disease, when the treatment might be most useful and efficient.Hence, mass screening of all diabetic patients (even without vision impairment) would help to diagnose this disease early enough for an optimal treatment. A significant contribution is done by Klein and team [24] on segmenting exudates using image processing techniques.

Monitoring

In order to assess the evolution of the disease, doctors have to compare images taken at different medical tests. This allows one to

- evaluate for each patient the efficiency of the opthalmologic and diabetic treatments.
- evaluate the efficiency of new therapeutics in a population of patients
- observe the development of single lesions (for example in order to study the turnover effect of microaneurysms).

However, the comparison of images taken at different moments is a very time-consuming task and open to human error due to the distortions between images that make superposition very difficult, and due to the large number of lesions that have to be compared. A computer assisted approach is needed. In addition to automatic detection of pathologies, such a tool needs a robust feature-based registration algorithm.

2.2.3 Diabetic Retinopathy Segmentation

In this section, a review is provided for how deep learning methods have been applied to DR segmentation. Widely Used datasets for this tasks are IDRiD, Messidor, DRIVE, STARE, EyePACs and so on. Various Features are needed to be segmented for detection of Diabetic Retinopathy. Some of them are discussed here:

Hemorrhages

Hemorrhages (HEs) are one of the visible pathological signs of DR. Accurate detection or segmentation of HEs is vital for DR diagnosis. In the task of detection/segmentation, patch-based methods are popular as a limited number of images in datasets and the need to reduce computational costs. Patch-based methods can generate tens of thousands of patches with only dozens of images, which can help improve performance and alleviate the problem of overfitting. However, HEs are typically relatively small in size, with their pixels only making up a small proportion of the whole image. This leads to an imbalance problem, where only a few patches contain hemorrhages and a large number do not contribute much to the detection/segmentation task. Imbalance is also common in other detection/segmentation tasks in this section, details of which will not be repeated for brevity. There are two main directions in the improvement of hemorrhage detection/segmentation; namely selective sampling and performing segmentation on coarsely-annotated datasets.

Selective Sampling : van Grinsven et al. [25] proposed a method called selective sampling to reduce the use of redundant data and speed up CNN training. They asked three experts to relabel the Messidor dataset and a subset of Kaggle. During the training process, weights of samples were dynamically adjusted according to the current iteration's classification results so that the informative samples were more likely to be included in the next training iteration. Inspired by VGG, they designed a nine-layer CNN as the classifier. On the Kaggle competition and Messidor datasets, experimental results showed that the CNN with selective sampling (SeS) outperformed the CNN without selective sampling (NSeS), and SeS reduced the number of training epochs from 170 to 60.

Segmentation and coarsely-annonated datasets : Huang et al. [26] proposed a bounding box refining network (BBRNet) to generate more accurate bounding box annotations for coarsely-annotated data. Then they utilized a RetinaNet [27] to detect hemorrhage. Rather than using the finely annotated IDRiD dataset, they performed hemorrhage detection on a private dataset with a coarsely annotated bounding box. They first established a dataset containing image pairs. One image was taken from

Madrid for each pair, and the other was obtained by simulating coarsely-annotated bounding boxes. BBR-Net took coarsely annotated patches as input and finely annotated patches as targets. After training, the authors introduced their private data to obtain more accurate bounding box annotations and then sent the results to the RetinaNet for hemorrhage detection.

Microaneurysms

MAs are the earliest clinical sign of DR and have thus captured more research interests. There are several barriers affecting the segmentation of MAs, including the existence of other lesions with similar color, extremely low contrast, and variation in image lighting, clarity and background texture. Two-stage multiscale architectures and guidance from clinical reports are some successful strategies for MAs detection.

Two-stage Multiscale networks : Sarhan et al. [28] proposed a two-stage deep learning approach embedding a triplet loss for microaneurysm segmentation. The first stage is called the hypothesis generation network (HGN), in which multi-scale FCNs are employed to generate a region of interest (ROI). The second stage is known as the patch-wise refinement network (PRN), in which patches extracted from around ROIs are passed to a modified ResNet-50 for classification. The authors introduced the triplet loss into the PRN to extract discriminative features. Further, the previously mentioned selective sampling method [25] is utilized to reduce the computational cost and solve the data imbalance issue.

Clinical report guided CNNs. : Dai et al. [29] proposed a technique called clinical report guided multi-sieving convolutional neural network (MS-CNN) for the detection of MAs. They first trained a weak image-to-text model from clinical reports and fundus images to generate a rough segmentation of microaneurysms. Then the proposed MS-CNN was used to generate final high quality segmentation using the rough segmentation as guidance. In order to tackle the data imbalance problem, MS-CNN adopts a method similar to boosting. Specifically, MS-CNN is composed of multiple CNNs, where the false positives from the previous CNN are fed into the following CNN as negative examples.

Exudates

Soft and hard exudates are usually the basis for the diagnosis of DR. Accurate detection of SEs and EXs are thus crucial for timely treatment. Like other lesion detection/segmentation tasks, there are several challenges. The barriers include low contrast, varied sizes and similarity to other lesions. There are several approaches for exudate detection, most of which can be divided into CNN with circular Hough conversion and modifications to the loss function.

CNN with circular Hough conversion : Adem [30] introduced a three-layer CNN architecture for the binary classification of exudated and exudate-free fundus images. During pre-processing, the OD region was removed by applying several methods, including adaptive histogram equalization, Canny edge detection and circular Hough conversion.

Modification to loss function. : Guo et al. [31] proposed a top-k loss and a bin loss to enhance performance for exudate segmentation. The class balanced cross entropy (CBCE) loss (Xie and Tu, 2015) solved the class imbalance problem to some extent. However, this introduced the new problem of loss imbalance, where background similar to exudate tends to be misclassified. The main reason is that with the different weights for background and foreground pixels in CBCE loss, the loss for misclassifying a background pixel is much smaller than that for misclassifying a hard exudate pixel. To solve this loss imbalance problem, top-k loss is proposed, which considers all hard exudate pixels but only top-k background pixels with the larger loss. They also proposed a fast version of top-k loss named bin loss with consideration of efficiency.

2.2.4 Diabetic Retinopathy Grading using Deep Learning

Diabetic retinopathy (DR) is a vascular disease that affects normal blood vessels in the eye and is the leading cause of preventable blindness worldwide. There is a unified standard for DR classification, namely the International Clinical Diabetic Retinopathy Scale (ICDRS). According to this standard, the severity of DR can be graded into five classes, namely 0 (no apparent DR), 1 (mild DR), 2 (moderate DR), 3 (severe DR), 4 (poliferative DR).

David et al. [32] proposed a system that can automatically detect DR, called IDx-DR X2.1. It applies a set of CNN-based detectors for each image in the detection. Their CNN structure is inspired by AlexNet and VGG and is able to predict four labels, namely negative (no or mild DR), referable DR (rDR), vision-threatening DR (vtDR), and low exam quality (protocol errors or low-quality images). CNN-based anatomy detectors can further detect hemorrhages, exudates, and other lesions.

Gulshan et al. [33] used a CNN for binary classification of with/without DR. They used a dataset of 128,175 images, which were annotated three to seven times by 54 experts. The specific network uses the structure of Inception-v3, and an ensemble of ten networks trained with the same data. The final result is the average of all network outputs.

Krause et al. [34] used a CNN for the five-class classification of DR. Their improvements over Gulshan et al. [33] include: using Inception-v4 instead of Inception-v3, using a larger dataset during training, and using higher-resolution input images. Their network structure is also an ensemble of ten networks.

Zhang et al. [35] established a high-quality labeled dataset, and adopted an ensemble strategy to perform two-class and four-class classifications. Features extracted from different CNN models are passed through the corresponding SDNN modules, which are defined as component classifiers. Then, the features are combined and fed into a FC layer to generate the final results.

Researchers at Budapest University of Technology and Economics, Hungary designed a modified version of U-Net called Spatial Attention U-Net for Retinal Vessel Segmentation. Guo et al. [9] designed model using encoder of of model that includes a structured dropout convolutional block and 2×2 max pooling operation.The convolutional layer of each convolutional block is followed by a DropBlock, a batch normalization layer and a rectified linear unit followed by max pooling operation is utilized fro downsampling with a stride of 2.In each down-sampling step, the number of feature channels are doubled. Each step in the decoder includes a 2×2 transposed convolution operation for upsampling and halves the number of feature channels, a concatenates with the corresponding feature map from the encoder, which then followed by a structured dropout convolutional block. The spatial attention module is added between the encoder and the decoder. At the final layer, a 1×1 convolution and Sigmoid activation function is used to get the output segmentation map.The model has accuracy of 96.98% on DRIVE Dataset. The structure of SA-UNet is shown in Fig. 2.3.

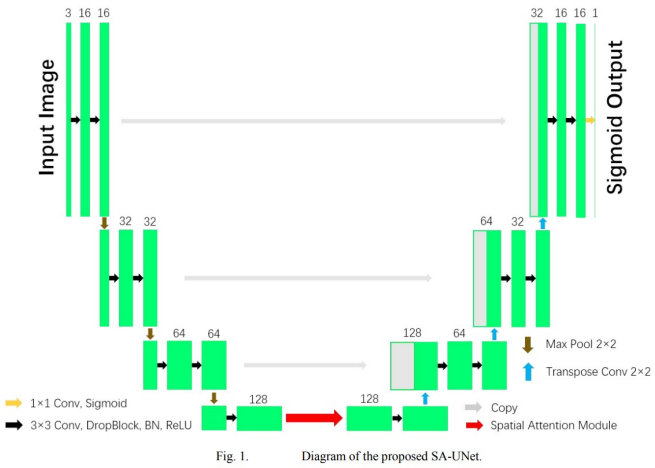

Figure 2.3: Structure of SA-UNet Model [9]

To further improve the performance of vessel segmentation,Researchers at Osaka University proposed IterNet, a new model based on UNet. Li et al. [10] designed a model that has the ability to find obscured details of the vessel from the segmented vessel image itself, rather than the raw input image. IterNet consists of multiple iterations of a mini-UNet, which can be $4\times$ deeper than the common UNet. IterNet also adopts the weight-sharing and skip-connection features to facilitate training; therefore, even with such a large architecture, IterNet can still learn from merely 1020 labeled images, without pre-training or any prior knowledge. IterNet achieves AUC of 0.9816 on DRIVE dataset.The network consists of two slightly different architectures: One is UNet, and the other is a simplified version of UNet, referred to as mini-UNet. UNet is used as base module because of its superior performance in various performance in various segmentation tasks, especially in the medical applications. The output of UNet is the one-channel map of the probabilities of pixels being on a vessels. The refinery modules' architecture is mini-UNet, and they use the output of the second last layer of its precedent module, which is a 32-channel feature map and thus can have more information, compared with the one-channel vessel probability map. The mini-UNet actually is a light-weight version of the UNet architecture with fewer parameters because the input to the refinery modules is a feature map that we consider is simpler than the raw retinal images with all the background and noises. The structure of IterNet is shown in Fig. 2.4

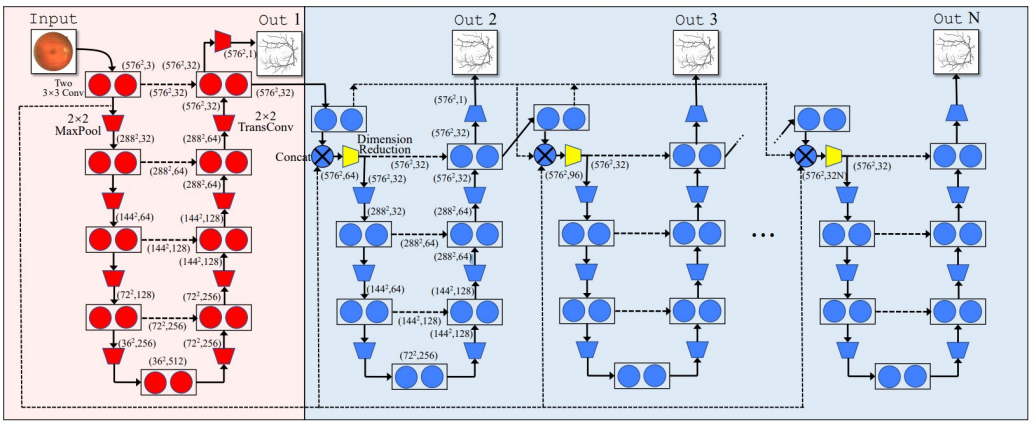

Figure 2.4: Structure of IterNet Model, which consists of one UNet and iteration of N-1 mini-UNets [10]

U-Net and its variants in the literature all have an encoder-decoder structure. However, the number of paths for information flow in U-Net is limited. LadderNet [11], a multi-branch convolutional neural network for semantic segmentation as shown in Fig. 2.5, which has more paths of information flow. Features in different spatial scales are

named with letters A to E, and columns are named with numbers 1 to 4.Column 1 and 3 as encoder branches, and column 2 and 4 as decoder branches.Convolution with a stride of 2 to go from small-receptive-field features to large-receptive field features (e.g., A to B) is used.LadderNet can be viewed as a chain of U-Nets. Columns 1 and 2 can be viewed as a U-Net, and Columns 3 and 4 can be viewed as another U-Net. Between two U-Nets, there are skip connections at levels A-D. Different from U-Net, where features from encoder branches are concatenated with features from decoder branches, we sum features from two branches. The LadderNet provides AUC of 0.9793.

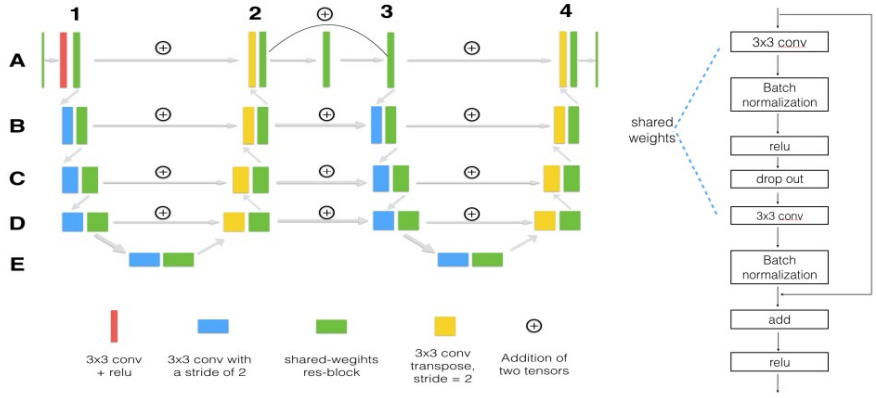

Figure 2.5: From left to right: Structure of LadderNet and shared-weights residual block. [11]

Edge-aTtention guidance Network (ET-Net) [12]embeds edge-attention representations to guide the segmentation network. Specifically, an edge guidance module is utilized to learn the edge attention representations in the early encoding layers, which are then transferred to the multi-scale decoding layers, fused using a weighted aggregation module. The experimental results on four segmentation tasks (i.e., optic disc/cup and vessel segmentation in retinal images, and lung segmentation in chest X-Ray and CT images) demonstrate that preserving edge-attention representations contributes to the final segmentation accuracy. ET-Net is primarily based on an encoder-decoder network, with the EGM and WAM modules appended on the end. The ResNet-50 is utilized as the encoder network, which comprises of four Encoding-Blocks (E-Blocks), one for each different feature map resolution. For each E-Block, the inputs first go through a feature extraction stream, which consists of a stack of 1×1 - 3×3 - 1×1 convolutional layers, and are then summed with the shortcut of inputs to generate the final outputs. With this residual connection, the model can generate class-specific high-level features. The decoder path is formed from three cascaded Decoding-Blocks (D-Blocks), which are used to maintain the characteristics of the high-level features from the E-Blocks and

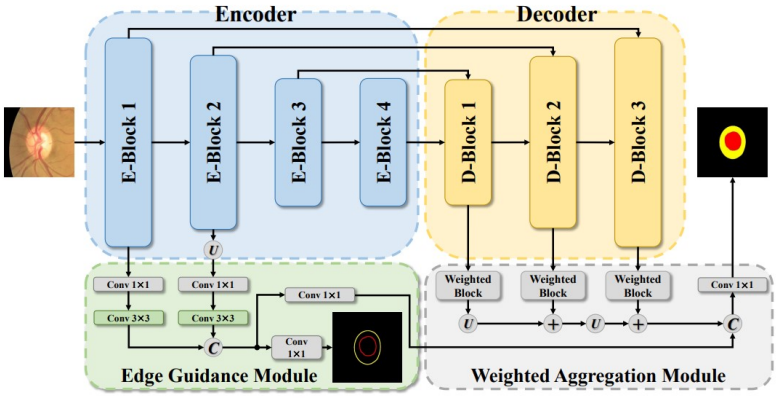

Figure 2.6: ET-Net includes the main encoder-decoder network with an edge guidance module and weighted aggregation module. 'U', 'C', and '+' denote the upsampling, concatenation and addition layers, respectively [12]

enhance their representation ability. As shown in Fig. 2.6, the D-Block first adopts a depth-wise convolution to enhance the representation of the fused low-level and high-level features. Then, a 1×1 convolution is used to unify the number of channels. ETNet achieved the accuracy of 95.60% on DRIVE [36] dataset.

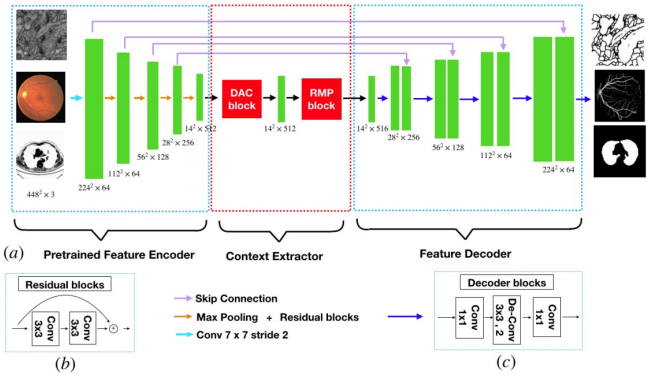

Figure 2.7: CENet with context extractor containing a dense atrous convolution (DAC) block and a residual multi-kernel pooling (RMP) block [13]

Zaiwang et al. [13] proposed a context encoder network (referred to as CE-Net) to capture more high-level information and preserve spatial information for 2D medical image segmentation. CENet mainly contains three major components: a feature encoder module, a context extractor and a feature decoder module.Pretrained ResNet block was used as the fixed feature extractor. The context extractor module is formed by a newly proposed dense atrous convolution (DAC) block and residual multi-kernel

pooling (RMP) block. They proposed a DAC block and RMP block to capture more high-level features and preserve more spatial information as shown in Fig.2.7.They integrated the proposed DAC block and RMP block with encoder-decoder structure for medical image segmentation and applied the method in different tasks including optic disc segmentation, retinal vessel detection, lung segmentation, cell contour segmentation and retinal OCT layer segmentation. Results show that the proposed method outperforms the state-of-the-art methods in these different tasks. CE-Net achieved the accuracy of 95.53% and AUC of 0.9786 on DRIVE dataset.

2.3 Research Gap

After performing rigorous literature survey following points arises:

- Diabetic Retinopathy is a disease which requires continuous medical examinations. However, the current datasets available are not continuous in the nature but still provides significantly good results for grading the DR. It is important to consider various other factors like age, gender and duration of time from which a patient is suffering from Diabetic retinopathy to avoid the vision loss.

- The Datasets that are currently used for grading and segmenting DR images are limited and in addition to that the same datasets are used for lesion detection, pathological myopia and glucoma.

- Apart from traditional 2-D Segmentation techniques, It is important to build a 3-D segmentation tool for performing the surgeries to remove hard exudates. A real time 3-D rendered model of eye can be created using cutting edge technologies like GANs etc.

- Various U-Net based architectures such as CE-Net [13], ET-Net [12], Ladder Net [11] etc. have been implemented for retinal vessel segmentation. However, their accuracy is limited and finer vessels do net get detected. Hence, we require a model which can improve the accuracy of retinal vessel segmentation.

- Diabetic Retinopathy stage detection at earlier stage can help to control the effects of the disease. A classification approach such as CNN is required for predicting the stage of the disease.

To summarize, a vast contribution by researchers is done in the field of Diabetic Retinopathy using automated techniques. However, significant amount of work is still required to be done in the form of scaling, optimizing the framework and algorithm as

number of patients suffering from Diabetic Retinopathy is significantly increasing day by day.

In this project, an effort is made to obtain better results by using Modified UNet Model which provides segmentation results more accurately using DRIVE dataset.

Chapter 3
Novel Method for Diabetic Retinopathy

3.1 Overview

UNET is an architecture developed by Olaf Ronneberger et al. for Biomedical Image Segmentation. It is one of the most popularly used approaches in any semantic segmentation task today. It is a fully convolutional neural network that is designed to learn from fewer training samples. It is an improvement over the existing FCN — "Fully convolutional networks for semantic segmentation". It is a U-shaped encoder-decoder network architecture, which consists of four en- coder blocks and four decoder blocks that are connected via a bridge.

3.2 Implementation Details

3.2.1 Data Augmentation

Data Augmentation is employed to increase the number of input samples for better performance of the overall algorithm. Higher the number of input images, better is the segmentation as the model can train over a large input dataset. Slightly modified slides of the input slides from the DRIVE dataset are generated to increase the number of samples in the input dataset. We have employed cropping, rotation and reflection operations for data augmentation.

Rotation

We rotated the retinal vessel slices in order to increase the number of slices.

Reflection

We reflected the retinal vessel slices along with mask to increase the number of infected slices. We performed horizontal, vertical and diagonal reflections.

3.2.2 UNET Architecture description

The encoder network half the spatial dimensions and double the number of filters at each encoder block. Likewise, the decoder network doubles the spatial dimensions and half the number of feature channels. The architecture of UNET is as shown Fig. 3.1.

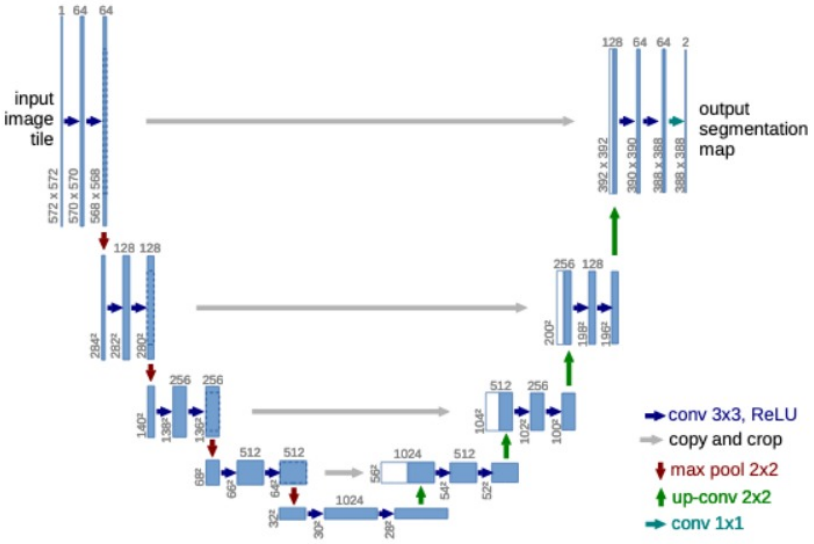

Figure 3.1: UNET Architecture [6]

- **Encoder Network**: Through a series of encoder blocks, the encoder network serves as a feature extractor, learning an abstract representation of the input image. Each encoder block is made up of two 3x3 convolutions, with a ReLU activation function after each convolution. The ReLU's output serves as a skip connection for the decoder block [37].

  Next, follows a $2x2$ max-pooling, where the spatial dimensions of the feature maps are reduced by half. This reduces the computational cost by decreasing the number of trainable parameters.

- **Skip Connection:** These skip connections provide additional information that helps the decoder to generate better semantic features. They also act as a shortcut connection that helps the indirect flow of gradients to the earlier layers without any degradation.

- **Bridge:** The encoder and decoder networks are connected by the bridge, which completes the information flow. It consists of two 3x3 convolutions, with a ReLU activation function after each convolution.

- **Decoder Network:** The abstract representation is utilised to produce a semantic segmentation mask using the decoder network. A $2x2$ transpose convolution is used to begin the decoder block. It is then concatenated with the encoder block's associated skip connection feature map. After that, two $3x3$ convolutions are utilised, followed by a ReLU activation function for each convolution [37].

The output of the last decoder passes through a $1x1$ convolution with sigmoid activation. The sigmoid activation function gives the segmentation mask representing the pixel-wise classification.

3.2.3 Pros and Cons of UNET

Advantages

- UNET model eases the training of CNN due to skip connections.
- UNET decreases the parameters of CNN.
- Can be easily scaled to have multiple classes.
- Relatively easy to understand how the architecture works, if we have basic understanding of how convolutions work.
- UNET model shows promising results for other datsets of DR.

Disadvantages

- UNET does not exploit spatial information.
- Because of many layers takes significant amount of time to train.
- Relatively high GPU memory footprint for larger images.
- Learning may slow down in the middle layers of deeper models, so there is some risk of the network learning to ignore the layers where abstract features are resprepresented.

3.3 Datasets

In the last decade benchmarks have played an increasingly important role in the medical image analysis community. Between 2010 and 2017 overall more than 90 medical imaging benchmarks have been organized. The most prominent tasks were segmentation, detection, classification and prediction tasks. Well established benchmarks often set guiding principles by the provided dataset and evaluation methods and therefore define a baseline for future improvements for certain medical imaging problems. Diabetic Retinopathy is detected and analysed using fundus images of the eye. However, a fully-automated segmentation of eye and its lesion remains still an open problem because different acquisition protocols, differing contrast-agents, varying levels of contrast enhancements and dissimilar scanner resolutions lead to unpredictable intensity differences between tissues. So, in order to judge and evaluate equivocally which method

performs best irrespective of the aforementioned problems, we need a common benchmark and for retinal vessel segmentation. Following are the datasets used for retinal vessel segmentation and diabetic retinopathy diagnosis:
1. DRIVE [36]
2. STARE [38]
3. CHASE-DB1 [39]
4. IDRiD (Indian Diabetic Retinopathy Image Dataset) [40]
5. DIARETB1 [41]

3.3.1 DRIVE

The photographs for the DRIVE database were obtained from a diabetic retinopathy screening program in The Netherlands. The screening population consisted of 400 diabetic subjects between 25-90 years of age. Forty photographs have been randomly selected, 33 do not show any sign of diabetic retinopathy and 7 show signs of mild early diabetic retinopathy.
The images were acquired using a Canon CR5 non-mydriatic 3CCD camera with a 45 degree field of view (FOV). Each image was captured using 8 bits per color plane at 768 by 584 pixels. The FOV of each image is circular with a diameter of approximately 540 pixels. For this database, the images have been cropped around the FOV. For each image, a mask image is provided that delineates the FOV [36].

3.3.2 STARE

The STARE (Structured Analysis of the Retina) dataset is a dataset for retinal vessel segmentation. It contains 20 equal-sized (700×605) color fundus images. The images were acquired using a TopCon TRV-50 fundus camera, FOV35° [38].

3.3.3 CHASE-DB1

The dataset contains 28 eye fundus images with a resolution of 1280 x 960. Two sets of ground-truth vessel annotations are available. The first set is commonly used for training and testing. The second set acts as a “human” baseline [39].

3.3.4 IDRiD

This dataset [40] was available as a part of ”Diabetic Retinopathy: Segmentation and Grading Challenge” organised in conjuction with IEEE International Symposium on Biomedical Imaging (ISBI-2018), Washington D.C.
The dataset is divided into three parts:

A. Segmentation: It consists of 81 original color fundus images divided into train and test set. It also contains groundtruth images for the Lesions (Microaneurysms, Haemorrhages, Hard Exudates and Soft Exudates divided into train and test set) and Optic Disc (divided into train and test set).
B. Disease Grading: It consists of 516 original color fundus images divided into train set (413 images) and test set (103 images). It also contains groundtruth labels for Diabetic Retinopathy and Diabetic Macular Edema Severity Grade divided into train and test set.
C. Localization: It consists of 516 original color fundus images divided into train set (413 images) and test set (103 images). It also contains groundtruth labels for Optic Disc Center Location divided into train and test set and Groundtruth Labels for Fovea Center Location divided into train and test set.

3.3.5 DIARETB1

This is a public database for benchmarking diabetic retinopathy detection from digital images. The main objective of the design has been to unambiguously define a database and a testing protocol which can be used to benchmark diabetic retinopathy detection methods.
The database consists of 89 colour fundus images of which 84 contain at least mild non-proliferative signs (Microaneurysms) of the diabetic retinopathy, and 5 are considered as normal which do not contain any signs of the diabetic retinopathy according to all experts who participated in the evaluation. Images were captured using the same 50 degree field-of-view digital fundus camera with varying imaging settings. The data correspond to a good (not necessarily typical) practical situation, where the images are comparable, and can be used to evaluate the general performance of diagnostic methods [41] .

We used DRIVE dataset for segmentation as it contains a set of high-resolution retina images taken under a variety of imaging conditions. The ground truths of all the training images are also available which could be used for testifying the obtained segmentation results.

3.3.6 Kaggle dataset for Classification

It contains large set of high resolution retinal images taken under a variety of imaging conditions. A left and right field is provided for every subject. Images are labeled with a subject id as well as either left or right (e.g. 1_left.jpeg is the left eye of patient id 1) [42].

A clinician has rated the presence of diabetic retinopathy in each image on a scale of 0 to 4, according to the following scale:
0 - No DR
1 - Mild
2 - Moderate
3 - Severe
4 - Proliferative DR

3.4 Modified UNET Architecture for Diabetic Retinopathy

Some modifications are made to the existing UNET architecture to get better results. The modified architecture is: Attention UNET(Fig. 3.2). The segmented images are passed through CNN to classify them among 5 stages of diabetic retinopathy based on severity.

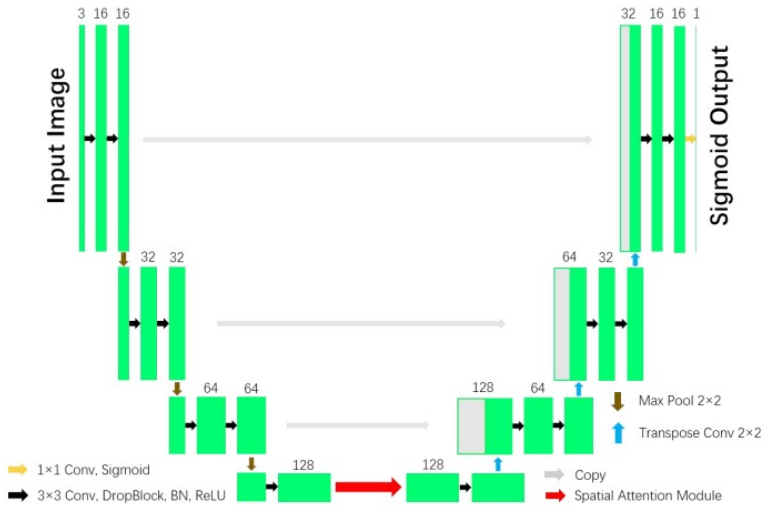

Figure 3.2: SA UNET Architecture [9]

3.4.1 Retinal Vessel Segmentation

In the context of image segmentation, attention is a technique for highlighting only the activations that are relevant during training. This saves computational resources by reducing the number of activations that aren't relevant, giving the network more

generalisation power. In other words, the network can ”pay attention” to certain areas of the image.

Types of Attention

Attention comes in two forms, hard and soft.

- **Hard Attention** Hard attention is based on cropping the image or iterative region proposal to highlight relevant sections. Hard attention is non-differentiable and requires reinforcement learning to train because it can only choose one part of an image at a time.

  Because it is non-differentiable, it means that the network can either pay ”attention” to a specific region in an image or not, with no in-between. As a result, ordinary backpropagation is not possible, and Monte Carlo sampling is required to calculate the accuracy of backpropagation at various stages. [43] Because the accuracy of the model is dependent on how well the sampling is done, other approaches such as reinforcement learning are required to make the model effective.

- **Soft Attention** Soft attention works by giving different sections of an image distinct weights. Those of high significance are given a higher weight, whereas areas of low relevance are given a lower weight. As the model is refined, the regions with greater weights receive more attention. These weights, unlike hard attention, can be given to several areas in the image.

  Soft attention is differentiable and can be trained using traditional backpropagation because of its deterministic nature. As the model is refined, the weighting is refined as well, allowing the model to better choose which portions to focus on.

Why is attention needed in the unet?

The spatial information produced during upsampling in the expanding path is imperfect. To address this issue, the U-Net employs skip connections, which mix spatial data from the downsampling and upsampling paths. However, because feature representation is weak in the earliest layers, this results in many redundant low-level feature extractions.

Soft attention implemented at the skip connections will actively suppress activations in irrelevant regions, reducing the number of redundant features brought across [43].

How is attention implemented?

The attention gates uses additive soft attention. Attention gate schematic and how AGs are implemented at every skip connection are shown in Fig. 3.3.

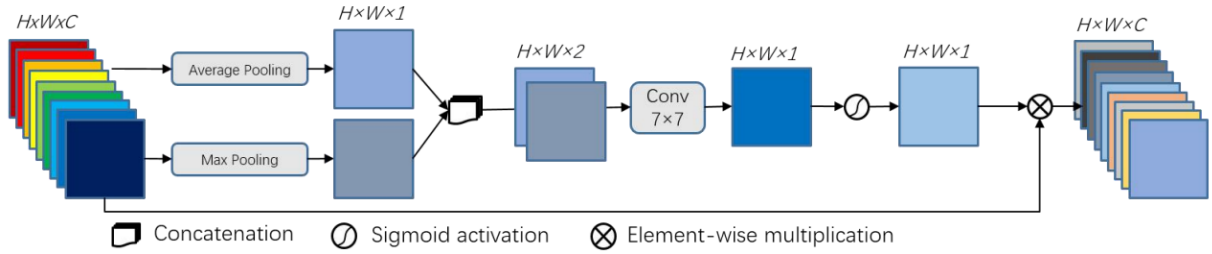

Figure 3.3: Attention module [9]

- A convolutional block attention module was introduced for classification and detection.
- It uses the spatial relationship between features to produce a spatial attention map.
- To calculate spatial attention, the module first applies max- pooling and average-pooling operations along the channel axis and concatenate them to produce an efficient feature descriptor as shown in Fig. 3.3.
- The input feature is fowarded through the channel-wise max-poling and average-pooling to generate outputs.
- Then a convolutional layer followed by the Sigmoid activation function on the concatenated feature descriptor is used to generate a spatial attention map.
- A convolution operation with a kernel size of 7 was performed as shown in Fig. 3.3.

As training progresses, the network learns to focus on the desired location. Because the attention gate is differentiable, it may be trained during backpropagation, making the attention coefficients better at highlighting significant regions.

In the total Dice Coefficient Score, Attention U-Net outperformes a simple U-Net by a significant margin in the total Dice Coefficient Score. While the Attention U-Net has more parameters, the difference isn't significant, and the inference time is just slightly longer.

In conclusion, attention gates provide a straightforward technique to improve the U-Net consistently across a wide range of datasets without incurring a major computational expense.

3.4.2 Classification

We have several coloured fundus images of size 264 *x* 264 in Diabetic Retinopathy Detection Challenge datset taken from Kaggle. We have developed an eight layer convolutional neural network consisting of several convolutional layers with maxpooling size of 2 and dropout of 0.25 followed by 2 dense layers to classify the images into 5 stages of diabetic retinopathy based on severity (0 - No DR, 5 - proliferative DR). The architecture of the classification model is shown in Fig. 3.4.

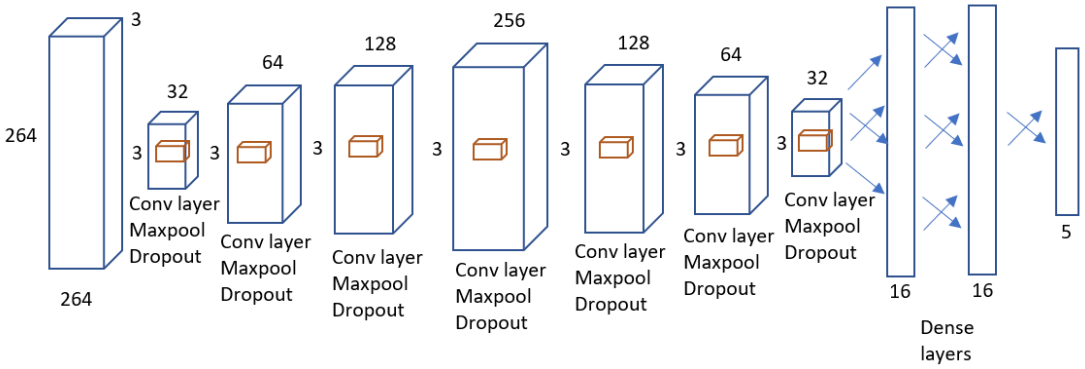

Figure 3.4: CNN architecture

Chapter 4
Simulation and Results

4.1 Overview

This chapter deals with various quantitative metrices that are used to compare the proposed method with several other techniques attempting the task of diabetic retinopathy detection. The quantitative metrices used are mentioned. Further, we have done data augmentation on DRIVE data and used it for training Attention-UNet. The training parameters for the U-net based approach are shown. The results of the classification model are also mentioned. We present a qualitative and quantitative comparison of our method with other available methods.

4.2 Metrices used for Evaluation

Crucial diagnostic tools that help in the interpretation of probabilistic forecast for binary (two-class) classification predictive modeling problems are ROC Curves, Precision-Recall curves and Dice Co-efficient.

The evaluation of the methods was done with Area Under Curve for Receiver Operating Characteristic (ROC AUC), Area Under Curve for Precision and Recall Curve (PR AUC), dice similarity coefficient or F1 measure, Accuracy and Sensitivity. AUC ROC is a graph which plots the trade-off between true positive rate and false positive rate. The more the value of TPR and the lesser the value of FPR is for individual threshold, the better and so the model that has graphs which is more top-left-side such that it has more area under the curve is superior. As for the AUC of PR, it is a curve that fuses precision and Recall in one graph. For each value of recall, the value of precision is calculated and a graph is obtained. The higher on vertical axis your graph is, the better is the performance of the model.

There exist certain terms that are used to define accuracy, sensitivity and other metrices which are as follows:
i. True positive: a result in which the model correctly predicts the positive class.
ii. True Negative: a result in which the model correctly predicts the negative class.
iii. False Positive: a result in which the model incorrectly predicts the positive class.
iv. False negative: a result in which the model incorrectly predicts the negative class.

Based on these terms, accuracy, sensitivity and Dice similarity co-efficient or F1 score are as given by Eq. 4.1, Eq. 4.2 and Eq. 4.3:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.2}$$

$$DSC = \frac{2TP}{FP + FN + 2TP} \tag{4.3}$$

4.3 Simulation of the Approach

4.3.1 Retinal Vessel Segmentation

The DRIVE [36] dataset used for the training of the model consisted of 20 images for training and 20 images for testing purposes. Such small number of images are not sufficient for achieving optimal performance in the deep learning model. So, these images are resized to the size 128 x 128 and they divided into small patches of size 128 × 128 with the overlap of 64 pixels between adjacent patches. In this way, we obtained around 4200 images which are sufficient for training of the model. We have also applied augmentations to the image such as reflection, rotation and added noise to the images for better results. The images are normalized and then, they are fed to the attention-unet model for training. To test the model performance, the test image is divided into patches for segmenting the vessels and they are then fused together to obtain final output.
Input images, and the ground truth have been used to train the model with the following parameters:

- Number of epochs = 10
- Batch Size = 5
- Optimizer: Adam
- Learning Rate = 0.001

The training results are shown in Fig. 4.1. This approach was implemented using TensorFlow library in Python. The model was trained on NVIDIA's K80 GPU on Google Colab.

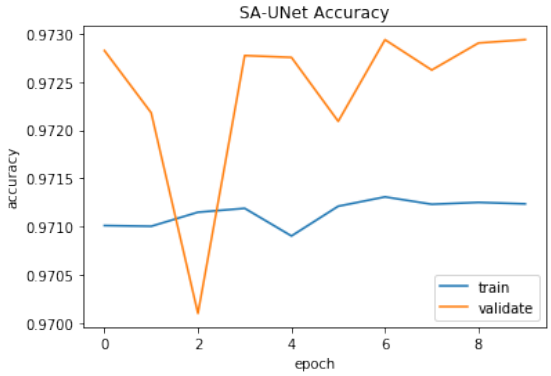

Figure 4.1: Attention-UNET Accuracy

4.3.2 Classification

The Diabetic Retinopathy Detection dataset consisting of a large set of high-resolution retina images taken under a variety of imaging conditions was used for building a classifier model to identify the stage of Diabetic Retinopathy (0-4). A left and right field is provided for every subject. We have used a 8 layer CNN model for training the images with the following parameters:

- Number of epochs = 50
- Batch Size = 16
- Optimizer: Adam
- Learning Rate = 0.001

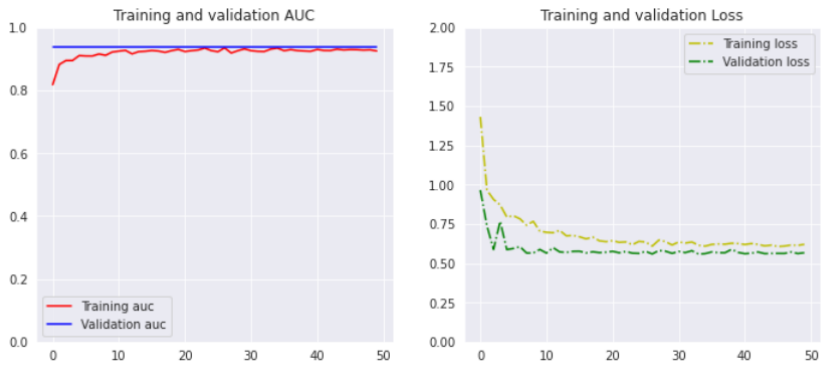

Figure 4.2: CNN Training Results

The training results are shown in Fig. 4.2. This approach was implemented using Keras library in Python. The model was trained on NVIDIA's GPU available in Kaggle notebook.

4.4 Model Performance Evaluation

The evaluation and comparison of the methods were done with the Dice similarity coefficient or F1 measure, Area Under Curve - Receiver Operating Characteristic (ROC) and Precision-Recall (PR).

4.4.1 Quantitative Evaluation

The quantitative comparison and evaluation of the developed model with other state-of-the-art methods is shown in Table 4.1. The attention block implemented at the skip connections actively suppresses activations in irrelevant regions, reducing the number of redundant features brought across and hence improves the accuracy of UNet architecture as is evident from the results.

Table 4.1: Quantitative Comparison of various approaches on DRIVE dataset

Approach	**Accuracy**
CE-Net [13]	0.9545
ET-Net [12]	0.956
U-Net [6]	0.9658
Ladder Net [11]	0.9698
Attention UNET	**0.9712**

The performance summary of the CNN classifier model for identifying the stage of Diabetic Retinopathy is given in Table 4.2.

Table 4.2: CNN Classifier Results on Diabetic Retinopathy Challenge dataset

Evaluation Metric	**Result**
Accuracy score	0.785
Precision score	0.616
Recall score	0.785
F1 Score	0.6904

4.4.2 Qualitative Evaluation

The segmented results on DRIVE dataset are shown in Fig. 4.3. It is compared with the U-Net result from other approaches as shown in Fig. 4.3. It can be inferred that the attention Unet architecture achieves superior qualitative performance as it can detect finer retinal vessels.

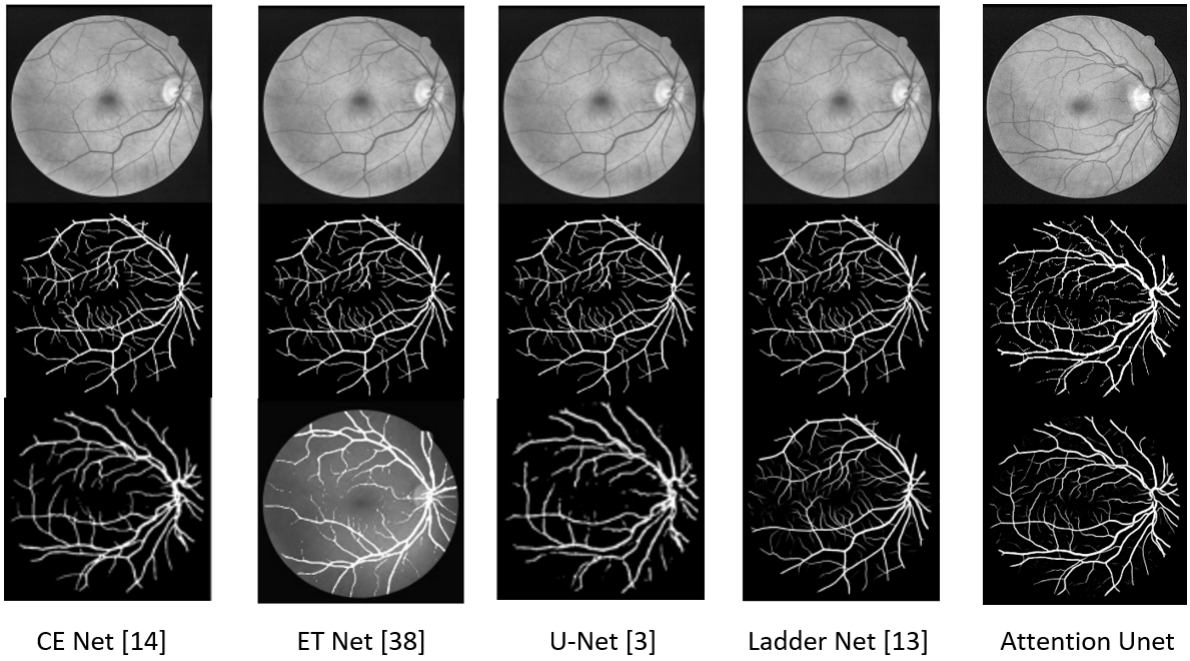

Figure 4.3: Qualitiative Comparison of Attention U-net with other approaches on DRIVE dataset. From top to bottom: input image, ground truth and prediction

The output of the CNN model for classification of Diabetic Retinopathy stages is shown in Fig. 4.4. For a coloured fundus image input, we can accurately predict the stage of Diabetic Retinopathy.

Figure 4.4: CNN classifier result

Chapter 5
Conclusion

5.1 Summary

Detection and diagnosis of retinal-related diseases is a crucial problem in the field of ophthalmology. Segmentation of blood vessels forms the basis for the identification of several ailments. Manual inspection of the retinal blood vessels necessitates utmost expertise and focus. Moreover, it is a strenuous and cumbersome process. It becomes imperative to develop a reliable automatic retinal vessel segmentation algorithm that can reduce the burden on the medical fraternity. Overview of image segmentation algorithms including the threshold based, wavelet-based approaches, artificial neural network-based approaches and unsupervised algorithms are discussed briefly in this report.

The massive upsurge in the volume and complexity of data in the biomedical field pivots all analysis techniques towards machine learning, in particular, deep learning. Not only is it able to handle the massiveness of these kinds of data but also handle the chaotic nature of it, thereby exploring a lot of new opportunities in challenging biomedical applications such as retinal vessel segmentation.

This report leverages attention blocks with U-Net to deploy an accurate deep learning architecture for efficacious retinal vessel segmentation. Additionally, a CNN classifier has been used to identify the stage of Diabetic Retinopathy. The segmentation model achieves an accuracy of 97.12% on DRIVE dataset and the CNN classfication model achieves an accuracy of 79%. The CNN Quantitative and qualitative comparison with several state-of-the-art methods indicate the improved performance of the modified method and establish the developed architecture's clinical reliability.

5.2 Future Scope

We have developed a novel method for the detection of Diabetic Retinopathy. The developed model has performed better than well-known algorithms and we have improved representation of existing network up to certain extent. Still the prediction of our model is not perfect and there is scope for improvement, certain post processing algorithms can be used to boost the final score. To further improve the results, unsupervised learning techniques such as multi-scale GANs can be used. A Study Group Learning scheme

can also be employed to improve the robustness of the model trained on noisy labels. Apart from that LSTM based techniques can also be incorporated as this task has spatial dimensions. The accuracy of the classifier model can also be further improved by using advanced algorithms such Inception V3 CNN model.

References

[1] "Eye structure and functions." [Online]. Available: https://en.excimerclinic.ru/press/stroenieglaza/

[2] G. Ghan, S. Chavan, and A. Chaudhari, "Diabetic retinopathy classification using deep learning," in *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*, 2020, pp. 761–765.

[3] M. Naeemabadi, B. Ordoubadi, A. Dehnavi, and K. Bahaadinbeigy, "Comparison of serpent, twofish and rijndael encryption algorithms in tele- ophthalmology system," *Advances in Natural and Applied Sciences*, vol. 9, pp. 137–149, 01 2015.

[4] M. S. Ajay, "Introduction to artificial neural networks." [Online]. Available: https://towardsdatascience.com/introduction-to-artificial-neural-networks-ac338f4154e

[5] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling, "M2det: A single-shot object detector based onmulti-level feature pyramid network," *AAAI*, 2019.

[6] O. Ronneberger, P. Fischer, and T. Brox., "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention, Springer*, p. 234–241, 2015.

[7] Z. Z, S. MM, T. N, and L. J., "Unet++: "a nested u-net architecture for medical image segmentation"," *Deep learning in medical image analysis and bi- learning for clinical decision support*, pp. 3–11, 2018.

[8] G. G. Gardner, D. Keating, T. H. Williamson, and A. T. Elliott, "Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool." *British Journal of Ophthalmology*, vol. 80, no. 11, pp. 940–944, 1996. [Online]. Available: https://bjo.bmj.com/content/80/11/940

[9] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "Sa-unet: Spatial attention u-net for retinal vessel segmentation," *ArXiv*, vol. abs/2004.03696, 2020.

[10] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 3656–3665.

[11] J. Zhuang, "Laddernet: Multi-path networks based on u-net for medical image segmentation," 2018. [Online]. Available: https://arxiv.org/abs/1810.07810

[12] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, "Et-net: A generic edge-attention guidance network for medical image segmentation," 2019. [Online]. Available: https://arxiv.org/abs/1907.10936

[13] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, oct 2019.

[14] J. Garrity, "Structure and function of the eyes," Jan 2020. [Online]. Available: https://www.msdmanuals.com/en-in/home/eye-disorders/biology-of-the-eyes/structure-and-function-of-the-eyes

[15] R. E. W. Rafael C. Gonzalez, "Digital image processing," *Pearson Education Limited England*, p. 699–804, 2018.

[16] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Pearson/Prentice Hall, NY, 2018.

[17] B. A and M. A., "An efficient watershed algorithm based on connected components. pattern recognition," vol. 33, no. 6, 2020, pp. 907–916.

[18] R. Khandelwal, "Machine learning : Gradient descent," [Visited on: 2022-05-08]. [Online]. Available: https://arshren.medium.com/gradient-descent-5a13f385d403

[19] R. Das, "Guide to optimizers for machine learning," [Visited on: 2022-05-08]. [Online]. Available: https://analyticsindiamag.com/guide-to-optimizers-for-machine-learning/

[20] H. C. Lau, Y. O. Voo, K. T. Yeo, S. L. Ling, and A. Jap, "Mass screening for diabetic retinopathy–a report on diabetic retinal screening in primary care clinics in singapore." *Singapore medical journal*, vol. 36, no. 5, pp. 510–513, 1995.

[21] L. B. Bäcklund, P. V. Algvere, and U. Rosenqvist, "New blindness in diabetes reduced by more than one-third in stockholm county," *Diabetic Medicine*, vol. 14, no. 9, pp. 732–740, 1997.

[22] T. Walter, J.-C. Klein, P. Massin, and A. Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina," *IEEE Transactions on Medical Imaging*, vol. 21, no. 10, pp. 1236–1243, 2002.

[23] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "The preprocessing of retinal images for the detection of fluorescein leakage," *Physics in Medicine and Biology*, vol. 44, no. 1, pp. 293–308, jan 1999.

[24] T. Walter and J.-C. Klein, "Automatic detection of microaneurysms in color fundus images of the human retina by means of the bounding box closing," ser. ISMDA '02. Springer-Verlag, 2002, p. 210–220.

[25] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1273–1284, 2016.

[26] Y. Huang, L. Lin, M. Li, J. Wu, P. Cheng, K. Wang, J. Yuan, and X. Tang, "Automated hemorrhage detection from coarsely annotated fundus images in diabetic retinopathy," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 1369–1372.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2999–3007.

[28] M. H. Sarhan, S. Albarqouni, M. Yigitsoy, N. Navab, and A. Eslami, "Multi-scale microaneurysms segmentation using embedding triplet loss," 2019.

[29] L. Dai, R. Fang, H. Li, X. Hou, B. Sheng, Q. Wu, and W. Jia, "Clinical report guided retinal microaneurysm detection with multi-sieving deep learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 5, pp. 1149–1161, 2018.

[30] K. Adem, "Exudate detection for diabetic retinopathy with circular hough transformation and convolutional neural networks," *Expert Systems with Applications*, vol. 114, pp. 289–295, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417418304780

[31] S. Guo, K. Wang, H. Kang, T. Liu, Y. Gao, and T. Li, "Bin loss for hard exudates segmentation in fundus images," *Neurocomputing*, vol. 392, pp. 314–324, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231219304783

[32] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning," *Investigative Ophthalmology Visual Science*, vol. 57, no. 13, pp. 5200–5206, 10 2016.

[33] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "Development and Validation of a Deep

Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 12 2016. [Online]. Available: https://doi.org/10.1001/jama.2016.17216

[34] J. Krause, V. Gulshan, E. Rahimy, P. Karth, K. Widner, G. S. Corrado, L. Peng, and D. R. Webster, "Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy," *CoRR*, vol. abs/1710.01711, 2017. [Online]. Available: http://arxiv.org/abs/1710.01711

[35] W. Zhang, J. Zhong, S. Yang, Z. Gao, J. Hu, Y. Chen, and Z. Yi, "Automated identification and grading system of diabetic retinopathy using deep neural networks," *Knowledge-Based Systems*, vol. 175, pp. 12–25, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705119301303

[36] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[37] N. Tomar, "What is UNET? - analytics vidhya - medium," https://medium.com/analytics-vidhya/what-is-unet-157314c87634, Jan. 2021, accessed: 2021-12-6.

[38] V. K. A. Hoover and M. Goldbaum, "Locating blood vessels in retinal images by piece-wise threhsold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, March 2000.

[39] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012.

[40] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research," p. 25, Jul 2018. [Online]. Available: https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid

[41] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, R. A., V. R., H. Uusitalo, H. Kälviäinen, and J. Pietilä, "Diaretdb1 diabetic retinopathy database and evaluation protocol," in *11th Conf. on Medical Image Understanding and Analysis (Aberystwyth, Wales, 2007)*, 2007. [Online]. Available: https://www.it.lut.fi/project/imageret/diaretdb1/

[42] "Diabetic retinopathy detection," https://www.kaggle.com/competitions/diabetic-retinopathy-detection, accessed: 2022-5-10.

[43] R. Vinod, "A detailed explanation of the attention U-Net," https://towardsdatascience.com/a-detailed-explanation-of-the-attention-u-net-b371a5590831, May 2020, accessed: 2022-5-10.

Publication

Manuscript of the Research Paper

We have prepared Manuscript for submission in IEEE International Conference. Moksha Sood, Sindhu Guntha, Dhruvkumar Patel and Dr. Jignesh Sarvaiya, “Attention-Unet for Retinal Vessel Segmentation”.