# SEGMENTATION OF MULTI CLASS RETINAL LESIONS FROM FUNDUS IMAGES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ELİF KÜBRA ÇONTAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

FEBRUARY 2022

Approval of the thesis:

**SEGMENTATION OF MULTI CLASS RETINAL LESIONS FROM FUNDUS IMAGES**

submitted by **ELİF KÜBRA ÇONTAR** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. İlkay Ulusoy
Head of Department, **Electrical and Electronics Engineering** _____

Prof. Dr. Gözde Bozdağı Akar
Supervisor, **Electrical and Electronics Engineering, METU** _____

**Examining Committee Members:**

Prof. Dr. İlkay Ulusoy
Electrical and Electronics Engineering, METU _____

Prof. Dr. Gözde Bozdağı Akar
Electrical and Electronics Engineering, METU _____

Prof. Dr. Gözde Ünal
Artificial Intelligence and Data Engineering, ITU _____

Prof. Dr. Ziya Telatar
Electrical and Electronics Engineering, Ankara University _____

Assoc. Prof. Dr. Yeşim Serinağaoğlu Doğrusöz
Electrical and Electronics Engineering, METU _____

Date:

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Surname: Elif Kübra Çontar

Signature :

# ABSTRACT

## SEGMENTATION OF MULTI CLASS RETINAL LESIONS FROM FUNDUS IMAGES

Çontar, Elif Kübra
M.S., Department of Electrical and Electronics Engineering
Supervisor: Prof. Dr. Gözde Bozdağı Akar

February 2022, 68 pages

Diabetic retinopathy is a leading cause of preventable blindness among adults. Detection of diabetic retinopathy-related retinal lesions is essential for automatic detection of DR. There are different kinds of lesions related to the disease, namely microaneurysm, hemorrhage, hard exudate, and soft exudate. Each lesion has different characteristics: color, size, and shape.

In the literature, the detection of retinal lesions has been examined as a localization or segmentation problem. Besides traditional image processing methods, machine learning-based and neural network-based methods have been proposed widely in the last years. Most of the works focused on detecting only one type of lesion. These methods can not be transferred to detect another kind of lesion because of the different characteristics of the lesions. Additionally, segmentation of retinal lesion task is an imbalanced classification problem. Task includes both foreground-background imbalance and imbalance between positive classes.

In this study, we developed a new instance-based intersection over union(IB_IoU) objective function to segment multi-class retinal lesions from fundus images. The

loss has targeted the following two problems. Firstly, it aims to solve the imbalance problem by averaging intersection over union(IoU) scores across the classes. Secondly, IoU score is calculated separately for every instance with a closed contoured shape. The aim is to improve the detection performance of lesions with small pixel areas. The connected component analysis is applied to find instances on a union of prediction and ground truth labels.

The results show that the proposed algorithm is comparable to state-of-the-art methods focused on detecting single lesions. Additionally, the proposed loss function has improved detection performance of microaneurysm and exudate lesions over other loss functions used in multi-class retinal lesion segmentation.

Keywords: diabetic retinopathy, segmentation, retinal lesion, deep learning, fundus image

# ÖZ

## FUNDUS GÖRÜNTÜLERİNDEN ÇOK SINIFLI RETİNA LEZYONLARININ SEGMENTASYONU

Çontar, Elif Kübra
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Gözde Bozdağı Akar

Şubat 2022 , 68 sayfa

Diyabetik retinopati, yetişkinlerde görülen önlenebilir körlüğün en önemli nedenlerinden biridir. Diyabetik retinopati(DR) ile ilişkili retina lezyonlarının tespiti, DR'nin otomatik teşhisi için önemli bir basamaktır. Hastalıkla ilgili birden fazla farklı lezyon tipi bulunmaktadır. Bunlar sırasıyla: mikroanevrizma, hemoraji, sert eksüda ve yumuşak eksüdadır. Lezyon tipleri renk, boyut ve şekil açısından farklı görsel özelliklere sahiptir.

Literatürde retina lezyonlarının tespiti lokalizasyon veya segmentasyon problemi olarak incelenmiştir. Geleneksel görüntü işleme yöntemlerinin yanı sıra, makine öğrenmesi tabanlı ve sinir ağları tabanlı yöntemler son yıllarda yaygın olarak geliştirilmiştir. Çalışmaların bir çoğu tek lezyon tipini tespit etmeye odaklanmıştır. Tek tip üzerinde geliştirilmiş yöntemler, lezyonların farklı özellikleri nedeniyle başka bir lezyon türünü saptamakta kullanılamıyor. Ek olarak, retina lezyonu segmentasyonu veri dağılımı sebebiyle dengesiz bir sınıflandırma problemidir. Görev, hem ön plan-arka plan dengesizliğini hem de pozitif sınıflar arasındaki dengesizliği içermektedir.

Bu çalışmada, fundus görüntülerinden çok sınıflı retina lezyonlarının bölütlemesi için örnek temelli kesişim bölü bileşim(IB_IoU) hata fonksiyonunu geliştirdik. Hata fonksiyonu, aşağıdaki iki sorunu hedef almıştır. İlk olarak, IoU puanı her bir sınıf için ayrı ayrı hesaplanarak ortalaması alınır. Bu şekilde dengesizlik sorununu çözmek amaçlanmıştır. İkinci olarak, kapalı çevreli şekle sahip her örnek için IoU puanı hesaplanır. Amaç, küçük piksel alanlarına sahip örneklerin de tespit edilebilmesi ve kaçırılmaması için hata fonksiyonuna katkı yapmalarını sağlamaktır. Kapalı çevreli şekle sahip örnekleri bulmak için bağlantılı bileşen analizi uygulanmıştır.

Sonuçlar, geliştirilen algoritmanın performansının, tek lezyon tipini tespit etmeye odaklanan son teknoloji yöntemlerle karşılaştırılabilir bir seviyede olduğunu göstermektedir. Ek olarak, önerilen hata fonksiyonu, çok sınıflı retina lezyonu segmentasyonunda kullanılan diğer hata fonksiyonlarına göre mikroanevrizma ve eksüda lezyonlarının tespit performansını iyileştirmiştir.

Anahtar Kelimeler: diyabetik retinopati, segmentasyon, derin öğrenme, fundus, retina lezyonu

To my dearest family

## ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Prof. Dr. Gözde Bozdağı Akar, for her support and guidance. In the light of her extensive knowledge and experience, it was much easier to write this thesis.

I would like to present my gratitude to Adem Günesen. He was always there for me with his endless support and help. Our technical discussions and his comments are the cornerstones of this research. It's nice to know that he'll be there whenever I need him.

My dear family. All my life, they are the ones who supported me the most and believed in me the most. Thank you for giving me the freedom I need while always wishing the best for me.

## TABLE OF CONTENTS

# LIST OF TABLES

TABLES

## LIST OF FIGURES

FIGURES

# LIST OF ALGORITHMS

ALGORITHMS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| DR | Diabetic Retinopathy |
| CNN | Convolutional Neural Network |
| CAD | Computer Aided Diagnosis |
| MA | Microaneurysm |
| EX | Exudate |
| SE | Soft Exudate |
| HE | Hemorrhage |
| KNN | K-Nearest Neighbor |
| SVM | Support Vector Machine |
| ROC | Receiver Operating Characteristic |
| AUC | Area Under Curve |
| IDRiD | Indian Diabetic Retinopathy Image Dataset |
| DRIVE | Digital Retinal Images for Vessel Extraction Dataset |
| STARE | Structured Analysis of the Retina Dataset |
| NPDR | Nonproliferative Diabetic Retinopathy |
| PDR | Proliferative Diabetic Retinopathy |
| TTA | Test Time Augmentation |
| GPU | Graphics Processing Unit |
| CPU | Central Processing Unit |
| IoU | Intersection Over Union |
| IB_IoU | Instance Based Intersection Over Union |
| MC | Multi Channel Loss |
| MCB | Multi Channel Bin Loss |
| MSE | Mean Squared Error |

| | |
|---|---|
| FCN | Fully Convolutional Network |
| TP | True Positive |
| TN | True Negative |
| FP | False Positive |
| FN | False Negative |
| AUPR | Area Under Precision Recall Curve |

# CHAPTER 1

# INTRODUCTION

According to the World Health Organization, 85 percent of blindness is preventable worldwide, and diabetic retinopathy(DR) is a leading cause of preventable blindness in working-age adults[9]. DR may result in damage in irreversible visual acuity without appropriate treatment. Therefore, regular ophthalmic examination sessions are essential for early diagnosis. DR causes abnormal retina patterns such as exudate, microaneurysm, hemorrhage, and vascularity in retinal blood vessels. These abnormalities are shown in Figure 1.1. Each anomaly has its own distinct visual characteristics, such as color, size, and shape. The existence of different abnormalities and their number of instances state the level of disease[10]. According to the international clinical diabetic retinopathy severity scale[11] DR is examined under five stages:

- 0: Healthy
- 1: Mild NPDR
- 2: Moderate NPDR
- 3: Severe NPDR
- 4: Proliferative DR.

The way of treatment and treatment cost changes according to the disease stage. The early stages of DR are less severe and clinically managed. It is important to identify early indicators for DR. DR is diagnosed by examining the fundus image by an ophthalmologist who is an expert. Fundus image is taken by using fundus cameras. The resolution, lighting, and quality of the image vary according to the camera. Retina image consists of disease-related sections called pathological lesions such as

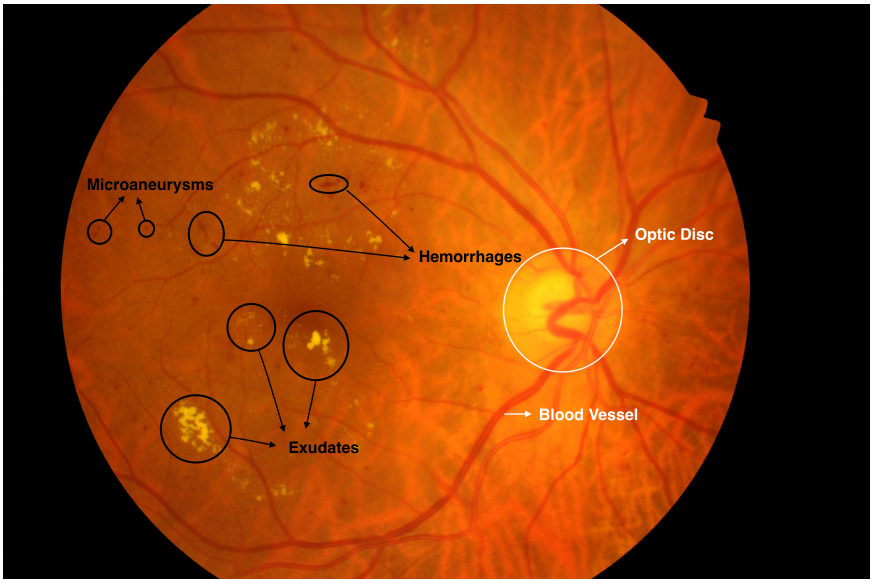



Figure 1.1: Retinal image with abnormalities(black) and natural structures(white).

exudate, microaneurysms, hemorrhage, and vascularity and natural structures such as blood vessels, optic disc, optic cup, and macula. There are 422 million diabetics globally, and the World Health Organization recommends annual eye examinations for diabetic patients. Considering that the average number of ophthalmologists per million population is changing from 9 to 79 for low-income countries to high-income countries[12], Computer-Aided Diagnosis(CAD) applications are important for an accessible eye examination. Early research uses different machine learning based methods like support vector machines and k-nearest neighbors to extract or segment retinal pathologies[13]. Additionally, traditional image processing methods like mathematical morphological operations, region growing methods, and ensemble methods[14] have been used[14, 15, 16]. Niemeijer applied a mathematical morphology based candidate extraction and pixel classification system. After that, the final candidate regions are decided by hybrid candidate classification. However, all traditional methods require manual feature extraction, and domain knowledge is mandatory. Also, deep learning methods have been developed with the increasing popularity of neural networks. Different methods have been developed for extracting visual attributes of the retina. FCN, U-Net, SegNet, and MaskRCNN architecture are popular in medical image segmentation and used with some modifications in various retinal applications[17, 18]. CNN based methods are used for red lesion

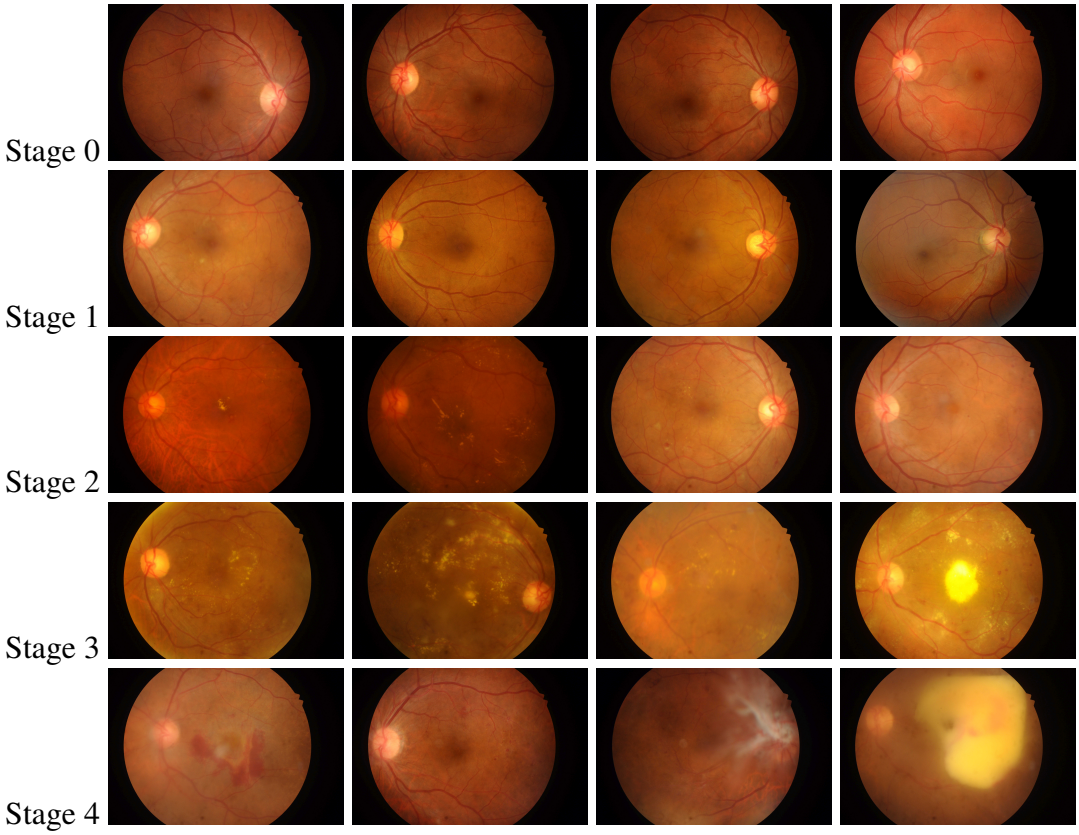



Figure 1.2: Retinal images at different stages of diabetic retinopathy.

segmentation[19, 20, 21] and autoencoder methods have been used too[22].

## 1.1 Contributions and Novelties

This thesis focuses on the deep learning methods for the segmentation of retinal lesions from color fundus images. Our contributions are as follows:

- A new objective function is proposed based on intersection over union(IoU) metric. This objective function aims to solve the class imbalance between positive classes and increase the detection performance of small-scale lesions. The objective function calculates IoU score for each instance individually. Experiments show that the proposed objective function increases the segmentation performance of small and scattered lesions compared to other loss functions.
- The two-stage model is proposed to decrease the background-foreground im-

balance effect on training. Patches can be divided into two. The ones include lesions structure, and the other all pixels belong to the background. Even positive patches, which include lesion areas, have imbalanced background-foreground distribution. Only the positive patches are used to train the multi-class segmentation model to decrease the imbalance effect. Also, another model is trained to classify patches, whether it includes any lesion structure or not. Combining these two models perform better compared to the single-stage segmentation model.

- Lesions close to blood vessels in the distance and lesions similar in color to the blood vessel, such as MA and HE, are more likely to be confused with blood vessels. Blood vessels are removed from images to investigate this relation, and the segmentation model is trained with blood vessel removed images. However, the IDRiD dataset does not include labels for blood vessels. Therefore, pseudo labels of blood vessels are obtained. The results are compared to understand the effect of blood vessel removal on MA and HE detection performance.

## 1.2 The Outline of the Thesis

The thesis is organized as follows. Problem definition and the contributions are given in Chapter 1: Introduction. In Chapter 2, literature review is given. Methods are divided into two parts according to the number of detected lesions. In Chapter 3, background information about the IDRiD database and the objective functions are given respectively. Detailed evaluation of the dataset, characteristics of lesion data, and its use in this research are given under dataset investigation. Patch creation and distribution of patched dataset are mentioned. Background information about the objective functions is given with their explanation and formulations. The proposed method is explained in Chapter 4. It starts with explaining the motivation of the method. The flow of the proposed method and individual parts are explained in this chapter. Combination of the patch classification model and segmentation model is given. The proposed objective function IB_IoU is explained in detail. It is explained how pseudo labels for blood vessels are obtained at the end of the section. Experimental results are given in Chapter 5. Also, hardware specifications and evaluation metrics

are explained in this section too. Both positive and negative findings are discussed in Chapter 5. At the final, Chapter 6 closed with conclusion.

# CHAPTER 2

# LITERATURE REVIEW

The studies on diabetic retinopathy-related lesion detection are divided into two sections according to the number of detected lesion types in the literature. Most of the works develop a model which detects a single lesion type[14, 15, 16, 23, 24, 24, 25, 26, 27, 22, 28, 2]. These works include traditional image processing methods, machine learning-based, and deep learning-based methods. In image processing methods, algorithms are developed based on the visual characteristics of the lesions, such as shape and color [14, 15, 16]. Machine learning algorithms consist of feature extractor and detection parts. For the feature extraction, both hand-crafted features[23, 24] and CNN-based extractors are used[24, 25]. For classifier, Naive Bayesian classifier and support vector machines(SVM) are used[26, 27]. In deep learning-based methods, encoder-decoder networks[22, 28, 2] and generative models are the most common. For the background-foreground imbalance problem, the following loss functions are used in the literature: dice loss, IoU loss, and weighted binary cross-entropy. There is no solution required for the class imbalance between positive classes because of the nature of the problem. Segmentation of one lesion is a binary classification problem. There are two main problems of the methods that detect a single type of lesion. Firstly, their inference time is larger than the methods which segment multiple lesions simultaneously. Secondly, these methods can not be transferred to detect another kind of lesion because of the different characteristics of the lesions. Some lesions give better results with global information, while local information is more informative for others. So models are developed accordingly. Moreover, the subsequent study separately trains the similar model with minor modifications for more than one lesion. The difficulty in this study is that some lesions work better with global information, while others work better with local information.

Yan[28] proposed a network that considers local and global information to overcome this challenge. The model receives both the resized whole image and the patch as input. However, the inference time problem persists. The difficulty of the segmentation of each lesions type is explained in the section 2.3.

In the second part, studies that work on detecting multi-class lesions are investigated. Badar proposed a lightweight version of SegNet[29] to detect three lesion types in 2018[30]. SegNet is a convolutional encoder-decoder architecture for image segmentation. The number of images in the dataset is increased using patching. Their result cannot be compared since they used the Messidor dataset with private annotation. Table 2.1 compares the results of existing methods. The table includes both methods: detecting a single type of lesion and detecting multi-class lesions. Given evaluation metric is area under precision curve. In Chapter 5 the proposed algorithm is compared with the last two methods because the last two methods focus on detecting multi-class lesions.

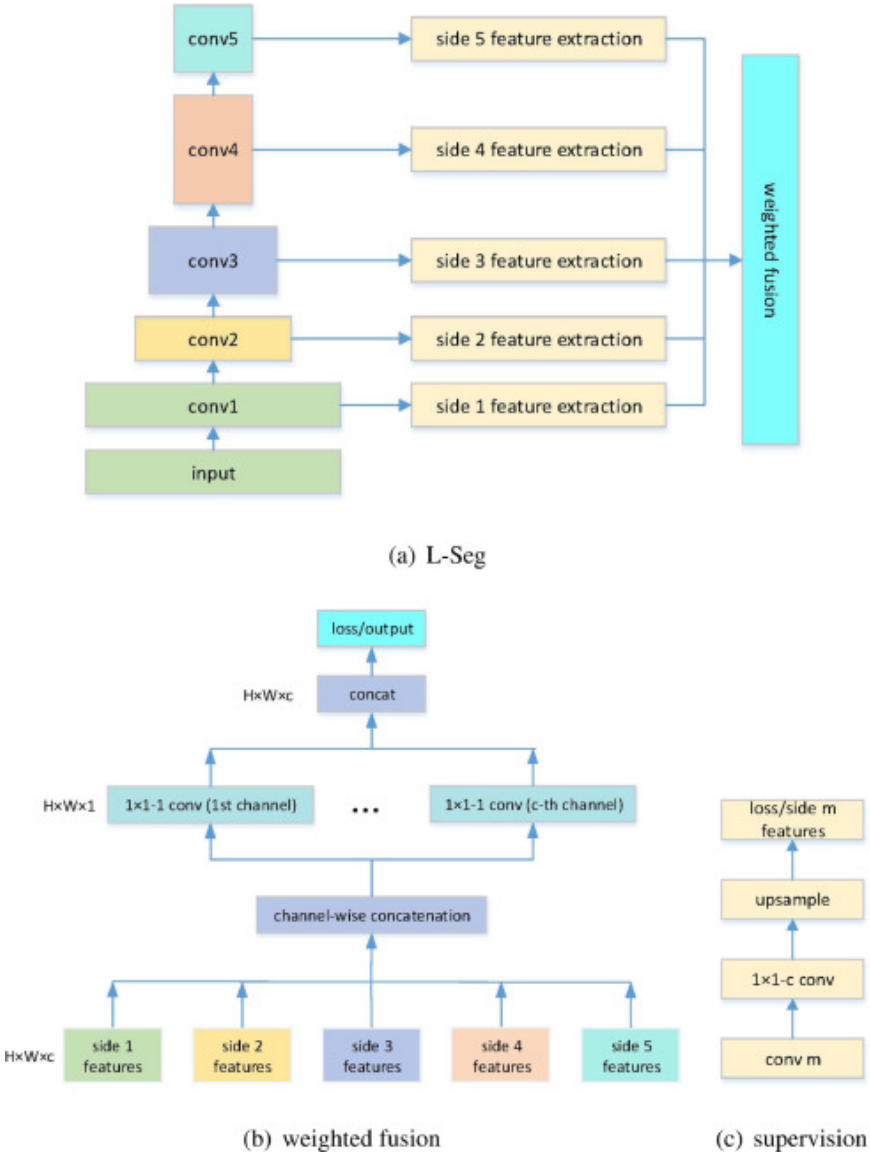
Table 2.1: Comparison of Existing Lesion Detection Methods

| Method | Number of Lesions | MA | HE | EX |
|---|---|---|---|---|
| VRT(1st of IDRiD competition[31]) | Single | 0.4951 | 0.6804 | 0.7127 |
| Yan[28] | Single | 0.525 | 0.703 | 0.0.889 |
| Wan[2] | Multi | 0.2408 | 0.5649 | 0.6083 |
| Guo[1], MCB | Multi | 0.4627 | 0.6374 | 0.7945 |
| Guo, MCB | Multi | 0.4710 | 0.5808 | 0.7410 |

## 2.1 L-Seg: An End-to-End Unified Framework for Multi-Lesion Segmentation of Fundus Images[1]

Guo proposed the L-Seg architecture for the segmentation of multi-class retinal lesions. Their approach is to create a shallow network using the middle layer's output of VGG16[3] architecture. The outputs of five convolutional blocks of VGG16 are used as a feature extractor. After that, these features are concatenated and used to obtain the final prediction. Let us say the shape of the input image is $W \times H \times C$,

and then convolutional blocks output is $A \times B \times C$. Then these outputs are upsampled with $1 \times 1$ convolution filter to the size $W \times H \times C$. Both upsampled features and the final output are used when calculating loss. The whole image is resized to $1440 \times 960$. Their network is lightweight, and the results are comparable with the literature. Additionally, multi-channel bin loss function is used. It assigns weights to the positive and background classes inversely proportional to the pixel-wise class ratio.




(a) L-Seg

(b) weighted fusion

(c) supervision

Figure 2.1: Overview of L-Seg architecture.

## 2.2 EAD-Net: A Novel Lesion Segmentation Method in Diabetic Retinopathy Using Neural Networks[2]

In 2021, Wan proposed the EAD-Net, encoder-decoder model with the attention module. The whole image is resized to 512-pixel size. They have come up with an evaluation metric that considers each closed contoured instance instead of pixels. An instance is counted as a true positive if the overlap ratio of ground truth and prediction of an instance is larger than a threshold. Dice loss is used to solve the imbalance problem. Even dice loss solves the background-foreground imbalance problem; it is inefficient to solve the imbalance between positive classes.

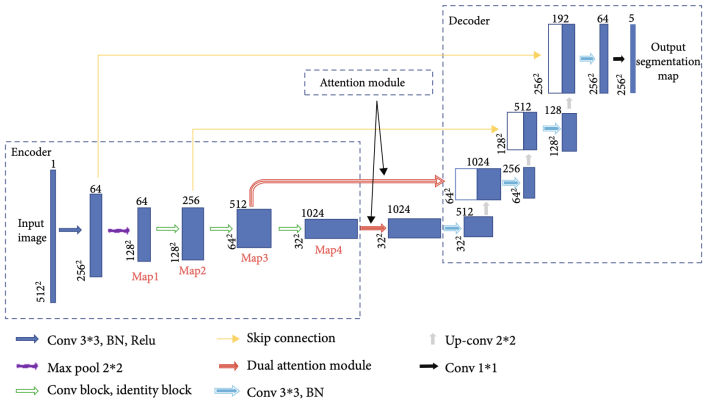



Figure 2.2: Overview of EAD-Net architecture.

## 2.3 Definition of Lesion Types

### 2.3.1 Microanerysm

Microaneurysm is one of the earliest signs of DR. Difficulty of the segmentation of the MA relies on three reasons: contrast, color, and size of the lesion. Microaneurysmand hemorrhage pathologies have similar color contrast with each other and also with blood vessels. Therefore, it is critical to distinguish them and blood vessels. However, they do not have regular shapes and are harder to detect by morphological operations. Lastly, the biological diameter of a MA lesion varies between 15-60 um [23] and

forms 0.15 percent of the total pixel area in the digital fundus image [32]. As a result, the imbalance problem should be considered too under the detection of MA.

### 2.3.2 Exudate

Exudate, the blood fluid that comes out of the tissue, can be described as yellow or white-colored and sharply shaped lesions. It is investigated under two types: hard and soft exudates. Hard exudates are bright yellow-colored and have sharp edges, while soft exudates are white or light yellow-colored lesions with poorly defined edges. Its contrast with the background is high compared with other types of lesions.

### 2.3.3 Hemorrhage

A hemorrhage is bleeding that occurs in blood vessels. It can be observed in the form of a dot, blot, or extensive subhyaloid hemorrhage[33]. Its color range has similarities with the microaneurysm and blood vessel. Therefore, its tiny dot form tends to be confused with MA. Another difficulty is that HE lesions size varies in a wide range, which can be seen in Figure 3.1. Developing a model that detects very small and large lesions that cover a significant part of the image is the difficulty of the task.

# CHAPTER 3

# BACKGROUND INFORMATION

## 3.1 Datasets for Diabetic Retinopathy Related Lesion Segmentation

There are several databases for pixel-wise retinal lesion segmentation in diabetic retinopathy. E_optha database includes two sub-datasets, namely E_optha_EX and E_optha_MA. E_optha_EX consists of 47 images with exudates and 35 images with no lesion. E_optha_MA consists of 148 images with microaneurysms or small hemorrhages and 233 images with no lesion. DiaretDB1 database consists of 89 images labeled for hemorrhages, soft exudates, hard exudates, and small red dots. DDR database has 757 images with pixel-wise lesion segmentation labels for microaneurysms, hemorrhages, hard exudates, and soft exudates. Although it has a large number of images, only a small part of it has positive labels for lesions. Additionally, its pixel resolution is $2592 \times 1728$. Lastly, there is an IDRiD database. It is explained in detail in subsection 3.1.1 Moreover, there are datasets such as Messidor, Messidor2, and CLEOPATRA, and several segmentation algorithms have been developed using them. However, Messidor and Messidor2 do not publish labels of the segmentation lesions. It includes only fundus images. Also, CLEOPATRA is not a public dataset. In this thesis, IDRiD dataset is used because of two reasons. One is that it includes labels for four types of lesions. Secondly, it has high resolution compared to other datasets.

### 3.1.1 Overview of Indian Diabetic Retinopathy Image Dataset (IDRiD) Dataset

Indian Diabetic Retinopathy Image Dataset (IDRiD) dataset was presented in 2018 at the Diabetic Retinopathy: Segmentation and Grading Challenge workshop held in

IEEE International Symposium on Biomedical Imaging[6]. Images are captured using Kowa VX-10 fundus camera. Images have $4288 \times 2848$ pixel resolution and a 50-degree field of view. The dataset consists of three sections: lesion segmentation, disease classification, and optic disc and fovea detection. The dataset includes 81 images labeled by experts considering four lesions type under the lesion segmentation. 81 images are divided into train set and test set respectively 54 and 27 images. Each image has a pixel-level annotation for microaneurysm, hemorrhage, hard exudate, and soft exudate. In this research, three lesions, microaneurysm, hemorrhage, and hard exudate, are examined for segmentation problem.

Table 3.1: Number of images per lesion for IDRiD dataset.

| Lesion Type | Training Set | Test Set |
|---|---|---|
| MA | 54 | 27 |
| HE | 53 | 27 |
| EX | 54 | 27 |
| SE | 26 | 14 |

Table 3.2: Pixelwise lesion area percentage for IDRiD dataset [1].

| Lesion Type | Percentage (%) |
|---|---|
| MA | 0.10 |
| HE | 1.01 |
| EX | 0.90 |
| SE | 0.19 |

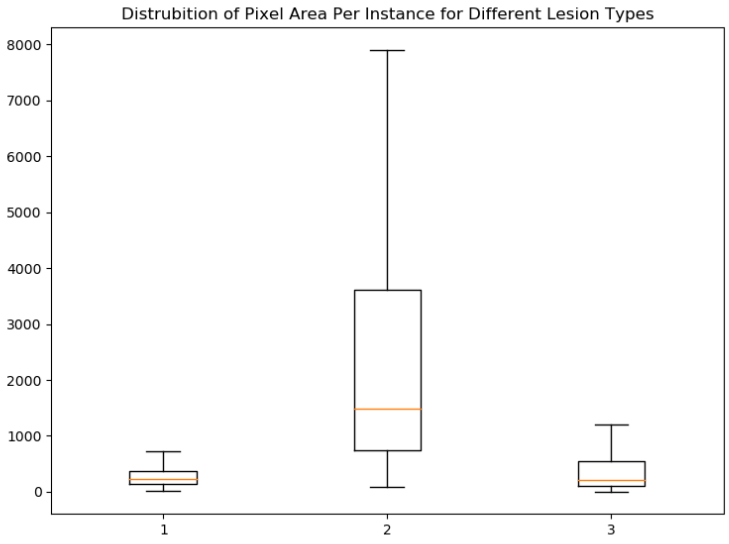



Figure 3.1: Distribution of pixel area per instance for different lesion types. From left to right microaneurysm, hemorrhage and exudate lesions types are given.

### 3.1.2 Evaluation of IDRiD Original Dataset

It is evident that the task has a background-foreground imbalance. The total percentage of lesion classes is less than three percent. As can be seen from Figure 3.2, MA is an underrepresented class. Its pixel area is about one-tenth of hemorrhage and hard exudate. This points to another imbalance problem between positive classes.

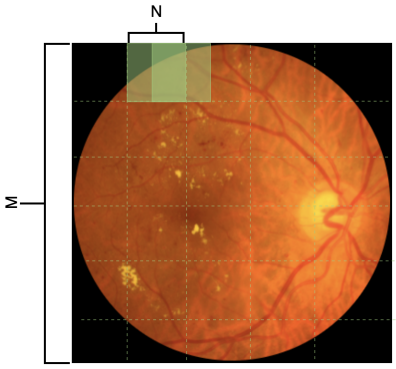
Figure 3.1 gives a piece of information about the visual characteristic of lesions. In the following part of the thesis, every shape with a closed area will be named as an instance. MA has the minimum average pixel area per instance. MA and EX lesions are generally small and scattered lesions. While the average pixel area of the EX instance is small, its total lesion area in the image, 0.9 percent, is significant. So, exudate lesions are composed of many small and scattered instances. There is one more important piece of information about the HE lesion from this table. Its average pixel area per instance is larger than the MA and EX. So, it can be called a relatively large-sized lesion. On the other hand, the distribution of pixel area per HE instance varies in a large range from 78-pixel to 8000-pixel size. This variation poses challenges to the detection of HE.

### 3.1.3 Patch Creation

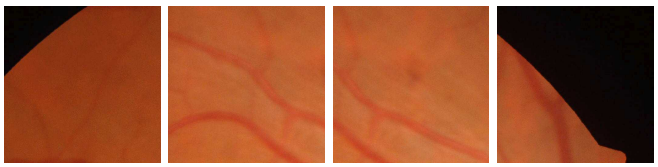
Original image size $4288 \times 2848$ of the dataset is too large to train a neural network. In the literature, resizing the original image size to a smaller size or patching[34, 30, 28, 35] is the solution for this problem. Some methods resize the the original images to smaller size such as $512 \times 512$ or $256 \times 256$ and fed into neural network[2, 1]. The advantage of the resized image is that it has global information. However, information about the very small size lesions is lost when resizing an image. For example, the pixel area of the MA lesions varies between 13 and 2051. After resizing the operation, smaller lesions are lost. On the other side, several detection algorithms working on high-resolution medical data processed the medical image patch by patch. Deep learning applications from different areas, including computational pathology, computer tomography, and retinal imaging, contains an example of this [36, 37]. The disadvantage of the patches is the lack of global information. However, it preserves

the details and enlarges the image number in the dataset. As a result, patching is chosen. The patch size has been determined considering the following. While size is getting smaller, valuable structures are divided into small parts on every patch. This situation ends up with a difficult learning process. Limitation about the larger size comes from the fact that training time and parameters of the model was increasing gradually with the patch size. Also, GPU ram size is another limitation in front of the larger size. There is a tradeoff between patch size and batch size because GPU ram size is constant.

Patches are created with size $N \times N$ by sliding window with stride size $N/2$ over the original image. $N$ is chosen as 512. A few examples of patches are visualized in Figure 3.2.




(a) Patch creation by window[6]



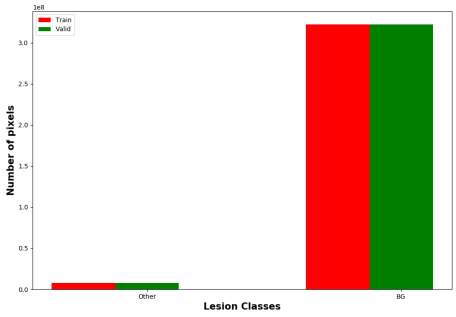
(b) Sample of patches

Figure 3.2: Patch creation from whole fundus image

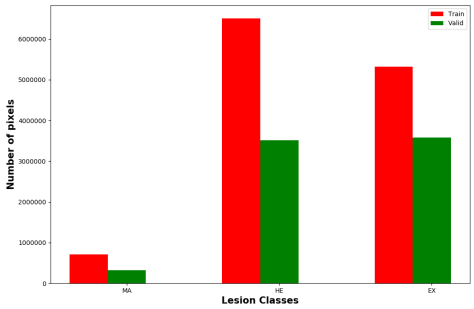
### 3.1.4 Evaluation of Patched Dataset

After the patch creation 11016 image is obtained for training and 5508 image is obtained for testing. Pixel-wise distribution of the positive lesion classes and background class is examined under train set. Following Table 3.3a shows the distribution in a bar graph. To understand the graph, distribution of the positive lesion classes was plotted seperately. Table 3.3b shows the distribution of the positive classes: MA, HE, and EX. Besides, patch distribution is examined in Figure 3.3c in addition to pixel wise distribution. It is shown that, number of patches where all pixels belong to the

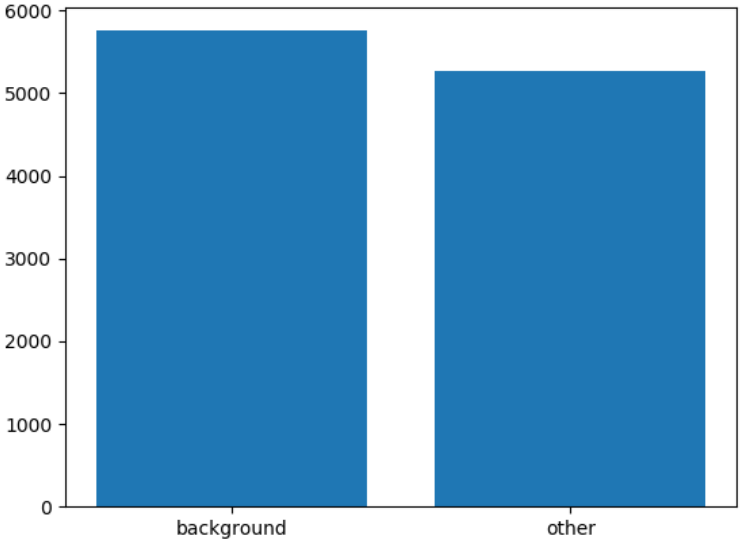
background and the number of remaining patches which has at least one pixel belongs to any class except for background are almost equal.




(a) Patch creation by window




(b) Patch creation by window




(c) Patch distribution of train set shows whether a patch includes any positive class or not. Background label means that all the pixel that belongs to the patch is background pixel. The other label represents patches with at least one pixel belonging to any positive class except the background.

Figure 3.3: Evaluation of patch class distribution. (a) shows the pixel-wise distribution of background vs. positive classes. (b) examines the pixel-wise distribution of positive classes among themselves. (c) shows the distribution of patches by examining the class of each patch, not by the number of pixels

## 3.2 Objective Functions

As neural networks are a class of optimization problems, they have cost functions. In the training phase, the optimizer algorithm seeks the point where cost is minimum. Loss functions for neural networks are used to calculate the cost, typically a single scalar value. There are several functions to calculate the loss of neural networks for given labels and predictions. Although it is possible to implement custom loss functions, only the widely used and well-known loss functions are focused on in this section. Since the loss itself depends on the labels, the loss function should be selected according to the type of labels. The labels can be continuous real values or discrete class categories. For example, in the case of continuous valued labels, the task is a regression problem.

If the task is a regression problem, mean squared error, mean squared logarithmic error, or mean absolute error losses can be selected as loss functions. When the nature of the problem is binary classification, i.e., either it is one class or the other class, the following loss functions are suitable to use: binary cross-entropy Loss or hinge loss. Often there are problems where outputs should have been mapped to more than two classes. In such cases, multi-class classification losses should be used. These are multi-class cross-entropy loss, weighted cross-entropy loss, and dice loss functions.

### 3.2.1 Mean Squared Error

Mean squared error is calculated by averaging the squared differences between the predicted and actual values. For a data point $x_i$, $y_i$ is the true value and $\hat{y}_i$ is its predicted value. The MSE is defined as:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.1}$$

Where $y_i$ is the true value, $\hat{y}_i$ is the predicted value and $n$ is the total number of data points in the dataset. It simply measures average magnitude of error.

### 3.2.2 Mean Squared Logarithmic Error Loss

When the desired output values are in broad ranges or scales, mean square error could be dominated by one or more large errors. Mean Squared Logarithmic Error Loss mitigates this problem by using logarithm function. Firstly it takes logarithm of error, then calculate MSE using this logarithmic error.

$$\frac{1}{n}\sum_{i=1}^{n}(\log(y_i) - \log(\hat{y}_i))^2 \tag{3.2}$$

Where $y_i$ is the true value, $\hat{y}_i$ is the predicted value and $n$ is the total number of data points in the dataset.

### 3.2.3 Mean Absolute Error Loss

This loss function is more robust to outliers since it does not amplifies huge errors to their squared values. It is calculated by averaging the absolute difference between the actual and predicted values.

$$\frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n} \tag{3.3}$$

Where $y_i$ is the true value, $\hat{y}_i$ is the predicted value and $n$ is the total number of data points in the dataset.

### 3.2.4 Binary Cross-Entropy Loss

Binary cross entropy can be used where labels are either 1 or 0. The function make use of cross entropy value and make it negative in order to comply with minimization task. The mathematical formula is the following.

$$-\frac{1}{n}\sum_{i=1}^{n}(y_i\log(p) + (1-y_i)\log(1-p)) \tag{3.4}$$

Where $y_i$ is the true label and the value of model's prediction output for the input $x_i$ is defined as $\hat{y}_i$. $n$ is the output size. For a binary classification case $y_i$ is either 1 or 0, which means in the above formula either $y_i$ or the $(1-y_i)$ becomes zero.

### 3.2.5 Hinge Loss

Hinge loss can be used where labels are either -1 or 1. The main idea behind the function is to penalise sign difference between the actual and predicted class values.

$$\frac{1}{n}\sum_{i=1}^{n} max(0, 1 - y_i \cdot \hat{y}_i) \tag{3.5}$$

Where $y_i$ is the true value, $\hat{y}_i$ is the predicted value.

### 3.2.6 Multi-Class Cross-Entropy Loss

It is the generalised version of the binary cross entropy loss for variable class numbers. The equation for the function is:

$$-\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{3.6}$$

Where $y_i$ is the true value,$\hat{y}_i$ is the predicted value and $n$ is the total number of classes. This formula intended to use with one hot encoding scheme for classes.

### 3.2.7 Weighted Cross-Entropy Loss

Although the function is derived from the cross entropy loss function, this function have distinction on class imbalance problems. It handles contribution of infrequent classes according to their respective weight. Therefore if the weights are selected properly, the function lessen the impact of class imbalances. The equation for the function is:

$$-\sum_{i=1}^{n} w_i y_i \log(\hat{y}_i) \tag{3.7}$$

Where $y_i$ is the true value, $\hat{y}_i$ is the predicted value and $n$ is the total number of classes. The $w_i$ which is not explicitly set in the cross entropy loss, is the weight of the respective class.

### 3.2.8 Dice Loss

Dice loss is also used in cases where class imbalance is an issue such as medical image segmentation. Dice loss is obtained by using dice coefficient and altering it's formula to serve as a loss function. The formula of the function is:

$$1 - \frac{2(Y \cap \hat{Y})}{Y + \hat{Y}} \tag{3.8}$$

Where $Y$ is a ground truth set, and $\hat{Y}$ is a prediction set.

# CHAPTER 4

# PROPOSED METHOD

## 4.1 Motivation

In this chapter we are going to give the details of the proposed algorithm to classify multiple retinal lesions simultaneosly. In this algorithm we are solving following three problems namely:

- Background-foreground imbalance.
- Class imbalance between positive classes. Pixel count of the microaneurysm class over dataset is almost 10th one of the hemorrhage and exudate class.
- Improve the detection of small lesions while preserving the detection rate of other lesions.

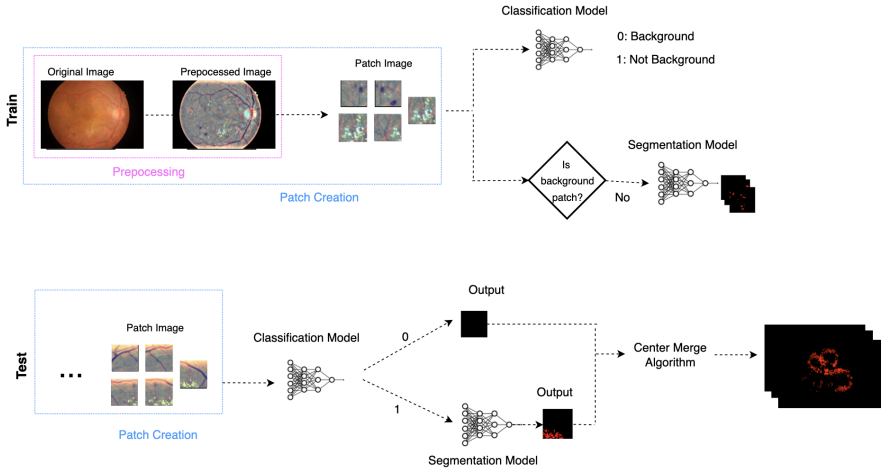



Figure 4.1: Flow of proposed method.

Two stage algorithm is proposed the decrease the effect of background-foreground imbalance. As described in Chapter 2, number of patches that were purely background and the others were almost equal in half. Moreover, even patches with lesion structures have imbalanced background-foreground distribution. Therefore, not all patches were used in the training of segmentation model. Patches, including positive classes, are used to feed the segmentation neural network. However, this segmentation model is blind to edges around the eye shape and optic disc. This kind of sharp edges may result in false positive prediction. Therefore, two stage algorithm is used similar to [38, 35]. Two stage algorithm is consist of two model for patch classification and lesion segmentation. Patch classification model classifies the patches whether patch includes any lesions structure or not. Segmentation model segments the multi-class lesions simultaneously.

A new objective function called instance-based class averaged IoU is introduced. In this loss function, the IoU score is calculated for every instance and averaged. Background imbalance is solved using the IoU score because it counts only positive classes. Each classes IoU score is calculated separately and averaging. Since we calculate the IoU score on an instance basis, lesions with small areas are also considered and included in the calculation.

In addition, it is aimed to increase the segmentation performance with the addition of pseudo labels of blood vessels. Lesions close to blood vessels in the distance and lesions similar in color to the blood vessel, such as MA and HE, are more likely to be confused with blood vessels. However, no dataset includes both lesion labels and blood vessel labels. Therefore, the pseudo label of blood vessel is used to remove blood vessels from images in the IDRiD dataset. Then, a neural network is trained for segmentation of multi-class lesions using blood vessel removed images.

## 4.2 Simultaneous Segmentation of Retinal Lesions using Pseudo Labeling

The flow of the algorithm is given in Figure 4.1. Original images are preprocessed. As a preprocessing, modified version of the algorithm used by Van[21] is used. Then, patches are created by sliding window over the whole preprocessed images. Patches,

including positive classes, are used to feed the segmentation neural network. Segmentation network is trained using instance-based class-averaged IoU loss function. Another neural network is trained for binary classification problem. It distinguishes patches if it includes any positive class or all pixels belonging to the background.

Test images are passed into the classification network. If it is classified as not a background patch, it is then passed into the segmentation network. Test patches are merged using the center merge algorithm, and a final test image is obtained. Center merge used the center of the patches to form final prediction image. In the second step of the proposed method, the same algorithm was tried with blood vessel removed images.

### 4.2.1 Preprocessing

In the literature, preprocessing methods used for analysis of retinal images included one or more steps of the followings:

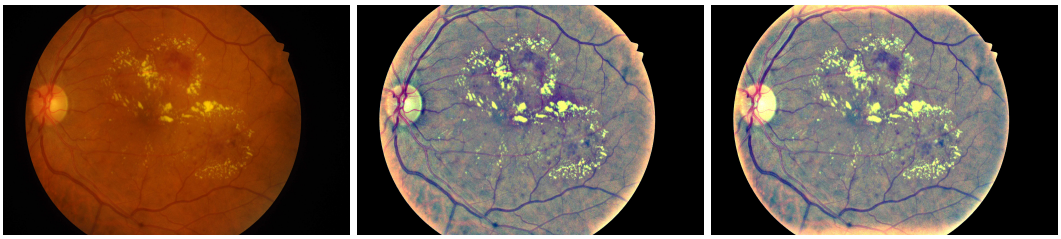
- Channel selection [26, 39]
- Contrast enhancement [26, 39, 24, 40]
- Background elimination [35, 24, 40]

Channel selection is generally preferred in problems where a single lesion or class is segmented because each lesion has maximum contrast with a different channel. In this research, the following preprocessing technique 4.1 is used with some modifications. The selected technique is used in several kinds of research with different forms [40, 21]. Same technique is used with parameters $\alpha = 4$, $\beta = -4$ and $\gamma = 128$. Let $I$ be the original pixel values of an image, and let $G$ be the Gaussian filtering function. Filtered image is obtained by the transformation $G$ applied to I. Filter size was selected as one 32th of the image size and each channel is filtered separately. Sample visualization of preprocessing method is shown in Figure 4.2.

When looking closely at the images, redundant darkening is seen around bright structures such as exudate and optic disc. Since patches do not have global information,

these darkening is easily confused with hemorrhage lesion. Reason of this redundant darkening is that, there are some big flat bright yellow areas which are also related to pathology. Therefore they are not related to background, and should be discarded from background calculations. Otherwise, the areas between these bright parts become extremely darker at the output of the preprocessing algorithm due to high contrast. In order to prevent that a simple threshold applied to the original image, and extremum pixels are found. These pixels are not used as input to the Gaussian filter. The original image, preprocessed output image using 4.1, and modified preprocessed output image are shown in Figure 4.2.

$$\widetilde{\mathbf{I}} = \alpha * I + \beta * G(I) + \gamma \tag{4.1}$$



(a) Original image with pathologic tissues.

(b) Output of preprocessing algorithm given equation 4.1.

(c) Output of modified preprocessing algorithm.

Figure 4.2: Visualization of fundus image after preprocessing.

### 4.2.2 Patch Classification

In this part aim is to develop a model which classifies patches of the fundus images into two class:

- 0: Patch includes at least one pixel belong to any class but not background
- 1: All the pixels in the patch belong to the background class

. Top scored convolutional neural network models such as VGG16, EfficientNet-B1, and EfficientNet-B3 are trained for binary classification problem using transfer

learning method. Threshold was decided using the best point of ROC curve. Best model had reached 0.955 area under ROC curve score.

### 4.2.3 Segmentation Model

Aim of the segmentation model is to segment multi-class lesions using one neural network. Three type of lesions, microaneurysm, exudate and hemorrhage, are segmented. U-Net[17], DeepLab v3[41] and FCN[42] are the most common and successful neural networks used for retinal image segmentation problem[43, 31, 44]. Therefore, Fast FCN[5], fully convolutional network, is selected as model architecture. Model had three output nodes for each lesion type.

Labels are encoded in the following way. Output size is $B \times W \times H \times C$ where $B$ is the batch size, $W$ is the height, $H$ is the height and $C$ is the class number. Consider that batch size is one and we investigate output corresponds to single patch image. Then output dimension will be $W \times H \times C$. Every pixel $(i, j)$ has a label with size $L_{i,j} = 1 \times C$. Each column in the label $L_{i,j}$ represents one class. Let us assume that, a pixel belongs to exudate lesion. Label of that pixel will be $[0, 0, 1]$. If a pixel is annotated as belonging to several classes, then cell of label array will be 1 for those classes and 0 for others. This is not a common case for the problem but dataset includes this type of annotation. For example, one pixel is annotated as belonging to both exudate and hemorrhage. So, its label will be $[0, 1, 1]$.

Segmentation model is trained using instance-based class-averaged IoU(IB_IoU) loss function which will be described in the next section.

#### 4.2.3.1 Instance Based Intersection Over Union(IB_IoU) Loss Function

Segmentation of retinal lesions is an imbalanced classification problem. Several solutions have been proposed to unravel the imbalanced classification problem. These solutions can be investigated under three subsections: data sampling approach, objective function-based methods, and generative methods. In this research, objective function based solutions were investigated. This problem includes two types of im-

balance: foreground-background imbalance and imbalance between positive classes. The most widely used loss function for multi class classification is the categorical cross-entropy function. Dice loss, focal loss, and weighted categorical cross-entropy loss functions have been suggested to resolve the imbalance problem. Also, some evaluation metrics specially conceived for imbalance classification can be converted to a loos function. IoU is an evaluation metric used as a loss function. Let us assume that A and B are two sample sets, and their IoU can be calculated by division of their intersection $A \cap B$ to their union $A \cup B$ which can be shown in Formula 4.2. Formula 4.3 and 4.4 show two loss functions generated by using this IoU metric. Formula 4.4 is found to be more useful in this research because it maps the output to a larger scale.

$$\mathbf{IoU} = \frac{A \cap B}{A \cup B} \tag{4.2}$$

$$\mathbf{IoULoss1} = 1 - \frac{A \cap B}{A \cup B} \tag{4.3}$$

$$\mathbf{IoULoss2} = -\ln(\frac{A \cap B}{A \cup B}) \tag{4.4}$$

$$\mathbf{IoU} = \frac{\sum_{n=1}^{k} i_n}{\sum_{m=1}^{l} u_m} = \frac{(i_1 + i_2 + i_3 + ... + i_k)}{(u_1 + u_2 + u_3 + ... + u_l)} \tag{4.5}$$

$$\mathbf{IB\_IoU} = \frac{1}{l} * \sum_{m=1}^{l} \frac{\sum_{n=1}^{k} (i_n \mid i_n \subseteq u_m)}{u_m} \tag{4.6}$$

It takes a lot of experimentation to find the optimum weights with a brute force search in the class-weighted categorical cross-entropy function3.7. The aim is to come up with a loss which solves the both foreground-background imbalance and imbalance between positive classes without need for weight arrangement. As a result IB_IoU loss is introduced.

Let us define intersection set $I = (i_1, i_2, i_3, .., i_k)$ where $i$ defines any closed contoured instance in the intersection set. And union set is defined as $U = (u_1, u_2, u_3, .., u_l)$ where $u$ defines any closed contoured instance in the union set. IoU is found to be

formula 4.5. It is known that the average size of the lesions varies for each class, and the lesions belonging to the same class can be of different sizes. Given this information, we can say that any chosen $i_j$ can be negligibly smaller than any other chosen $i_k$ where $j \neq k$ from the set $I$. In this case, any chosen $i$ contributes insignificantly to the loss function defined in equation 4.5. A new IB_IoU loss function is introduced to overcome this problem.

Let us define class-averaged IoU score first. As defined in Algorithm 1, IoU score is calculated for every class. Final score is found by averaging IoU score of each class. This way, imbalance between classes overcomed. Value of final score is between 0 and 1. Loss is found by taking negative logarithm of the score.

In the second step, challenge is to improve the detection of small size lesions. Therefore, IB_IoU is defined. In this loss function IoU score of each class is found separately likewise class-averaged IoU score. As defined in formula 4.5, number of intersected pixels written in the numerator, and number of union pixels written in the denominator in classical IoU score calculation. In this loss function IoU score is calculated for each instance. Instance is refers to each closed contoured shape. Instances are found from union matrix. Connected component analysis is applied on a union matrix to find instances. After calculating IoU score of each instance for given class, class score is found by taking average of these scores. This way, score is kept between 0 and 1. Final score is found by taking average of all class scores. The flow of algorithm is explained in algorithm 2.

**Algorithm 1:** Class based IoU objective function. IoU is calculated for every class and average is taken. Loss is calculated by taking negative logarithm of average IoU score.

Prediction matrice $\widetilde{Y}$, ground truth matrice $Y$ are a matrices with dimension of $(BatchSize, Width, Height, ClassNumber)$ and $C = \{c_1, c_2, \ldots, c_i\}$ denotes class set where $i$ is the class number and $I$ is the total number of classes.

**Initialization:**

Let $TotalIoU = 0$

Let $smooth = 1e - 7$ be a negligible small number to avoid zero division

**for** *each,* $c \leq C$ *except* $c_i$ *belongs to background class* **do**

    // Calculate

    $intersection = \widetilde{Y}_c * Y_c$ where $\widetilde{Y}_c$ is the $c$ channel of the matrice $\widetilde{Y}$. And $\{*\}$ denotes elementwise matrix multiplication

    $union = \widetilde{Y}_c + Y_c - intersection$

    $IoU = \frac{(intersection+smooth)}{(union+smooth)}$

    $TotalIoU = TotalIoU + IoU$

**end**

**return** $-log\frac{TotalIoU}{(ClassNumber-1)}$

**Algorithm 2:** IB_IoU objective function. IoU of every closed contour shaped instance in the union set is calculated. Class scores are found by taken average of IoU scores of instances. Then, final score is average of the class scores. Loss is calculated as negative logarithm of final IoU score.

**Inputs:** Prediction matrice $\widetilde{Y}$, ground truth matrice $Y$ with dimension of $(BatchSize, Width, Height, ClassNumber)$.
$C = \{c_1, c_2, \ldots, c_i\}$ is a class set where $i$ is the class number.
**Initialization:**
Let $TotalIoU = 0$
Let $smooth = 1e-7$ be a negligible small number to avoid zero division
**for** *each,* $c \in C$ *except* $c_i$ *belongs to background class* **do**
    **Initialization:**
    Let $IoU_{class} = 0$ as total class iou score
    Let $Num = 0$ as total number of instance for one batch
    **for** *each,* $b$ *in* $BatchSize$ **do**
        **Definition:**
        Let $\mathbf{M_I}$ and $\mathbf{M_U}$ as 2D intersection and union matrice
        $M_I = \widetilde{Y}_c * Y_c$
        $M_U = \widetilde{Y}_c + Y_c - M_I$
        Apply: Connected component analysis to $M_U$ to get instances.
        **for** *each, instance in instances* **do**
            **Definition:**
            Let $\mathbf{M_{mask}}$ as 2D mask matrice which has pixel value 1 for the corresponding instance's pixel positions and 0 for remaining pixel positions.
            $IoU_{class} = IoU_{class} + \frac{M_I * M_{mask}}{M_U * M_{mask}}$
        **end**
        $Num = Num + instances$
    **end**
    $TotalIoU = TotalIoU + \frac{IoU_{class}}{Num}$
**end**
**return** $-log\frac{TotalIoU + smooth}{ClassNumber - 1}$

Let us define ground truth label is a channel last matrix $Y_{B\times W\times H\times C}$ where $B$ is a batch size. Prediction label is defined as $\widetilde{Y}_{B\times W\times H\times C}$. Intersection matrix $I_{B\times W\times H\times C}$ is obtained by pixel-wise matrix multiplication of $Y$ and $\widetilde{Y}$. Union matrix $U_{B\times W\times H\times C}$ is obtained by summing $Y$, $\widetilde{Y}$ and negative of $I$ matrices. For the easy understanding of the algorithm, let us define batch size as one and apply the algorithm on a selected channel. So, $U_c$ and $I_c$ are union and intersection matrices representing channel $c$ with dimension of $(W \times H)$. Instances are found by applying connected component analysis on a union matrix $U_c$. After that, following IoU score is calculated for each instance. A mask matrix $M_{W\times H}$ which has the same size with $U_c$ and $I_c$ is created. This is a fixed shaped matrix. However, its active cells are changed for each instance. For each instance, cells of mask matrix M corresponding to that instance set to 1. Other cells set to 0. Union and intersection matrices, $\{U_c, I_c\}$, are pixel-wise multiplied by the mask matrix $M$. Then classical IoU score is calculated for that instance. Score is calculated for each instance in the union matrix $U_c$. Finally, average of the scores are returned as score of the class $c$. Same process applied for each channel and average is returned as resultant score. Resultant score is limited between 0 and 1. If the ground truth matrix $Y$ and prediction matrix $\widetilde{Y}$ are matched perfectly, resultant score will be 1. Loss value is calculated by taking negative logarithm of the resultant score.

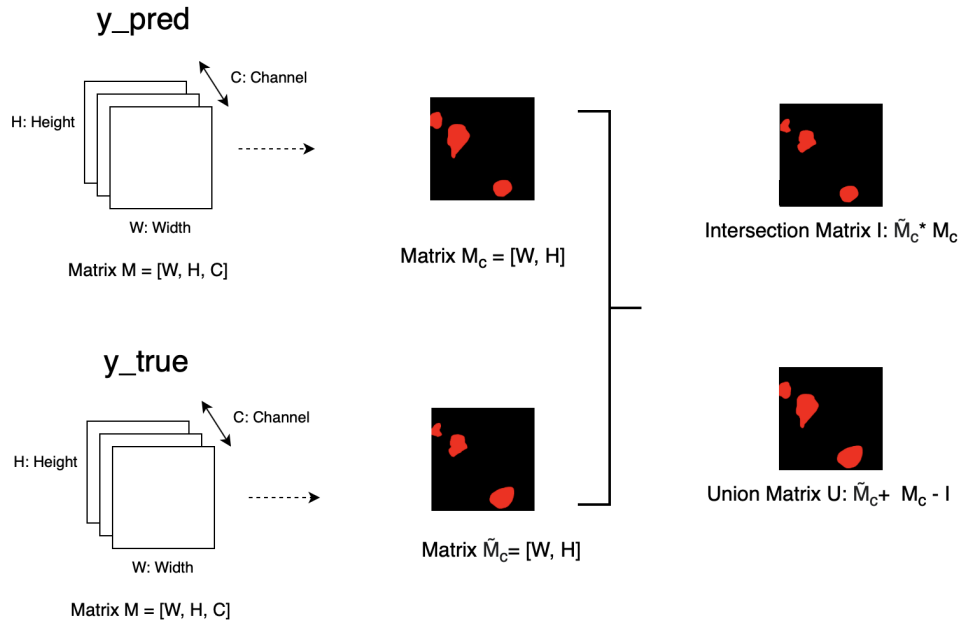
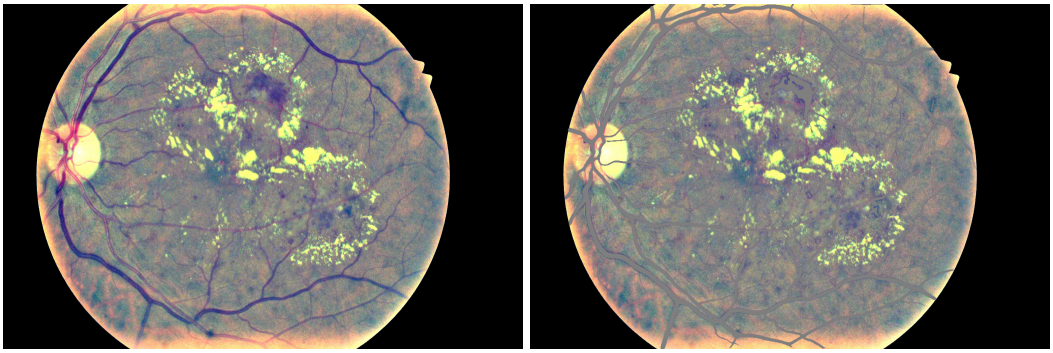



Figure 4.3: Flow of IB_IoU objective function. Prediction and ground truth matrices are shown as $y_{pred}$ and $y_{true}$. Then, a channel is selected and obtained 2D matrix is shown. Finally, Intersection and union matrices are obtained using element wise matrix multiplication and matrix summation.

### 4.2.4 Removing Blood Vessel from Images

Pseudo label of the blood vessels for the images in IDRiD dataset is obtained by the way described in section 4.2.4.1. Pseudo labels were used as masks and pixels that appeared as vessels were removed from the image. By removing, these pixels values are changed with new value. Assigning zeros to these pixels, that is, black, created a great contrast with the rest of the image. Therefore, these places are assigned gray color value $(128, 128, 128)$, which is the same color as the background of the pre-processed image. Figure 4.4 shows the blood vessel removed images.



(a) Pre-processed image. (b) Blood vessel removed image.

Figure 4.4: Visualization of fundus image after removing blood vessels using pseudo label.

#### 4.2.4.1 Obtaining Pseudo Label for Blood Vessel

Blood vessel segmentation algorithms can be investigated under three subsections called machine learning based methods, deformable models and tracking algorithms [45]. This work will focus on deep learning methods. There are several blood vessel segmentation datasets published publicly. Digital Retinal Images for Vessel Extraction database(DRIVE), STructured Analysis of the Retina database(STARE) and Digital Retinal Images for Vessel Extraction(CHASE_DB1) are the most common ones used as benchmarks. DRIVE has 20 images for training and 20 images for test[46]. STARE[47] has a total of 20 images and no division for train and test. CHASE_DB1[48] has 28 images annotated by two different experts.

Table 4.1: Dataset for blood vessel segmentation

| Dataset | Number of images | Resolution |
|---|---|---|
| DRIVE | 40 | 584x565 |
| STARE | 20 | 700x605 |
| CHASE_DB1 | 28 | 999x960 |

U-Net [17] outperforms on DRIVE and STARE datasets with 0.9855 and 0.9898 AUC in 2015 when it was first introduced. As can be seen from the Table 4.2,

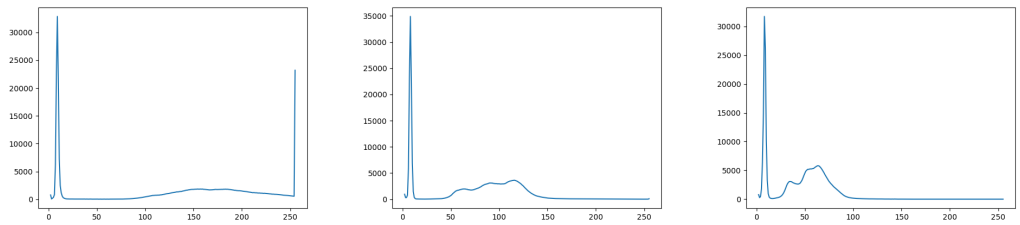
Table 4.2: Comparison of deep learning based blood vessel segmentation methods

| Dataset | Methods | Sensitivity | Specificity | Dice | AUC | Mean-IOU | Accuracy |
|---|---|---|---|---|---|---|---|
| DRIVE | R2U-Net[49], Alom, 2018 | 0.7792 | 0.9813 | 0.8171 | 0.9784 | - | 0.9556 |
| | IterNet[50], Li, 2019 | 0.7791 | 0.9831 | 0.8218 | 0.9813 | - | 0.9574 |
| | SA-UNet[51], Guo, 2020 | 0.8212 | 0.9840 | 0.8263 | 0.9864 | - | 0.9698 |
| | Study Group Learning[52], Zhou, 2021 | 0.8380 | 0.9834 | 0.8316 | 0.9886 | - | 0.9705 |
| | RV-GAN[53], Kamran, 2021 | 0.7927 | 0.9969 | - | 0.9887 | 0.9762 | 0.9790 |
| STARE | R2U-Net, Alom, 2018 | - | - | - | - | - | - |
| | IterNet, Li, 2019 | 0.7715 | 0.9886 | - | 0.9881 | - | 0.9701 |
| | SA-UNet, Guo, 2020 | - | - | - | - | - | - |
| | Study Group Learning, Zhou, 2021 | - | - | - | - | - | - |
| | RV-GAN, Kamran, 2021 | 0.8356 | 0.9864 | - | 0.9887 | 0.9754 | 0.9754 |
| CHASE_DB1 | R2U-Net, Alom, 2018 | 0.7756 | 0.9712 | 0.7928 | 0.9815 | - | 0.9634 |
| | IterNet, Li, 2019 | 0.7969 | 0.9881 | 0.8072 | 0.9899 | - | 0.9760 |
| | SA-UNet, Guo, 2020 | 0.8573 | 0.9835 | 0.8153 | 0.9905 | - | 0.9755 |
| | Study Group Learning, Zhou, 2021 | 0.8690 | 0.9843 | 0.8271 | 0.9920 | - | 0.9771 |
| | RV-GAN, Kamran, 2021 | 0.8199 | 0.9806 | - | 0.9914 | 0.9705 | 0.9697 |

various applications based on U-Net have been published and shown top rate area under ROC scores in later years. Additionally, in 2021 generative adversarial network based method RV-GAN is proposed and shown state-of-the art performance on DRIVE dataset. Their approach is to use two generative networks to predict coarse and fine vessels.

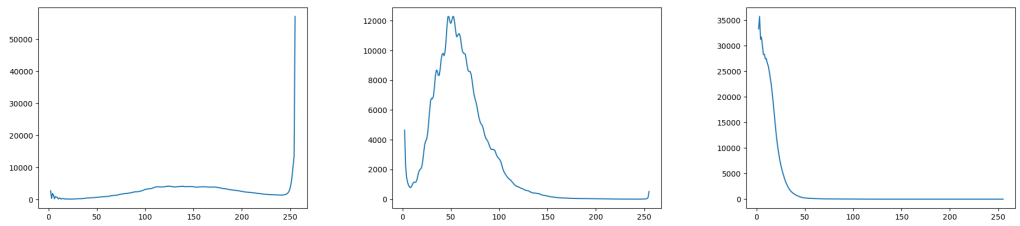
There are some important challenges to be aware of in obtaining pseudo-labels. The resolution of the images in each dataset is different. Secondly, all the mentioned networks are trained and tested on specific datasets. Stability is an important point as the prediction will be made for an unseen set that has never been trained on. The chosen method was decided by considering these difficulties and the performances of the networks. RV-GAN showed state-of-art performance however adversarial networks still have stability issues. Additionally, the method shows little improvement over the previous best result using a complicated algorithm. Since stability is important for clinical usage and our problem, this method was found unsuitable. On the other hand, another top-performing method Study Group Learning focuses on noisy data and its results are very close to RV-GAN on the DRIVE dataset and better on the CHASE_DB1 dataset. As a result, the Study Group Learning method is chosen to obtain a pseudo-label on the IDRiD dataset. To examine performance of the method, cross testing is done. By cross-testing, a model trained on the CHASE_DB1 dataset

was evaluated for the success of the other datasets on both the training and test partitions. With all of these, the histogram distributions of both datasets were examined and observed that there was a difference in RGB channel distributions. Histogram distribution of datasets can be found in Figure 4.5 and Figure 4.6. Since fundus images has a lot of zero pixels coming from the circular mask around the retina image, another histogram distribution was generated by removing zero pixels. Pixel value of the black mask is observed to be 0-10. So, values from 0 to 10 are removed from the histogram calculation. Final histogram graphs can be found in Figure 4.7 and Figure 4.8.



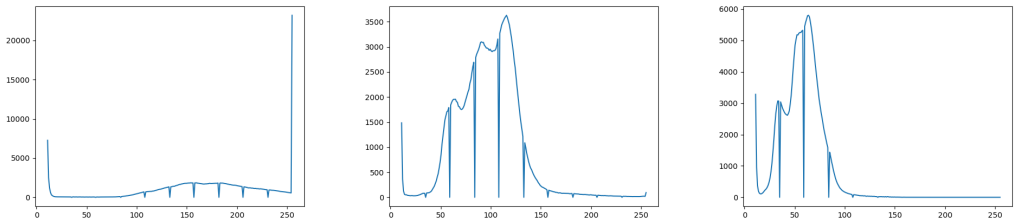
(a) Red channel (b) Green channel (c) Blue channel

Figure 4.5: Histogram of channel distribution of DRIVE dataset.



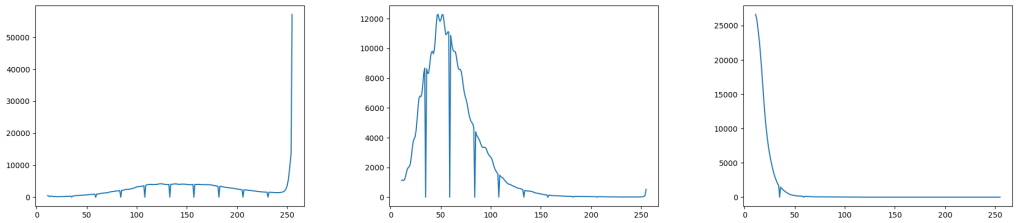
(a) Red channel (b) Green channel (c) Blue channel

Figure 4.6: Histogram of channel distribution of CHASE_DB1 dataset.



(a) Red channel (b) Green channel (c) Blue channel

Figure 4.7: Histogram of channel distribution of DRIVE dataset after removing black pixels.



(a) Red channel (b) Green channel (c) Blue channel

Figure 4.8: Histogram of channel distribution of CHASE_DB1 dataset after removing black pixels.

As a result, histogram matching is applied in order to investigate its effect on accuracy. The images in the dataset to be tested were preprocessed by histogram matching using the histogram of the training dataset. The improvement in test results after histogram matching was approximately three percent. Consequently histogram matching algorithm is applied on IDRiD dataset before obtaining the pseudo-label.

Final and important point about obtaining pseudo label for blood vessel is the fact that image resolution of the different datasets. As can be seen from the Table 4.1, resolution of the datasets published for blood vessel segmentation are smaller than the resolution of the IDRiD dataset. Images in the IDRiD dataset are downsampled to the $712 \times 1072$ by multiplier k=0.25. Then, these images are fed into blood vessel segmentation network pre-trained on CHASE_DB1. Output binary images had $712 \times$

1072 resolution. Output is upsampled to $2848 \times 4288$ by copying every pixel 4 times. Flow of the algortihm and example images of the upsampling is given in Figure 4.9.

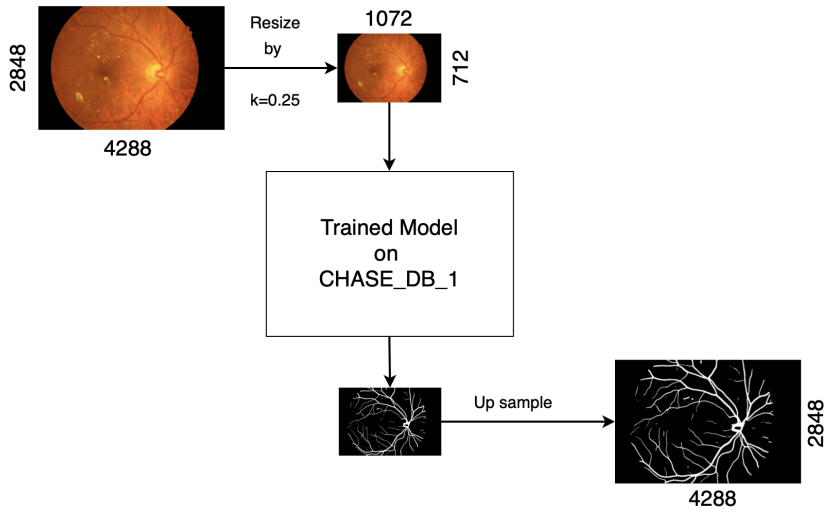



Figure 4.9: Retinal image with abnormalities(black) and natural structures(white).

# CHAPTER 5

# EXPERIMENTAL RESULTS

## 5.1 Hardware Specifications

All experiments, including training and testing the model, were performed on PC with NVIDIA GeForce RTX 2080 TI with 11 GB Memory Graphic Card and Intel Core i9-9900KS CPU. The training time of the segmentation model per one epoch was about 40 minutes. The inference time of the segmentation model for each patch was about 0.21 sec and the inference time of the two-staged model, including patch classification and segmentation model, per patch was about 0.3 sec. The average inference time of the whole image was about 12 sec.

## 5.2 Evaluation Metrics

The following metrics are used to evaluate the results.

**Area Under Precision-Recall Curve(AUPR):**

The area under the precision-recall curve is used as an official evaluation metric in the Segmentation and Grading Challenge workshop held at ISBI, 2018. 33 equally divided point between [0,1] or [0,255] is decided as threshold[31]. Precision and recall pairs are calculated for each threshold. Finally, a curve is drawn using precision, recall pairs. The best threshold is found by calculating the F-1 score of each pair. Pairs that give the best F-1 score are chosen as a threshold.

**Sensitivity,Specificity, F-1 Score:**

Sensitivity is a ratio that shows how many of the diseased pixels the model correctly predicts. Specificity is a ratio that shows how many of the healthy pixels the model correctly predicts. Precision is the percentage of pixels that the model predicts as diseased are correct. Finally, the F-1 score is the harmonic mean of precision and recall.

Sensitivity, specificity, and F-1 score were calculated using the best threshold from the precision-recall curve. Their formula are given in from equation 5.1 to 5.4. TP, FP, TN, and FN represents true positive, false positive, true negative, and false negative respectively.

$$\textbf{Precision} = \frac{TP}{TP + FP} \tag{5.1}$$

$$\textbf{Sensitivity} = \textbf{Recall} = \frac{TP}{TP + FN} \tag{5.2}$$

$$\textbf{Specificity} = \frac{TN}{TN + FP} \tag{5.3}$$

$$\textbf{F1} = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{5.4}$$

**Area Under ROC Curve:**

Prediction value is thresholded to obtain an accuracy score in the binary classification problem. There are plenty of ways to obtain this threshold value. However, the accuracy score or some other performance metrics are changed according to the threshold value. Therefore, the ROC curve is proposed to solve this problem. ROC curve is plotted using multiple threshold values. X-axes represent the false positive rate found by 1-specificity. Y-axes represent the sensitivity. To obtain ROC curve, the threshold is swapped from 0 to 1. Moreover, false positive rate and sensitivity score is calculated for each threshold. Finally, the area under the ROC curve is calculated.

This evaluation metric is used to compare the performance of patch classification models in this research.

## 5.3 Performance of Patch Classification

The total number of 11.016 patches were obtained from the original training set of the IDRiD dataset. These images are divided 20 percent for validation and 80 percent for training. The test set of the IDRiD dataset results in the number of 5.508 image patches. Class distribution of the dataset is given in Figure 3.3c. The model was trained using the transfer learning method, which means it used pre-trained imagenet weights.

The following parameters were searched for each model architecture, batch size, and learning rate. The batch size was searched for $\{16, 8, 4\}$. The limitation of the maximum batch size is the GPU memory. The learning rate was swapped from $10^{-3}$ to $10^{-5}$. Each combination of these parameters was tried for each model. Models were compared using test accuracy and area under the ROC curve. Results of the trials can be found in Table 5.1, Table 5.2, and Table 5.3 for VGG16[3], EfficientNet-B1[4], and EfficientNet-B3[4] models.

Table 5.1: Results of VGG16[3] architecture with different configurations.

| Model | Batch Size | Learning Rate | Accuracy | Area Under ROC |
|---|---|---|---|---|
| VGG16 | 16 | $10^{-3}$ | 0.8331 | 0.8703 |
| VGG16 | 16 | $10^{-4}$ | **0.8423** | **0.8994** |
| VGG16 | 16 | $10^{-5}$ | 0.8101 | 0.8710 |
| VGG16 | 8 | $10^{-3}$ | 0.7935 | 0.8667 |
| VGG16 | 8 | $10^{-4}$ | 0.7748 | 0.8635 |
| VGG16 | 8 | $10^{-5}$ | 0.7759 | 0.8546 |
| VGG16 | 4 | $10^{-3}$ | 0.7846 | 0.8690 |
| VGG16 | 4 | $10^{-4}$ | 0.7789 | 0.8608 |
| VGG16 | 4 | $10^{-5}$ | 0.7902 | 0.8601 |

Table 5.2: Results of EfficientNet-B1[4] architecture with different configurations.

| Model | Batch Size | Learning Rate | Accuracy | Area Under ROC |
|---|---|---|---|---|
| EfficientNet-B1 | 16 | $10^{-3}$ | 0.861 | 0.950 |
| EfficientNet-B1 | 16 | $10^{-4}$ | 0.853 | 0.949 |
| EfficientNet-B1 | 16 | $10^{-5}$ | 0.853 | 0.947 |
| EfficientNet-B1 | 8 | $10^{-3}$ | 0.851 | 0.950 |
| EfficientNet-B1 | 8 | $10^{-4}$ | **0.861** | **0.951** |
| EfficientNet-B1 | 8 | $10^{-5}$ | 0.849 | 0.946 |
| EfficientNet-B1 | 4 | $10^{-3}$ | 0.860 | 0.946 |
| EfficientNet-B1 | 4 | $10^{-4}$ | 0.854 | 0.943 |
| EfficientNet-B1 | 4 | $10^{-5}$ | 0.852 | 0.948 |

Table 5.3: Results of EfficientNet-B3[4] architecture with different configurations.

| Model | Batch Size | Learning Rate | Accuracy | Area Under ROC |
|---|---|---|---|---|
| EfficientNet-B3 | 16 | $10^{-3}$ | **0.877** | 0.952 |
| EfficientNet-B3 | 16 | $10^{-4}$ | 0.876 | **0.955** |
| EfficientNet-B3 | 16 | $10^{-5}$ | 0.864 | 0.952 |
| EfficientNet-B3 | 8 | $10^{-3}$ | 0.873 | 0.952 |
| EfficientNet-B3 | 8 | $10^{-4}$ | 0.872 | 0.953 |
| EfficientNet-B3 | 8 | $10^{-5}$ | 0.873 | 0.954 |
| EfficientNet-B3 | 4 | $10^{-3}$ | 0.869 | 0.950 |
| EfficientNet-B3 | 4 | $10^{-4}$ | 0.875 | 0.954 |
| EfficientNet-B3 | 4 | $10^{-5}$ | 0.867 | 0.954 |

VGG16 is a neural network architecture which scores 92.7% top-5 test accuracy in the ImageNet challenge and is widely used as a feature extractor. Additionally, the model implemented on open source deep learning libraries such as Tensorflow, Pytorch, Keras, Caffe etc. Therefore, it is selected as base model. EfficientNet is proposed in 2019 and it is one of top performed models in ImageNet challenge with reasonable

number of parameters[4]. The best performance has been reached by EfficientNet-B3 model architecture. As given in Figure 5.3 the best model has a performance of 95.5 percent AUC score and 87.7 percent accuracy.

## 5.4 Segmentation Model

In this section data augmentation, implementation details of training, hyper parameters are explained. Center merge algorithm applied end of the prediction to composed of the whole image is mentioned. Finally, performance of model is explained.

### 5.4.1 Data Augmentation

Following augmentation techniques are used to avoid overfitting: horizontal flip, vertical flip, clockwise rotation and anticlockwise rotation. Augmentation is applied to images in the train set during training. No augmentation is applied to test set. Augmentation is applied to the images and the labels together. Example of the augmented images and related labels are shown in Figure 5.1

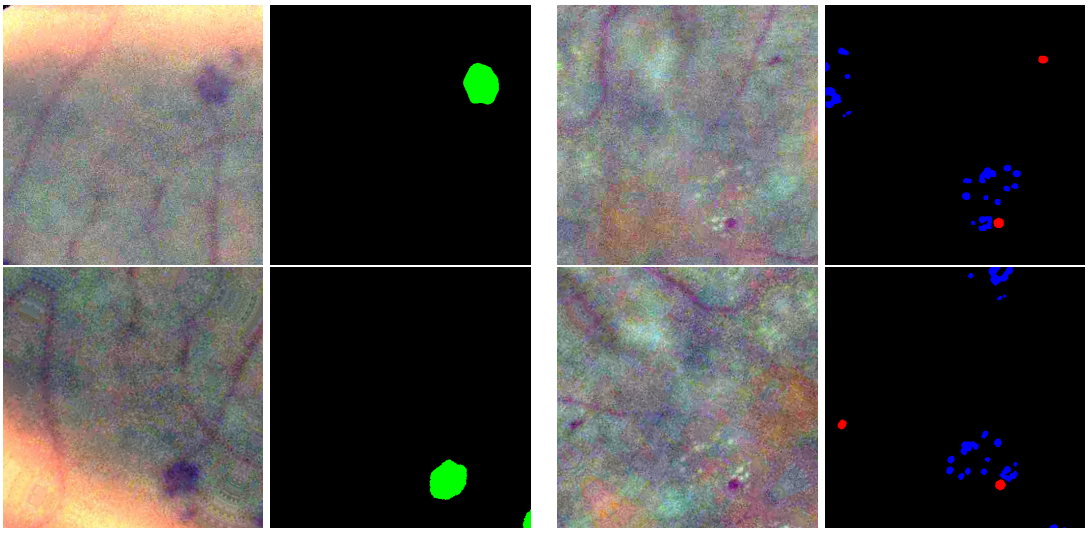


Figure 5.1: Augmented image patches and related labels from train set. Left and right quarter are the two different patch samples. First row show the original images patches and colorized binary label. The second row the shows the augmented version of above image and colorized binary labels.

### 5.4.2 Training and Implementation Details

FastFCN architecture is used as segmentation model. Adam optimizer is used[54]. Learning rate is selected as $10^{-4}$ and it is reduced to one-tenth until it reaches $10^{-9}$. It is reduced where the validation loss is flat for three epochs. Batch size, image size, epoch number are 2, 512x512 and 40 respectively. Overview of the FastFCN can be seen in figure 5.2

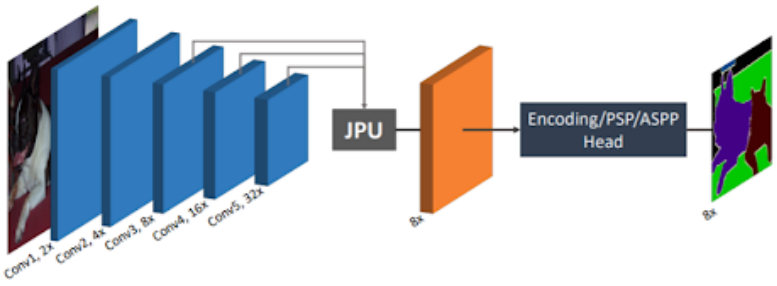



Figure 5.2: Framework Overview of FastFCN[5]

### 5.4.3 Center Merge Algorithm

Patches were created by sliding window over the whole image as described chapter 2. Test patches were needed to merged to create whole prediction image. According to stride size, consecutive patches have overlapping regions. Test patches either created with stride size equal to window size. Then, there would be no overlapping region. Another option is to merge overlapping areas similar to [55, 56]. In this way, only the center region of the patch is used to compose of whole prediction image. Center merge algorithm and the active region used by the algorithm is shown in Figure 5.3.

Detection capability of the model is weaker at the edge of the image because lesions can partially exist at the edges. Pixels at the center have spatially four adjacent pixels. Model can use those nearby pixel values while predicting the class of the center pixel. However, pixels near to the edge of the image have less adjacent pixel points compared to center pixels. So, the patch has a lack of information about pixels near the edge compared to the center pixels. Additionally, every lesion instance is closed

shaped. These closed shaped lesion instances are generally separated at the edge of the patch. As a result, patch includes only a part of these lesion structures and it does not have information of whole structure. To overcome this weakness of the model, only the center part of the patches was used.

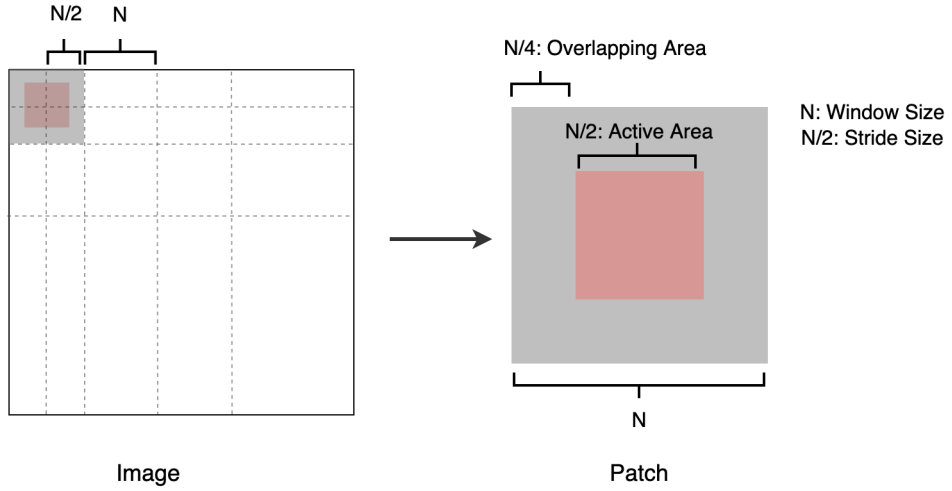



Figure 5.3: Center merge algorithm applied on patches to composed of whole image. Left image is the whole image and gray area is the patches. Red area centered over the gray area is the active apart used to composed of whole image.

### 5.4.4 Performance of Segmentation Model

Performance of the segmentation model is evaluated using area under precision recall curve. All metrics given in following tables are the area under precision recall curve. Result of the our segmentation models with different loss functions are shown in Table 5.4. Also, results are compared with the existing multi-class segmentation methods from the literature in the Table 5.4.

MC loss is the loss used by Guo in the L-Seg paper[1]. Our method is trained with our IB-IoU loss and the MC loss to compare the effenct of loss function to the performance. FastFCN is the model used in this study. FastFCN model trained with IB-IoU loss function showed better performance in MA and EX segmentation compared to same model trained with MC loss function. Its segmentation performance in HE lesions is lower. Additionally, proposed method's segmentation performance on

MA and EX classes have 2.36 and 0.46 percentage improvement over Guo's method. However, its performance in HE segmentation was lower compared to existing methods.

Both findings show that the proposed loss function improves performance in the MA and EX classes. However, there is no improvement in the HE lesion class. One goal of the proposed function was to improve the detection performance of small lesions. Looking at the results, there was an improvement in the MA and EX classes, that is, in the lesion classes consisting of small scattered fragments. However, its success is low in lesions that cover large areas such as HE. As a result, it can be said that the proposed loss function improves to the detection of small lesions.

Table 5.4: Result of segmentation models with different loss functions and methods from literature

| Method | MA | HE | EX |
|---|---|---|---|
| Proposed, FastFCN, IB-I0U | **0.4946** | 0.5133 | **0.7991** |
| Proposed, FastFCN, MC[1] | 0.4732 | 0.5510 | 0.7642 |
| Guo, MC[1] | 0.4710 | 0.5808 | 0.7410 |
| Guo, MCB | 0.4627 | **0.6374** | 0.7945 |
| Feng[2] | 0.2408 | 0.5649 | 0.6083 |

## 5.5 Performance of Two Stage Model Trained Using Instance Based IoU Loss Without Pseudo Label

Segmentation model and patch classification model are combined to increase the performance. Segmentation model is trained with patches that has lesions parts to avoid background imbalance. Because even those patches has large number of background pixels. However, segmentation model has weakness to correctly classify pixels located at the edges around the eye shape and prominent structures such as optic disc.

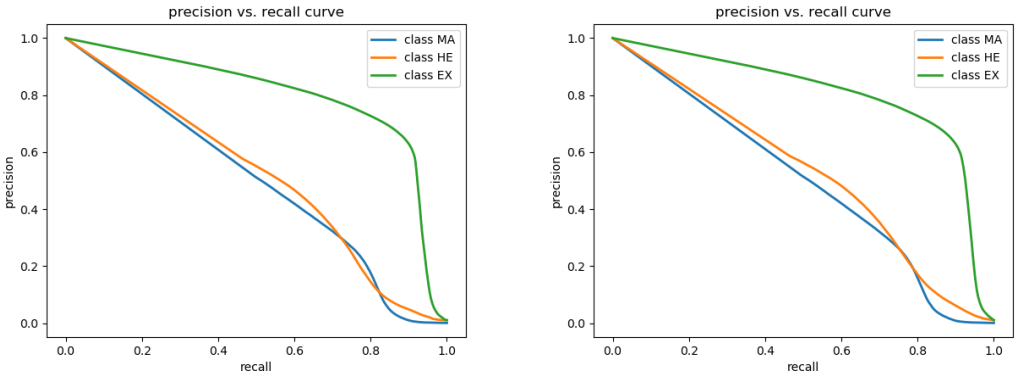
Therefore, patches are classified by the patch classification model and only the positive predictions are sent to segmentation model. Combination of the patch classifica-

tion and the segmentation model is called two staged model.

Performance of the two staged model is compared with segmentation model in Table 5.5. As can be seen from the Table, combination of two model made an small improvement. Improvement is 0.7 , 1.06 ,and 0.15 percentage for microaneurysm, hemorrhage and exudate classes. The improvement in hemorrhage is the most compared to the others, while the improvement in exudate is quite small. From this, it can be said that the two-stage model has reduced the number of false positives in the hemorrhage class. Overall, the two-stage model provided a small improvement by reducing false positive predictions, as we expected.

Table 5.5: Comparison of segmentation model and two stage model results

| Method | MA | HE | EX |
|---|---|---|---|
| Segmentation model | 0.4946 | 0.5133 | 0.7991 |
| Two-stage model | **0.5016** | **0.5239** | **0.8006** |




(a) Precision recall curve of segmentation model. (b) Precision recall curve of two staged model.

Figure 5.4: Precision recall curve of segmentation model and two staged model.

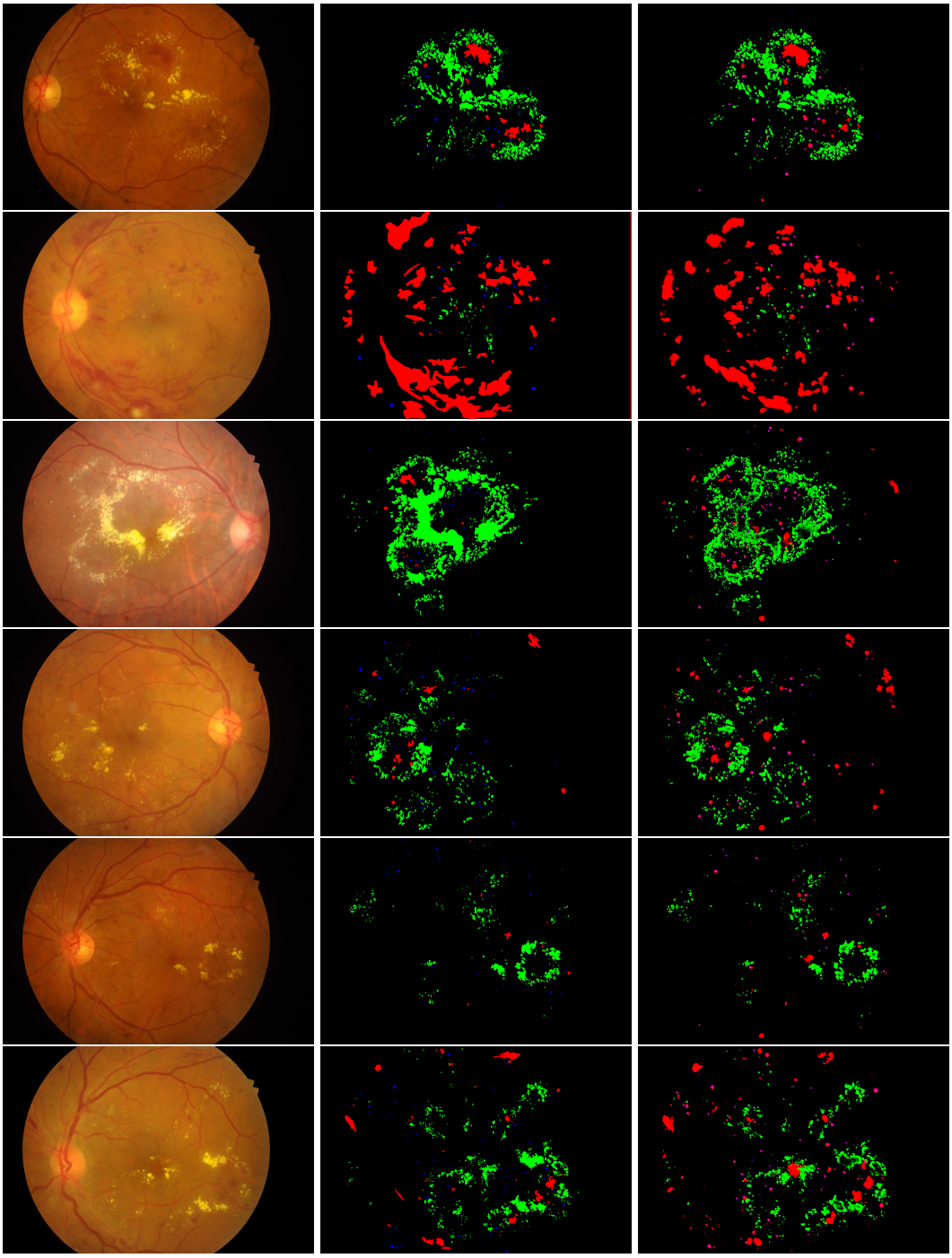


Figure 5.5: Visualization of ground truths and prediction results. From left to right, original fundus image[6], ground truth label and prediction result are shown. Ground truth and predictions are colorized using a color for each channel. Red, blue and green colors represent hemorrhage, microaneurysm and exudate lesions respectively.

Proposed two stage model is tested on healthy images to investigate performance.

Healthy images are selected from grading sub-challenge of IDRiD dataset. Grading sub-challenge has a disease level of images for diabetic retinopathy. Images with zero level, healthy, are selected from the test set. Total number of 35 images are used to test the performance of the method. Following Table 5.6 shows the pixel-wise prediction result of two staged model tested on healthy data. In addition to information given in the table, false predictions are investigated in terms of instance number. Model correctly predicts no lesion for 4 number of images. It falsely predicts 1, 2, and 3 lesion instance for 12, 10 and 1 number of images. False positive instances belongs to the hemorrhage and exudate classes.

Table 5.6: Pixel-wise result of two stage model on healthy data.

| Lesion Type | TN | FP |
|---|---|---|
| MA | 415,252,616 | 0 |
| HE | 415,212,101 | 3515 |
| EX | 415,214,985 | 631 |

## 5.6 Performance of Two Stage Model Trained Using Instance Based IoU Loss With Pseudo Labeling

In the second step of the proposed method, blood vessels were removed from pre-procesed images using pseudo labels. After the removal, patches are used to train the segmentation model. Results are compared in the table 5.7. Removing blood vessels do not contribute the performance of the model. It has insignificantly affected performance. Removing pseudo label of blood vessels decreased the performance by 0.03, 0.12, 0.11 percent.

Low quality of the pseudo labels could be a reason for this result. It is mentioned that, pseudo labels have redundant positive results. It marked some lesion pixels as blood vessel. We examined which lesion was most confused with false positive blood vessel labels. In the Figure 5.6, blood vessel prediction and the pixels confused with the lesions are shown. Total number of pixels predicted as blood vessel on the test is 23898096. Number of pixels predicted as blood vessel but belongs to lesions

according to ground truth is 2206279. Distribution of these confused pixels are 28262, 1168623, and 1009394 for microaneurysm, hemorrhage and exudate classes. Ratio of the confusion for each class is calculated by divison of number of confused pixels belongs to that class by number of total pixels predicted as blood vessel. Confusion ratio for MA, HE and EX are 0.001, 0.049 and 0.042 respectively. Important ratio of the confused pixels belongs to HE and EX lesions.

Table 5.7: Comparison result of two stage model with and without blood vessel removal

| Method | Pseudo Label | MA | HE | EX |
|---|---|---|---|---|
| Two staged | without pseudo label | 0.5016 | 0.5239 | 0.8006 |
| Two staged | with pseudo label | 0.5013 | 0.5227 | 0.7995 |

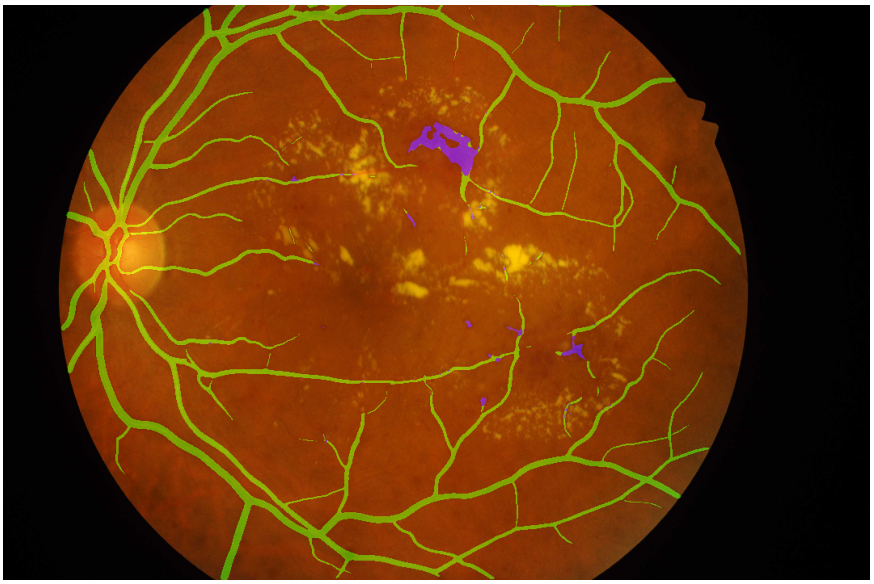
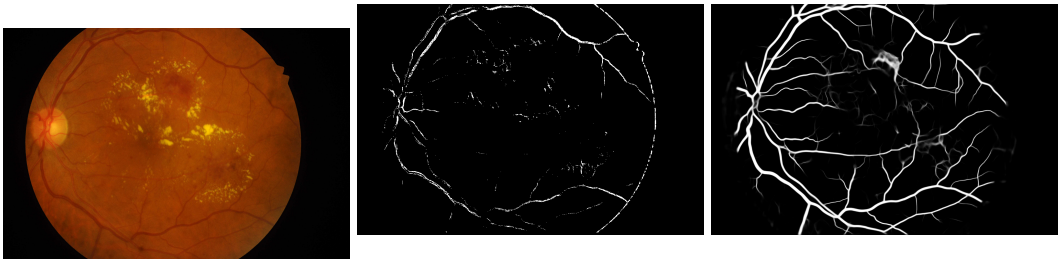


Figure 5.6: Retinal image with pseudo label of blood vessel. Green part show the areas predicted as blood vessel. Purple regions are predicted as blood vessel but belongs to the other lesion types according to ground truth labels.

### 5.6.1 Performance of Obtaining Pseudo Label for Blood Vessel

Two top performed methods tried to obtain pseudo label of blood vessels. RV-GAN proposed by Kamran and Study Group Learning proposed by Zhou are the methods. Original trained models published by the authors are used for both methods. RGB images from IDRiD dataset are given to the model to predict blood vessel output. Visual result of the both methods are shown in Figure 5.7. As can be shown from the figure, results obtained using Zhou's method is better than compared to Kamran's method. Result at the figure 5.7b has redundant extra labels around edge of the eyeshape and the around exudate lesions. Additionally, it missed the fine veins. Only the coarse vessels were founded. On the other hand, result at the figure 5.7c has captured vessels better. However, there is redundant extra labels around the hemorrhage lesions. This situation leads to removing lesions from pre-processed images. This situation would affect the performance of the segmentation model directly.

As mentioned in the Chapter 4.2.4.1, methods were cross tested on different datasets. Additionally, histogram matching is applied on test sets. Dice score of the model tested on different datasets is reported in Table 5.8. As expected, performance of the model dropped from 82.7 percent to around 73.9 percent when cross tested. And, histogram matching improves the performance as can be seen from the table. It made improvement on train and test set around 2.7 and 3.3 percent respectively. Therefore, histogram matching is applied to the images in the IDRiD dataset.



(a) Original image from test set of IDRiD[6] database.

(b) Gray level output of blood vessel segmentation model proposed by Kamran [53].

(c) Gray level output of blood vessel segmentation model proposed by Zhou [52].

Figure 5.7: Visualization of prediction results of blood vessel algorithms.

Table 5.8: Study Group Learning Blood Vessel Segmentation Algorithm Result Over Different Dataset

| Dataset: Model trained on | Dataset: Model Tested on | Result(Dice Score) |
|---|---|---|
| CHASE_DB1 | CHASE_DB1 | 0.827071 |
| CHASE_DB1 | DRIVE-Test | 0.739538 |
| CHASE_DB1 | DRIVE-Train | 0.723808 |
| CHASE_DB1 | DRIVE-Test$_{HistogramEqualized}$ | 0.772055 |
| CHASE_DB1 | DRIVE-Train$_{HistogramEqualized}$ | 0.750172 |

# CHAPTER 6

# CONCLUSION

In this study, we proposed a method to segment multi-class lesions simultaneously from fundus images. The method included the following steps: creating patches from the fundus image, classifying the patches according to whether they contain information about the lesion and segmentation of multi-class lesions from the patches. Under the patch segmentation, IB-IoU loss function and using the pseudo label of blood vessels is proposed.

This study aims to solve the class imbalance problem and increase the detection performance of small lesions. The proposed loss function has improved over other loss functions in microaneurysm and exudate lesions, which are characteristically small and scattered. At the same time, it has improved according to multi-class segmentation algorithms in the literature. However, it has not been successful in lesions that spread over large areas, such as hemorrhage. In addition, when the proposed method is compared to methods that distinguish a single lesion type, its performance did not surpass them but reached a comparable point. It did this without using a specialized network for each lesion and, accordingly, with a lower inference time for the test image.

In the second part of the method, pseudo-labels of blood vessels obtained for the IDRiD dataset were deleted from the images. The aim here is to increase the success of lesions similar in color to vessels and close to it in the distance. However, the result was not positive. The reason is that this result is tightly dependent on the quality of the pseudo label.

As future work, combination of two losses which are MC and IB_IoU will be tried.

It is observed that IB_IoU improves the performance of small lesions but does not perform well in the larger lesions such as HE. Combining the two loss functions can enable both small and larger lesions to be considered. Additionally, proposed loss function was compared with MC losses using FastFCN and has made improvement on performance. In order to show that this improvement provided by the proposed loss function is a general result, the proposed loss function can be compared on other models from the literature.

# REFERENCES


[1] S. Guo, T. Li, H. Kang, N. Li, Y. Zhang, and K. Wang, "L-seg: An end-to-end unified framework for multi-lesion segmentation of fundus images," *Neurocomputing*, vol. 349, pp. 52–63, 2019.

[2] S. Feng, W. Zhu, H. Zhao, F. Shi, Z. Li, and X. Chen, "Encoder-decoder attention network for lesion segmentation of diabetic retinopathy," in *International Workshop on Ophthalmic Medical Image Analysis*, pp. 139–147, Springer, 2019.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[4] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, pp. 6105–6114, PMLR, 2019.

[5] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation," *arXiv preprint arXiv:1903.11816*, 2019.

[6] 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," *Data*, vol. 3, no. 3, p. 25, 2018.

[7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

[8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

[9] G. Roglic *et al.*, "Who global report on diabetes: A summary," *International Journal of Noncommunicable Diseases*, vol. 1, no. 1, p. 3, 2016.

[10] L. M. Aiello, "Perspectives on diabetic retinopathy," *American journal of ophthalmology*, vol. 136, no. 1, pp. 122–135, 2003.

[11] C. Wilkinson, F. L. Ferris III, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, J. T. Verdaguer, *et al.*, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003.

[12] S. Resnikoff, W. Felch, T.-M. Gauthier, and B. Spivey, "The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200 000 practitioners," *British Journal of Ophthalmology*, vol. 96, no. 6, pp. 783–787, 2012.

[13] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE transactions on medical imaging*, vol. 26, no. 10, pp. 1357–1365, 2007.

[14] M. Niemeijer, B. Van Ginneken, J. Staal, M. S. Suttorp-Schulten, and M. D. Abràmoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Transactions on medical imaging*, vol. 24, no. 5, pp. 584–592, 2005.

[15] S. Ravishankar, A. Jain, and A. Mittal, "Automated feature extraction for early detection of diabetic retinopathy in fundus images," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 210–217, IEEE, 2009.

[16] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic medicine*, vol. 19, no. 2, pp. 105–112, 2002.

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

[18] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 618–624, 2017.

[19] J. H. Tan, H. Fujita, S. Sivaprasad, S. V. Bhandary, A. K. Rao, K. C. Chua, and U. R. Acharya, "Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network," *Information sciences*, vol. 420, pp. 66–76, 2017.

[20] M. Haloi, "Improved microaneurysm detection using deep neural networks," *arXiv preprint arXiv:1505.04424*, 2015.

[21] M. J. 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.

[22] J. Shan and L. Li, "A deep learning method for microaneurysm detection in fundus images," in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 357–358, IEEE, 2016.

[23] E. Imani, H.-R. Pourreza, and T. Banaee, "Fully automated diabetic retinopathy screening using morphological component analysis," *Computerized medical imaging and graphics*, vol. 43, pp. 78–88, 2015.

[24] H. Wang, G. Yuan, X. Zhao, L. Peng, Z. Wang, Y. He, C. Qu, and Z. Peng, "Hard exudate detection based on deep model learned information and multi-feature joint representation for diabetic retinopathy screening," *Computer methods and programs in biomedicine*, vol. 191, p. 105398, 2020.

[25] A. Sopharak, M. N. Dailey, B. Uyyanonvara, S. Barman, T. Williamson, K. T. Nwe, and Y. A. Moe, "Machine learning approach to automatic exudate detection in retinal images from diabetic patients," *Journal of Modern optics*, vol. 57, no. 2, pp. 124–135, 2010.

[26] R. Saha, A. R. Chowdhury, and S. Banerjee, "Diabetic retinopathy related lesions detection and classification using machine learning technology," in *Inter-*

*national Conference on Artificial Intelligence and Soft Computing*, pp. 734–745, Springer, 2016.

[27] R. Srivastava, L. Duan, D. W. Wong, J. Liu, and T. Y. Wong, "Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels," *Computer methods and programs in biomedicine*, vol. 138, pp. 83–91, 2017.

[28] Z. Yan, X. Han, C. Wang, Y. Qiu, Z. Xiong, and S. Cui, "Learning mutually local-global u-nets for high-resolution retinal lesion segmentation in fundus images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 597–600, IEEE, 2019.

[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[30] M. Badar, M. Shahzad, and M. Fraz, "Simultaneous segmentation of multiple retinal pathologies using fully convolutional deep neural network," in *Annual Conference on Medical Image Understanding and Analysis*, pp. 313–324, Springer, 2018.

[31] P. Porwal, S. Pachade, M. Kokare, G. Deshmukh, J. Son, W. Bae, L. Liu, J. Wang, X. Liu, L. Gao, *et al.*, "Idrid: Diabetic retinopathy–segmentation and grading challenge," *Medical image analysis*, vol. 59, p. 101561, 2020.

[32] L. Zhang, S. Feng, G. Duan, Y. Li, and G. Liu, "Detection of microaneurysms in fundus images based on an attention mechanism," *Genes*, vol. 10, no. 10, p. 817, 2019.

[33] V. M. Kanukollu and S. S. Ahmad, "Retinal hemorrhage," *StatPearls [Internet]*, 2021.

[34] C. Lam, C. Yu, L. Huang, and D. Rubin, "Retinal lesion detection with deep learning using image patches," *Investigative ophthalmology & visual science*, vol. 59, no. 1, pp. 590–596, 2018.

[35] N. Eftekhari, H.-R. Pourreza, M. Masoudi, K. Ghiasi-Shirazi, and E. Saeedi, "Microaneurysm detection in fundus images using a two-step convolutional neural network," *Biomedical engineering online*, vol. 18, no. 1, pp. 1–16, 2019.

[36] G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. W. K. Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.

[37] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th international conference on control automation robotics & vision (ICARCV)*, pp. 844–848, IEEE, 2014.

[38] S. Long, J. Chen, A. Hu, H. Liu, Z. Chen, and D. Zheng, "Microaneurysms detection in color fundus images using machine learning based on directional local contrast," *Biomedical engineering online*, vol. 19, no. 1, pp. 1–23, 2020.

[39] A. Z. H. Ooi, Z. Embong, A. I. Abd Hamid, R. Zainon, S. L. Wang, T. F. Ng, R. A. Hamzah, S. S. Teoh, and H. Ibrahim, "Interactive blood vessel segmentation from retinal fundus image based on canny edge detector," *Sensors*, vol. 21, no. 19, p. 6380, 2021.

[40] B. Graham, "Kaggle diabetic retinopathy detection competition report," *University of Warwick*, 2015.

[41] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[42] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[43] T. Li, W. Bo, C. Hu, H. Kang, H. Liu, K. Wang, and H. Fu, "Applications of deep learning in fundus images: A review," *Medical Image Analysis*, p. 101971, 2021.

[44] P. Furtado, "Multi-class segmentation of diabetic retinopathy lesions: effects of metrics, improvements and loss," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1410–1417, IEEE, 2020.

[45] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics," *Computer methods and programs in biomedicine*, vol. 158, pp. 71–91, 2018.

[46] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. 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.

[47] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[48] C. G. Owen, A. R. Rudnicka, R. Mullen, S. A. Barman, D. Monekosso, P. H. Whincup, J. Ng, and C. Paterson, "Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program," *Investigative ophthalmology & visual science*, vol. 50, no. 5, pp. 2004–2010, 2009.

[49] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," 2018.

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

[51] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "Sa-unet: Spatial attention u-net for retinal vessel segmentation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 1236–1242, IEEE, 2021.

[52] Y. Zhou, H. Yu, and H. Shi, "Study group learning: Improving retinal vessel segmentation trained with noisy labels," in *International Conference on Medi-*

*cal Image Computing and Computer-Assisted Intervention*, pp. 57–67, Springer, 2021.

[53] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, K. M. Sanders, and S. A. Baker, "Rv-gan: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 34–44, Springer, 2021.

[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[55] V. Iglovikov, S. Mushinskiy, and V. Osin, "Satellite imagery feature detection using deep convolutional neural network: A kaggle competition," *arXiv preprint arXiv:1706.06169*, 2017.

[56] Z. Feng, J. Yang, and L. Yao, "Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 1742–1746, IEEE, 2017.

[57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

# APPENDIX A

# CONVOLUTIONAL NEURAL NETWORKS

## A.1 VGG16

VGG16[3] is a neural network architecture which scores 92.7% top-5 test accuracy in the ImageNet challenge. The architecture became very popular after its succession in 2014. The model weights trained on that challenge are published and the model implemented on open source deep learning libraries such as Tensorflow, Pytorch, Keras, Caffe etc. Hence it is one of the most available models. Although there are deeper and more successful CNN models in the following years, VGG16 is still widely used for image classification tasks, especially as a baseline feature extractor. The architecture consists of basic CNN layers such as convolution, max pooling and fully connected layers. The detailed layer composition can be seen in the following figure A.1. The

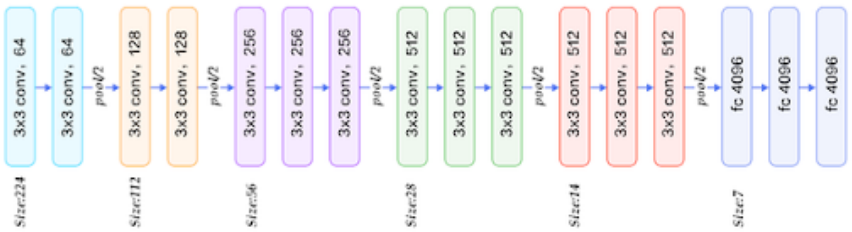



Figure A.1: Architecture of VGG16[3]

model with ImageNet weights is used as a feature extractor if top layers are omitted. Top layers can be modified for the requirements of any interested image classification task. It consists of 16 layers and 138 million parameters.

## A.2 EfficientNet

EfficientNet[4] offers a way to scale models while keeping the number of parameters low and accuracy gain high. As the model depth increases, models can extract more complex features from their inputs. However deeper models can suffer from vanishing gradients or they have too many parameters to train. Although ResNets[57] solve the vanishing gradient problem, deeper models saturate after some degree. On the other hand increasing the input resolution is another way of scaling which also saturates after some degree. Since higher resolution images require deeper networks to have the same receptive field for a filter responsible for a specific feature, the increase in resolution should be balanced with an increase in depth. Similarly an increase in width of the network should be balanced with an increase in resolution and an increase in depth for the optimum gain in return of complexity and the increase in number of parameters to train. The authors of EfficientNet developed the compound scaling method which offers a way to scale the models in a balanced way. As they experimentally showed that their scaling method indeed efficient in terms of accuracy vs. FLOPS (floating-point operations per second), they also provided a baseline model to scale up. Other CNN architectures can also be scaled by the compound scaling method. However their baseline model, namely EfficientNetB0 is optimised as to be a good starting point for scaling. It can be seen from the following figure that EfficientNetB0 has similar accuracy with other networks with just less than half the FLOPS they have. By using EfficientNetB0 as a baseline they scaled their networks several times and created a family of networks called EfficientNets. There are EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6 and EfficientNetB7 networks in the family.

The architecture of EfficientNetB0 can be seen in the figure which consists of mobile inverted bottleneck convolution MBConv combined with squeeze and excitation.

To further explain this main block, original mobile inverted bottleneck convolution can be described. In a residual block as in the ResNet the inputs shrink down to smaller channel size then outputs are expanded back to the same channel size again at the end of the block. Residual connection is established between bigger channel sized inputs and outputs. On the contrary, in an inverted residual block inputs are expanded

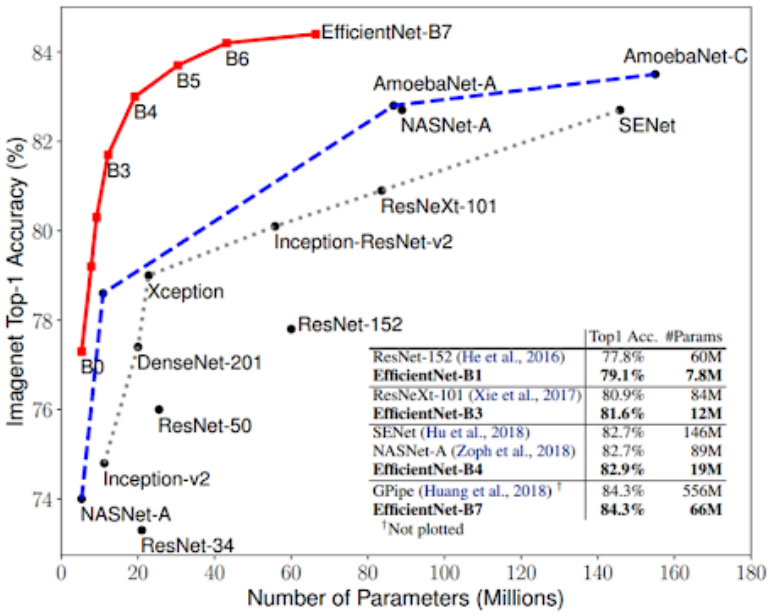



Figure A.2: ImageNet top-1 accuracy vs model parameters graph[4]

to higher channel sizes and outputs are shrink to original small size. Therefore residual connection is established between small channel sized inputs and outputs which effectively act as bottlenecks. The figure A.4 compares the two types of residual connections.

Another aspect of this block is that mentioned bottlenecks at the outputs are linear bottlenecks which means they do not have any nonlinear activation function. They experimentally show that the nonlinear activation on the output layer decreases performance as it may cause information loss.

Authors of EfficientNet combined the mobile inverted bottleneck convolution with squeeze and extraction optimization which first developed with SENets[8]. The building blocks of SENets are SE blocks. SE block basically rescales each channel of output with a respective weight. The rescaling weights are also calculated from the output itself using the squeezing function F_sq. F_sq can be global average pooling which embeds global information in a scalar value for every channel. These computed values are fed into a fully connected bottleneck block to produce excitation weights. This process is shown as F_ex in the figure below. Therefore SE blocks allow layers to use the information not only their receptive field but also the global information.

| Stage $i$ | Operator $\hat{\mathcal{F}}_i$ | Resolution $\hat{H}_i \times \hat{W}_i$ | #Channels $\hat{C}_i$ | #Layers $\hat{L}_i$ |
|---|---|---|---|---|
| 1 | Conv3x3 | $224 \times 224$ | 32 | 1 |
| 2 | MBConv1, k3x3 | $112 \times 112$ | 16 | 1 |
| 3 | MBConv6, k3x3 | $112 \times 112$ | 24 | 2 |
| 4 | MBConv6, k5x5 | $56 \times 56$ | 40 | 2 |
| 5 | MBConv6, k3x3 | $28 \times 28$ | 80 | 3 |
| 6 | MBConv6, k5x5 | $14 \times 14$ | 112 | 3 |
| 7 | MBConv6, k5x5 | $14 \times 14$ | 192 | 4 |
| 8 | MBConv6, k3x3 | $7 \times 7$ | 320 | 1 |
| 9 | Conv1x1 & Pooling & FC | $7 \times 7$ | 1280 | 1 |

Figure A.3: The architecture of EfficientNetB0.

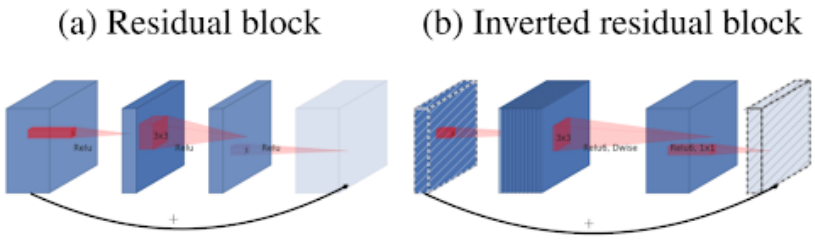



Figure A.4: Two types of residual connections[7]

The figure A.5 shows the mentioned functions on a SE block.

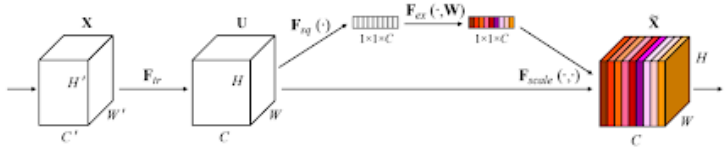



Fig. 1. A Squeeze-and-Excitation block.

Figure A.5: Squeeze and excitation block[8]

Following figure A.6 shows how any type of residual block can be combined with a SE block.

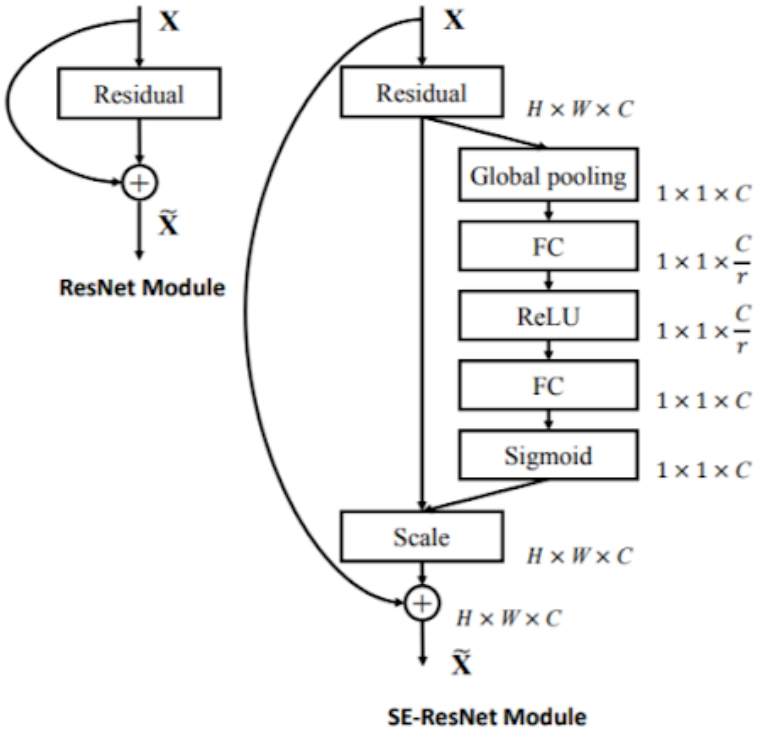



Figure A.6: Residual block can be combined with a squeeze and excitation block[8]

### A.3 FastFCN

Among the semantic segmentation models Fully Convolution Network (FCN)[42] is the most basic model as it just consists of regular convolutional blocks. However the output resolution is much lower than the input image due to pooling operations. Such problems are mitigated by using encoder decoder networks. Encoder is the basically same thing as the FCN while decoder does the reverse function. Decoder uses all spatial information from encoding layers to generate a full resolution output image. Recently, Dilated FCN was developed to avoid complex decoder models. Dilated FCN takes advantage of dilated convolutions to keep the resolution unchanged while increasing the receptive field. The last layers of FCN are replaced with dilated convolutions in order to generate high resolution images. Although resolution is dropped due to previous convolutional blocks, it still has better resolution than FCN. The mentioned models can be seen on figure A.7. However, the FastFCN[5] model removes dilated convolutional blocks as they are computationally costly. Instead it proposes

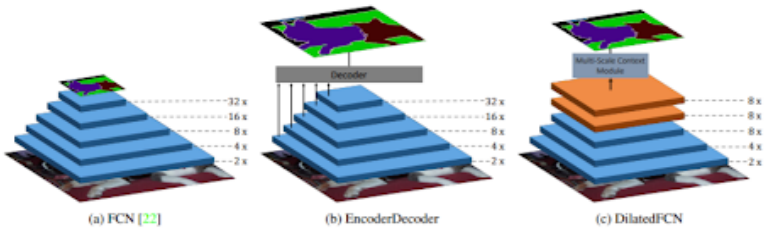



Figure A.7: From left to right: original FCN, encoder-decoder style and dilated convolutions to obtain high-resolution final feature maps[5].

a structure like encoder decoder network, which makes use of the last three previous layer outputs as input to a special decoder named Joint Pyramid Upsampling JPU. JPU combines several upsampling steps and generate high resolution feature map. At the end the high resolution feature map converted into a label outputs using a head encoding module. The described structure is shown on figure A.8.

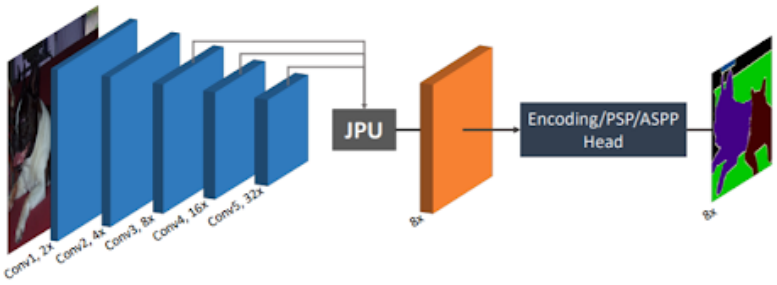



Figure A.8: Framework Overview of FastFCN[5]