

RETINAL VESSEL SEGMENTATION IN FUNDUS IMAGES  
USING DEEP LEARNING

CHEN CHUN HUI

FACULTY OF ENGINEERING  
UNIVERSITY OF MALAYA  
KUALA LUMPUR

2021

**RETINAL VESSEL SEGMENTATION IN FUNDUS  
IMAGES USING DEEP LEARNING**

**CHEN CHUN HUI**

**THESIS SUBMITTED IN FULFILMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF INSUSTRIAL  
ELECTRONICS AND CONTROL ENGINEERING**

**ENGINEERING FALCULTY  
UNIVERSITY OF MALAYA  
KUALA LUMPUR**

**2021**

**UNIVERSITY OF MALAYA**  
**ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: **CHEN CHUNHUI**

Matric No: **KQC190001/17202431**

Name of Degree: **Master Degree of Industrial Electronics and Control Engineering**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

**Retinal Vessel Segmentation in Fundus Images Using Deep Learning**

Field of Study:

**Artificial Intelligence and Deep Learning**

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature

Date: 2021-06-07

Subscribed and solemnly declared before,

Witness's Signature

Date: 10/06/2021

Name:

Designation:

# **RETINAL VESSEL SEGMENTATION IN FUNDUS IMAGES USING DEEP LEARNING**

## **ABSTRACT**

Retinal vessel is the only microvascular system that can be viewed from digital fundus cameras directly and non-invasively, they are closely relative with human blood circulation, so the appearance and change of retinal blood vessels can reflect cardiovascular and cerebrovascular diseases, such as diabetic and hypertensions. In practice, observing retinal blood vessels has become a crucial step for ophthalmologist to make diagnosis and conduct timely treatment. However, retinal vessels in fundus images are difficult, tedious and time-consuming to recognize for ophthalmologist. Hence, computer-aided diagnosis was introduced to make automatic retinal vessels segmentation. Deep learning is used as it is a promising technique since its high efficiency and accuracy. In this project, we proposed a new method to conduct automatic retinal vessel segmentation. Firstly, we enhanced the quality of raw fundus images by using image processing technique, which the pre-processed images present a better quality than before. Secondly, we proposed a deep-learning based model to make predictions. Inspired by U-net and ensemble learning, our model is comprised of two cascaded U-shaped networks, and each of the sub-network is composed of CBR (Conv., BatchNormalization, ReLU) blocks. The second sub-network aims to fine-tune coarse vessel maps produced by the first sub-network, since it learns features from combination of coarse vessel map and raw input. To enlarge the receptive field, dilation convolution was adopted with whose dilated rates arranged deliberately to make dense sampling. In addition, residual learning was adopted to ease the optimization and sufficient skip connections were added between the two sub-networks to make full use of feature maps. Finally, three public databases were chosen to verify the proposed model and compared its performance with other recent publications. The model produced an accuracy of 0.9552/0.9699/0.9642, an

AU\_ROC of 0.9787/9852/9846, a sensitivity of 0.8211/0.8466/0.8395 on DRIVE, STARE and CHASE\_DB1 databases, respectively. Cross-validation was also conducted to evaluate the generalization capacity of the proposed model. In these intensive experiments, the proposed model can produce a good performance after training, so it can provide a good reference for ophthalmologist to perform diagnosis.

Keywords: Retinal vessel segmentation, fundus images, deep learning, convolutional neural network, computer-aided diagnosis.

Universiti Malaya

# SEGMENSI VESEL RETINAL DALAM GAMBAR FUNDUS MENGGUNAKAN PEMBELAJARAN DEEP

## ABSTRAK

Saluran retina adalah satu-satunya sistem mikrovaskular yang dapat dilihat dari kamera digital fundus secara langsung dan tidak invasif, ia sangat berkaitan dengan peredaran darah manusia, jadi penampilan dan perubahan saluran darah retina dapat mencerminkan penyakit kardiovaskular dan serebrovaskular, seperti diabetes dan hipertensi. Dalam praktiknya, memerhatikan saluran darah retina telah menjadi langkah penting bagi pakar oftalmologi untuk membuat diagnosis dan melakukan rawatan tepat pada masanya. Walau bagaimanapun, kapal retina dalam gambar fundus sukar, membosankan dan memakan masa untuk dikenali oleh pakar oftalmologi. Oleh itu, diagnosis berbantuan komputer diperkenalkan untuk membuat segmentasi kapal retina automatik. Pembelajaran mendalam digunakan kerana merupakan teknik yang menjanjikan kerana kecekapan dan ketepatannya yang tinggi. Dalam projek ini, kami mencadangkan kaedah baru untuk melakukan segmentasi kapal retina automatik. Pertama, kami meningkatkan kualiti gambar fundus mentah dengan menggunakan teknik pemprosesan gambar, yang mana gambar yang diproses sebelumnya mempunyai kualiti yang lebih baik daripada sebelumnya. Kedua, kami mencadangkan model berasaskan pembelajaran mendalam untuk membuat ramalan. Diilhamkan oleh pembelajaran U-net dan ensemble, model kami terdiri daripada dua rangkaian berbentuk U lata, dan masing-masing sub-rangkaian terdiri dari blok CBR (Konv., BatchNormalization, ReLU). Sub-rangkaian kedua bertujuan untuk menyempurnakan peta kapal kasar yang dihasilkan oleh sub-rangkaian pertama, kerana ia mempelajari ciri-ciri dari kombinasi peta kapal kasar dan input mentah. Untuk memperbesar bidang penerimaan, konvolusi pelebaran diadopsi dengan kadar dilatasi yang disusun dengan sengaja untuk membuat persampelan padat. Di samping itu, pembelajaran sisa diadopsi untuk memudahkan pengoptimuman dan sambungan langkau

yang memadai ditambahkan di antara dua sub-jaringan untuk memanfaatkan peta ciri sepenuhnya. Akhirnya, tiga pangkalan data awam dipilih untuk mengesahkan model yang dicadangkan dan membandingkan prestasinya dengan penerbitan lain yang baru-baru ini. Model menghasilkan ketepatan 0.9552/0.9699/0.9642, AU\_ROC 0.9787/9852/9846, kepekaan 0.8211/0.8466/0.8395 pada pangkalan data DRIVE, STARE dan CHASE\_DB1. Pengesahan silang juga dilakukan untuk menilai kapasiti penjanaan model yang dicadangkan. Dalam eksperimen intensif ini, model yang dicadangkan dapat menghasilkan prestasi yang baik setelah latihan, sehingga dapat memberikan rujukan yang baik bagi pakar oftalmologi untuk melakukan diagnosis.

Kata kunci: Segmentasi kapal retina, gambar fundus, pembelajaran mendalam, rangkaian saraf konvolusional, diagnosis berbantuan komputer.

## ACKNOWLEDGEMENTS

Firstly, I want to express sincere appreciation to my supervisor, Ir. Dr. Chuah Joon Huang. I acquired sufficient knowledge about machine learning and deep learning from his course, and this project is also completed under his warm guidance and encouragement. Working in the same lab, I want to express my special gratitude to my PhD senior, Raza Ali as he gave me much help in reading and writing paper. Thanks to Raza and my classmate, Sriramk, for helping me revise my paper several times.

Secondly, I would like to thank my lectures in master program, Prof. Dr. Saad Mekhilef (Dean of Faculty of Engineering), Assoc. Prof. Ir. Dr. Chow Chee Onn (Head of Electrical Department), Dr. Marizan Binti Mubin, Assoc. Prof. Ir. Dr. Jeevan Kanesan. They shared their professional knowledge unreservedly and gave me many warm suggestions. I also would like to extend my gratitude to all faculty members, they also gave me much help in master program.

Thirdly, I would like to thank my friends while studying in Malaysia, Tang Ying Feng, Feng Jia Hui, Zhang Zi Mu and Wang Tao. We helped each other to overcome obstacles in study and daily life. They also encouraged me a lot when I feel disappointed and frustrated.

Finally, I would like to express heartfelt thanks to my parents and sister. My journey in completing a Master program in Malaysia is a beautiful and unforgettable experience, and I cannot have pursued my master program without their support and caring.

## TABLE OF CONTENTS

Abstract .....	iii
Abstrak .....	v
Acknowledgements .....	vii
Table of Contents .....	viii
List of Figures .....	xi
List of Tables.....	xiii
List of Symbols and Abbreviations.....	xiv
List of Appendices .....	xv
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>16</b>
1.1 Background.....	16
1.2 Problem statement .....	18
1.3 Objectives of research.....	18
1.4 Scope of study.....	18
1.5 Thesis organization.....	19
<b>CHAPTER 2: LITERATURE REVIEW.....</b>	<b>20</b>
2.1 Introduction.....	20
2.2 Overview of deep learning.....	20
2.2.1 Convolutional neural networks (CNNs).....	21
2.2.1.1 CNN architecture components .....	22
2.2.1.2 Loss function .....	26
2.2.2 Fully convolutional networks (FCNs).....	27
2.2.3 U-net.....	27
2.3 Retinal vessel segmentation.....	28

2.4	Hardware and software .....	34
2.5	Summary .....	35
<b>CHAPTER 3: METHODS .....</b>		<b>36</b>
3.1	Introduction.....	36
3.2	Database.....	36
3.3	Evaluation metrics .....	38
3.4	Image pre-processing.....	39
3.4.1	Grayscale conversion .....	40
3.4.2	Normalization .....	41
3.4.3	Contrast limited adaptive histogram equalization .....	41
3.4.4	Gamma correction .....	42
3.5	Data augmentation.....	43
3.6	Model architecture .....	43
3.7	Loss and optimizer.....	46
3.8	Implementation details.....	47
3.9	Summary.....	47
<b>CHAPTER 4: RESULTS AND DISCUSSION .....</b>		<b>48</b>
4.1	Introduction.....	48
4.2	Training process.....	48
4.3	Overall performance .....	48
4.4	Comparison with the state-of-art .....	51
4.5	Limitation .....	56
4.6	Summary.....	58
<b>CHAPTER 5: CONCLUSION AND FEATURE WORK.....</b>		<b>60</b>

5.1 Conclusion .....	60
5.2 Future work.....	60
References .....	62
Appendix A Network implementation.....	69

Universiti Malaya

## LIST OF FIGURES

Figure 1.1: Examples of retinal fundus images with annotations. Left: image form DRIVE (Staal, Abràmoff, Niemeijer, Viergever, & Van Ginneken, 2004) dataset, annotated by us. Right: image from IDRiD (Porwal et al., 2018), annotated by T. Li et al. (2021). ...	17
Figure 2.1: Architecture of LeNet-5 (LeCun et al., 1998). C: convolution layer, S: sub-sampling layer, F: fully connected layer. ....	21
Figure 2.2: Architecture of AlexNet (Krizhevsky et al., 2012). ....	22
Figure 2.3: Convolution operation in convolutional layer. Stride=1 and assume bias=0. ....	23
Figure 2.4: Pooling operations, with 2*2 filter and stride=2. ....	26
Figure 2.5: Architecture of a FCN (Long et al., 2015). ....	27
Figure 2.6: Architecture of U-net (Ronneberger et al., 2015). ....	28
Figure 3.1: Image-label-mask show. ....	38
Figure 3.2: Images and their channel splits. From top to bottom: DRIVE, CHASE_DB1, STARE. ....	41
Figure 3.3: Contrast Limited Adaptive Histogram Equalization. ....	42
Figure 3.4: Result of image preprocessing technique. From top to bottom: DRIVE, CHASE_DB1, STARE. From left to right: green channel, normalization, CLAHE and gamma correction. ....	43
Figure 3.5: Architecture of segmentation model. ....	45
Figure 3.6 Architecture of building blocks. (a): CBR block, (b): residual-dilated CBR block. ....	45
Figure 3.7: Sampling of cascaded dilated convolution. d means dilation rate. ....	45
Figure 4.1: Training curves in public databases. From top to bottom: CHASE_DB1, DRIVE, STARE. From left to right: loss, accuracy, AU_ROC. ....	48
Figure 4.2: Retinal vessel segmentation examples for CHASE_DB1 database. (a): pre-processed images, (b): ground truth images, (c): segmentation results. ....	49
Figure 4.3: Retinal vessel segmentation examples for DRIVE database. (a): pre-processed images, (b): ground truth images, (c): segmentation results. ....	50

Figure 4.4: Retinal vessel segmentation examples for STARE database. (a): pre-processed images, (b): ground truth images, (c): segmentation results. .... 50

Figure 4.5: ROC curve and PR curve on testing databases. From top to bottom: CHASE\_DB1, DRIVE, STARE. From left to right: ROC curve, PR curve. .... 51

Figure 4.6 Pixel histogram of image samples from three databases. .... 54

Figure 4.7: Enlarged image patches of image in DRIVE database. From left to right: preprocessed image, enlarged image patches, ground truth image patches, segmentation result patches. .... 57

Figure 4.8: Enlarged image patches of image in STARE database. From left to right: preprocessed image, enlarged image patches, ground truth image patches, segmentation result patches. .... 58

Figure 4.9: Enlarged image patches of image in CHASE\_DB1 database. From left to right: preprocessed image, enlarged image patches, ground truth image patches, segmentation result patches. .... 58

Universiti Malaysia

## LIST OF TABLES

Table 2.1: Overview of papers for retinal vessel segmentation. ....	33
Table 3.1: Pixel measures in vessel segmentation. ....	39
Table 3.2: Evaluation metrics for image segmentation.....	39
Table 4.1: Confusion matrix of proposed model on three databases. ....	52
Table 4.2: Performance comparison on DRIVE database. ....	52
Table 4.3: Performance comparison on STARE database. ....	52
Table 4.4: Performance comparison on CHASE_DB1 database. ....	53
Table 4.5: Cross-validation on DRIVE database. ....	55
Table 4.6: Cross-validation on STARE database.....	55
Table 4.7: Cross-validation on CHASE_DB1 database.....	56

## LIST OF SYMBOLS AND ABBREVIATIONS

CNN	:	Convolutional Neural Network
FCN	:	Fully Convolutional Network
RVS	:	Retinal Vessel Segmentation
CAD	:	Computer-aided Diagnosis
FOV	:	Field of View
HE	:	Histogram Equalization
CLAHE	:	Contrast Limited Adaptive Histogram Equalization
CBR	:	Convolutional, Batch Normalization, Relu
TP	:	Truth Positive
FP	:	False Positive
TN	:	Truth Negative
FN	:	False Negative
GT	:	Ground Truth
Pre	:	Precision
Sen	:	Sensitivity
Spe	:	Specificity
Acc	:	Accuracy
PR_curve	:	Precision-Recall curve
ROC	:	Receiver Operating Characteristic Curve
AUC	:	Area Under ROC
DRIVE	:	Digital Retinal Images for Vessel Extraction
STARE	:	Structured Analysis of the Retina
CHASE_DB1	:	Child Heart and Health Study in British
HRF	:	High Resolution Fundus Image

## LIST OF APPENDICES

Appendix A: Implementation of network .....	69
---------------------------------------------	----

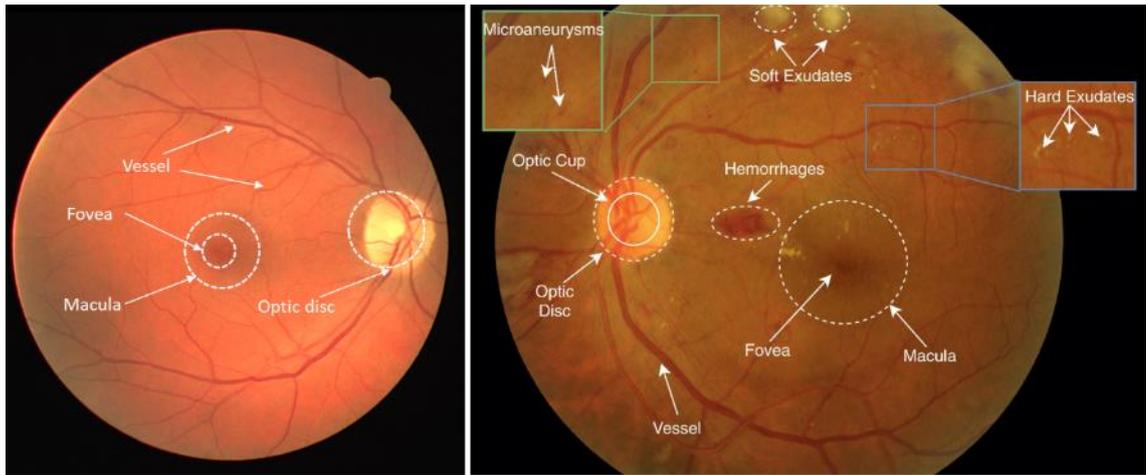
Universiti Malaya

## CHAPTER 1: INTRODUCTION

### 1.1 Background

Eye is one of the most important organs of human-beings since it provides sufficient environment information. Its health has a crucial effect on our life.

As the most important part of eyes, fundus retina is the only deeper microvascular system that can be observed non-invasively from the digital fundus camera. Retina locates in the inner layer of the eyeball and it is the most sensitive area of visual information. Retinal fundus images can be acquired by a digital fundus camera attached with a low-power microscope. Pupil of the human eye is used as entry/exit point for fundus camera illumination and imaging light beams on the retina. Retinal fundus images can also be acquired by using Scanning Laser Ophthalmoscopy (SLO) (Webb & Hughes, 1981). The fundus image of the retina illustrates its structure such as retinal blood vessel tree, optic disk (OD), fovea, macula, and abnormal structures such as microaneurysms (MAs), haemorrhages, exudates, cotton wool spots, if exist (Franklin & Rajan, 2014). Figure 1.1 presents 2 examples of retinal fundus images with structure annotations. Retinal blood vessel tree is composed of central retinal artery and vein, and their branches. Figure 1.1 shows the annotated structure of retinal in a retinal fundus image from HRF dataset (Köhler et al., 2013)



**Figure 1.1: Examples of retinal fundus images with annotations. Left: image from DRIVE (Staal, Abràmoff, Niemeijer, Viergever, & Van Ginneken, 2004) dataset, annotated by us. Right: image from IDRiD (Porwal et al., 2018), annotated by T. Li et al. (2021).**

Changes in the structure and morphology of retinal blood vessels are the most common form of fundus disease. Because the blood circulation of the retinal blood vessels is affected by the local tissues around it, many other cardiovascular and cerebrovascular diseases are often the cause of the structural variation of the retinal blood vessels in the human eye. For example, it is common that the lens of the fundus image is cloudy and the boundary is unclear for cataract patients, glaucoma patients may have symptoms such as optic atrophy and reduced visual field, atherosclerosis and cotton wool spots often exists in fundus images of hypertension patients, diabetic patients may have edema and hemorrhage in the retina, even new blood vessels in some severe cases.

Since geometric characteristics of retinal vessels such as vessel diameter, branch angles, and branch lengths reflect clinical and pathological features such as hypertension, diabetes, and atherosclerosis (Fathi & Naghsh-Nilchi, 2013; Group, 1991; Kanski & Bowling, 2011). Retinal image assessment has been an indispensable step for identification of retinal pathology. Precise identification and diagnosis of eye abnormalities and their timely medication are vital in preventing blindness.

## **1.2 Problem statement**

Geometric characteristics of retinal blood vessels such as width, curve, and length reveal important healthy status of patients such as diabetic retinopathy (DR), diabetic maculopathy (MD) and hypertension. Traditionally, retinal blood vessels in fundus images are segmented by ophthalmologists manually, which is tedious and time-consuming. At the same time, the segmentations of different experts maybe inconsistent since the complex image condition. Unsupervised methods such as matching filter methods, vascular tracing-based segmentation methods and model-based segmentation methods still need handcrafted features which limit their generalization capacity, and they do not produce satisfying performance since they cannot learn from existing ground truth.

## **1.3 Objectives of research**

The objectives of this research are:

1. To propose a model based on deep learning that can segment vessels in color fundus images with the help of image processing technique.
2. To evaluate the performance and accuracy of the proposed model in fundus images.

## **1.4 Scope of study**

The scope of this research is limited to the design of deep learning-based model and its implement, training and evaluation. To be specific, the model is based on convolutional neural networks (CNNs) and implemented using Python, it is cooperated with Google Tensorflow framework based on GPU. The analysis is based on recognition capacity for pixels in fundus images, several public databases are used in this project.

## **1.5 Thesis organization**

The thesis is organized as follows. Chapter 1 illustrates the background of this project, as well as the problem statement, project objectives and scope of research. Chapter 2 gives an overview of deep learning and a detailed literature review of retinal vessel segmentation. Chapter 3 indicates the research methods to conduct this task. Chapter 4 presents the test results of the proposed method and result discussion. Chapter 5 concludes the whole thesis and points out possible future works for this project.

Universiti Malaya

## CHAPTER 2: LITERATURE REVIEW

### 2.1 Introduction

This chapter presents an overview of deep learning, including learning types, basic terms, the history of deep learning and some prevalent models. In addition, this chapter also illustrates others researches about retinal vessel segmentation. It also introduces some prevalent software and hardware to implement the task.

### 2.2 Overview of deep learning

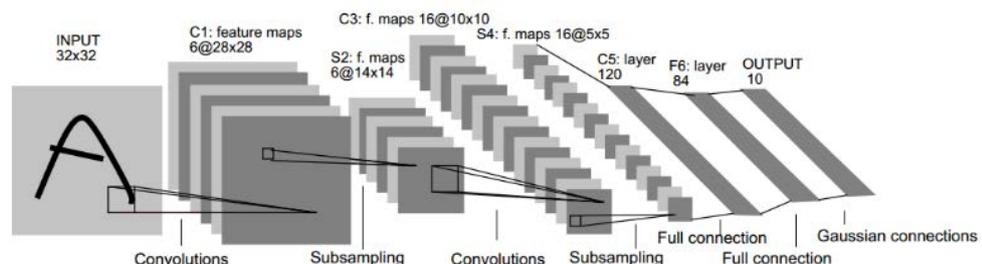
Deep learning is a sub-field of machine learning, which mainly uses hierarchical structured neural layers to translate input information into a meaningful output. Deep learning has been developed a rich family since 1990 (Schmidhuber, 2015), such as deep neural networks (DNNs) (Schmidhuber, 2015), auto-encoders (AEs) (G. E. Hinton & Salakhutdinov, 2006) and stacked auto-encoders (SAEs) (Vincent et al., 2010) neural networks, deep believe network (DBNs) (Bengio, Lamblin, Popovici, & Larochelle, 2007; G. E. Hinton, Osindero, & Teh, 2006), restricted Boltzmann machines (RBMs) (G. Hinton, 2010), convolution neural networks (CNNs) (LeCun et al., 1990), recurrent neural networks (RNNs) (Bengio, Simard, & Frasconi, 1994; Hochreiter & Schmidhuber, 1997).

Deep learning is categorized into supervised learning and unsupervised learning. In supervised learning, a model is trained to learn features from input and produces a result that is desired to equal to corresponding label pair of that input. While in unsupervised learning, we train a model with only input data and do not feed corresponding label pair to it. Generally, AEs, SAEs, DBNs and RBMs are categorized as unsupervised learning, while DNNs, CNN and RNNs are regarded as supervised learning.

In this section, we will discuss the most widely used CNNs architectures for image computer vision tasks.

### 2.2.1 Convolutional neural networks (CNNs)

Convolutional Neural Networks (CNNs) originate from multi-layered perceptrons (MLPs), and have been widely used for image processing such as classification, segmentation and localization. Hubel and Wiesel (1968) conducted a first experiment based on CNN, which indicated that cells in cat's visual cortex were responsible for detecting light in corresponding receptive fields. LeCun et al. (1990) proposed another CNN based network that recognized handwritten digits. The network was comprised of convolution operation and pooling operation, the network was trained by using back-propagation algorithm. Later, LeCun, Bottou, Bengio, and Haffner (1998) proposed the LeNet-5 for document recognition. However, these architectures were not widely used since the lack of training data and computation power at that time. Krizhevsky, Sutskever, and Hinton (2012) proposed a powerful deep CNN for image classification, called AlexNet. The model showed significant improvement and outperformed all existing methods, thus won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) (Deng et al., 2009). AlexNet has a deeper architecture than LeNet-5 and utilizes ReLU function as activation function. Figure 2.1 and Figure 2.2 show the architecture of LeNet-5 and AlexNet, respectively.



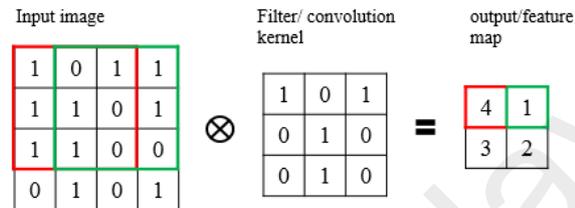
**Figure 2.1: Architecture of LeNet-5 (LeCun et al., 1998). C: convolution layer, S: sub-sampling layer, F: fully connected layer.**



the model and make training process easier. Mathematically, the feature map  $z$  generated by convolution kernel can be expressed as:

$$z = W * x + b \quad (2.1)$$

where  $x$  is the input image,  $W$  is the convolution kernel, while  $b$  is the bias for the convolution layer.



**Figure 2.3: Convolution operation in convolutional layer. Stride=1 and assume bias=0.**

*(b) Batch normalization*

The input or feature maps generated by convolutional layer may vary greatly, so for large or small values sent to activation function they face a problem of vanishing/exploding gradients, which hamper the training process (Bengio et al., 1994; Glorot & Bengio, 2010). To address this problem, batch normalization (Ioffe & Szegedy, 2015) was proposed to accelerate the training process, which scaled the input of activation function and reduced internal covariate shift by applying normalization operation to each mini-batch. Generally, batch normalization is performed before activation function, but the function can also be used after activation function based on application.

*(c) Activation function*

Activation function are non-linear functions that map inputs into outputs non-linearly, they are applied to improve the feature representation ability. It often follows convolutional layers and uses feature maps as input in neural networks. Sigmoid function

(Mhaskar & Micchelli, 1994) is a prevalent alternative for activation function, which is defined as:

$$y = \text{sigmoid}(x) = \frac{1}{1+e^{-x}} \quad (2.2)$$

where  $x$  is the input and  $y$  represents the output. The sigmoid function faces the vanishing gradient problem for very large or very small input.

ReLU (Nair & Hinton, 2010) is another frequently used activation function, which preserves the positive part in the feature maps and prunes the negative part to 0. It is expressed as:

$$y = \text{ReLU}(x) = \max(x, 0) \quad (2.3)$$

where  $x$  is the input of ReLU function and  $y$  represents its output. ReLU can alleviate the problem of vanishing gradient since its gradient is 1 when the input is positive, no matter how large it is.

However, when the input is negative, the output of ReLU and its gradient is always assigned to 0. It does reduce over-fitting, but it also obstacles CNNs to learn in some cases because 0 means disconnection of neurons. LReLU was proposed to address the problem of zero gradient when the input is negative for ReLU function (Maas, Hannun, & Ng, 2013). LReLU preserves the negative part by scaling it in a ratio  $\lambda$  (range from 0 to 1) as well as preserves the positive part completely. It is expressed as:

$$y = \text{LReLU}(x) = \max(x, 0) + \lambda \min(0, x) \quad (2.4)$$

when the input is negative, both output and gradient are non-zero value.

Softmax function is able to produce a normalized probability distribution of input, and it is often used as the last activation in CNNs for K-classes classification task. It is expressed as:

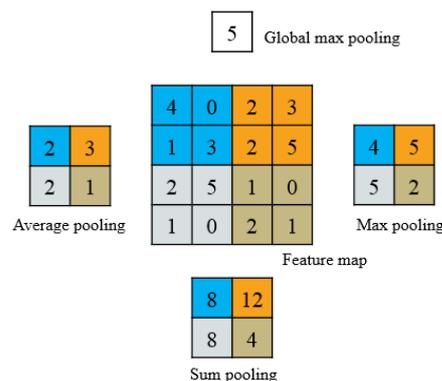
$$y(x)_i = \frac{e^{x_i}}{\sum_{i=1}^K e^{x_i}} \quad (2.5)$$

where  $x$  is input vector with  $K$ -dimension,  $x_i$  is its components.  $y(x)_i$  is the output that represents the probability of the input vector to be classified into the  $i$ th class.

(d) **Pooling layer**

The feature map produced by convolutional layer records position of pixels precisely, so it is very sensitive to the location of features, which means a small movement of the feature position, such as rotation, zoom and shift, will cause a different map, thus decreases the robustness of CNNs. Considering this, pooling layer is applied to CNNs and inserted after convolutional layer. Pooling layers conduct pooling operation, which can reduce specific feature positioning reliance and increase the shift-invariance of CNNs. At the same time, pooling operation can also decrease the computation burden by reducing the resolution of feature maps.

Pooling operation can be divided as max pooling (Boureau, Ponce, & LeCun, 2010), average pooling (T. Wang, Wu, Coates, & Ng, 2012) and sum pooling. Figure 2.4 shows pooling operations: a sliding window is placed upon feature maps, then the max value, average value or sum value in this window is calculated as output according to the pooling type. Specially, if the size of pooling window equals to that of feature map, it is referred to as global pooling, otherwise it is local/regional pooling.



**Figure 2.4: Pooling operations, with 2\*2 filter and stride=2.**

(e) **Fully connected layer**

Fully connected layers (FCs) are flattened layers like MLP. Each neuron in fully connected layer has connection with all the neurons in previous layer, then all activations can be computed with matrix multiplication followed by biases.

**2.2.1.2 Loss function**

The loss function is used to evaluate the difference between predicted result and desired result. An appropriate loss function can measure the difference between result and label properly and guide a fast and correct training process. Here we introduce some popular loss functions that are widely used in CNN architectures.

Cross-entropy function loss is the mostly used in multi-classes classification task. It is given as::

$$\mathcal{L} = -\sum_{i=1}^M c_i \log(p_i) \quad (2.6)$$

where M is the number of classes,  $c_i$  is the practical label of an input belongs to  $i$ th class so it is 0 or 1,  $p_i$  is the probability of the input predicted by networks.

Cross-entropy can also be used in binary classification task, then Softmax function is degraded to sigmoid function. The binary cross-entropy loss function is expressed as:

$$\mathcal{L} = -(y \log(p) + (1 - y) \log(1 - p)) \quad (2.7)$$

where y is ground truth and p is predicted value.

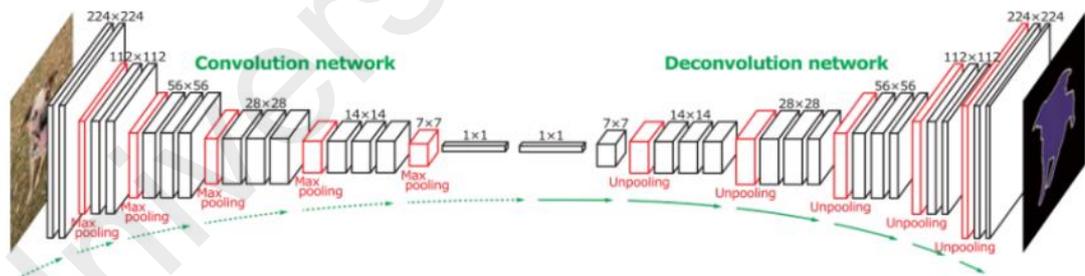
Dice loss is widely adopted in image segmentation tasks. It is represented as:

$$Dice = \frac{2|GT \cap SR|}{|GT| + |SR|} \quad (2.8)$$

where  $|GT|$  represents the ground truth magnitude while  $|SR|$  represents the segmentation result magnitude,  $|GT \cap SR|$  represents the common elements between  $GT$  and  $SR$ .

### 2.2.2 Fully convolutional networks (FCNs)

In image segmentation task, we need to predict each pixel in an image using contextual information, however, traditional CNNs predict the center pixel of an image or image patch once a time, so it is slow and heavy computation burden. Long, Shelhamer, and Darrell (2015) proposed fully convolutional networks (FCNs) to make dense prediction. They replaced fully connected layer in CNNs by up-sampling layers, then the feature maps were resized to the same size with the input images by up-sampling. The proposed FCN architecture is shown in Figure 2.5. FCNs are able to predict each pixel in an image or image patch, so it is more suitable and fast for image segmentation task than traditional CNNs.

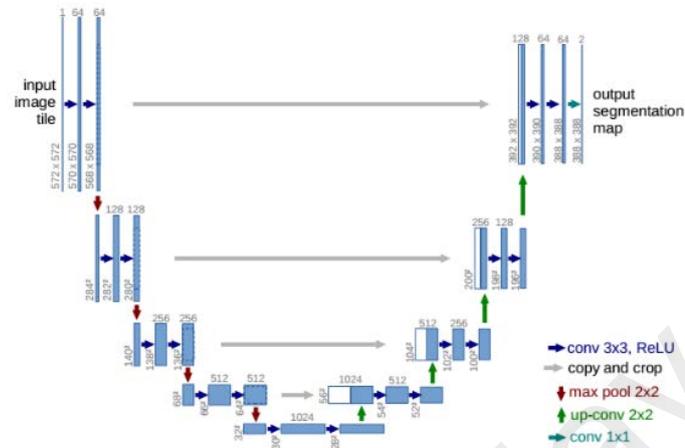


**Figure 2.5: Architecture of a FCN (Long et al., 2015).**

### 2.2.3 U-net

Ronneberger, Fischer, and Brox (2015) proposed U-net which has symmetrical encoder-decoder structure and skip connections from encoding path to decoding path. In the U-net, Features were extracted in encoder and images were reconstructed in decoder. Skip connections sent low level feature maps generated in encoder to decoder directly. Since low-level feature maps contained local information while high-level feature maps

contained global information, the proposed U-net integrated low-level and high-level feature map and thus made better prediction. Figure 2.6 shows the architecture of U-net.



**Figure 2.6: Architecture of U-net (Ronneberger et al., 2015).**

### 2.3 Retinal vessel segmentation

We need to assign a label, that is, vessel or non-vessel, for each pixel in the field of view (FOV) of retinal fundus images. Initially, ophthalmologist would manually segment vessel, but that is time-consuming and tedious, so computer-assist automatic segmentation came into being. Earlier, unsupervised methods are the most common approach for automatically segments the retinal vessels, which do not rely on any annotation for segmentation (M. M. Fraz et al., 2012). These methods are roughly divided into matching filter (Al-Rawi, Qutaishat, & Arrar, 2007), vascular tracing based segmentation (Yin, Adel, & Bourennane, 2012) and model-based segmentation methods (Kaba, Salazar-Gonzalez, Li, Liu, & Serag, 2013). Unsupervised methods show some defects in their performance because they cannot benefit from the hand-labelled ground truth.

Later, supervised models were introduced into this task. Supervised models retinal vessel segmentation in two stages: feature extraction and pixels classification. Features can be further divided into handcrafted features or automatically learned features. In machine

learning, the process of feature extraction from fundus images is manual, and some typical classifiers are adopted, such as k-nearest neighbour classifier (KNN) (Staal et al., 2004) and support vector machine (SVM) (You, Peng, Yuan, Cheung, & Lei, 2011). However, selecting features manually lacks of generalization ability and may miss some useful patterns (Mo & Zhang, 2017).

Currently, deep learning, especially CNNs, has gained much attention for images analysis (Litjens et al., 2017; Pouyanfar et al., 2018). Deep learning methods learn features automatically by using massive data without human inference. They have better generalization ability and recognition capability because they can learn different level patterns automatically, thus they are not application specific. In this task, many researchers have proposed their networks to segment retinal vessels.

Khalaf, Yassine, and Fahmy (2016) constructed a CNN with 7 layers. They divided pixels in an image into 3 classes: background, large vessel and small vessel to reduce the intra-classes variance. They extracted green channel of images and applied adaptive histogram equalization (AHE) and top-hat filtering to green channel in pre-processing phase. The green channel and AHE increased image contrast and suppressed noise, and top-hat filtering enhanced vessels in training images.

Oliveira, Pereira, and Silva (2018) proposed an FCN and added skip connection to propagate features from shallow layer to deeper layer. They also explored the multiscale nature of the vascular system by using stationary wavelet transform (SWT) which added extra channels to input. Their result illustrated that deep learning methods can benefit from domain knowledge.

Z. Jiang, Zhang, Wang, and Ko (2018) used a network based on the fully convolutional version of AlexNet. In pre-processing phase, they replaced the black ring in fundus

images and refilled these areas with average value from FOV, then they applied Gaussian smooth to reduce the discontinuity between FOV and the replaced region. After they performed contrast enhancement and placed a black ring back again to reduce training complexity. They extracted image patches and enlarged it to magnify the details. The segmented vessels were thicker than ground truth, so Z. Jiang et al. (2018) applied a 9\*9 filter to refine the result and reduce noise in post-processing phase.

Since the consecutive down-sampling operations in encoder lead to loss of information, which is critical to determine vessel boundaries and thin vessels. Y. Luo, Cheng, and Yang (2016) proposed a size-invariant fully convolutional neural network (SIFCN) to reduce its effect. They hold the size of feature maps in each layer by padding and assigning strides and thus reduces loss of information.

Zhang and Chung (2018) adopted a U-shaped network with long-term and short-term skip-connection. They regarded retinal vessel segmentation as multi-class classification task and introduced an edge-aware mechanism. They divided pixels into 5 classes: background, thick vessels, thin vessels, background near thick vessels and background near thin vessels. The network can pay more attention to the boundary areas of vessels in this way. They leveraged deep supervision to ease optimization.

Y. Jiang, Tan, Peng, and Zhang (2019) proposed a U-shaped network integrated residual block with dilated convolution for retinal vessel segmentation. They used three residual sub-blocks to construct dilated convolution block and replaced several regular convolution blocks in U-net by it. They arranged the dilated convolution block with different dilated rates deliberately to obtain dense sampling of input and thus avoid chessboard effect. They introduced multi-scale information fusion module to capture different sized feature maps, where they connected several depthwise separable

convolution parallelly (Howard et al., 2017) to reduce the computation cost and number of parameters.

Mou et al. (2019) proposed a dense dilated U-net which used residual block as building block to extract image features. They added dense dilated blocks between encoder and decoder to compute an initial map of vessels. They fused multi-level feature maps in decoder to generate a coarse segmentation and adopted multi-level dice loss to train the network. Finally, they introduced probability regularized walk (PRW) algorithm to reconnect fractured vessels. PRW is an extension of random walk algorithm on probability map.

Attention mechanism (Vaswani et al., 2017) has been applied to locate region of interest (ROI) and strengthen feature representations in retinal vessel segmentation. Lian et al. (2019); Z. Luo et al. (2019); Lv, Ma, Li, and Liu (2020) made attention masks manually with the same size of original images and multiplied it by final feature maps in element-wise. Lian et al. (2019); Z. Luo et al. (2019) assigned elements as 1 in FOV and 0 in other region, while Lv et al. (2020) assigned  $\theta$  range from 0 to 1 for pixels outside the FOV. X. Li, Jiang, Li, and Yin (2020); B. Wang et al. (2020) designed attention modules to strengthen feature representations. The proposed modules took intermediate feature maps as input and generated masks by a set of element-wise addition and multiplication of feature maps. The attention matrixes were learned by networks instead of assigned by experts. B. Wang et al. (2020) also introduced a structure loss by adding spatial weights to cross-entropy loss to guide the network to focus more on the thin vessels and boundaries.

Inspired by U-net and feature pyramid network (T.-Y. Lin et al., 2017), Liu et al. (2021) proposed a feature pyramid U-net which adopted U-net as backbone and added two extra

feature pyramid pathways. The feature pyramid pathways are used to extract multi-level features. Their proposed network can reduce the effect of optic discs and capture more details.

Many researchers had found the limited prediction capability of a single model, so they proposed multi-model networks for stronger prediction ability. A simple U-Net proposed in (Yan, Yang, & Cheng, 2018) could not performed well, therefore a three stage segmentation network is proposed for retinal vessels using three sub-networks (Yan, Yang, & Cheng, 2019). The segmentation of the whole vessel tree was divided into three sub-tasks: thick vessel segmentation using FCN, thin vessel segmentation using U-net and fusion of segmentations. The proposed three-stage segmentation can be regarded as coarse-and-fine segmentation because it separated thick and thin vessels separately.

More researchers proposed coarse-to-fine segmentation by cascading several sub-networks.-The following sub-network can inherit the learning experiences of previous sub-models (Budak, Cömert, Çibuk, & Şengür, 2020; Francia, Pedraza, Aceves, & Tovar-Arriaga, 2020; Hu et al., 2019; L. Li, Verma, Nakashima, Nagahara, & Kawasaki, 2020; K. Wang, Zhang, Huang, Wang, & Chen, 2020; Y. Wu, Xia, Song, Zhang, & Cai, 2020; Xia, Zhuge, & Li, 2018). Generally, they added intra- and inter- skip connections to send low-level feature maps and learned knowledge to deeper layers and sub-networks. The followed sub-network segmented vessels coarsely and the following sub-network refined vessel maps. The following sub-model used segmented result of previous sub-models and original images as input. Therefore, it can learn correct knowledge that had been learned and not get stuck into incorreced knowledge. Y. Wu et al. (2020) added an auxiliary layer to the followed network to get an auxiliary loss, so their model was trained by main supervision and auxiliary supervision.

GAN (Goodfellow et al., 2014) is a type of deep unsupervised learning model, composed of a generator and a discriminator elements. C. Wu, Zou, and Yang (2019) adopted U-net as generator and improved it by dense block. Additionally, they introduced attention gate to each block in decoder for focusing on target structure. They used a CNN as discriminator and inserted several dense blocks to reuses features. Park, Choi, and Lee (2020) chained two U-nets in generator and inserted multi-kernel pooling block between these 2 U-nets to support the scale-invariance. The proposed M-GAN used residual convolution block as building block in both generator and discriminator. They utilized automatic color equalization (ACE) to enhance images in pre-processing phase with Lanczos resampling method to smooth the vessel branches and reduce false negatives in post-processing phase.

Table 2.1 summarizes the existing methods briefly.

**Table 2.1: Overview of papers for retinal vessel segmentation.**

Reference	Method	Remarks
Al-Rawi et al. (2007)	Unsupervised method	Matching filter
Yin et al. (2012)		Vascular tracing based segmentation
Kaba et al. (2013)		Model based segmentation
Staal et al. (2004)	Machine learning	KNN as classifier
You et al. (2011)		Used SVM as classifier
Khalaf et al. (2016)	CNN	3-class classification, grayscale conversation, adaptive histogram equalization and top-hat filtering
Oliveira et al. (2018)	FCN	Skip connection, stationary wavelet transform
Z. Jiang et al. (2018)	FCN	Transfer learning, matching filter in post-processing phase
Y. Luo et al. (2016)	FCN	Size-invariant feature map to reduce of information
Zhang and Chung (2018)	U-net	Long- and short-term skip connections, multi-class classification, edge-aware mechanism. deep supervision
Y. Jiang et al. (2019)	U-net	Residual learning, dilated convolution, depth-wise separable convolution

Reference	Method	Remarks
Mou et al. (2019)	U-net	Dense connection, dilated convolution, probability regularized walk algorithm, coarse-to-fine segmentation
Lv et al. (2020)	U-net	Attention mechanism to locate region of interest
B. Wang et al. (2020)	U-net	Attention mechanism to strengthen features
T.-Y. Lin et al. (2017)	U-net	Feature pyramid pathways
Yan et al. (2019)	U-net	Three stage segmentation, coarse-and-fine segmentation
Y. Wu et al. (2020)	U-net	Ensemble learning, deep supervision
C. Wu et al. (2019)	GAN	U-net as generator, dense block, attention gate
Park et al. (2020)	GAN	chained two U-nets in generator, residual learning, Lanczos resampling method in post processing phase

#### 2.4 Hardware and software

To implement the model and complete the task, hardware and software are required. Central processing units (CPU) and graphics processing units (GPU) are the most common hardware to deploy deep learning models. GPUs are highly parallel computing engines, which can afford more execution threads than CPUs. With current hardware, deep learning on GPUs is generally faster than on CPUs.

In addition to hardware, software is also necessary to implement various models. There are several open source software packages widely used in deep learning, which are:

1. Tensorflow: developed by Google and widely used in industry. It provides C++ and Python interface.
2. Pytorch: developed by Facebook. It can be regarded as a Python version of torch, which provides a Lua interface.

There are third-party packages written upon these frameworks, such as Keras and Lasagne.

## **2.5 Summary**

This chapter provides an overview of deep learning, including some prevalent models used for image segmentation. It also reviewed some publications for the specific task, from the review we can see, CNNs were adopted to segment vessels in the earlier, then FCNs were introduced to this task to make dense predictions. Currently, U-net has gained much attention because their excellent performance, at the same time lots of researchers introduced various improvement methods to U-net, such as dilated convolution, residual learning and attention mechanism. Some researchers constructed multi-model networks to improve the performance, they connected several sub-networks serially or in parallel. Finally, we introduced hardware and software to implement and deploy deep learning models.

## CHAPTER 3: METHODS

### 3.1 Introduction

This chapter illustrates the method used to segment retinal vessels, including the database, metrics to evaluate model performance, the image processing technique, the proposed model and its implementation details.

### 3.2 Database

For retinal vessel segmentation, DRIVE (Staal et al., 2004) and STARE (Soares, Leandro, Cesar, Jelinek, & Cree, 2006), CHASE\_DB1 (Muhammad Moazam Fraz et al., 2012) and HRF (Köhler et al., 2013) are mostly used databases which are also publicly available.

DRIVE (Digital Retinal Images for Vessel Extraction) database comes from the Netherlands-initiated DR screening project. It contains 40 images with a unified resolution of 584x565, and each image is color fundus image with three 8-bit channels. Each image is assigned a binary field of view (FOV) mask and ground truth (GT) with respect to manual annotations of 2 independent experts. Actually, these 40 images are selected from 400 diabetic subjects randomly, and 33 images are from healthy persons while other 7 images are from persons that has shown symptoms of early DR. Usually, the manual annotations of the first expert are used as ground truth to train networks while the annotations of the second expert are used to evaluate the performance of networks.

The STARE (Structured Analysis of the Retina) database composes of 20 color fundus images with a unified resolution of 605x705. The first 10 images are pathological images while the last 10 images are healthy images. Each image is also assigned with manual annotations of two observers, but without any FOV mask. The performance of networks is usually evaluated by the manual segmentation of the first observer.

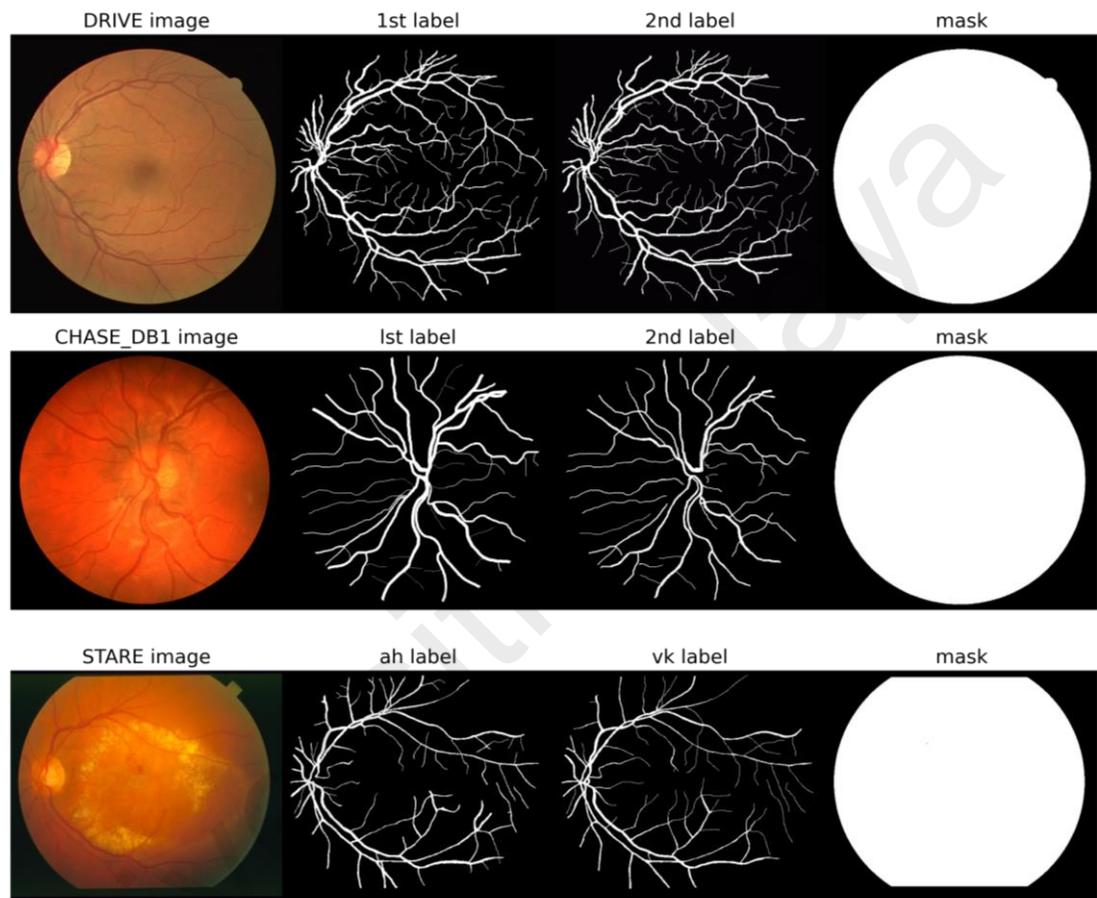
The CHASE\_DB1 (Child Heart and Health Study in British) database is a subset of retinal images of multi-ethnic children. It is comprised of 28 color fundus images and each image has a resolution of 960x996. Each image is assigned with manually segmented annotations but without FOV mask. The training and testing data are not partitioned by experts, so the first 20 images are usually used for training while the last 8 images for testing of networks. Compared with images in DRIVE and STARE database, images of CHASE\_DB1 has uneven background illumination, low contrast between background and vessels, and broader arterioles.

ARIA (FARNELL, 2006; Farnell et al., 2008) has 143 images for vessel segmentation and is not split into training and test dataset. The images were captured at a resolution of 768x584 pixels, and each image was annotated by two experts. 61 images were captured from healthy subjects, 59 from diabetic patients and 23 from patients with age-related macular degeneration.

DRiDB (Prentašić et al., 2013) contains 50 colorful images with a uniformed resolution of 720x576 pixels and a 45° FOV. 36 images contain signs of the diabetic retinopathy and 14 images do not contain any signs of the diabetic retinopathy.

In addition, Scanning Laser Ophthalmoscopy (SLO) (Webb & Hughes, 1981) technique can also be leveraged to capture retinal images, with the advantages of lower light exposure and better contrast (LaRocca, Nankivil, Farsiu, & Izatt, 2014). By the means of SLO, IOSTAR and RC-SLO datasets were established. IOSTAR dataset contains 30 images with a uniformed resolution of 1024x1024 pixels and a 45° FOV. The RC-SLO dataset contains 40 images with a resolution of 360x320 pixels. Both IOSTAR and RS-SLO datasets provide FOV masks and contain a ground truth of vessel trees for each image, but did not split samples into training and test dataset.

This project chose DRIVE, STARE and CHASE\_DB1 to evaluate model performance. Since only DRIVE has FOV masks, we generated FOV masks for CHASE\_DB1 and STARE database by using binary-threshold method. Figure 3.1 shows image samples and corresponding labels, masks from these databases.



**Figure 3.1: Image-label-mask show.**

### 3.3 Evaluation metrics

Generally, pixels in FOV of fundus images are classified as vessel pixel (positive) or non-vessel pixel (negative). To measure the identification of pixels, ground truth label is compared with pixel identification. On the basis, there are four basic pixel measures i.e., TP (true positives), FP (false positives), FN (false negatives), and TN (true negatives). Table 3.1 shows the measures of these elements through pixels.

**Table 3.1: Pixel measures in vessel segmentation.**

Classification result		Ground truth	
		Vessel	Non-vessel
Segmentation result	Vessel	TP	FP
	Non-vessel	FN	TN

Several evaluation metrics are defined to evaluate the performance of segmentation networks. Some of the prevalent metrics are listed in Table 3.2.

**Table 3.2: Evaluation metrics for image segmentation.**

Matric	Expression
Sensitivity	$Sen = \frac{TP}{TP + FN}$
Specificity	$Spe = \frac{TN}{TN + FP}$
Precision	$Pre = \frac{TP}{TP + FP}$
Accuracy	$Acc = \frac{TP + TN}{TP + FP + TN + FN}$
F1-score	$F1 = \frac{2TP}{2TP + FP + FN}$

Meanwhile, AUC\_ROC (the area under the ROC curve) and AUC\_PR (the area under the PR curve) are available to evaluate the performance of the networks. The ROC curve illustrates the relationship between sensitivity and specificity by changing the threshold on the generated probability map image, while PR curve illustrates the relationship between precision and sensitivity.

### 3.4 Image pre-processing

The low quality of existing image samples hinders models to learn better feature representations. Image noise, uneven illumination, low contrast especially for thin vessels decrease the performance of proposed models. In this project, we applied some image

processing technique to enhance raw mages, which includes grayscale conversion, contrast limited adaptive histogram equalization and gamma correction. Finally, we normalize the images to adjust its variance and mean value.

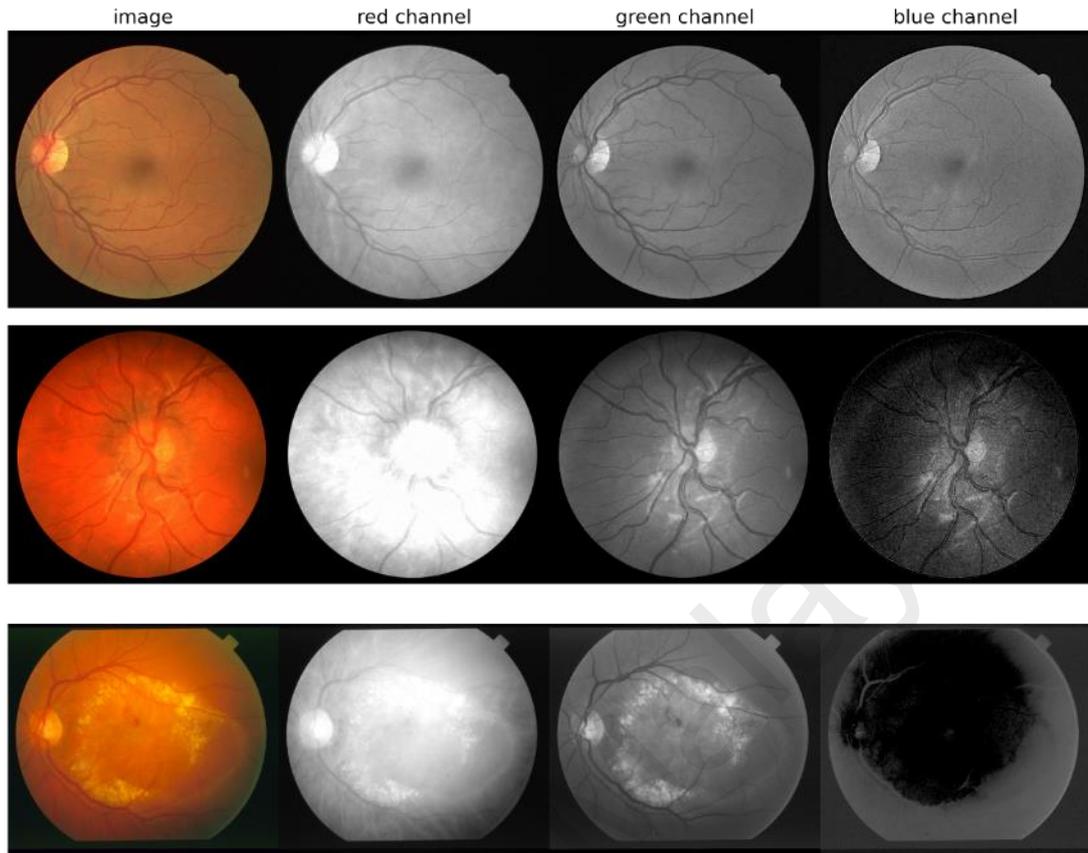
### 3.4.1 Grayscale conversion

The original images from databases are colorful with low contrast, so we can use grayscale conversion to enhance the contrast and reduce computation burden of networks.

Images can be converted to grayscale by combining pixel intensity of RGB channels:

$$\text{grayscale} = \alpha * R + \beta * G + \gamma * B \quad (3.1)$$

Here, we set  $\alpha = \gamma = 0$  while  $\beta = 1$ , which means we just extracted green channel because it has the best contrast (as shown in Figure 3.2).



**Figure 3.2: Images and their channel splits. From top to bottom: DRIVE, CHASE\_DB1, STARE.**

### 3.4.2 Normalization

We need to normalize images since they were captured at different time, different illumination and even by different devices, then there exists great variance in the grayscale distribution. Generally, min-max normalization and mean-standardization normalization. We adopted Mean-standardization in this experiment, which can be represented by:

$$I = \frac{I - \sigma}{\mu} \quad (3.2)$$

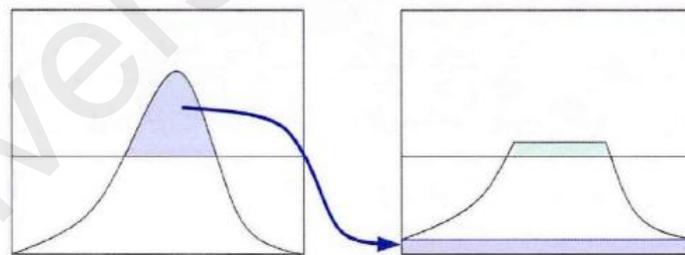
### 3.4.3 Contrast limited adaptive histogram equalization

Histogram equalization (HE) is a common method used to enhance the contrast of image. It transforms grayscale of images by histogram of grayscale distribution. It adjusts

contrast of images based on the global image, so it cannot enhance local contrast efficiently. It shows unsatisfactory performance when there exist obviously dark regions.

Adaptive histogram equalization (AHE) is improved based on histogram equalization. It split the whole image into several regions, then applies histogram equalization to each region. It can enhance contrast of images locally, but it also causes over amplification of noise, which leads to increase local contrast too much (Mosaic effect) and distortion of background.

We applied contrast limited adaptive histogram equalization (CLAHE) in this experiment because CLAHE can increase contrast as well as alleviate Mosaic effect. As shown in Figure 3.3, we set a threshold (the horizontal line in Figure 3.3) to the cumulative distribution function of each region and then limit the local contrast. It cuts the histogram that exceeds the set height and distribute the upper area to the low-end gray distribution histogram.



**Figure 3.3: Contrast Limited Adaptive Histogram Equalization.**

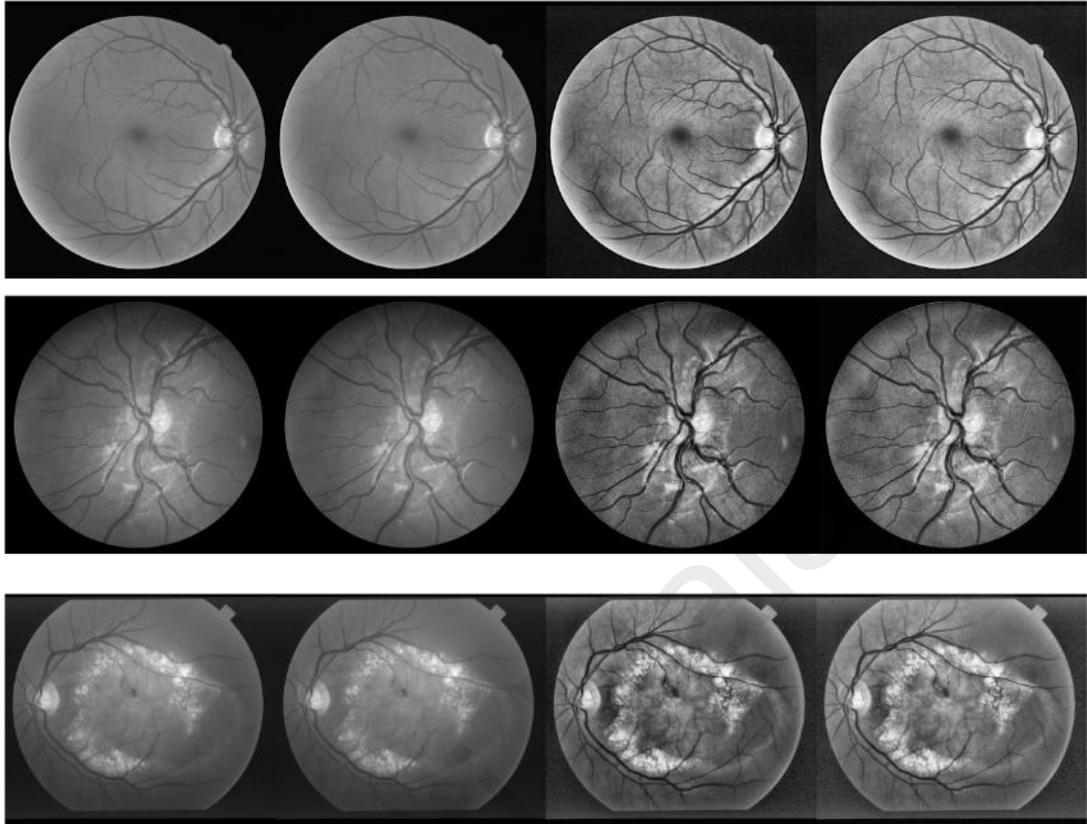
#### 3.4.4 Gamma correction

Gamma correction is used to adjust the brightness of images. Pixel intensity is reflected by using a non-linear function, which is expressed:

$$I = I^\gamma \quad (3.3)$$

where  $I$  is normalized pixel intensity.

Figure 3.4 indicates the result of image preprocessing techniques.



**Figure 3.4: Result of image preprocessing technique. From top to bottom: DRIVE, CHASE\_DB1, STARE. From left to right: green channel, normalization, CLAHE and gamma correction.**

### 3.5 Data augmentation

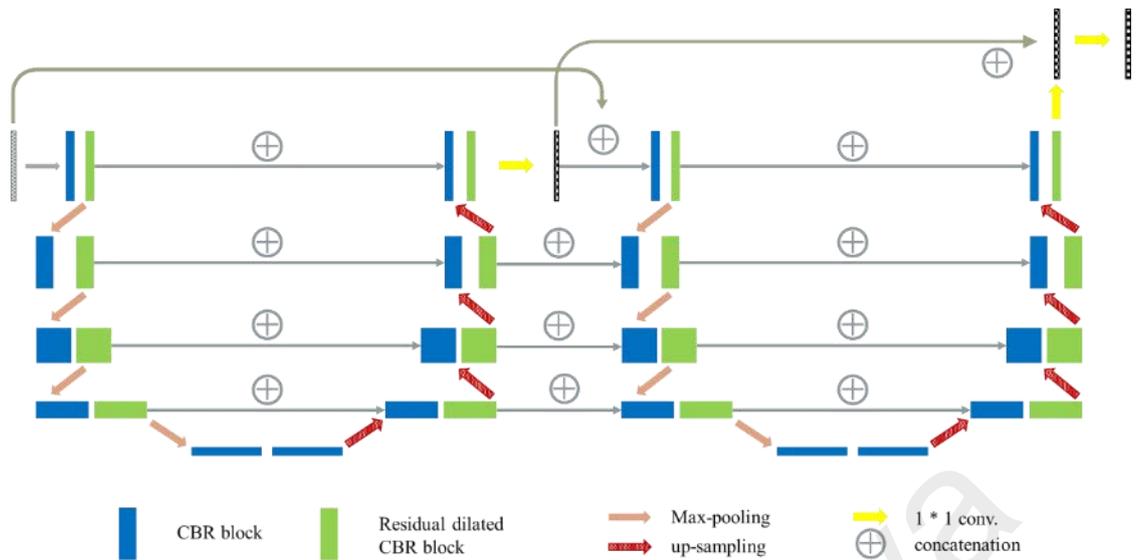
There are 40 images in DRIVE databases, 20 images in STARE database and 28 images in CHASE\_DB1 database. It is not enough to train the model, so we adopted data augmentation method to enlarge databases. We flipped images horizontally or vertically and rotate them then we other seven times images, then we cropped images into image patches.

### 3.6 Model architecture

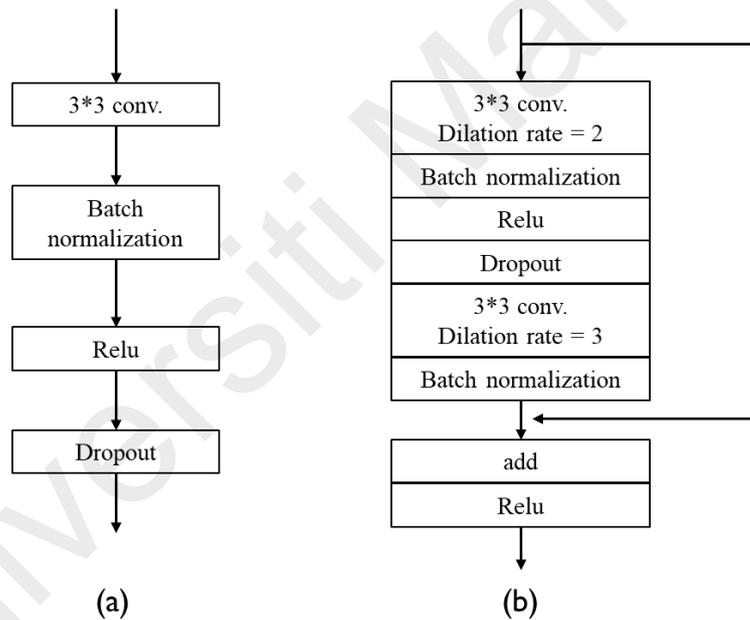
In this experiment, we chained 2 U-nets and added more skip-connections between these 2 sub-networks to reuse low-level information. Figure 3.5 shows the architecture of our

model. The first network makes coarse prediction and the second network fine-tunes the coarse map to make fine segmentation. We generate first vessel maps using feature maps produced in the front network and send it to the second network together with original input, the vessel map produced in the second network is combined with the intermediate vessel map to generate the final vessel map. We added more skip connection between decoder of first sub-network and encoder of second sub-network to reuse features better.

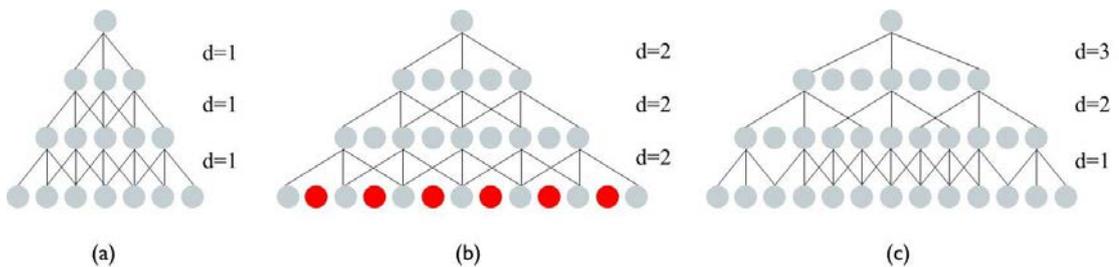
Each sub-network is composed of CBR (convolution-BatchNormalization-ReLU) block and residual-dilated-CBR block, whose architectures are shown in Figure 3.6. Residual learning could ease the optimization of network and improve its performance. In addition, we adopted dilated convolution and arranged dilation rate as 2 and 3 deliberately to avoid chessboard effect. The combination of dilated convolutions can enlarge the receptive field as well as making dense sampling. Figure 3.7 illustrates the effect of cascaded dilated convolutions. There are 3 convolution layers in each level totally, if we set all dilation rates as 1, which means no dilation, the receptive field is 7. If we set all dilation rates as 2, the receptive field is enlarged but sparse sampling also appears, which is called as chessboard effect. If we set dilation rates as 1, 2 and 3, then the receptive field is enlarged to 14 and the sampling is still dense.



**Figure 3.5: Architecture of segmentation model.**



**Figure 3.6 Architecture of building blocks. (a): CBR block, (b): residual-dilated CBR block.**



**Figure 3.7: Sampling of cascaded dilated convolution. d means dilation rate.**

### 3.7 Loss and optimizer

In this experiment, we adopted weighted binary cross-entropy loss function to alleviate the unbalance problem between classes. In addition, we added regularization term to reduce overfitting, so the final loss function can be expressed as:

$$\mathcal{L} = -(\alpha * y \log(p) + (1 - y) \log(1 - p)) + \frac{\lambda}{2n} \sum ||w||^2 \quad (3.4)$$

where  $\alpha$  is the weight coefficient for positive samples,  $w$  represents weight factors and  $\lambda$  is coefficient of regularization term. we set  $\alpha = 2$  and  $\lambda = 0.01$  in the experiment.

An appropriate optimizer could speed up the training process and reduce local optimal. We adopted Adam optimizer in this experiment, which can calculate adjustment of weight factors adaptively. The adjustment of weight factors by Adam optimizer is shown as:

$$g = \frac{1}{m} \nabla_{\theta} \sum \mathcal{L}(f(x_i, \theta), y_i) \quad (3.5)$$

$$s = \rho_1 * s + (1 - \rho_1) * g \quad (3.6)$$

$$r = \rho_2 * r + (1 - \rho_2) g^2 \quad (3.7)$$

$$\hat{s} = \frac{s}{1 - \rho_1 * t} \quad (3.8)$$

$$\hat{r} = \frac{r}{\rho_2^t} \quad (3.9)$$

$$\Delta\theta = -\eta \frac{\hat{s}}{\sqrt{\hat{r} + \delta}} \quad (3.10)$$

$$\theta = \theta + \Delta\theta \quad (3.11)$$

where  $\theta$  represents weight factors in neural network,  $m$  is the number of samples,  $x_i$  and  $f(x_i, \theta)$  represent input and outputs of model,  $y_i$  represents ground truth.  $g$  is local

gradient calculated by loss function  $\mathcal{L}$ ,  $s$ ,  $r$  and  $\eta$  are first moment estimate, second-row moment estimate and learning rate respectively.  $\rho_1$  and  $\rho_2$  are hyper-parameters. In Adam optimizer,  $\rho_1 = 1.0$ ,  $\rho_2 = 0.999$ ,  $\eta = 0.001$ ,  $\delta = 10^{-8}$ .

### **3.8 Implementation details**

The experiment was implemented with Keras (Tensorflow as backend) in Google Colab. Images in DRIVE databases have been split into training samples or testing samples after download. In STARE database, we selected 15 images as training images and the rest 5 images (the 1<sup>st</sup>, 9<sup>th</sup>, 12<sup>th</sup>, 17<sup>th</sup>, 18<sup>th</sup> images) as testing images. In CHASE\_DB1 database, we chose the first 20 images as training images and the rest 8 images as testing images. We extracted 400,000 image patches from every database to train the network. All of these image patches were extracted randomly and were ensured that at least half of pixels in patches locate in FOV. In testing phase, the whole test images were split into patches in sequence with overlap, then we recomposed vessel maps after prediction.

We adopted Adam optimizer to optimize the model and set a decayed learning rate. We set the initial learning rate as 0.0001 and decay it with a factor of 0.8 when the valid loss does not decrease after 3 epochs. We set batch size as 512 and reduce training time by using EarlyStopping method.

### **3.9 Summary**

This chapter presents the method for retinal vessel segmentation. Firstly, we introduced databases used to do this experiment, then we introduced some metrics to evaluate the model performance. We conducted image preprocessing and data augmentation since the complex condition and limited number of fundus images. We proposed a cascaded U-net to learn features and segment retinal vessels, the model is composed of CBR and residual-dilated CBR blocks. Finally, we showed our implementation details of this task.

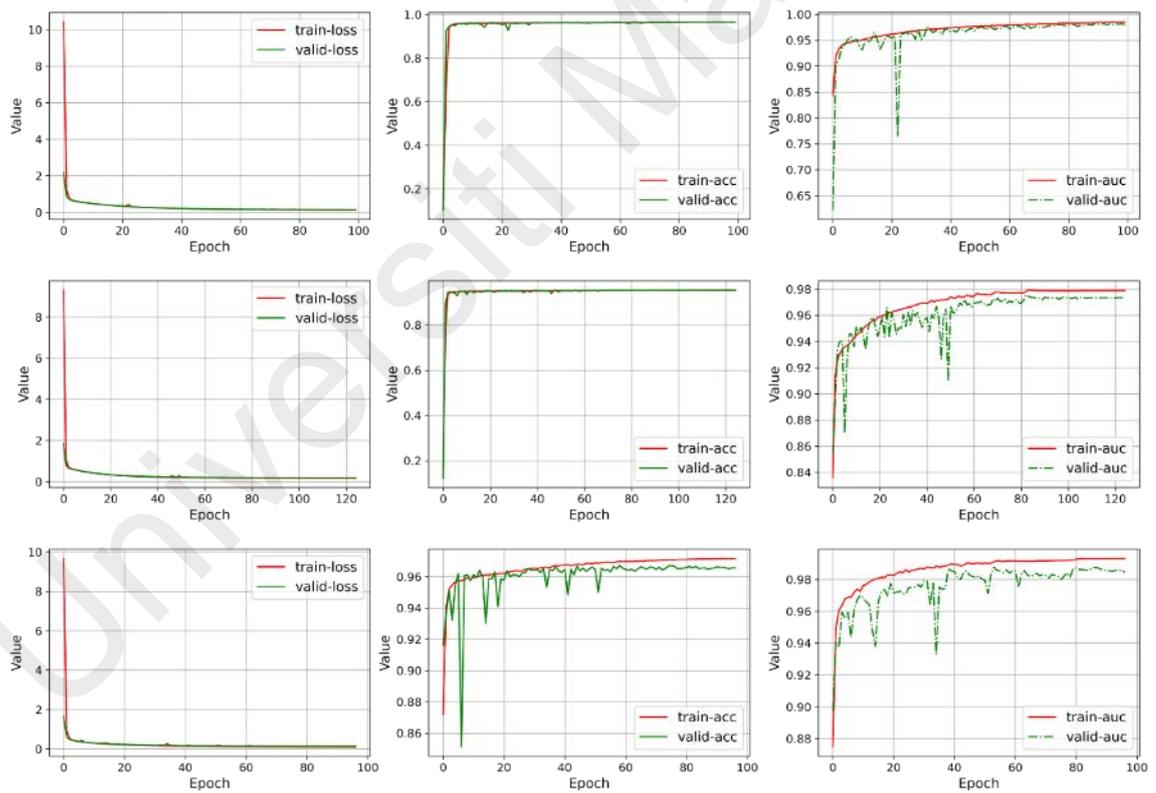
## CHAPTER 4: RESULTS AND DISCUSSION

### 4.1 Introduction

This chapter represents the test results of the proposed model on the forgoing databases. This chapter includes the characteristic curves in training phase, the predictions in testing samples and the performance of the model. At last, we compared our results with the state-of-the-art.

### 4.2 Training process

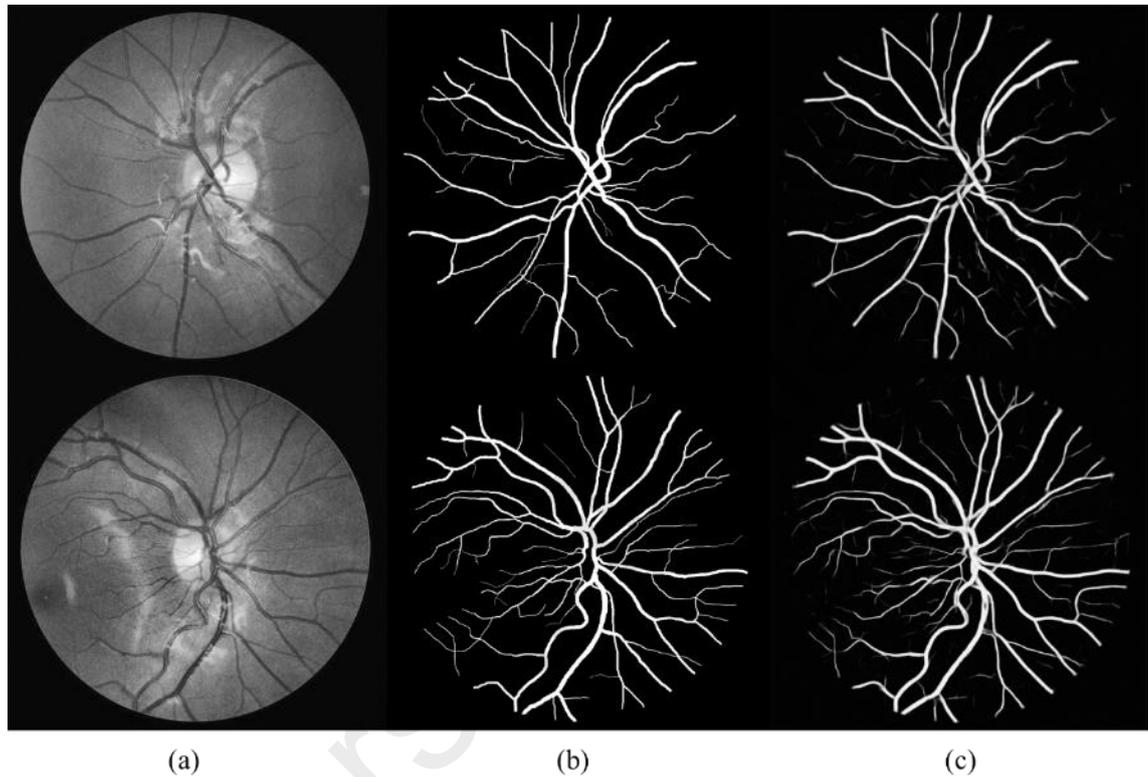
We conducted intensive experiments on these databases. Figure 4.1 indicates the accuracy and loss in training phase. It is obvious that the characteristic curves converge rapidly.



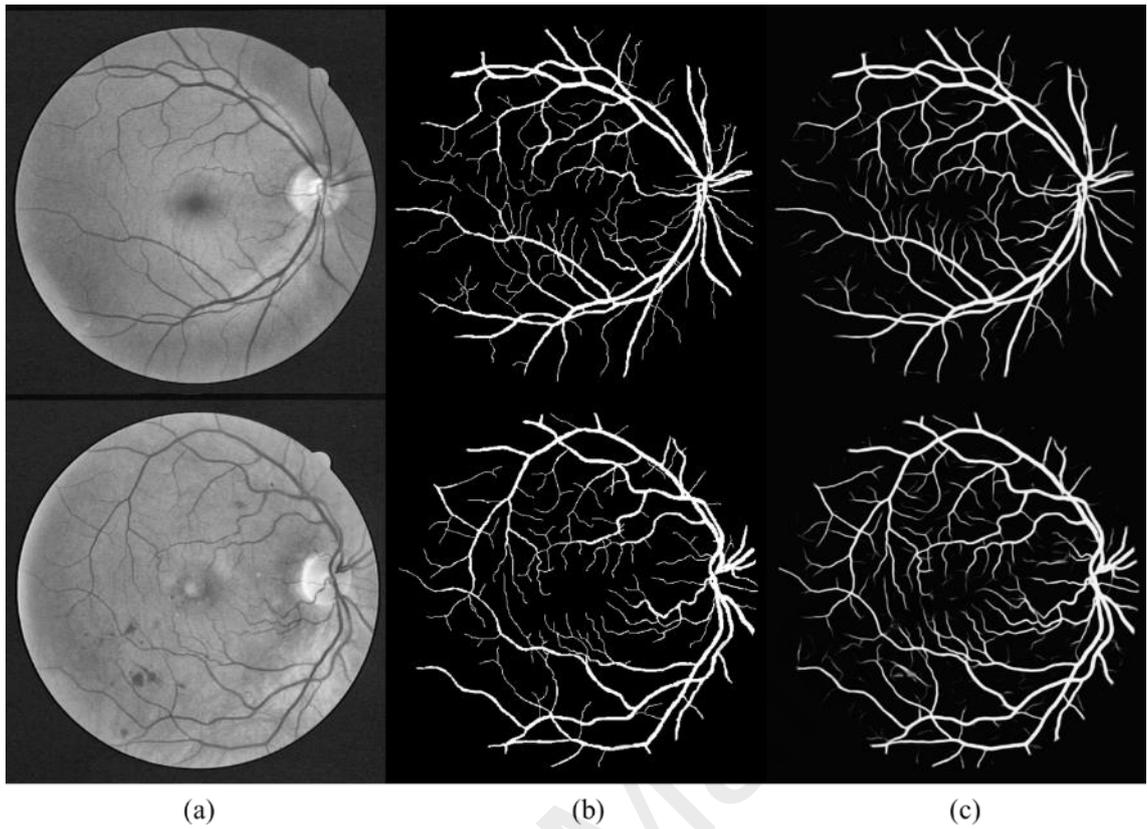
**Figure 4.1: Training curves in public databases. From top to bottom: CHASE\_DB1, DRIVE, STARE. From left to right: loss, accuracy, AU\_ROC.**

### 4.3 Overall performance

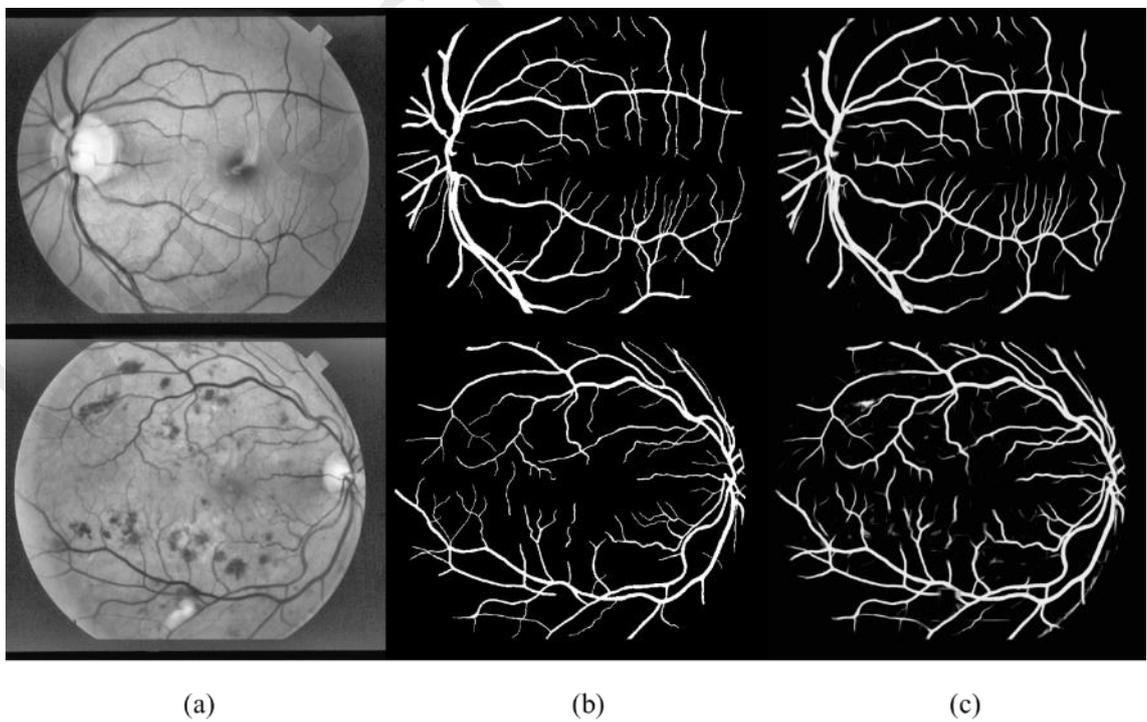
Figure 4.2, Figure 4.3 and Figure 4.4 indicate the overall segmentation results of part samples of the three databases. The overall segmentation results show an accurate segmentation and the segmented structure is completed. Thick vessels and most thin vessels can be recognized accurately.



**Figure 4.2: Retinal vessel segmentation examples for CHASE\_DB1 database. (a): pre-processed images, (b): ground truth images, (c): segmentation results.**



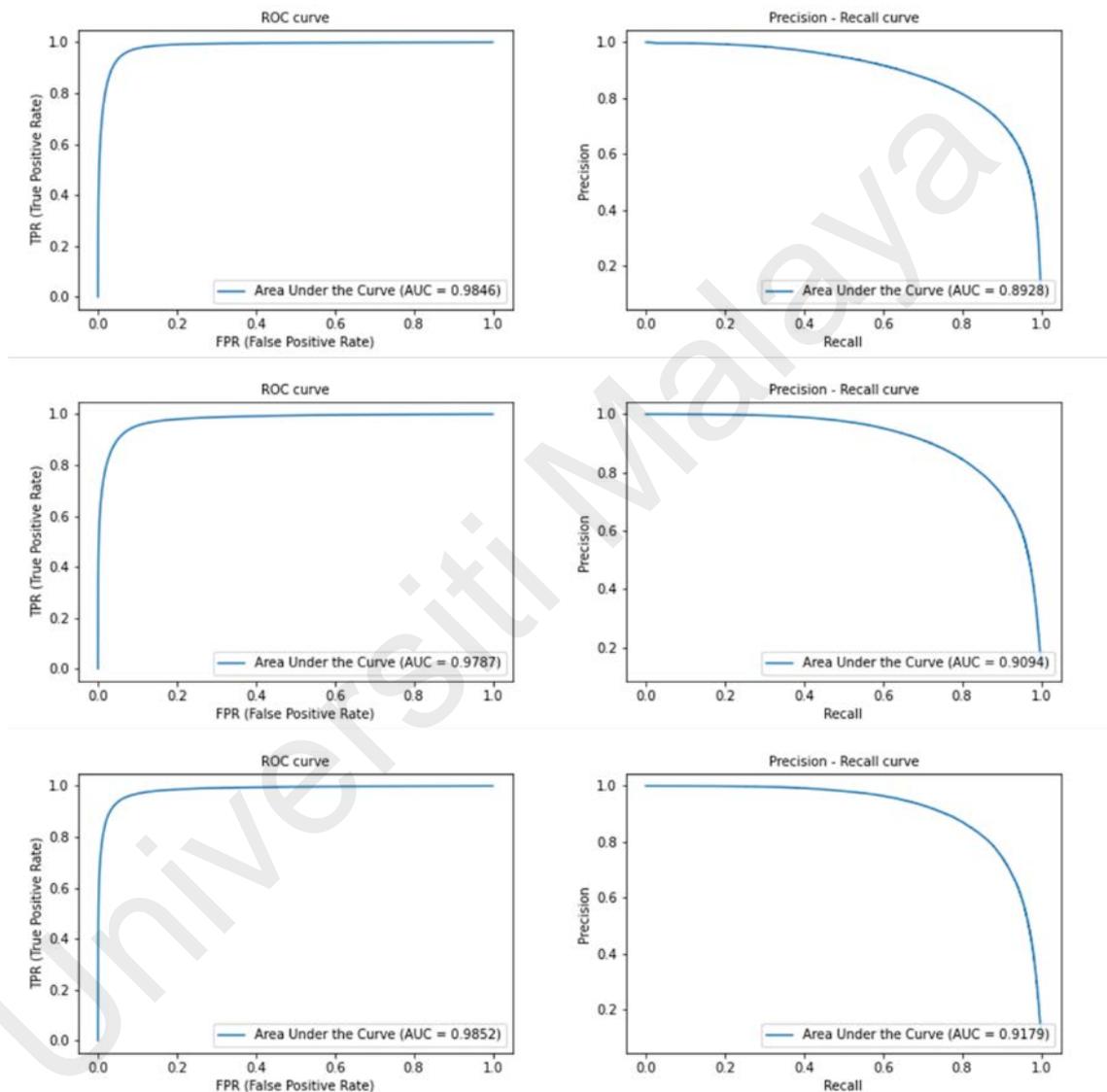
**Figure 4.3: Retinal vessel segmentation examples for DRIVE database. (a): pre-processed images, (b): ground truth images, (c): segmentation results.**



**Figure 4.4: Retinal vessel segmentation examples for STARE database. (a): pre-processed images, (b): ground truth images, (c): segmentation results.**

#### 4.4 Comparison with the state-of-art

Figure 4.5 shows the Receiver Operating Characteristic curve (ROC) and Precision-Recall curve (PR\_curve) of the proposed model on the testing phase. Table 4.1 indicates the confusion matrix of the proposed method on three testing databases.



**Figure 4.5: ROC curve and PR curve on testing databases. From top to bottom: CHASE\_DB1, DRIVE, STARE. From left to right: ROC curve, PR curve.**

Table 4.4, Table 4.3 and Table 4.4 list the performance evaluations of the proposed method and compare them with other recent publications. In the comparison, we can see that our proposed model can produce a good performance although not the best performance in all metrics. Most publications always obtain better performance in some

metrics instead of all metrics, because the metrics are correlated and traded off. In addition, we obtained 0.8928, 0.9094 and 0.9179 area under PR curve on CHASE\_DB1, DRIVE, STARE databases, respectively. We do not list them in tables since most publications did not adopted this metrics.

**Table 4.1: Confusion matrix of proposed model on three databases.**

Prediction result		Ground truth	
		Vessel	Non-vessel
Prediction (DRIVE)	Vessel	3860503	99991
	Non-vessel	103345	474304
Prediction (STARE)	Vessel	1352381	24452
	Non-vessel	21196	117026
Prediction (CHASE_DB1)	Vessel	4721480	112900
	Non-vessel	77608	405812

**Table 4.2: Performance comparison on DRIVE database.**

Methods	Performance					
	Pre	Sen	Spe	Acc	AUC	F1_score
Lv et al. (2020)	/	0.7941	0.9798	0.9558	0.9847	0.8216
Mishra, Chen, and Hu (2020)	/	<b>0.8916</b>	0.9601	0.9540	0.9724	/
X. Li et al. (2020)	/	0.7921	<b>0.9810</b>	<b>0.9568</b>	<b>0.9806</b>	/
Upadhyay, Agrawal, and Vashist (2021)	0.760	0.8840	0.9730	0.9660	0.980	0.817
Liu et al. (2021)	<b>0.8745</b>	/	/	0.9503	0.9650	0.7949
Ours (2021)	0.8259	0.8211	0.9748	0.9552	0.9787	<b>0.8235</b>

**Table 4.3: Performance comparison on STARE database.**

Methods	Performance					
	Pre	Sen	Spe	Acc	AUC	F1_score
Lv et al. (2020)	/	0.7598	<b>0.9878</b>	0.9640	0.9824	0.8142
Mishra et al. (2020)	/	0.8805	0.9651	0.9601	0.9763	/

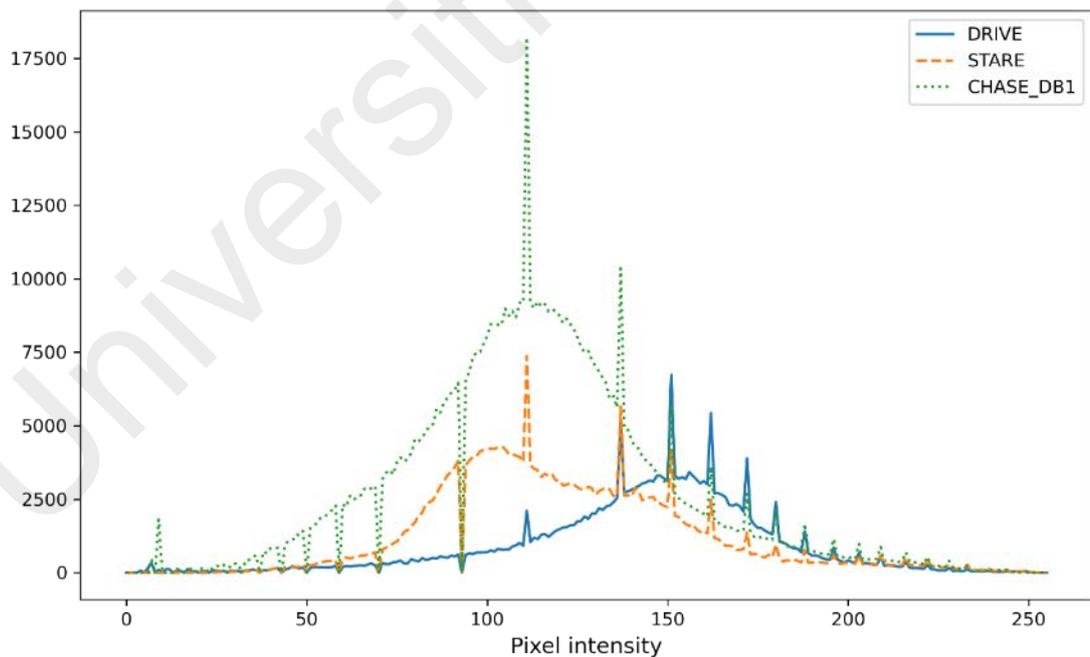
Methods	Performance					
	Pre	Sen	Spe	Acc	AUC	F1_score
X. Li et al. (2020)	/	0.8352	0.9823	0.9678	<b>0.9875</b>	/
Upadhyay et al. (2021)	0.6800	<b>0.8960</b>	0.9630	0.9580	0.9820	0.7750
Liu et al. (2021)	<b>0.8525</b>	/	/	0.9686	0.9702	0.7817
Ours (2021)	0.8272	0.8466	0.9822	<b>0.9699</b>	0.9852	<b>0.8368</b>

**Table 4.4: Performance comparison on CHASE\_DB1 database.**

Methods	Performance					
	Pre	Sen	Spe	Acc	AUC	F1_score
Lv et al. (2020)	/	0.8167	0.9704	0.9608	<b>0.9865</b>	0.7892
B. Wang et al. (2020)	/	0.8427	<b>0.9836</b>	<b>0.9706</b>	0.9824	<b>0.8105</b>
Mishra et al. (2020)	/	<b>0.8771</b>	0.9634	0.9571	0.9742	/
X. Li et al. (2020)	/	0.7818	0.9819	0.9635	0.9810	/
Liu et al. (2021)	0.7372	/	/	0.9585	0.9594	0.6976
Ours (2021)	<b>0.7823</b>	0.8395	0.9766	0.9642	0.9846	0.8099

To verify the generalization capacity of our proposed model, we also conducted cross-validation on these databases, which means train the model in one database while test it in other databases without retraining. Table 4.5, Table 4.6 and Table 4.7 show the result of cross-validation on three databases. From the tables we can see the proposed model can still make good prediction on DRIVE/ STARE database although there is slightly performance degradation between DRIVE and STARE databases. However, the proposed model obtained a very bad result in cross-validation between CHASE\_DB1 and DRIVE/STARE databases, since images in CHASE\_DB1 database have quite different background condition. They have more background noise than images in DRIVE/STARE databases, as shown in Figure 3.1, Figure 3.2 and Figure 3.3.

We suppose the inconsistent distribution of data (pixel intensity) is the main cause of performance degradation in cross-validation between DRIVE and STARE, but we do not regard it as the main cause of that between CHASE\_DB1 and STARE/DRIVE. Figure 4.6 indicates the pixel histogram of images shown in Figure 3.1 (only in FOV), from where we can see that the distribution of pixel intensity between DRIVE and STARE database are different, but there is only a slight performance degradation between these two databases (Table 4.5, Table 4.6). The pixel distribution between STARE and CHASE\_DB1 are similar, but there still exists great performance degradation. We attribute the performance degradation between CHASE\_DB1 and STARE/DRIVE to the different background conditions. Images in CHASE\_DB1 have more background noise than images in DRIVE/STARE databases, which can be seen in Figure 3.1, Figure 3.2 and Figure 3.3.



**Figure 4.6 Pixel histogram of image samples from three databases.**

**Table 4.5: Cross-validation on DRIVE database.**

Train dataset	Methods	Pre	Sen	Spe	Acc	AUC	F1-score
STARE	Mo and Zhang (2017)	/	0.7412	0.9799	0.9492	0.9653	/
	Oliveira et al. (2018)	/	0.6706	<b>0.9916</b>	0.9505	<b>0.9748</b>	/
	Guo et al. (2019)	/	<b>0.7446</b>	0.9784	0.9502	0.9709	/
	Yan et al. (2019)	/	0.7443	0.9814	0.9509	0.9720	/
	Y. Wu et al. (2020)	/	0.7187	0.9881	0.9538	0.9761	/
	Zhuo, Huang, Lu, Pan, and Feng (2020)	/	/	/	0.9499	/	<b>0.8010</b>
	Liu et al. (2021)	0.8149	/	/	0.9359	0.9392	0.6962
	Ours (2021)	<b>0.8823</b>	0.7315	0.9858	<b>0.9534</b>	0.9687	0.7998
CHASE_DB1	Mo and Zhang (2017)	/	<b>0.7315</b>	0.9778	<b>0.9460</b>	<b>0.9650</b>	/
	Guo et al. (2019)	/	0.6960	0.9699	0.9377	0.9523	/
	Liu et al. (2021)	0.7061	/	/	0.9222	0.9010	0.6454
	Ours (2021)	<b>0.9135</b>	0.5034	<b>0.9930</b>	0.9307	0.9453	<b>0.6490</b>

**Table 4.6: Cross-validation on STARE database.**

Train dataset	Methods	Pre	Sen	Spe	Acc	AUC	F1-score
DRIVE	Mo and Zhang (2017)	/	0.7009	<b>0.9843</b>	0.9570	0.9751	/
	Oliveira et al. (2018)	/	<b>0.8453</b>	0.9762	0.9597	<b>0.9846</b>	/
	Guo et al. (2019)	/	0.7188	0.9816	0.9548	0.9686	/
	Yan et al. (2019)	/	0.7319	0.9840	0.9580	0.9678	/
	Y. Wu et al. (2020)	/	0.7378	0.9785	0.9540	0.9635	/

Train dataset	Methods	Pre	Sen	Spe	Acc	AUC	F1-score
	Zhuo et al. (2020)	/	/	/	0.9569	/	0.7866
	Liu et al. (2021)	<b>0.8263</b>	/	/	<b>0.9628</b>	0.9641	0.7536
	Ours (2021)	0.7437	0.8410	0.9709	0.9590	0.9756	<b>0.7894</b>
CHASE_DB1	Mo and Zhang (2017)	/	<b>0.7387</b>	0.9787	0.9549	0.9620	/
	Guo et al. (2019)	/	0.6799	0.9808	0.9501	0.9517	/
	Liu et al. (2021)	0.6871	/	/	0.9496	0.9416	0.6582
	Ours (2021)	<b>0.8428</b>	0.6676	<b>0.9875</b>	<b>0.9583</b>	<b>0.9724</b>	<b>0.7451</b>

**Table 4.7: Cross-validation on CHASE\_DB1 database.**

Train dataset	Methods	Pre	Sen	Spe	Acc	AUC	F1-score
DRIVE	Mo and Zhang (2017)	/	<b>0.7003</b>	0.9750	<b>0.9478</b>	<b>0.9671</b>	/
	Guo et al. (2019)	/	0.6980	0.9715	0.9441	0.9511	/
	Liu et al. (2021)	0.5132	/	/	0.9309	0.9104	<b>0.5135</b>
	Ours (2021)	<b>0.6292</b>	0.3922	<b>0.9769</b>	0.9237	0.7361	<b>0.4832</b>
STARE	Mo and Zhang (2017)	/	<b>0.7032</b>	0.9794	<b>0.9515</b>	<b>0.9690</b>	/
	Guo et al. (2019)	/	0.6726	0.9710	0.9411	0.9511	/
	Liu et al. (2021)	0.5949	/	/	0.9478	0.9183	0.5338
	Ours (2021)	<b>0.7143</b>	0.3672	<b>0.9853</b>	0.9291	0.8071	0.4850

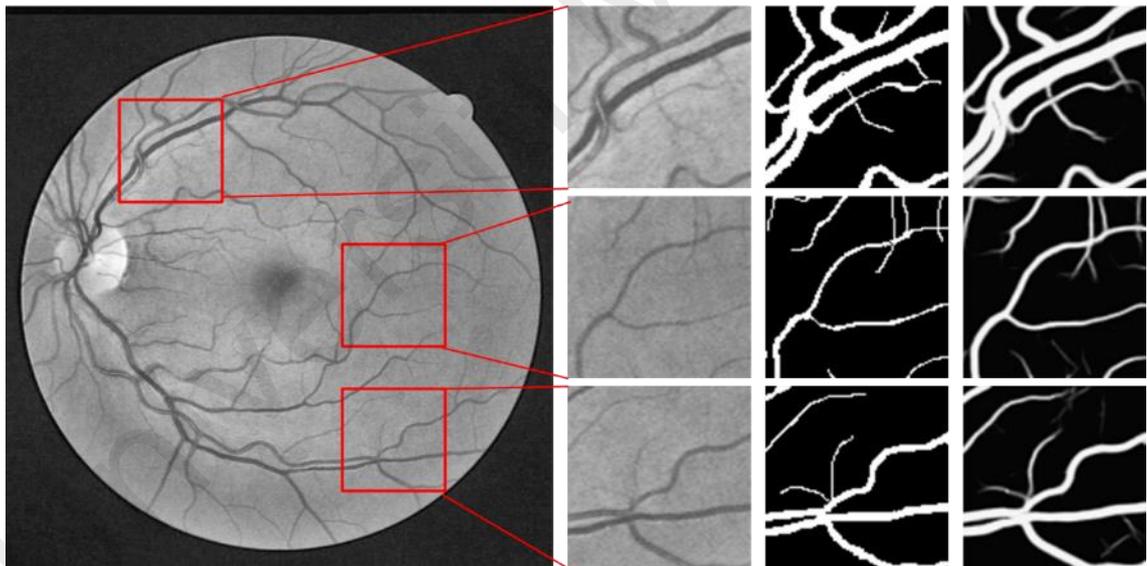
## 4.5 Limitation

Figure 4.7, Figure 4.8 and Figure 4.9 show some enlarged image patches, where we can view more details. We can see that although the basic structure of vessels can be recognized completely, some thin vessels and vessel boundary still cannot be identified

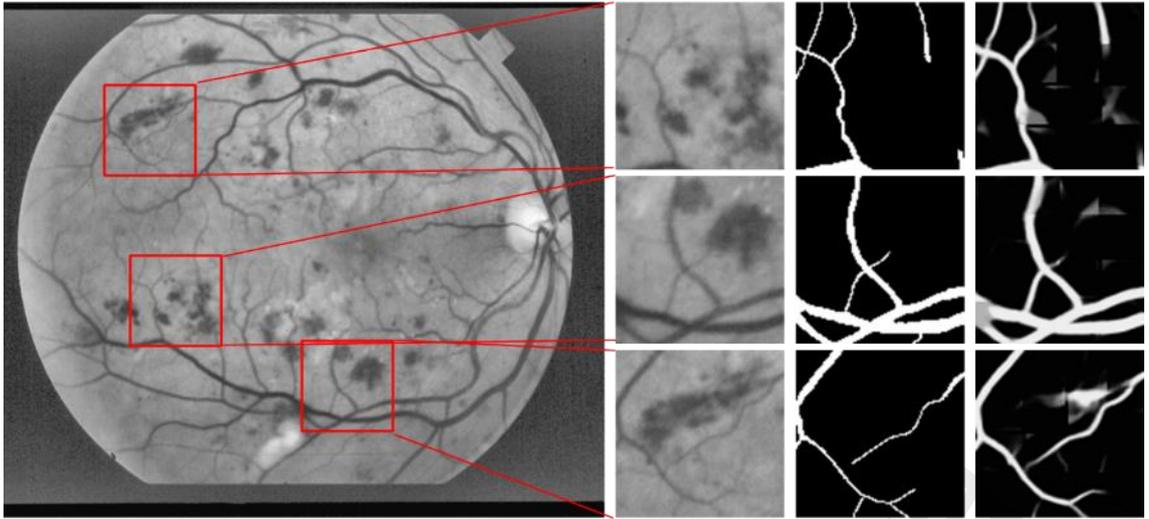
correctly. At the same time, some vessels are fractured, i.e. they are not connected with other vessels.

Figure 4.8 indicates a sample with many abnormalities. There are many non-vessel pixels are classified as vessel pixels due to the abnormality, thus there are some isolated bright blocks in the segmentation results. Figure 4.9 indicates a sample from CHASE\_DB1 database, which has a terrible background condition in the preprocessed image, hence some background pixels are also identified as vessel pixels due to the disturbance of background noise, which lower the precision of segmentation results.

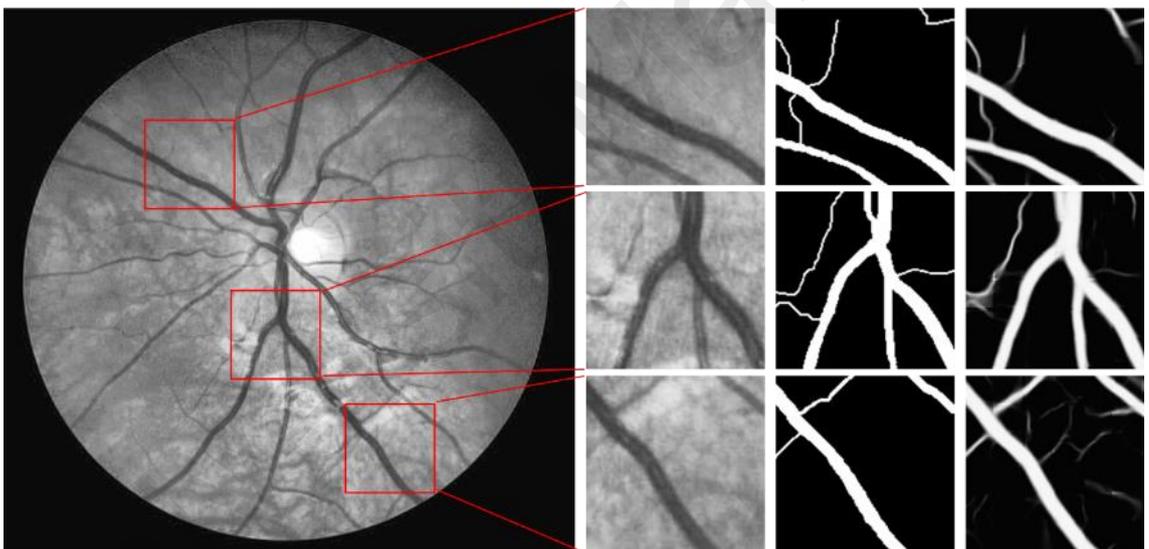
The generalization capacity still should be improved, especially between CHASE\_DB1 and DRIVE/STARE databases.



**Figure 4.7: Enlarged image patches of image in DRIVE database. From left to right: preprocessed image, enlarged image patches, ground truth image patches, segmentation result patches.**



**Figure 4.8: Enlarged image patches of image in STARE database. From left to right: preprocessed image, enlarged image patches, ground truth image patches, segmentation result patches.**



**Figure 4.9: Enlarged image patches of image in CHASE\_DB1 database. From left to right: preprocessed image, enlarged image patches, ground truth image patches, segmentation result patches.**

#### 4.6 Summary

This chapter indicates the test results of the proposed performance on three public databases. We showed the overall segmentation of part samples in these databases. We can see a completed vessel structure in these overall segmentation, thick vessels and most thin vessels are identified correctly. We also conducted quantitative analysis to the performance of the proposed model, we calculated several import evaluation metrics and

compared them with other recent works. The proposed model produced a good performance compared with other publications.

Universiti Malaya

## CHAPTER 5: CONCLUSION AND FEATURE WORK

### 5.1 Conclusion

Retinal fundus images can be captured easily and non-invasively. Retina abnormality can appear in various types, but the change of retinal vessel is the most common. Retinal vessel is the only structure that can be observed intensively and directly. It is also closely relative with human blood circulation, so the status of retinal vessels can reflect many diseases, such as diabetic retinopathy (DR) and hypertension. Extracting retinal vessels is a crucial procedure for ophthalmologist to make diagnosis.

In this project, we proposed a deep learning-based model for automatic retinal vessel segmentation. It is a challenging task since the retinal fundus images are under low contrast and uneven illumination. At the same time, various structures, image noise and uneven texture also hinder the classification of pixels, thus we applied image processing technique to enhance the quality of raw images. We adopted residual learning and dilation convolution to construct basic building block, then we leveraged these building blocks to format a U-net and obtained the final model by cascading two U-nets. We selected three public databases and conducted intensive experiments to evaluate the performance of the proposed model. The model can produce an over 95% accuracy on all three databases.

The experiments indicate the proposed model can be used to provide a good reference for ophthalmologist, and it can also be applied to other tasks after appropriate adjustment and improvement.

### 5.2 Future work

The proposed model can produce good performance in public databases, but there still exist some limitations for the method. In the future, the project can be extended for these research works.

1. The model makes predictions according to contextual information, but the model can only leverage less information due to limitation of computation capacity and receptive field, although we have adopted dilation convolution. There are still some vessels that cannot be segmented well, especially for thin vessels and boundary pixels. Hence, we should introduce a new method to leverage more information as well as reducing the computation burden.
2. Ophthalmologist makes diagnosis according to not only vessel appearance but also geometric parameters. Thus, we could measure parameters such as vessel width after vessel segmentation in the future.
3. We also should consider multi-task segmentation/detection, since there often exists other abnormalities in fundus images, such as soft/hard exudates and hemorrhages. We can also consider joint segmentation of optic disc and blood vessels.
4. The proposed method is supervised method, which requires massive data to train the model. However, the labeled training sample are rare, although we leveraged data augmentation to enlarge the databases. Therefore, we can consider semi-supervised method, combining supervised and unsupervised methods, to make full use of massive unlabeled data.

## REFERENCES

- Al-Rawi, M., Qutaishat, M., & Arrar, M. (2007). An improved matched filter for blood vessel detection of digital retinal images. *Computers in Biology and Medicine*, 37(2), 262-267.
- Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007). *Greedy layer-wise training of deep networks*. Paper presented at the Advances in neural information processing systems.
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2), 157-166.
- Boureau, Y.-L., Ponce, J., & LeCun, Y. (2010). *A theoretical analysis of feature pooling in visual recognition*. Paper presented at the Proceedings of the 27th international conference on machine learning (ICML-10).
- Budak, Ü., Cömert, Z., Çıbuk, M., & Şengür, A. (2020). DCCMED-Net: Densely connected and concatenated multi Encoder-Decoder CNNs for retinal vessel extraction from fundus images. *Medical Hypotheses*, 134, 109426.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *Imagenet: A large-scale hierarchical image database*. Paper presented at the 2009 IEEE conference on computer vision and pattern recognition.
- FARNELL, D. J. (2006). Automated Retinal Image Analysis (ARIA) Data Set. Retrieved from [http://www.damianjifarnell.com/?page\\_id=276](http://www.damianjifarnell.com/?page_id=276)
- Farnell, D. J., Hatfield, F., Knox, P., Reakes, M., Spencer, S., Parry, D., & Harding, S. P. (2008). Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. *Journal of the Franklin institute*, 345(7), 748-765.
- Fathi, A., & Naghsh-Nilchi, A. R. (2013). Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation. *Biomedical Signal Processing and Control*, 8(1), 71-80.
- Francia, G. A., Pedraza, C., Aceves, M., & Tovar-Arriaga, S. (2020). Chaining a U-Net With a Residual U-Net for Retinal Blood Vessels Segmentation. *IEEE Access*, 8, 38493-38500.
- Franklin, S. W., & Rajan, S. E. (2014). Computerized screening of diabetic retinopathy employing blood vessel segmentation in retinal images. *Biocybernetics and Biomedical Engineering*, 34(2), 117-124.
- Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., & Barman, S. A. (2012). Blood vessel segmentation methodologies in retinal images – A survey. *Computer Methods and Programs in Biomedicine*, 108(1), 407-433. doi:<https://doi.org/10.1016/j.cmpb.2012.03.009>

- Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., & Barman, S. A. (2012). An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9), 2538-2548.
- Glorot, X., & Bengio, Y. (2010). *Understanding the difficulty of training deep feedforward neural networks*. Paper presented at the Proceedings of the thirteenth international conference on artificial intelligence and statistics.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). *Generative adversarial nets*. Paper presented at the Advances in neural information processing systems.
- Group, E. T. D. R. S. R. (1991). Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS report number 12. *Ophthalmology*, 98(5), 823-833.
- Guo, S., Wang, K., Kang, H., Zhang, Y., Gao, Y., & Li, T. (2019). BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation. *Int J Med Inform*, 126, 105-113. doi:10.1016/j.ijmedinf.2019.03.015
- Hinton, G. (2010). A Practical Guide to Training Restricted Boltzmann Machines. *Momentum*, 9, 1.
- Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786), 504-507.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., . . . Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Hu, J., Wang, H., Gao, S., Bao, M., Liu, T., Wang, Y., & Zhang, J. (2019). S-UNet: A Bridge-Style U-Net Framework With a Saliency Mechanism for Retinal Vessel Segmentation. *IEEE Access*, 7, 174167-174177. doi:10.1109/access.2019.2940476
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1), 215-243.
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
- Jiang, Y., Tan, N., Peng, T., & Zhang, H. (2019). Retinal Vessels Segmentation Based on Dilated Multi-Scale Convolutional Neural Network. *IEEE Access*, 7, 76342-76352. doi:10.1109/access.2019.2922365

- Jiang, Z., Zhang, H., Wang, Y., & Ko, S. B. (2018). Retinal blood vessel segmentation using fully convolutional network with transfer learning. *Comput Med Imaging Graph*, 68, 1-15. doi:10.1016/j.compmedimag.2018.04.005
- Kaba, D., Salazar-Gonzalez, A. G., Li, Y., Liu, X., & Serag, A. (2013). *Segmentation of retinal blood vessels using gaussian mixture models and expectation maximisation*. Paper presented at the International Conference on Health Information Science.
- Kanski, J. J., & Bowling, B. (2011). *Clinical ophthalmology: a systematic approach*: Elsevier Health Sciences.
- Khalaf, A. F., Yassine, I. A., & Fahmy, A. S. (2016, 25-28 Sept. 2016). *Convolutional neural networks for deep feature learning in retinal vessel segmentation*. Paper presented at the 2016 IEEE International Conference on Image Processing (ICIP).
- Köhler, T., Budai, A., Kraus, M. F., Odstrčilik, J., Michelson, G., & Hornegger, J. (2013). *Automatic no-reference quality assessment for retinal fundus images using vessel segmentation*. Paper presented at the Proceedings of the 26th IEEE international symposium on computer-based medical systems.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks*. Paper presented at the Advances in neural information processing systems.
- LaRocca, F., Nankivil, D., Farsiu, S., & Izatt, J. A. (2014). True color scanning laser ophthalmoscopy and optical coherence tomography handheld probe. *Biomedical optics express*, 5(9), 3204-3216.
- LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. (1990). *Handwritten digit recognition with a back-propagation network*. Paper presented at the Advances in neural information processing systems.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- Li, L., Verma, M., Nakashima, Y., Nagahara, H., & Kawasaki, R. (2020). *Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks*. Paper presented at the The IEEE Winter Conference on Applications of Computer Vision.
- Li, T., Bo, W., Hu, C., Kang, H., Liu, H., Wang, K., & Fu, H. (2021). Applications of deep learning in fundus images: A review. *Medical image analysis*, 101971.
- Li, X., Jiang, Y., Li, M., & Yin, S. (2020). Lightweight Attention Convolutional Neural Network for Retinal Vessel Segmentation. *IEEE Transactions on Industrial Informatics*.
- Lian, S., Li, L., Lian, G., Xiao, X., Luo, Z., & Li, S. (2019). A Global and Local Enhanced Residual U-Net for Accurate Retinal Vessel Segmentation. *IEEE/ACM Trans Comput Biol Bioinform*. doi:10.1109/TCBB.2019.2917188

- Lin, M., Chen, Q., & Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). *Feature pyramid networks for object detection*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., . . . Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60-88.
- Liu, Y. P., Rui, X., Li, Z., Zeng, D., Li, J., Chen, P., & Liang, R. (2021). Feature pyramid U - Net for retinal vessel segmentation. *IET Image Processing*.
- Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Luo, Y., Cheng, H., & Yang, L. (2016, 13-16 Dec. 2016). *Size-Invariant Fully Convolutional Neural Network for vessel segmentation of digital retinal images*. Paper presented at the 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA).
- Luo, Z., Zhang, Y., Zhou, L., Zhang, B., Luo, J., & Wu, H. (2019). Micro-Vessel Image Segmentation Based on the AD-UNet Model. *IEEE Access*, 7, 143402-143411. doi:10.1109/access.2019.2945556
- Lv, Y., Ma, H., Li, J., & Liu, S. (2020). Attention Guided U-Net With Atrous Convolution for Accurate Retinal Vessels Segmentation. *IEEE Access*, 8, 32826-32839. doi:10.1109/access.2020.2974027
- Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). *Rectifier nonlinearities improve neural network acoustic models*. Paper presented at the Proc. icml.
- Mhaskar, H. N., & Micchelli, C. A. (1994). *How to choose an activation function*. Paper presented at the Advances in Neural Information Processing Systems.
- Mishra, S., Chen, D. Z., & Hu, X. S. (2020). *A Data-Aware Deep Supervised Method for Retinal Vessel Segmentation*. Paper presented at the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI).
- Mo, J., & Zhang, L. (2017). Multi-level deep supervised networks for retinal vessel segmentation. *Int J Comput Assist Radiol Surg*, 12(12), 2181-2193. doi:10.1007/s11548-017-1619-0
- Mou, L., Chen, L., Cheng, J., Gu, Z., Zhao, Y., & Liu, J. (2019). Dense Dilated Network with Probability Regularized Walk for Vessel Detection. *IEEE transactions on medical imaging*, 39(5), 1392-1403.
- Nair, V., & Hinton, G. E. (2010). *Rectified linear units improve restricted boltzmann machines*. Paper presented at the ICML.

- Oliveira, A., Pereira, S., & Silva, C. A. (2018). Retinal vessel segmentation based on Fully Convolutional Neural Networks. *Expert Systems with Applications*, 112, 229-242. doi:10.1016/j.eswa.2018.06.034
- Park, K.-B., Choi, S. H., & Lee, J. Y. (2020). M-GAN: Retinal Blood Vessel Segmentation by Balancing Losses Through Stacked Deep Fully Convolutional Networks. *IEEE Access*, 8, 146308-146322.
- Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabudhe, V., & Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. *Data*, 3(3), 25.
- Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., . . . Iyengar, S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, 51(5), 1-36.
- Prentašić, P., Lončarić, S., Vatauvuk, Z., Benčić, G., Subašić, M., Petković, T., . . . Tadić, R. (2013). *Diabetic retinopathy image database (DRiDB): a new database for diabetic retinopathy screening programs research*. Paper presented at the 2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA).
- Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Cham.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Soares, J. V., Leandro, J. J., Cesar, R. M., Jelinek, H. F., & Cree, M. J. (2006). Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE transactions on medical imaging*, 25(9), 1214-1222.
- Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., & Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4), 501-509.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). *Going deeper with convolutions*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Upadhyay, K., Agrawal, M., & Vashist, P. (2021). *U-Net based Multi-level Texture Suppression for Vessel Segmentation in Low Contrast Regions*. Paper presented at the 2020 28th European Signal Processing Conference (EUSIPCO).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). *Attention is all you need*. Paper presented at the Advances in neural information processing systems.

- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., & Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12).
- Wang, B., Wang, S., Qiu, S., Wei, W., Wang, H., & He, H. (2020). CSU-Net: A Context Spatial U-Net for Accurate Blood Vessel Segmentation in Fundus Images. *IEEE J Biomed Health Inform*, PP. doi:10.1109/JBHI.2020.3011178
- Wang, K., Zhang, X., Huang, S., Wang, Q., & Chen, F. (2020). *CTF-Net: Retinal Vessel Segmentation via Deep Coarse-To-Fine Supervision Network*. Paper presented at the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI).
- Wang, T., Wu, D. J., Coates, A., & Ng, A. Y. (2012). *End-to-end text recognition with convolutional neural networks*. Paper presented at the Proceedings of the 21st international conference on pattern recognition (ICPR2012).
- Webb, R. H., & Hughes, G. W. (1981). Scanning laser ophthalmoscope. *IEEE Transactions on Biomedical Engineering*(7), 488-492.
- Wu, C., Zou, Y., & Yang, Z. (2019). *U-GAN: Generative Adversarial Networks with U-Net for Retinal Vessel Segmentation*. Paper presented at the 2019 14th International Conference on Computer Science & Education (ICCSE).
- Wu, Y., Xia, Y., Song, Y., Zhang, Y., & Cai, W. (2020). NFN+: A novel network followed network for retinal vessel segmentation. *Neural networks*.
- Xia, H., Zhuge, R., & Li, H. (2018). *Retinal vessel segmentation via a coarse-to-fine convolutional neural network*. Paper presented at the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
- Yan, Z., Yang, X., & Cheng, K. (2018). Joint Segment-Level and Pixel-Wise Losses for Deep Learning Based Retinal Vessel Segmentation. *IEEE Transactions on Biomedical Engineering*, 65(9), 1912-1923.
- Yan, Z., Yang, X., & Cheng, K. T. (2019). A Three-Stage Deep Learning Model for Accurate Retinal Vessel Segmentation. *IEEE J Biomed Health Inform*, 23(4), 1427-1436. doi:10.1109/JBHI.2018.2872813
- Yin, Y., Adel, M., & Bourennane, S. (2012). Retinal vessel segmentation using a probabilistic tracking method. *Pattern Recognition*, 45(4), 1235-1244.
- You, X., Peng, Q., Yuan, Y., Cheung, Y.-m., & Lei, J. (2011). Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. *Pattern Recognition*, 44(10-11), 2314-2324.
- Zhang, Y., & Chung, A. C. (2018). *Deep supervision with additional labels for retinal vessel segmentation task*. Paper presented at the International conference on medical image computing and computer-assisted intervention.

Zhuo, Z., Huang, J., Lu, K., Pan, D., & Feng, S. (2020). A size-invariant convolutional network with dense connectivity applied to retinal vessel segmentation measured by a unique index. *Computer Methods and Programs in Biomedicine*, 105508.

Universiti Malaya