



Retinal Image Analysis Based on Deep Learning

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy

by

Baidaa Al-Bander

June 2018

Abstract

The interpretation of ophthalmic images is typically performed by trained clinical experts. However, due to the volume and complexity of these images, and the large variation in pathology, in addition to the variation among experts, there has been increasing interest in computer-assisted assessment and diagnosis of such images. There has been particular interest in finding a cost-effective approach with high sensitivity and specificity, independent of human intervention, and robust enough to be applied to large populations in a timely manner to identify retinal diseases.

This thesis introduces novel deep learning methodologies based on convolutional neural networks (CNNs) to address key challenges in different retinal image analysis tasks. Three retinal image analysis objectives have been considered in this research project: fovea and optic disc (OD) localisation, choroid and optic disc/cup segmentation, and disease and lesion classification tasks.

In the first retinal image analysis task, simultaneous detection of the centres of the fovea and the optic disc from colour fundus images is considered as a regression problem. A deep multi-scale sequential CNN is designed and trained. The proposed method achieves an accuracy of 97%, 96.7% for the detection of the OD centre, and 96.6%, 95.6% for the detection of the foveal centre of the MESSIDOR and Kaggle test sets respectively. These promising results demonstrate the excellent performance of the proposed CNNs in simultaneously detecting the centres of both the fovea and OD without human intervention or hand-crafted features. Moreover, the landmarks of an image can be localised in 0.007 seconds.

In the second phase, two segmentation schemes are presented to identify interesting regions in fundus and optical coherence tomography (OCT) images. In the first scheme, a new framework that features three stages: partitioning OCT images into super-pixels combining spatial, intensity and texture information; classification using CNNs; and refinement of the segmentation has been proposed. The proposed framework achieves state-of-the-art accuracy of 0.986%, demonstrating its effectiveness in segmenting the choroid which can pave the way for more accurate diagnoses of a range of diseases. The second scheme defines the method to segment the OD and the optic cup (OC) in fundus images. The core of the proposed method is DenseNet incorporated with a fully convolutional network for

accurate and robust segmentation. The predicted boundaries of the OD and OC are then used to estimate the cup-to-disc ratio (CDR) for glaucoma diagnosis. Our method is generalised to segment five datasets of fundus images taken from different devices, outperforming the state of the art on three and achieving comparable results on the remaining two.

Finally, based on the image classification concept, three new feature learning approaches for detecting glaucoma using colour retinal fundus images, grading the severity of diabetic macular oedema (DMO) on colour retinal fundus images and identifying retinal lesions in OCT volumes have been presented. Developing these automated diagnosis systems based on the feature learning approach helps early diagnosis of the disease and thus averts (or delays) its progression. The proposed systems relying on CNN and transfer learning concept are capable of identifying and extracting features that are characteristic of glaucoma and DMO diseases, and retinal lesions automatically without the need of any kind of user intervention.

Overall, new automation methods based on deep learning have been developed and demonstrated to be effective in addressing the main weaknesses of traditional systems where image features are extracted manually. This kind of automated systems could provide ophthalmologists with novel strategies to identify retinal diseases from fundus and OCT images. These presented deep feature learning based approaches will be valuable tools to be used for these and other challenging medical problems. Furthermore, the integration of automated grading systems will also significantly reduce the manual grading workload and thus help towards reducing unnecessary costly referrals.

Contents

Abstract	i
Notations	xiii
Acknowledgements	xv
1 Introduction	1
1.1 Background	1
1.2 Problem Statement and Motivation	2
1.3 Aim and Objectives	5
1.4 Novelty and Contribution	6
1.5 Thesis Structure	9
2 Background	11
2.1 The Eye and Retinal Diseases	11
2.1.1 Anatomy of the Eye	11
2.1.2 Retinal Diseases	14
Age-related Macular Degeneration	15
Glaucoma	17
Diabetic Retinopathy and Maculopathy	17
2.2 Retinal Imaging Modalities	20
2.2.1 Fundus Photography	20
2.2.2 Optical Coherence Tomography	21
2.3 Medical Image Analysis	23
2.3.1 Low-level Image Analysis	24
Enhancement	24
Noise Reduction	25
2.3.2 High-level Image Analysis	25
Segmentation	26
Classification/Detection	27
2.4 Artificial Intelligence, Machine Learning and Deep Learning	28
2.4.1 Feature Engineering and Learning	29
2.4.2 Deep Learning Architectures	31
2.4.3 General Components of Deep CNNs	33

2.4.4	Deep Convolutional Neural Network Architectures	45
2.5	Applications of Deep CNNs in Medical Image Analysis	50
2.6	Summary	53
3	Detection of Optic Disc and Fovea	54
3.1	Introduction	54
3.2	Related Work	55
3.3	Novelty	62
3.4	Materials and Methods	63
3.4.1	Materials	63
3.4.2	Pre-processing	64
3.4.3	Landmarks Detection by CNN	65
3.4.4	Performance Evaluation	68
3.5	Results	69
3.6	Discussion	73
3.7	Summary	78
4	Choroid and Optic Disc/Cup Segmentation	79
4.1	Choroid Segmentation in OCT Images	80
4.1.1	Introduction	80
4.1.2	Related Work	81
4.1.3	Materials and Method	85
	Dataset	86
	ROI Extraction	87
	Supapixel Segmentation	88
	CNN for Patch Labelling	90
	Refining	91
4.1.4	Results and Discussion	93
4.1.5	Conclusion	100
4.2	Optic Disc/Cup Segmentation in Fundus Images	102
4.2.1	Introduction	102
4.2.2	Related Work	103
4.2.3	Materials and Methods	107
	Image Datasets	107
	Methods	108
4.2.4	Results and Discussion	111
4.2.5	Conclusion	122
4.3	Summary	122
5	Retinal Disease and Lesion Classification	123
5.1	Automated Glaucoma Grading	123
5.1.1	Introduction	123
5.1.2	Related Work	124
5.1.3	Material and Method	125
	Data	125

	Pre-processing	125
	Feature Extraction using CNN	125
	Training SVM using CNN Features and Predication	126
5.1.4	Results and Discussion	128
5.1.5	Conclusion	129
5.2	Automated Diabetic Macular Oedema Grading	130
5.2.1	Introduction	130
5.2.2	Related Work	130
5.2.3	Material and Method	133
	Data	133
	Pre-processing	133
	Network Design and Training	134
	Testing	136
5.2.4	Results and Discussion	137
5.2.5	Conclusion	138
5.3	Retinal Lesion Classification	139
5.3.1	Introduction	139
5.3.2	Related Work	140
5.3.3	Materials and Methods	142
	Data	142
	Pre-processing	142
	Multi-label Classification with CNN	143
	Ensemble Learning and Parameter Selection	149
5.3.4	Results and Discussion	150
5.3.5	Conclusion	152
5.4	Summary	153
6	Conclusions and Future Work	154
6.1	Summary	154
6.2	Discussion	156
6.3	Future Work	157
6.4	Conclusion	159

Bibliography	160
---------------------	------------

Illustrations

List of Figures

2.1	The anatomical structure of the human eye [12].	12
2.2	Cross section of the human retina showing the anatomical structure [15].	13
2.3	An example fundus image illustrating the key retinal anatomical structures.	14
2.4	Vision with Age-related macular degeneration (AMD). AMD causes loss of sight in the centre of the field of vision [17].	15
2.5	Fluid leakage affecting the retina [19].	16
2.6	Fluids as visualised in OCT images [19]. IRF: Intraretinal fluid, SRF: Subretinal fluid, PED: Pigment Epithelial Detachment.	16
2.7	Visual loss in glaucoma. Visual field loss begins at the periphery and moves towards the centre [22].	17
2.8	Fundus image shows the optic nerve head from a healthy subject [24].	18
2.9	Abnormal vision effected by diabetic macular oedema condition associated with diabetic retinopathy [17].	19
2.10	Colour fundus image captured by Zeiss camera.	21
2.11	Normal retina imaged with OCT from three different devices. (a) Cirrus (b) Spectralis (c) Topcon [19].	23
2.12	An example of contrast image enhancement. (a) Original (b) Enhanced.	25
2.13	Image denoising by filtering technique. (a) Original image (b) Image corrupted with Gaussian noise (c) Denoised image by Wiener filter. . .	26
2.14	Deep learning is a subset of representation learning which aims at finding an appropriate representation of input data in order to perform a machine learning task which is in turn a subset of artificial intelligence. Adapted from [59].	29
2.15	CPUs versus GPUs. Each green square represents a core in the GPU while the blue square refers to CPU [60].	30
2.16	Deep learning versus machine learning [62].	31
2.17	Features are generated by a feature learning technique. From left to right: low-level feature, mid-level-feature, and high-level feature [63]. .	32
2.18	Diagram shows convolution operation between a sub-image (I) and kernel (K) [78].	34

2.19	An example of a convolutional layer. Let the input is 32 feature maps and the output is 64 feature maps, the size of the filter is 3×3 then the size of filter in the input space is $3 \times 3 \times 32$. Thus, the total number of weights is $3 \times 3 \times 32 \times 64$	34
2.20	Max and Average pooling layer.	35
2.21	ReLU activation function work	36
2.22	Comparison between two networks without and with dropout. (a) Without dropout (b) With dropout [82].	40
2.23	Training loss of CNN over 12 epochs with different weight initialisation schemes [87].	42
2.24	NVIDIA GPU which was used to run our experiments [100].	45
2.25	AlexNet architecture [69].	46
2.26	VGG architecture [103].	47
2.27	Performance of different CNN models in a classification task [112]. . .	47
2.28	Inception block [104].	48
2.29	Residual block [110].	49
2.30	Dense block [118].	50
2.31	Advanced (block-based) deep convolutional neural network architectures.	50
3.1	Image pre-processing stages. From left to right: Original image, Grayscale image, and Enhanced image.	65
3.2	Block diagram of the proposed system.	66
3.3	Block diagram of convolutional neural network.	68
3.4	Performance of the conventional neural network (ANN) during training. It shows that simple model suffers from an under-fitting problem where the complexity of the network is not sufficient to capture the import features of the landmarks.	69
3.5	Performance of the deep neural network during training. It shows that the RMSE for both training and validation data is lower than the conventional neural network with slightly overfitting and thus better landmark detection performance.	70
3.6	Example shows $1R$, $0.5R$, and $0.25R$ of OD.	71
3.7	Examples of correct joint OD-Fovea detection results from MESSIDOR. The green plus signs refer to the locations annotated by ophthalmologists while the blue ones indicate the results of our proposed method.	74

3.8	Examples of correct joint OD-Fovea detection results from Kaggle. The green plus signs refer to the locations annotated by ophthalmologists while the blue ones indicate the results of our proposed method.	75
3.9	Examples of incorrect OD and fovea detection results. From left to right (1) Incorrect detection from MESSIDOR; (2) Incorrect detection from Kaggle. The green plus signs refer to the locations annotated by ophthalmologists while the blue ones indicate the results of our proposed method	76
3.10	Examples of fundus images show the original centers (green plus), centres obtained from CNN1 (white plus), and centres from CNN2 (blue plus). It is clear that CNN2 improves the location accuracy.	76
4.1	Example of conventional OCT image shown in the first row is compared with EDI-OCT image shown in the second row. It shows how EDI-OCT provides better information and visualisation for the choroid region.	81
4.2	Block diagram shows the main stages of the proposed framework. . . .	85
4.3	An EDI-OCT image of the choroid (first row) is segmented to extract the sub-RPE region which is partitioned with a superpixel technique (second row). Distance information and enhanced intensities are then calculated and combined with the image (third row) for the CNN classification of each superpixel. The CNN result is refined (forth row) [green contour] and smoothed to give the final result [red contour] which is close to the annotated segmentation [blue contour].	86
4.4	CNN Architecture	92
4.5	Results of our model for four test images comparing our result (red curve) with the expert grading (Grader1Round1)(blue curve).	96
4.6	Examples of our results (red contour) with grader one's first annotation (green), grader one's second annotation (blue) and grader two's annotation (cyan).	99
4.7	Example of segmentation of a three dimensional EDI-OCT choroid dataset.	100
4.8	An example fundus image showing the optic disc and cup with their boundary contours shown in blue.	103

4.9	Block diagram of the proposed optic disc and cup segmentation system. (a) Methodology and fully convolutional DenseNet architecture, (b) Dense Blocks (DB), (c) One layer in DB, (d) Transition Down block (TD), (e) Transition Up block (TU). The circle (C) refers to concatenation process. Note, red and blue represent the cropped rim and OC respectively in the segmentation.	109
4.10	Examples of joint OD-OC segmentation results. From the first row to the fifth row, the examples are from the Origa (OR), DRIONS-DB (DD), Drishti-GS (DG), ONHSD (ON), and RIM-ONE (RO) respectively. The green contour refers to the ground truth provided with the images while the blue one indicates the results of our proposed method. The DRIONS-DB and ONHSD show the contour of OD only because the ground truth for OC is not provided.	117
4.11	From left to right: ROC curves of the expert's glaucoma diagnosis and the proposed system.	119
5.1	Proposed system stages of glaucoma grading.	126
5.2	Confusion matrix shows prediction performance evaluation (0: normal; 1: glaucomas).	128
5.3	A block diagram showing the comparison between the traditional systems and our proposed system. The first row shows the proposed system and the second row is the traditional system. The proposed system does not depend on any kind of segmentation, hand-crafted feature and sums up many stages of the traditional system in only two automatic stages. CSMO: Clinically Significant Macular Oedema. . . .	133
5.4	Proposed multi-stage convolutional neural network architecture. . . .	135
5.5	Prediction ratio per class in confusion matrix, (0) refers to normal images, (1) non-CSMO and (2) CSMO. It shows good prediction results despite the rarity of classes 1 and 2.	137
5.6	OCT images with IRF fluid. The images on the first row are the original images while the images on the second row are the denoised and enhanced images. The desired histogram of the enhanced images is specified by the histogram of Spectralis for the devices Cirrus, Spectralis, and Topocon, respectively.	144

5.7	OCT images with SRF fluid. The images on the first row are the original images while the images on the second row are the denoised and enhanced images. The desired histogram of the enhanced images is specified by the histogram of Spectralis for the devices Cirrus, Spectralis, and Topocon, respectively.	144
5.8	OCT images with PED fluid. The images on the first row are the original images while the images on the second row are the denoised and enhanced images. The desired histogram of the enhanced images is specified by the histogram of Spectralis for the devices Cirrus, Spectralis, and Topocon, respectively.	145
5.9	Schematic diagram of Inception v3 network architecture [270].	146
5.10	Ensemble of CNN combined by weighted sum. DenH1: Denoising and Spectralis image histogram, DenH2: Denoising and Cirrus image histogram, DenH3: Denoising and Topcon image histogram, W: Weight.	150
5.11	Area under curve for retinal fluid detection.	151

List of Tables

2.1	Weight initialisation schemes. n_{in} , n_{out} is the number of nodes in the previous and next layers.	41
3.1	Existing optic disc (OD) and fovea (F) detection methods in the literature.	58
3.2	Architecture of deep neural network with the best detection performance	67
3.3	Performance of different networks: The networks are trained and tested on Messidor dataset (1R criterion). These reported results are obtained from the first stage of the proposed system.	71
3.4	Performance (in terms of accuracy) of the network trained on Kaggle and tested on Messidor. CNN1 refers to the first stage and CNN2 refers to the second stage.	72
3.5	Performance (in terms of accuracy) of the network trained and tested on kaggle. CNN1 refers to the first stage and CNN2 refers to the second stage.	73
4.1	CNN Architecture	91
4.2	The performance of patch classification in the CNN considering the patches size, the type and number of channels used in the image passed to CNN.	95
4.3	Results of our method and competing methods	97

4.4	Mean accuracy of our results compared to two different annotations from one expert grader (Grad 1) and an annotation from a second expert grader (Grad 2). From the calculations, each of the graders gave similar annotations and our results are very close.	97
4.5	Accuracies of our results (OM) compared to two different annotations from one expert grader (G1A1 and G1A2) and an annotation from a second expert grader (G2). The entries in bold indicate our results achieving closer segmentation than an expert grading.	98
4.6	Comparison with the existing methods in the literature for only OD segmentation on different datasets.	112
4.7	Comparison with the existing methods in the literature for only OC segmentation on different datasets.	115
4.8	Comparison with the existing methods in the literature for joint OC and OD segmentation on different datasets.	115
4.9	Results of the proposed method for OD and OC segmentation on Origa dataset compared with the existing methods in the literature. .	116
4.10	The optic disc segmentation performance on the DRIONS-DB dataset considering different data processing schemes. The network is trained on Origa dataset only.	116
4.11	The optic disc, cup, and rim segmentation performance on the Drishti-GS dataset considering different data processing schemes. The network is trained on the Origa dataset only.	118
4.12	The optic disc segmentation performance on the ONHSD dataset considering different data processing schemes. The network is trained on the Origa dataset only.	118
4.13	The optic disc, cup, and rim segmentation performance on the RIM-ONE dataset considering different data processing schemes. The network is trained on the Origa dataset only.	119
4.14	Results of the proposed method for OD and OC segmentation on Drishti-GS dataset compared with the existing methods in the literature.	121
4.15	Results of the proposed method for OD and OC segmentation on RIM-ONE dataset compared with the existing methods in the literature. .	121
4.16	Results of the proposed method for OD segmentation on DRIONS-DB dataset compared with the existing methods in the literature. . .	121
4.17	Results of the proposed method for OD segmentation on ONHSD dataset compared with the existing methods in the literature.	121

5.1	Architecture of AlexNet.	127
5.2	Accuracy detection in terms of B-scans of three fluids in each network. DenH1: Denoising and Spectralis image histogram, DenH2: Denoising and Cirrus image histogram, DenH3: Denoising and Topcon image his- togram, IRF: Intraretinal fluid, SRF: Subretinal fluid, PED, Pigment epithelium detachment.	149
5.3	Area under curve of each fluid in each method considering the vol- umes. B-scan scheme represents the AUC calculation without taking the volumes into consideration	151

Notations

The following notations and abbreviations are found throughout this thesis:

AMD: Age-Related Macular Degeneration
ANN: Artificial Neural Network
AUROC: Area Under The Receiving Operator Curve
BM: Bruch's Membrane
BN: Batch Normalisation
CADS: Computer-Aided Diagnosis Systems
CDR: Cup To Disc Ratio
CLAHE: Contrast-Limited Adaptive Histogram Equalisation
CNNs: Convolutional Neural Networks
CRF: Conditional Random Field
CSMO: Clinically Significant Macular Oedema
CSR: Coupled Shape Regression
CT: Computed Tomography
DBMs: Deep Boltzmann Machines
DBNs: Deep Belief Networks
DC: Dice's Coefficient
DenseNet: Dense Network
DL: Deep Learning
DM: Diabetic Maculopathy
DMO: Diabetic Macular Oedema
DR: Diabetic Retinopathy
EDI: Enhanced Depth Imaging
ETDRS: Early Treatment Diabetic Retinopathy Study
EXs: Exudates
FA: Fluorescein Angiography
FC: Fully Connected
FCN: Fully Convolutional Network
FOV: Field of View
GPUs: Graphical Processing Units

HA: Haemorrhage
HD-OCT: High Definition Optical Coherence Tomography
HE: Hard Exudate
IOP: Intra Ocular Pressure
IRF: Intraretinal Fluid
ISNT: Inferior-Superior-Nasal-Temporal
K-NN: K-Nearest Neighbours
LSTM: Long Short-Term Memory
MAs: Microaneurysms
ML: Machine Learning
MLPs: Multilayer Perceptron Neural Networks
MSE: Mean Squared Error
Non-CSMO: Non-Clinically Significant Macular Oedema
NPDR: Non-Proliferative Diabetic Retinopathy
NRR: Neuroretinal Rim
OC: Optic Cup
OCT: Optical Coherence Tomography
OD: Optic Disc
ODD: Optic Disc Diameter
ONH: Optic Nerve Head
PDR: Proliferative Diabetic Retinopathy
PED: Pigment Epithelial Detachment
ReLU: Rectified Linear Unit
ResNets: Residual Network
RMSE: Root Mean Squared Error
ROC: Receiver Operating Curve
ROI: Region of Interest
RPE: Retinal Pigment Epithelium
SGD: Stochastic Gradient Descent
SRF: Subretinal Fluid
SVM: Support Vector Machine
TC: Tanimoto Coefficient
TD-OCT: Time Domain Optical Coherence Tomography

Acknowledgements

My deep thanks and appreciations go to my supervisors, Dr. Waleed Al-Nuaimy and Dr. Yalin Zheng who supervised and guided this work. Dr. Al-Nuaimy's advice and guidance throughout my research have been invaluable. It has also been a great opportunity to work with Dr. Zheng in his research group (CRiA) in the Eye and Vision Science Department, University of Liverpool; his meticulous scrutiny, scholarly advice, dedication, and keen interest to help his students have motivated and helped me a lot to accomplish this work.

I am extremely thankful to my colleague Dr. Bryan M. Williams from CRiA for his prompt inspiration, timely suggestions with kindness, and advice in each trouble or difficulty I have had during research journey. I learned from him many great things, not only the knowledge but also critical thinking, research skills and passions for research which transformed me from an absolute newbie to a qualified researcher. Without his endless help and support, it would be very difficult to finish my thesis. Many thanks also go to the rest members of CRiA for their encouragement and support.

I gratefully acknowledge the graders David Parry and Sophie Leach, Liverpool Reading Centre and Dr Huiqi Lu and Dr Nattapon Boonarpa who provided the annotation of datasets. Special thanks also go to Dr. Majid Al-Tae and Dr. Hilal Al-Libawy for all their valuable advice and suggestions while writing up my papers and thesis. In addition, I would like to thank all my colleagues and the staff in the Department of Electrical Engineering and Electronics, University of Liverpool, for their academic and technical support.

My sincere gratitude to my family for their unconditional love, wise counsel and sympathetic ear. They have always been there for me. Thanks to my friends Theiab Alzahrani, Ola Younis, Ahmed Al-Tahmeesschi, Saeed Alzahrani, Bayan Al-Saaidah, Rwayda Al-Janabi, and Fatin Sultan for their unconditional support. Finally, I gratefully acknowledge The Higher Committee for Education Development in Iraq for the financial support of my research project.

Chapter 1

Introduction

1.1 Background

For a long time, the study of the anatomical structure of the human eye has been considered in order to avoid and treat the diseases of the eye. With the development of new systems and the invention of new techniques, research into eye disease diagnosis has accelerated and developed significantly. In particular, the advancement of medical imaging modalities has been significantly improved with the availability of new technologies to capture and process digital images. This improvement has required crucial innovation in computational technologies for the different types of image processing. In addition, medical image processing and analysis exhibit an excellent solution as a non-invasive procedure for the diagnosis and control of diseases.

Medical imaging research requires the analysis of images of different organs such as the brain, bones, kidneys, heart and eye, which involves different processing strategies. Furthermore, various types of images such as ultrasound, tomography and magnetic resonance images cannot all be manipulated by a single processing technique. Moreover, the type of image acquisition process is itself an independent and mature area of development. Research in medical imaging has contributed to the diagnosis and avoidance of diseases such as cerebral tumours, kidney dysfunctions, arthritis, retinopathy, glaucoma, age-related macular

degeneration and many others.

Several retinal eye diseases such as retinopathy, glaucoma and age-related macular degeneration have been extensively studied in clinical science as they are the most prevalent causes of blindness. Different image types such as optical coherence tomography (OCT), fundus photographs and angiography are used for the management of retinal diseases by ophthalmologists. Manual investigation of medical images is a time-consuming process and therefore developing automatic image analysis systems is a significant task. The analysis of retinal images, such as segmentation, localisation, identification, classification and so on, provides vital information about the health of the vision system by detecting pathological risk or damage. The automated analysis of images can be basically fulfilled by designing interactive and automated systems.

The purpose of this chapter is to provide an introduction to the research, showing the problem statement and motivation behind the research, and highlighting the aim and objectives. Finally, the novelty and publications emerging from this thesis are presented. The structure of the thesis is also described in order to create a clear road map and assist the reader in navigating the document.

1.2 Problem Statement and Motivation

The retina is vulnerable to microvascular changes as a result of many retinal diseases. To prevent the vision impairment caused by retinal conditions, periodic eye examination is recommended for patients under high risk [1]. Since diagnostic and investigation procedures involve a high attention of ophthalmologists, as well as regular monitoring of the condition, and the number of patients is constantly increasing, in addition to the shortage of physicians, these demands will eventually exceed the current healthcare capabilities. Due to the workload, human graders are supposed to grade images for several patients per day. Therefore,

they can become easily fatigued, causing a decrease in their examination accuracy. Furthermore, despite firm grading guidelines, human graders are subjective and therefore grades for certain image can vary considerably amongst different graders [2]. Automatic image analysis algorithms based on image processing and computer vision strategies have been gaining momentum in various medical applications and, in particular, retinal disease diagnosis. By automating the analysis process for retinal images, more patients can be screened and referred for further tests, allowing the ophthalmologists to have more time for patients who need their attention. In the literature, a large number of researchers have reported various analysis techniques for retinal images with a noticeable improvement in the performance. Nevertheless, researchers face several challenges and issues which cover different retinal image analysis aspects including localisation, segmentation and classification. Some of the issues that should be taken into consideration include:

1. **Autonomy and Independence:** existing automated retinal image analysis models have been based on supervised systems that work on features extracted from images manually [3], [4], [5], [6]. The performance of these systems highly depends on the type of features extracted from a specific dataset. Furthermore, those features might be applicable for a specific dataset but may not generalise well on other datasets. The-state-of-the-art retinal image analysis has the need for automated systems that perform the analysis of retinal images independent of any human interaction or pre-selection of features.
2. **Speed:** retinal image processing tasks often require complex computations. Although there are several retinal image analysis methodologies that achieve a good performance, but they cost a long processing time [7], [3], [8]. Hence, the need for time-efficient and cost-effective algorithms has become critical.
3. **Performance Improvement:** commercialisation of computer-aided approaches as clinical applications in the medical field is limited as it requires

much more accurate computerised analysis systems [9]. Therefore, to achieve the goal of considering the computerised systems as an important complement to the physicians professional knowledge and judgement in making decisions, highly sensitive and accurate methods should be developed and provided.

4. **Robustness and Generalisation:** retinal images usually contain noise due to interference, they also have inhomogeneous intensities, blurred edges and poorly defined boundaries. These characteristics have affected the process of image analysis in many methods in the literature, especially in image segmentation tasks [10], [11]. Therefore, development of robust methods that continue to work well even on poor-quality images is a significant requirement. In addition to that, the developed methods need to generalise well to various datasets which are captured from different imaging modalities.

Feature learning algorithms have emerged recently in the form of deep learning including many algorithms such as convolutional neural networks (CNNs). Deep feature learning algorithms require computationally intensive operations which are made up of multiple matrix multiplications during the learning process. The recent advancement of hardware resources represented by graphics processing units (GPUs) has enabled researchers to develop such automated feature learning algorithms. The availability of such hardware capabilities enables the mathematical operations to be processed as parallel computations; thus, computational performance has been significantly improved. This technology development has therefore made it possible to propose new ways of retinal image analysis that automate the feature extraction process at a low cost.

1.3 Aim and Objectives

The main aim of this research is to develop automated computer-aided systems based on deep learning that can help in image analysis to diagnose retinal diseases such as age-related macular degeneration, diabetic retinopathy/maculopathy and glaucoma using OCT and colour retinal fundus images. Therefore, in this project the feasibility of feature learning algorithms has been studied. Features are learnt from data automatically to analyse retinal images as opposed to the process in traditional feature engineering methods where the features are designed and selected manually.

In response to the limitations mentioned earlier, in Section 1.2, the research described in this thesis is focused on proposing novel solutions to address retinal image analysis-related problems based on deep learning techniques. Towards this end, the following specific objectives have been identified that are categorised according to the image analysis task, which comprises localisation, segmentation and classification:

1. The development of a methodology for the automated localisation of the most important retinal landmarks: optic disc and fovea.
2. The design of automated segmentation frameworks to segment the anatomic landmarks including the choroid, and the optic disc and cup areas.
3. The proposal of automated classification systems to detect the presence of glaucoma , grade the severity of diabetic maculopathy, and identify the retinal lesions (fluids).

1.4 Novelty and Contribution

Novelty and scientific contribution arising from the work presented in this thesis are reflected by developing novel approaches based on deep learning towards automating retinal disease diagnosis process. A number of high-quality publications have also been generated and published in peer-reviewed international journals and conferences. The main novelty and contribution aspects are outlined as follows:

1. A novel approach based on a multiscale sequential deep learning for the simultaneous detection of the centres of the optic disc and fovea in colour fundus images is proposed, designed and successfully implemented. This is achieved by extracting complex data representations from retinal images without the need for human supervision.
2. The development of a new framework for image segmentation which is targeted towards EDI-OCT images of the choroid region. The developed framework, which is based on clustering and deep learning algorithm, uses not only image intensities but also automatically enhanced intensities and distance information introducing a new energy function for partitioning the image into unlabelled clusters which should not cross the boundaries of the choroid and achieving improved choroid segmentation performance over previous work.
3. Development of a new segmentation method based on the most recent CNN architectures for the optic disc and cup regions in colour fundus images. This is achieved by adapting a fully-convolutional DenseNet. Most comprehensive study involving five publicly available datasets are carried out. This allows for evaluation with images from many different devices and conditions proving the robustness of the proposed method.
4. Investigating potential application of the transfer learning concept in detecting the presence of glaucoma in small dataset of colour fundus images. This has led to utilising pre-trained weights of the CNN to extract the features of

glaucoma and use them to train a support vector machine classifier (SVM) to detect the presence of the disease.

5. Development of an enhanced diabetic macular oedema grading system in which features of colour fundus images are automatically learnt through developing a multi-stage CNN.
6. Proposing a novel framework to estimate the probability of presence of the retinal fluids existing in OCT volumes. The proposed framework comprises three main stages; (i) preprocessing, (ii) deep learning, and (ii) ensemble learning and parameter selection. Furthermore, in the deep learning stage, identification of the retinal fluids is formulated as multi-label classification task. This is achieved through an appropriate adaptation of the Inception CNN.

The work contribution is also reflected through publication of the following list of papers.

Journal Publications

1. **B. Al-Bander**, W. Al-Nuaimy, B. M. Williams, Y. Zheng, “Multiscale Sequential Convolutional Neural Networks for Simultaneous Detection of Fovea and Optic Disc”. *Biomedical Signal Processing and Control*, 40:91-101, 2018.
2. **B. Al-Bander**, B. M. Williams, W. Al-Nuaimy, Y. Zheng, “Dense Fully Convolutional Segmentation of the Optic Disc and Cup in Colour Fundus for Glaucoma Diagnosis”. *Symmetry*, 10(4):87, 2018.
3. **B. Al-Bander**, B. M. Williams, W. Al-Nuaimy, Y. Zheng, “A Framework for Automatic Segmentation of the Choroid in EDI-OCT Images using Deep Learning and Clustering”, *IEEE Transactions on Medical Imaging*. Under Revision. 2018.

4. H. Pratt, B. M. Williams, J. Y. Ku, C. Vas, E. McCann, **B. Al-Bander**, Y. Zheng, “Automatic Detection and Distinction of Retinal Vessel Bifurcations and Crossings in Colour Fundus Photography”, *Journal of Imaging*, 4(1), 2017.
5. I. MacCormick, B. M. Williams, Y. Zheng, K. Li, **B. Al-Bander**, S. Czanner, R. Cheeseman, C. Willoughby, E. Brown, G. Czanner, “Accurate Glaucoma Diagnosis with Automated Spatial Analysis of the Cup to Disc Profile”, *Nature Scientific Reports*. Under Revision. 2018.

Conference Publications

1. **B. Al-Bander**, W. Al-Nuaimy, M. A. Al-Tae, Y. Zheng, “Diabetic Macular Edema Grading based on Deep Neural Network,” in 19th International Conference on Medical Image Computing and Computer Assisted Intervention Ophthalmic Medical Image Analysis International Workshop MICCAI OMIA, Greece, 2016, pp. 121-128.
2. **B. Al-Bander**, W. Al-Nuaimy, M. A. Al-Tae, Y. Zheng, “Automatic Feature Learning Method for Detection of Retinal Landmarks,” in 9th International Conference on the Developments on eSystems Engineering (DeSE), UK. IEEE, 2016, pp. 13-18.
3. **B. Al-Bander**, W. Al-Nuaimy, M. A. Al-Tae, Y. Zheng, “Automated Glaucoma Diagnosis using Deep Learning Approach,” in 14th International Multi-Conference on Systems, Signals & Devices (SSD), Morocco. IEEE, 2017, pp. 207-210.
4. **B. Al-Bander**, B. M. Williams, M. A. Al-Tae, W. Al-Nuaimy, Y. Zheng, “A Novel Choroid Segmentation Method for Retinal Diagnosis Using Deep Learning,” in 10th International Conference on the Developments on eSystems Engineering (DeSE), France. IEEE, 2017.

5. **B. Al-Bander**, W. Al-Nuaimy, D.G. Parry, S. Leach, Y. Zheng, “Automatic Detection of Optic Disc and Fovea using Deep Neural Network,” 26th EASDec abstract in European Journal of Ophthalmology, UK, 2016.
6. B. M. Williams, **B. Al-Bander**, H. Pratt, S. Lawman, Y. Zhao, Y. Zheng, Y. Shen, “Fast Blur Detection and Parametric Deconvolution of Retinal Fundus Images,” in Ophthalmic Medical Image Analysis. Springer, Cham, 2017, pp. 194-201

1.5 Thesis Structure

The remaining chapters of this thesis are organised as follows.

Chapter 2 - Background

This chapter presents the literature review, and an investigation of the clinical problems and relevant technical background. The anatomical components of the human eye including the anatomy of the retina are explained. A description of diseases related to the retina and an overview of the most common retinal imaging techniques are also provided. The deep learning concept, architectures, and the differences between traditional machine learning and feature learning techniques are also discussed. Finally, a review of the use of deep convolutional neural networks in medical image analysis is included.

Chapter 3 - Optic Disc and Fovea Detection

This chapter describes the developed retinal landmarks detection methodology. Initially, existing detection and localisation of optic disc and fovea techniques are described. Following this, the proposed method to overcome the limitations existing in the previous methods is presented. The performance evaluation technique and experimental results are next described, demonstrating the effectiveness of this work.

Chapter 4 - Choroid and Optic Disc/Cup Segmentation

This chapter presents two developed segmentation methodologies. The first method is for choroid segmentation in enhanced depth imaging optical coherence tomography images (EDI-OCT). The second method is proposed for the optic disc and cup segmentation from fundus images.

Chapter 5 - Retinal Disease and Lesion Classification

This chapter presents retinal disease diagnosis and fluid identification based on deep convolutional neural networks in two types of images: digital colour fundus and OCT. First, the system is designed to detect the presence of glaucoma in fundus images based on features extracted by CNN to train SVM classifier. Next, a multi-stage CNN to diagnose the diabetic maculopathy disease from fundus images is proposed. Finally, the concept of using pre-trained weights in the CNN and ensemble learning is considered to identify the retinal fluids: IRF, SRF, and PED in OCT images.

Chapter 6 - Conclusions and Future Work

This chapter summarises the main findings concluding the work presented in this thesis, and discusses the possible future research directions that can be explored based on this work.

Chapter 2

Background

In this chapter, the clinical problems and relevant technical background are presented for the benefit of the reader. The first section explains the anatomical structure of the human eye including the retina; this is followed by a description of diseases related to the retina in the second section. An overview of the most common retinal imaging techniques is provided in the third section. The deep learning concept, deep learning architectures, and the differences between traditional machine learning and feature learning techniques are then discussed, concluding with a review of deep convolutional neural networks in the medical image analysis in Section 5, followed by a short summary.

2.1 The Eye and Retinal Diseases

2.1.1 Anatomy of the Eye

The human eye is a specialised sense organ capable of receiving visual images, which are then carried to the brain. It is shaped like a sphere, giving us the sense of sight by reacting to light [12]. The human eye comprises the anterior segment, which is made up of the cornea, iris and lens, and the posterior segment, which is composed of the vitreous body, retina, choroid and the sclera. Figure 2.1 presents a cross section of the human eye with the anatomical structure labelled.

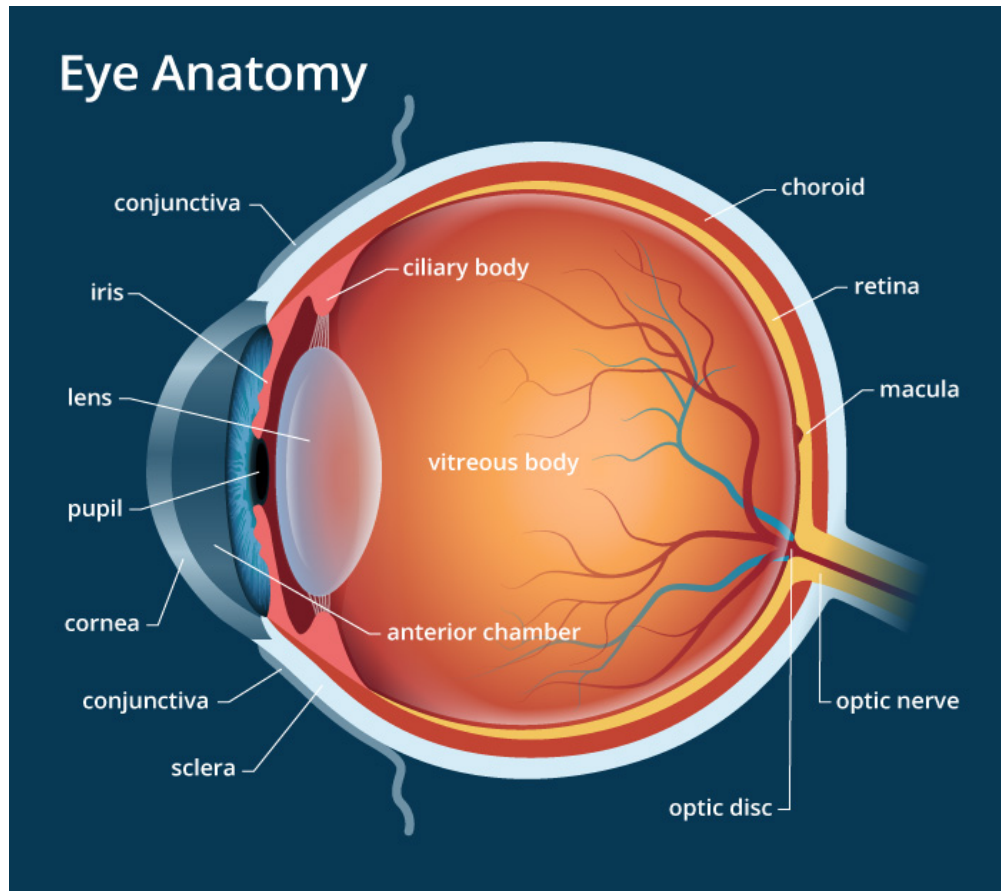


FIGURE 2.1: The anatomical structure of the human eye [12].

The eye processes data in a similar way to a camera. Similar to light passing through the camera's optics to a sensor, light reflected from an external medium passes through the eye's cornea, pupil and lens, and focuses onto the retina to be mapped into meaningful information which is interpretable by the brain.

The cornea is the clear front surface of the eye which is protective outer layer lying directly in front of the iris and pupil. It receives the incoming light which travels through the pupil. The pupil is the centre of the iris, a muscle located between the cornea and the lens. The iris can contract and relax according to the light conditions so as to allow a certain amount of the light to pass to the lens. The lens, which is shaped like a convex disc, shrinks or stretches the light to focus it on the retina. Subsequently, the light received by the retina is transformed into electrical signals and analysed by the brain [13].

The cavity of the eye between the lens and retina is filled with the vitreous body,

which is like a colourless and transparent gel. Anatomically, there are three layers of tissue wrapped around the vitreous body which are responsible for maintaining the shape of the eye. The outer layer is the sclera, a white tissue that covers the whole eye excluding the cornea region with muscles and protects the inner structure of the eye. The layer next to the sclera is the choroid, which is filled with blood vessels that are in charge of supplying the retinal cells with the necessary oxygen and nutrition. The inner-most layer is the retina, which comprises different layers of cells that contain the light-sensitive cells, photoreceptors (cones and rods) and neurons. These cells create signals that can be sent to the brain for further analysis. The thickness of the retina is about 0.5 mm where the centre of the retina, the fovea, has a thickness of less than $300\text{ }\mu\text{m}$ thick [14]. Figure 2.2 provides a deep view of the anatomical structure of the human retina in cross section.

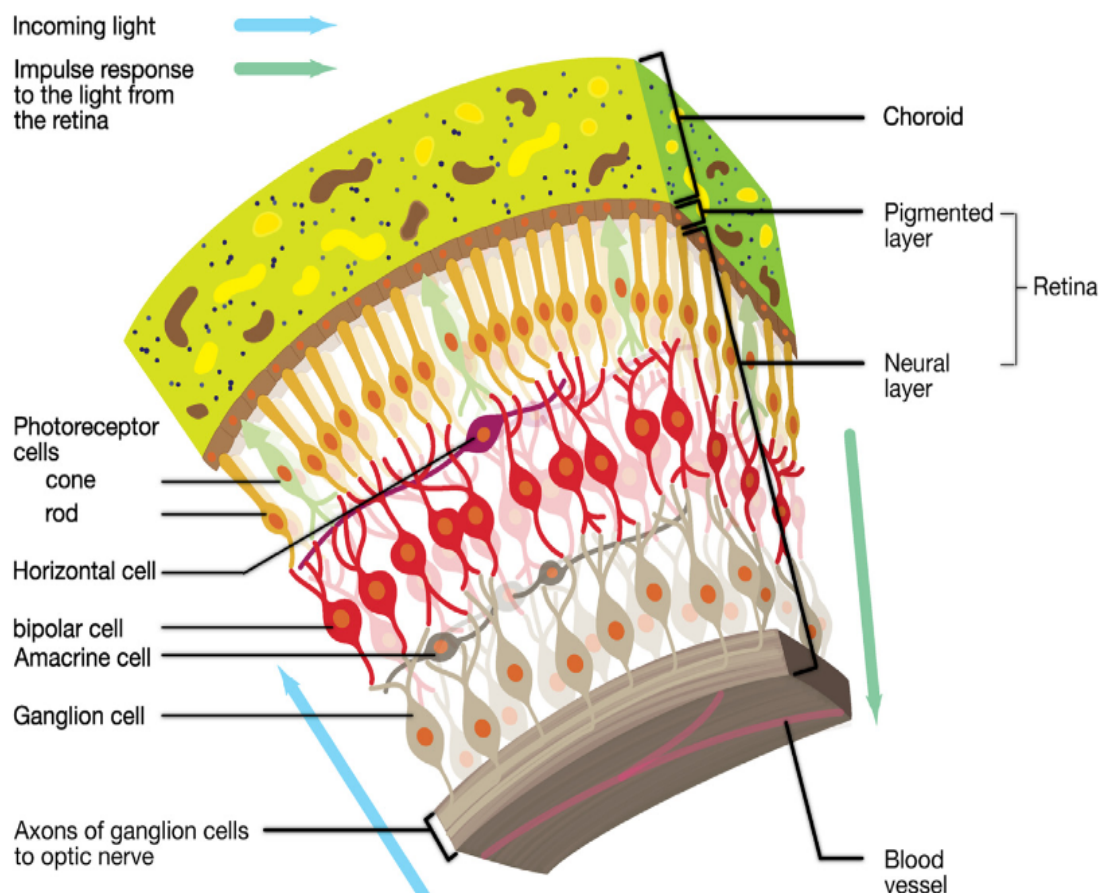


FIGURE 2.2: Cross section of the human retina showing the anatomical structure [15].

From the view of the fundus photography, it is apparent that there are three components that characterise the retina: the fovea (centre of the macula), the optic nerve head and the blood vessels. The macula (its centre is the fovea) is the spot where the eye has the ability to best recognise the visual details. Any damage in this region will cause loss of the central vision. All the photoreceptors (rods and cones) are linked to the brain through millions of nerves which leave the eye in the optic nerve. The blood vessels provide nutrients and oxygen to the inner layers of the retina [14]. Figure 2.3 shows a colour retinal fundus image with the key anatomical structures denoted.

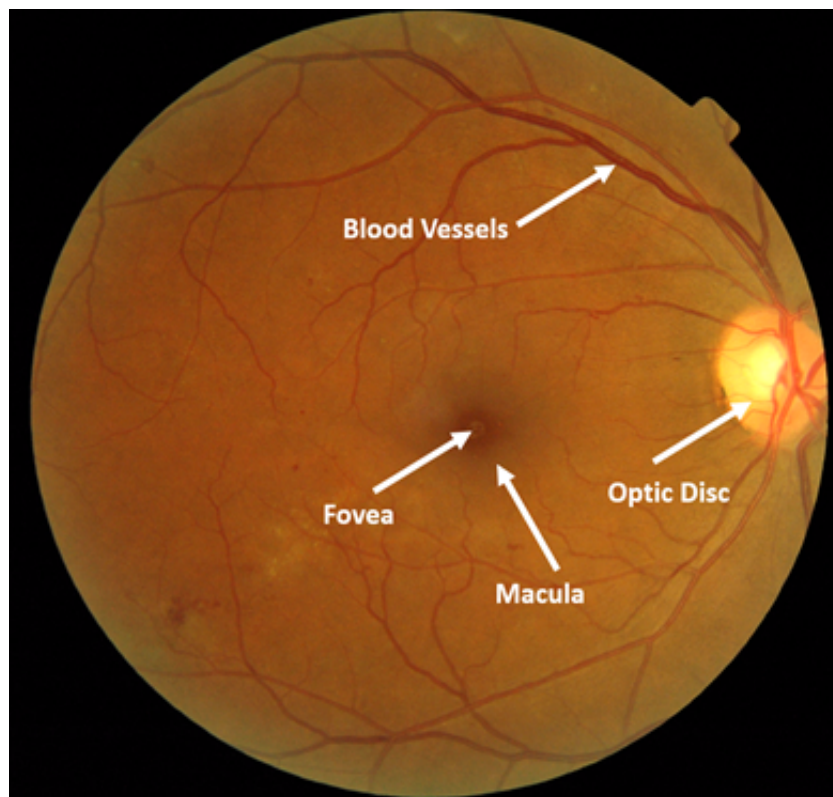


FIGURE 2.3: An example fundus image illustrating the key retinal anatomical structures.

2.1.2 Retinal Diseases

There are various disorders, diseases and age-related changes that may affect the vital tissue in the eye, the retina. These issues can affect the vision of the eye

and subsequently cause blindness. Some of the most common retinal diseases are discussed below.

Age-related Macular Degeneration

Age-related macular degeneration (AMD) is the leading cause of irreversible visual impairment worldwide. This condition is common in people over 50 years old and causes damage to the central vision of the eye. In the UK, it affects 600,000 people currently and it is estimated that almost 700,000 patients will suffer from AMD by 2020 because of ageing population [16].

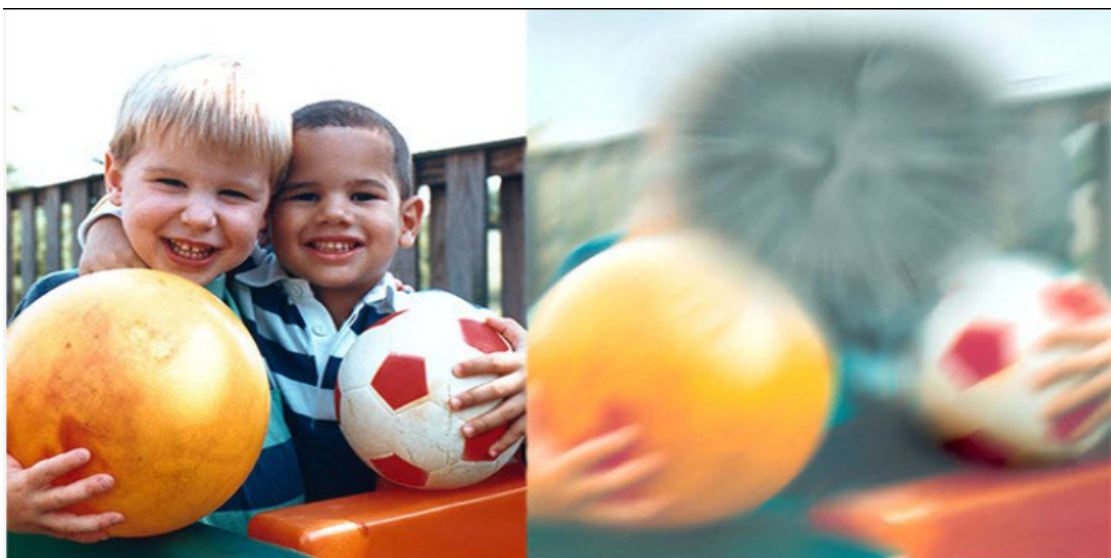


FIGURE 2.4: Vision with Age-related macular degeneration (AMD). AMD causes loss of sight in the centre of the field of vision [17].

It has a major effect on the centre of vision, the macula, where vision becomes increasingly blurred, as shown in Figure 2.4. AMD can be mainly classified as either dry AMD in which gradual loss of visual acuity can result, or wet AMD, which can be considered the type that most threatens the vision. Although in the dry form there are limitations in the vision that happen at night or under poor illumination, there is no leaking of blood or serum. In wet AMD, abnormal blood vessels start to appear under the macula and damage its cells [18].

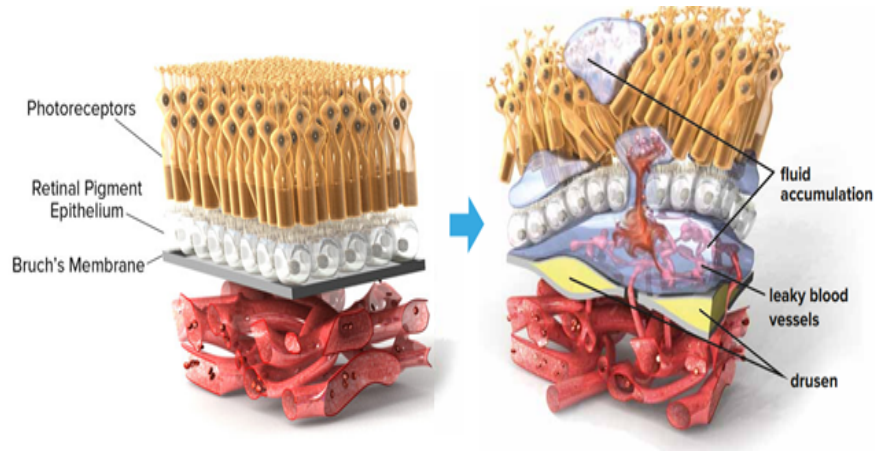


FIGURE 2.5: Fluid leakage affecting the retina [19].

One of the anatomic changes associated with wet AMD presence is accumulating fluids (shown in Figure 2.5) as a result of decreases in the outflow or increases in the inflow of fluid. Clinically, three types of fluids are distinguishable in medical images: Intraretinal fluid (IRF), Subretinal fluid (SRF) and Pigment Epithelial Detachment (PED) [20]. The fluids visualised using optical coherence tomography (OCT) images are shown in Figure 2.6.

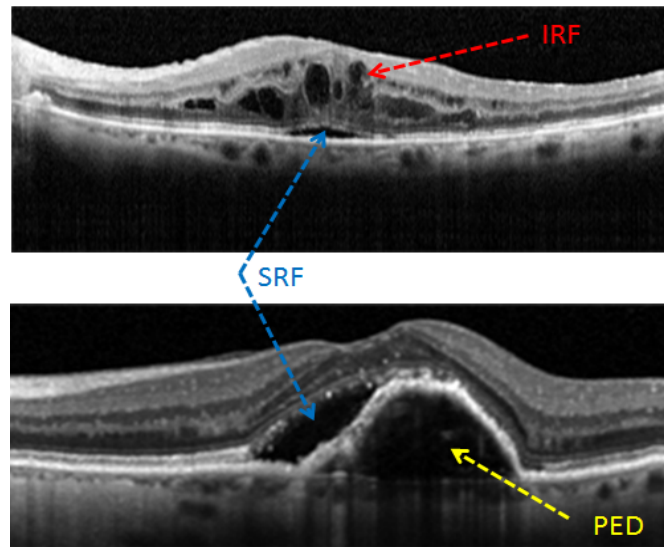


FIGURE 2.6: Fluids as visualised in OCT images [19]. IRF: Intraretinal fluid, SRF: Subretinal fluid, PED: Pigment Epithelial Detachment.

Glaucoma

Glaucoma is one of the chronic retinal diseases, which results from damage to the optic nerve head due to hypertensive intra-ocular pressure (IOP) of the eye. By 2020, it is predicted to affect around 80 million people worldwide. In the early phases of the disease, patients do not have symptoms of vision loss while as the disease progresses, the loss of peripheral vision occurs (as shown in Figure 2.7). In the advanced stage of glaucoma, patients may suffer total blindness [21].



FIGURE 2.7: Visual loss in glaucoma. Visual field loss begins at the periphery and moves towards the centre [22].

In fundus images, the optic nerve head (ONH) is divided into two regions: a peripheral zone called the neuroretinal rim and a central bright region called the optic cup, as shown in Figure 2.8. One of the indicators of glaucoma is the enlargement of the cup zone with respect to OD, which can be estimated by measuring the vertical cup to disc ratio (CDR) [23].

Diabetic Retinopathy and Maculopathy

Diabetic retinopathy (DR) is a condition that is associated with diabetes mellitus where high blood glucose levels cause damage to the vessels supplying blood to the retina. It can cause blood vessels in the retina to leak fluid or haemorrhage

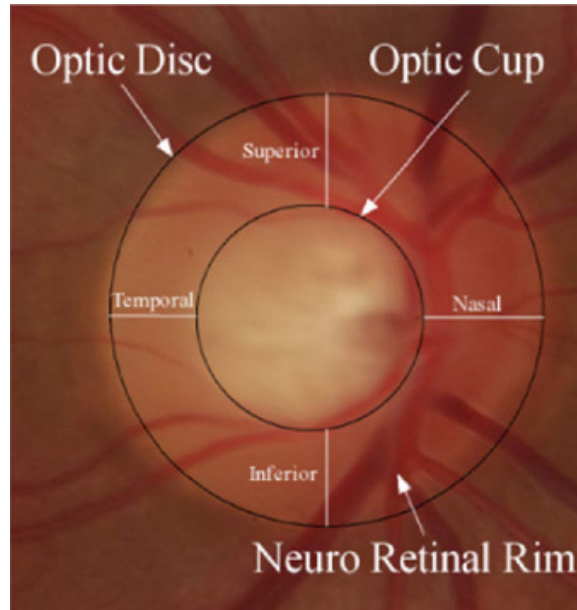


FIGURE 2.8: Fundus image shows the optic nerve head from a healthy subject [24].

(bleed), distorting vision. In its most advanced stage, new abnormal blood vessels appear on the surface of the retina, which can lead to scarring and cell loss in the retina. It is the main reason of poor vision in people with type 1 diabetes or type 2 diabetes. The major issue with this disease is that the patient has no signs of any degradation in the vision at the early stages of the disease. It can be broadly divided into two stages: non-proliferative DR (NPDR) and proliferative DR (PDR) where, with NPDR, damage in the retinal blood vessels occurs when, in PDR, new abnormal blood vessels start to appear. Many retinal lesions are taken into consideration while determining the severity of the DR such as microaneurysms (MAs), haemorrhages (HMs), exudates (hard and soft) (EXs) and intra-retinal microvascular abnormalities (IRMA) [25, 26].

It is worth pointing out that DR has been classically considered a microvascular disease of the retina. However, there is emerging evidence to suggest that retinal neurodegeneration is an early event in the pathogenesis of DR which could participate in the development of microvascular abnormalities [27–29]. In general, there are three possible relationships between microvascular DR and diabetic retinal neurodegeneration (DRN): i) microvasculopathy causes neurodegeneration; ii)

neurodegeneration causes microvasculopathy or iii) they are mutually independent [30].



FIGURE 2.9: Abnormal vision effected by diabetic macular oedema condition associated with diabetic retinopathy [17].

DR can cause a breakdown in the inner endothelial blood-retinal barrier (BRB) which results in diabetic macular oedema (DMO) disease [31]. DMO, also known as Diabetic Maculopathy (DM), usually develops at any time during the progression of DR, as shown in Figure 2.9. According to the Early Treatment Diabetic Retinopathy Study (ETDRS), DMO is characterised by the thickening of the macula, hard exudate (HE) and blot haemorrhage (HA) [25]. Clinically, the severity of DMO is mainly divided into two classes: non-clinically significant macular oedema (non-CSMO) and clinically significant macular oedema (CSMO). Non-CSMO is a mild class of maculopathy in which the distance between the lesions and the centre of the macula is greater than one optic disc diameter. It is characterised by retinal thickening and hard exudate as clinical features. CSMO is the severe form of maculopathy in which lesions (blot haemorrhage and exudate) occur within a distance of less than one optic disc diameter from the centre of macula [25, 32].

Early detection and treatment of the retinal conditions helps in preventing the

progression of the disease. Considering the increasing number of people suffering from these conditions, relying on medical devices to diagnose and detect the disease may become infeasible. Extensive research using various image processing techniques is being applied to overcome the aforementioned problem for early detection of retinal diseases.

2.2 Retinal Imaging Modalities

To enable ophthalmologists to diagnose retinal diseases accurately, a vast array of imaging cameras dedicated to capture retinal images have been developed. The principles of some imaging modalities are briefly introduced below.

2.2.1 Fundus Photography

Fundus photography is the earliest type of retinal imaging techniques. Its first appearance was in the mid-1800s when the ophthalmoscope was introduced, in 1861. In fundus photography, a photograph of the back of the eye (i.e. fundus) is captured. To image the retina, specialised fundus cameras are used where the central and peripheral retina, macula and optic disc are the main structures that can be visualised on the photo. Many manufacturer of fundus camera are currently available such as Topcon, Zeiss and Canon. Fundus retinal images can be mainly classified into colour fundus and monochromatic fundus.

Colour fundus photography is obtained using a customised camera attached to a specialised microscope with mirrors and lenses. The fundus cameras are designed to image the interior surface of the eye by illuminating the retina and screening it in full colour. The importance of the colour fundus photograph in imaging belongs to its crucial role in documenting and recording the presence and progress of diseases such as diabetic retinopathy and maculopathy, age related macular degeneration, glaucoma and retinal detachment in the optic nerve, retina blood

vessels, macula, optic disc and posterior pole (the fundus) [33]. Figure 2.10 shows a colour fundus image captured using a Zeiss FF450+ fundus camera.

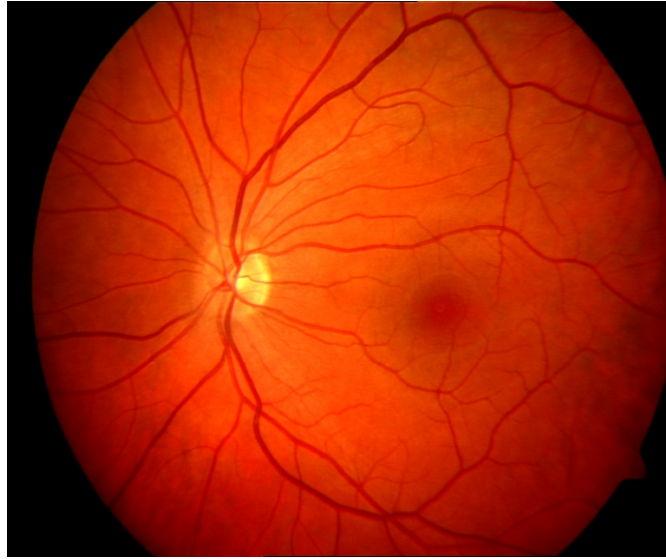


FIGURE 2.10: Colour fundus image captured by Zeiss camera.

The monochromatic/red free photography utilises infrared light to illuminate the retina instead of the white light used in colour fundus photography. In order to improve the level of observation and allow a good contrast of viewing for some lesions and abnormalities in the retina, a filter is used to prevent red wavelengths of the light in this type of image. Monochromatic images can be captured by a monochrome sensor which produces grayscale images as it is more light sensitive than colour sensors [34].

2.2.2 Optical Coherence Tomography

Optical Coherence Tomography (OCT) visualises the changes in the layers of the retina via cross-sectional tomographic imaging of the internal structure. OCT has emerged as an objective tool that provides detailed 2-dimensional and 3-dimensional images for the retina. The principal work of OCT is to provide non-invasive and non-contact tomographic imaging of the retina tissue by using

low-coherence interferometry. It has been widely used in clinical practice to measure the thickness and volume of the retina and carry out quantitative assessment with high resolution and scan speed [35].

The earliest use of OCT was in the mid-1980s when measurements of one-dimensional (axial information) which is equivalent to ultrasound A-scans were demonstrated. In 1991, the generation of two-dimensional images (cross-sectional) which is equivalent to ultrasound B-scans was demonstrated by [36]. Since that time, OCT has quickly developed as an optical medical diagnostic imaging technique. Acquiring three-dimensional OCT images and providing volumetric information is a crucial advance because it provides comprehensive structural details. These details and information can be used for generating cross-sectional images that precisely give retinal layer thicknesses and volumetric views of retinal structure analogous to magnetic resonance (MR) images. One of the earliest uses of 3D-OCT retinal images called en-face OCT was demonstrated using time domain detection [37].

Different types of OCT imaging modalities have been developed such as time domain (TD-OCT) and frequency domain (FD-OCT) techniques. In time domain OCT, the path length of the reference arm is varied in time where the most popular TD-OCT configurations are full-field OCT and A-scan acquisition while in FD-OCT (Fourier domain), swept-source and spectral are the most commonly utilised. Over the years, several OCT imaging techniques for both TD-OCT and FD-OCT have been created and developed such as polarisation-sensitive and spectroscopic OCT.

Nowadays, many commercial OCT devices are available such as Cirrus, Heidelberg's Spectralis, Topcon, Optovue and Stratus. To compare the OCT image variability among different manufactures, three spectral-domain OCT images, Cirrus, Spectralis and Topcon, for a normal retina are shown in Figure 2.11.

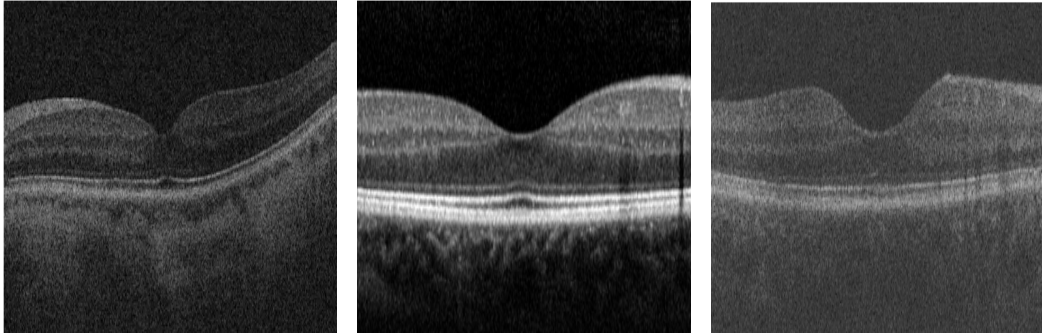


FIGURE 2.11: Normal retina imaged with OCT from three different devices.
(a) Cirrus (b) Spectralis (c) Topcon [19].

2.3 Medical Image Analysis

Following the description of some retinal imaging modalities in the previous section, the aim is now to explain some image processing techniques used to analyse and process the medical images. Recently, the automatic detection and analysis of retinal diseases from digital images has received increasing attention in the medical image processing community. This may support the development of computer-aided diagnosis (CAD) tools for the better management of eye disease.

For several decades, many techniques of digital image processing and analysis have been developed to extract meaningful information from the images. Generally, computer image processing and analysis comprises the fields of imaging and computer vision with heavy exploiting of signal processing and pattern recognition. Digital image processing facilitates the use and development of much more sophisticated algorithms which can be used for different purposes.

In this section, some low-level image analysis techniques such as enhancement and denoising for image pre-processing purposes are described and some high-level image analysis strategies such as segmentation, detection, and classification which are taken into consideration through the thesis.

2.3.1 Low-level Image Analysis

Low-level image analysis algorithms are often a pre-requisite step to make images more appropriate for further analysis by high-level analysis algorithms.

Enhancement

Image enhancement has been widely used in many applications of image processing where the principal objective is to improve the visual appearance of the images and make them more interpretable for humans and suitable for further analysis. Enhancement can be achieved by sharpening the image features such as boundaries, edges or contrast. The implementation of image enhancement techniques can be based on either spatial or frequency domain.

Typically, image enhancement algorithms are mainly classified into enhancement by point processing, enhancement by spatial operations, enhancement by transform operations and pseudo-colouring [38–42]. In point processing enhancement methods, an immediate manipulation is applied to the intensity image pixels such as contrast stretching, window slicing and histogram processing (equalisation, specification and local enhancement). However, in spatial operations, spatial masks such as smoothing and sharpening filters are utilised over the pixels of the images. In contrast to the spatial domain, the frequency domain operations are used by applying the Fourier transform on the image pixels using a filter transform function. An example of original and contrast-enhanced image by contrast-limited adaptive histogram equalisation (CLAHE) algorithm [43] is shown in Figure 2.12.

Different enhancement techniques such as contrast, histogram specification and filtering have been considered in this thesis to provide a better representation of the images.

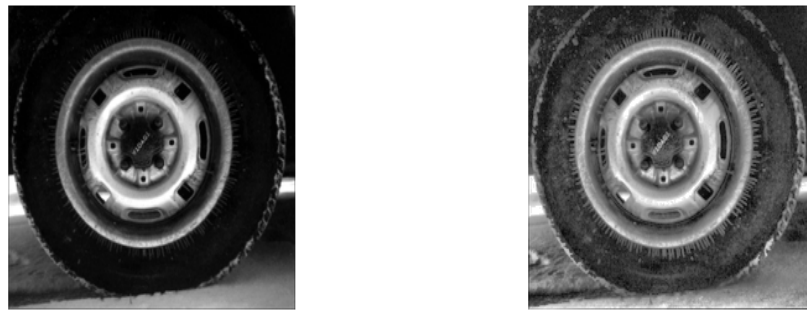


FIGURE 2.12: An example of contrast image enhancement. (a) Original (b) Enhanced.

Noise Reduction

Generally, medical images are acquired by instruments which are prone to be contaminated by noise such as Gaussian, salt and pepper, poisson, speckle noise, etc.; causing degradation to the images. To manipulate and compensate for image corruption, many denoising algorithms have been proposed to prepare the images before passing them for further analysis. However, implementation of accurate image denoising algorithms still remains a challenge because the noise removal causes blurring and artefacts to the images.

Image denoising approaches can be broadly grouped into two basic areas, filtering such as spatial and frequency domain filtering [44–46], and variational energy minimisation techniques [47, 48]. Figure 2.13 shows the use of a filtering technique in image denoising.

2.3.2 High-level Image Analysis

High-level image analysis algorithms are concerned with giving a higher level of interpretation to the images. This interpretation can be achieved by detecting an object in an image, classifying and diagnosing the severity of a disease by analysing the disease's characteristics in an image, or locating the area of interest by segmenting the desired object in an image.



FIGURE 2.13: Image denoising by filtering technique. (a) Original image (b) Image corrupted with Gaussian noise (c) Denoised image by Wiener filter.

Segmentation

Image segmentation is the process of dividing an image into regions in which the pixels have common characteristics such as colours, intensity, texture and any other common feature by assigning a label to every pixel such that pixels that share the same characteristics are allocated the same label. There has been a significant amount of research in developing segmentation techniques for problems in many areas including medical imaging, astronomical imaging and industry [49].

Segmentation algorithms can be widely categorised into five main techniques: thresholding [50], edge-based methods [51], region-based methods [52], machine learning [53] and model based methods [54]. Each of these segmentation techniques has distinctive advantages and disadvantages. Thus, to obtain better segmentation

for the objects in the medical images, a mixture of these techniques has been adopted and adapted in this thesis.

Classification/Detection

Image classification refers to the task of assigning an input image one label from a fixed set of categories/labels while object detection is defined as the process of finding a specific object in an image. Many techniques have been introduced and developed to accomplish these tasks. The techniques are based on machine learning algorithms by training a system to recognise or classify an image/object and then provide unseen input image to be classified/object recognised.

Depending on the goal behind designing machine learning algorithms, which is either for synthesis/generation or recognition/classification, machine learning techniques can be broadly categorised as generative, discriminative and hybrid methods. The main difference between generative and discriminative models is that the discriminative models learn the conditional probability distribution $p(y|x)$, i.e. learn the boundary between classes, while the generative models learn the joint probability distribution $p(x, y)$, i.e. model the distribution of the classes and how the data were generated [55].

Depending on the availability of true labels during the learning stage, machine learning algorithms fall under two main categories: supervised and unsupervised learning. In supervised learning and during the training stage, the real image labels of images are fed along with the images into the classifier while, in unsupervised learning, no labels are given to the learning algorithms [56]. More details about such concepts are given in the next section.

2.4 Artificial Intelligence, Machine Learning and Deep Learning

Artificial intelligence is implementing technologies that behave like a human, i.e. algorithms that do something deemed smart. While achieving such a goal still seems to be in the distant future, many important tools and strategies have been implemented, developed and successfully applied to a wide variety of scientific fields and problems such as self-driving cars, smart homes, personal assistants and many other emerging technologies. Machine learning (ML) is a branch of the artificial intelligence field that uses programs to learn from datasets. In fact, ML has now become widespread technology and there is a wide variety of algorithms for implementing such intelligent systems.

An emerging research trend in ML is deep learning (DL). The origin of the term deep learning term come from the new strategies implemented to generate deep hierarchies of non-linear features whilst overcoming the vanishing gradients problem. The vanishing gradient issue [57] appears in the very deep layers where the gradients become too small to provide a learning signal and get stuck in apparent local minima. This problem has been alleviated in the developed deep learning architectures so that architectures with many layers can be trained effectively [58]. Figure 2.14 shows a Venn diagram that explains the relation among artificial intelligence, machine learning and deep learning.

find an appropriate representation of data in order to perform a machine learning task

Many of the new computational techniques rely on the ability of Graphical Processing Units (GPUs) to quickly run complex algorithms in parallel form. GPUs have become the platform for training large and complex deep learning-based systems where the success of deep learning networks has been greatly accelerated. It is shown that, contrary to the raw processors which are not efficient, it is adequate

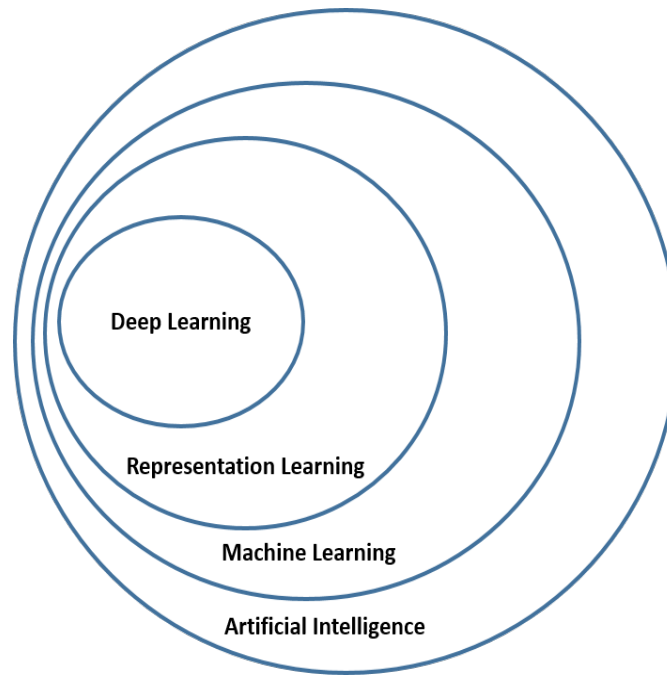


FIGURE 2.14: Deep learning is a subset of representation learning which aims at finding an appropriate representation of input data in order to perform a machine learning task which is in turn a subset of artificial intelligence. Adapted from [59].

to train deep architectures without major difficulties by utilising GPUs with activation functions that give better gradient flow. The main difference between the traditional multi-core processors and GPUs is that traditional processors typically contain 4-24 general purpose Central Processing Units (CPUs) but GPUs might have 1000-4000 specialised data processing cores as explained in Figure 2.15. This makes GPUs highly parallel and ideal for floating-point vector operations [60].

2.4.1 Feature Engineering and Learning

In traditional machine learning methods and for a long time, the features of images have been extracted manually through adopting various feature-engineering approaches. Feature engineering can be defined as the art of extracting useful and important patterns from data. Extracting features makes it easier for machine learning models to distinguish between classes. This feature is useful for the machine learning algorithms because it limits the number of labels that need to

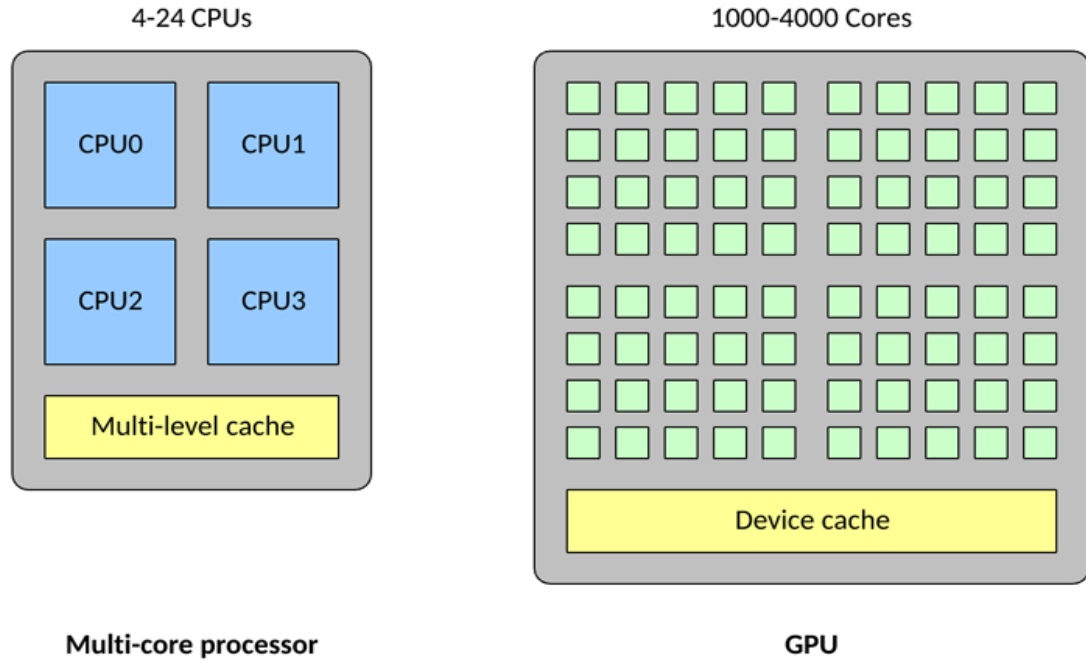


FIGURE 2.15: CPUs versus GPUs. Each green square represents a core in the GPU while the blue square refers to CPU [60].

be considered for good classification. Therefore, to achieve good results for most predications tasks, feature engineering is the most important skill [61].

Traditional image classification and prediction methods usually utilise hand-designed image features, such as colour, edges, shape, texture, local binary pattern (LBP) and histograms of the oriented gradients (HOG), and machine learning classifiers and regressors such as k-Nearest Neighbours (k-NN) algorithm, Multi-class Support Vector Machines (SVM) and Artificial Neural Network (ANN). After extracting the features, a classifier is trained to complete the classification, detection or prediction task for the images.

However, the existing generic guidelines for extracting features from different datasets are unable to fulfil the requirements of different datasets. For example, the features that are usable for one dataset are often not usable for other datasets. Therefore, the search for new algorithms that are capable of learning features automatically has become a key requirement for developing more accurate machine learning analytics.

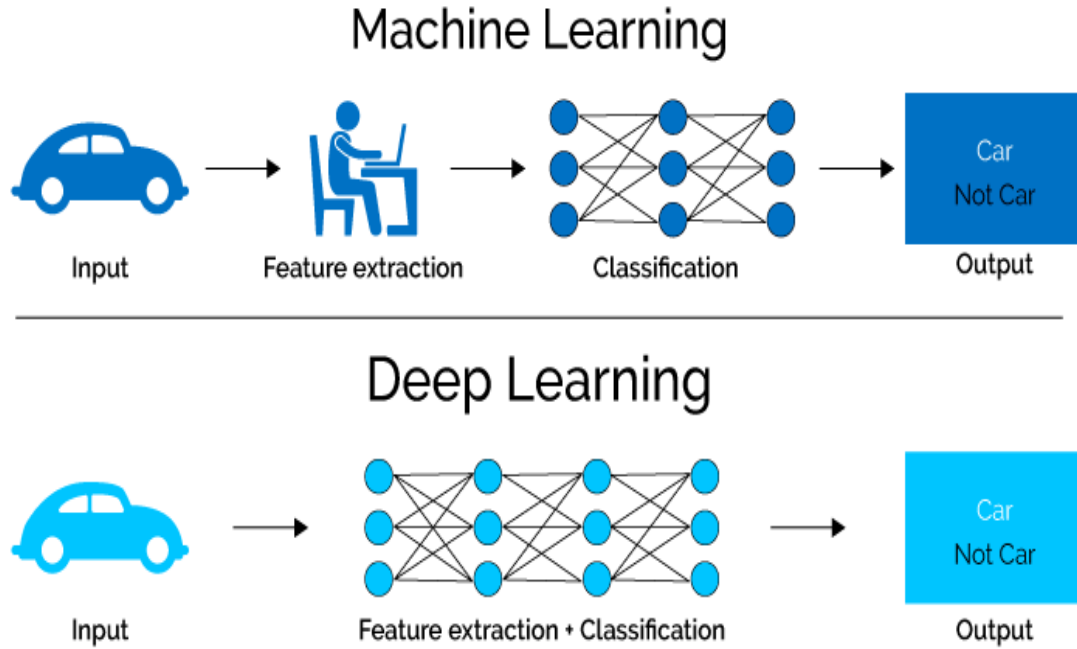


FIGURE 2.16: Deep learning versus machine learning [62].

To overcome the problems of the feature engineering techniques, feature learning from data methods have been proposed and developed. The feature learning can be defined as the process of finding the common patterns in the data which are significant to distinguish among classes. Feature learning technique is seen as engineering and extracting of the features automatically by algorithms to be used in the regression and classification tasks [63]. Figure 2.16 shows the concept of traditional machine learning versus feature learning algorithms.

Generally, the extracted features play an important role in many computer vision tasks such as object recognition/classification, localisation, object detection and image segmentation where the accuracy of the implemented algorithms greatly depends on the type and meaning of the features.

2.4.2 Deep Learning Architectures

Various deep learning architectures such as deep belief networks (DBNs) [64], deep Boltzmann machines (DBMs) [65], variational autoencoder [66], generative adversarial networks (GAN) [67] and Long short-term memory recurrent neural network

architectures (LSTM) [68] as unsupervised generative models, and deep convolutional neural networks (CNNs)[69] as supervised discriminative models have been introduced and developed. These architectures have been widely applied to many applications like automatic speech recognition [70], natural language processing [71], face generation [72], visual recognition and description [73], image-to-image translation [74], generation images [75] and 3D object detection [76] where they have been shown to produce state-of-the-art results for various tasks.

In this thesis,a supervised discriminative architecture- CNNs has been adopted as feature learning algorithm because the convolutional neural layers have credible ability in detecting good features in the images and forming hierarchies of non-linear features where their complexity grow in while going deeper through the network. The main idea of CNN is stacking such very deep hierarchies of non-linear features because complex features cannot be learnt from a few layers. For images, it can be mathematically shown that edges and blobs are the best features that can be extracted in the earlier layers. To generate features that contain more information, earlier features (edges and blobs) are transformed again in order to obtain more complex features that consist of more information to discriminate among the classes, as shown in Figure 2.17.

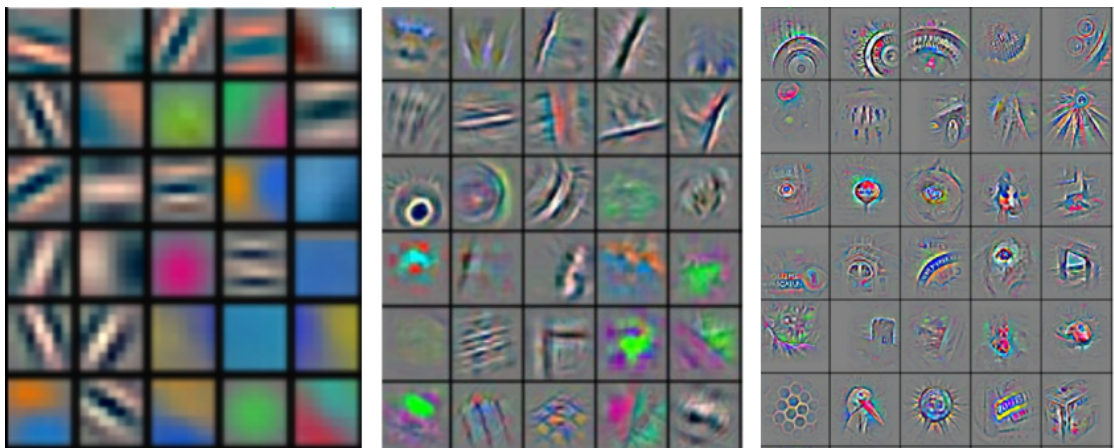


FIGURE 2.17: Features are generated by a feature learning technique. From left to right: low-level feature, mid-level-feature, and high-level feature [63].

2.4.3 General Components of Deep CNNs

A typical CNN comprises one or more convolutional layer alternated with pooling layers (subsampling layers) and then followed by fully connected layers (FC) and finally a classification/regression layer. CNNs can be considered as a special form of feed-forward multilayer perceptron neural networks (MLPs). However, the number of parameters that need to be tuned is reduced to a level that becomes tractable for the current computing power. For example, in convolution layers, a limited number of convolutional kernels is needed. Furthermore, whereas multilayer perceptron neural networks perform a composition of weighted sum of inputs and an activation function, CNNs perform a composition of convolution of inputs with trainable kernels and an activation function.

Below is a description to the main components of the CNN model including layers, function, and learning process and its set-up:

1. **Convolutional Layer:** the convolutional layer [77] represents the core building block of a deep CNN. The neurons in the convolutional layer connect to local regions of the input and compute their outputs based only on these local regions. This layer is parameterised by a set of learnable filters (kernels) convolved over the width and height of the input image and the result of each filter is called a feature map as shown in Figure 2.18.

Given an input volume size $N_i \times N_i \times D_i$, the filter or receptive field size F , the depth of the convolutional layer K , the stride parameter S and the amount of zero padding P_i , the number of neurons in the output volumes $N_o \times N_o \times D_o$ can be calculated by the formula

$$N_o = \frac{(N_i - F + 2P)}{S + 1}; D_o = K \quad (2.1)$$

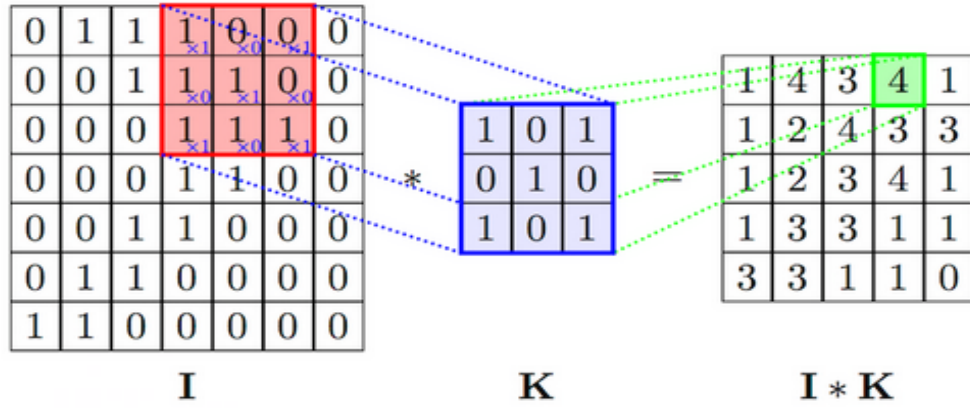


FIGURE 2.18: Diagram shows convolution operation between a sub-image (I) and kernel (K) [78].

where the value of the stride parameter S should be chosen such that N_o is an integer. The convolution layers use learned filters k over the input X_i by considering all possible offsets in the input image:

$$X_o = F\left(\sum_i X_i * k_i + b\right) \quad (2.2)$$

where F is a non-linear activation function. Figure 2.19 shows how the number of parameters is calculated in the convolutional layers.

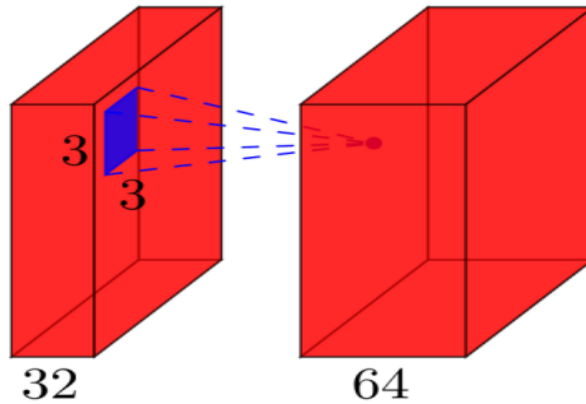


FIGURE 2.19: An example of a convolutional layer. Let the input is 32 feature maps and the output is 64 feature maps, the size of the filter is 3×3 then the size of filter in the input space is $3 \times 3 \times 32$. Thus, the total number of weights is $3 \times 3 \times 32 \times 64$.

2. **Pooling Layer:** the feature map resulting from a convolution layer is usually subsampled with $R \times R$ non-overlapped regions (windows), where R is a hyper-parameter that can be empirically defined by the user. This window is shifted over the feature map: each time, the value within this window that is most responsive (highest activation value) is selected while other values are neglected as Max-pooling scheme. In the Average-pooling scheme, while the window is shifted over the image, the elements located within the window are averaged and considered as the output of this layer. The purpose of this layer is to speed up convergence by reducing the number of parameters and amount of computation in the deep neural network, and to provide translation invariance [79]. Given an input volume of size $N_i \times N_i \times D_i$, max-pooling window size $R \times R$, and the stride parameter S , the number of neurons in the output volumes $N_o \times N_o \times D_o$ is calculated by the formula

$$N_o = \frac{(N_i - R)}{S + 1}; D_o = D_i \quad (2.3)$$

Figure 2.20 presents the two types of the pooling, Max and Average pooling, with a window size of 2×2 .

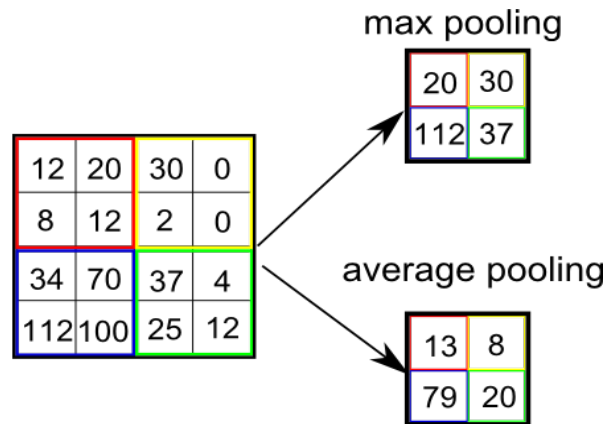


FIGURE 2.20: Max and Average pooling layer.

3. **Activation Functions:** this layer is also commonly referred to as the non-linearity layer. This layer follows the convolutional layer and is applied

to each kernel. Rectified function is the one most commonly used as an activation function for deep neural networks because it is less susceptible to vanishing gradient problems [80]. A unit employing the rectifier is called a Rectified Linear Unit (ReLU) which is defined by the formula

$$\theta : x \rightarrow \max(0, x) \quad (2.4)$$

The smoothed version of the ReLU is called the Softplus function, which can be represented as follows:

$$f(x) = \ln(1 + \exp^x) \quad (2.5)$$

Figure 2.21 explains the principal work of the ReLU activation function by setting the negative values into 0 and retaining the positive values.

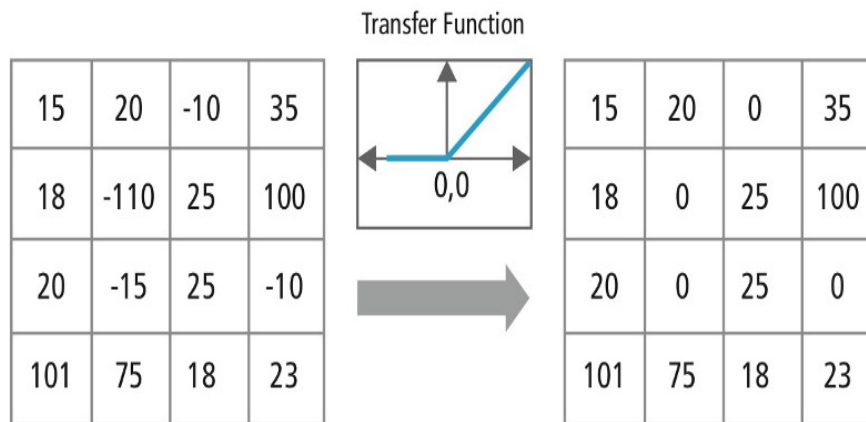


FIGURE 2.21: ReLU activation function work

Other non-linearities such as *tanh* or *sigmoid* can also be used rather than ReLU, but ReLU has been found to achieve better performance in most cases.

4. **Fully Connected Layer:** this usually represents the final layers of a deep neural network architecture. Each node in the fully connected layer is completely connected to all of the nodes in the previous layer and the weights

of these links are specific to each node. The number of neurons in the fully connected layers is considered as a hyper-parameter to be empirically chosen.

5. **Normalisation Layer:** the distribution of the input changes during the learning stage due to parameter updates which result in an internal covariate shift problem that slows down the converging process. To ensure that the model can continue training on the same input distributions (inputs are whitened, i.e. they have zero mean and unit variance, and decorrelated) and to make the network convergence faster, then normalisation becomes a crucial step.

In [81], it was shown that this idea can be simply achieved by performing standardisation of each input mini-batch and the statistics are collected from the input data during the learning process, which is called batch normalisation (BN). The collected statistics are then used for the processing, which is also has an advantage as a regulariser, keeping the weights of the deep network small.

Consider x is the layer that we want to normalise, d is the dimensions of layer x where $x = (x_1, \dots, x_d)$. Then, we can normalise the k^{th} dimension as follows [81]:

$$\hat{x}^k = \frac{x^k - E[x^k]}{\sqrt{Var[x^k]}} \quad (2.6)$$

So, in order to scale and shift the resulted normalised input, \hat{x} is transformed by:

$$y^k = \gamma^k \hat{x} + \beta^k \quad (2.7)$$

Where γ and β are parameters to be learned. The normalisation layer is usually added just before the non-linearity (activation function). Assume a model $z = f(Wx + b)$, where W , x , b and $f(\cdot)$ are the parameters to

be updated, the network input, the network bias and activation function, respectively. Thus, it becomes $z = f(BN(Wx))$ with batch normalisation, where β of the BN taking over the role of the bias b .

6. **Softmax Function:** for the classification task, given the input feature map, the Softmax function is a normalised exponential probability of class observations which assigns probability of input x belongs to class l . Softmax function is a generalisation of the binary form of Logistic Regression. This function is used widely in multiclass classification problems such as linear discriminant analysis, naive Bayes classifiers, and in the final layer of artificial neural networks. It can be given by:

$$\rho(z)_i = \frac{\exp^{z_i}}{\sum_{l=1}^l \exp^{z_l}} \quad (2.8)$$

where z is a vector of the inputs to the output layer, i indexes the output units, so $i = 1, 2, \dots, l$.

7. **Regularisation:** there are several regularisation methods used in CNNs to prevent or decrease the overfitting problem such as $L2$ norm regularisation, $L1$ norm regularisation and Dropout. The most common regularisation method is $L2$ norm, which can be implemented by penalising the squared magnitude of all parameters directly in the loss function. For every weight in the network, the term $\frac{1}{2}\lambda w^2$ is added to the cost function, where λ is the regularisation strength. In contrast, in $L1$ norm form, for each weight, the term $\lambda|w|$ is added to the objective function. $L1$ and $L2$ norm regularisation on least square loss function are defined by the following equations, respectively:

$$w^* = \underset{w}{\operatorname{argmin}} \sum_j \left(t(x_j) - \sum_i w_i h_i(x_j) \right)^2 + \lambda \sum_{i=1}^k |w_i| \quad (2.9)$$

$$w^* = \operatorname{argmin}_w \sum_j \left(t(x_j) - \sum_i w_i h_i(x_j) \right)^2 + \frac{\lambda}{2} \sum_{i=1}^k w_i^2 \quad (2.10)$$

A dropout layer [82] is an effective regularisation strategy that stochastically adds noise to the hidden layers of deep neural networks. More specifically, the overfitting problem can be alleviated by randomly dropping out the output of each hidden unit with a certain probability at each training step (i.e. multiplying hidden activations by Bernoulli distributed random variables that take the value $\frac{1}{p}$ with probability p and 0 otherwise; $p = 1$ means no drop out and low values of p imply more dropout). A deactivated unit will not take part in forward propagation or backpropagation in the training stage. At the testing stage, all of the units are re-enabled by multiplying them with one minus the probability p of masking.

Given a neural network with L hidden layers, $l \in 1, \dots, L$ is the index of the hidden layers in the network, and i is any unit in a hidden layer. Let $z^{(l)}$ denote the vector of inputs into layer l , $y^{(l)}$ denotes the vector of outputs from layer l , $W^{(l)}$ and $b^{(l)}$ are the weight and biases at layer l , and f is any activation function. The feed-forward operation can be represented by:

$$\begin{aligned} z_1^{(l+1)} &= w_i^{(l+1)} y^l + b_i^{(l+1)}, \\ y_i^{(l+1)} &= f(z_i^{(l+1)}), \end{aligned} \quad (2.11)$$

With dropout, the feed-forward operation becomes

$$\begin{aligned} r_j^{(l)} &\sim \operatorname{Bernoulli}(p), \\ \tilde{y}^{(l)} &= r^{(l)} * y^{(l)}, \\ z_i^{(l+1)} &= w_i^{(l+1)} \tilde{y}^{(l)} + b_i^{(l+1)}, \end{aligned}$$

$$y_i^{(l+1)} = f(z_i^{(l+1)}). \quad (2.12)$$

where $*$ denotes an element-wise product. Figure 2.22 explains the mechanism of deactivation of some nodes during the training after applying dropout.

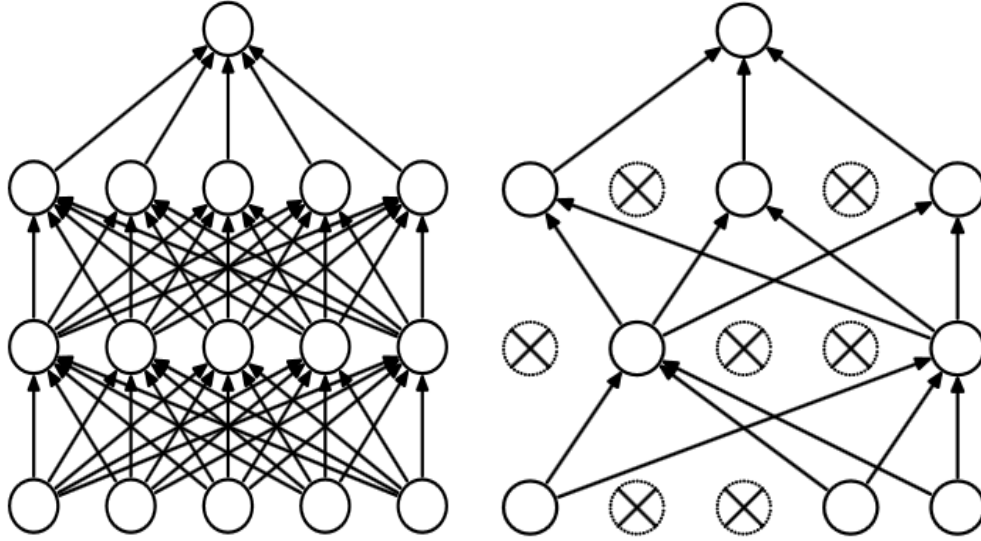


FIGURE 2.22: Comparison between two networks without and with dropout.
(a) Without dropout (b) With dropout [82].

8. **Loss Function:** the function that should be optimised and minimised through the network learning process. Given the true and predicted labels of an image, the loss function is responsible for guiding the training process by measuring the predication error for an input image. There are many types of loss function which are used in different machine learning problems like $L1$, $L2$, Hinge, (log) cross entropy, etc. For classification problems, mean squared error $L2$ (MSE) and cross entropy loss are widely used while for regression problems $L1$ loss is usually used. As an example, cross entropy loss function is given in the following equation:

$$H_{\hat{y}}(y) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}) \quad (2.13)$$

Where \hat{y} is the ground truth label and y is the prediction result of the classifier.

9. **Weight Initialisation:** after building the CNN and before starting the training, the parameters (weights) of the network should be initialised. Weight initialisation has a great impact on the convergence rate and on the final quality, and performance of a network. Many weight initialisation strategies for CNN have been introduced where they are usually chosen to be small and centred around zero. One way to characterise various initialisation strategies is given by

$$w \sim \alpha.v[-\delta, \delta] + \beta.\eta(0, \delta) + \gamma \quad \text{with} \quad \alpha, \beta, \gamma \geq 0 \quad (2.14)$$

Where $v[-\delta, \delta]$ refers to the uniform distribution in the range $[-\delta, \delta]$ and $\eta(0, \delta)$ is the normal distribution with mean zero and variance δ . Initialisation by Gaussian distributions, constant values [83], uniform distribution, orthogonal and sparse matrix [84], Xavier [85] and He [86] are some examples of weight initialisation schemes. TABLE 2.1 shows some commonly used weight initialisation schemes.

TABLE 2.1: Weight initialisation schemes. n_{in} , n_{out} is the number of nodes in the previous and next layers.

Method	α	β	γ	δ
Constant	$\alpha = 0$	$\beta = 0$	$\gamma \geq 0$	N/A
Xavier/Glorot uniform	$\alpha = 1$	$\beta = 0$	$\gamma = 0$	$\delta = \sqrt{\frac{6}{n_{in} + n_{out}}}$
Xavier/Glorot normal	$\alpha = 0$	$\beta = 1$	$\gamma = 0$	$\delta = \sqrt{\frac{2}{n_{in} + n_{out}}}$
He	$\alpha = 0$	$\beta = 1$	$\gamma = 0$	$\delta = \sqrt{\frac{2}{n_{in}}}$

To study the effect of the parameter initialisation method, a comparison of many initialisation techniques is presented in Figure 2.23 [87]. In the left-most plot of the figure, the parameters are initialised by setting them to zeros but the network does not have the ability to learn. In the middle case, the weights are initialised from a normal distribution with a standard deviation of 0.4, while in the third plot they are initialised from normal distributions

with variances that are inversely proportional to the number of inputs into each neuron. In the second case, the loss improves over time but the rate of convergence is very low. In the third plot, the convergence has been highly accelerated.

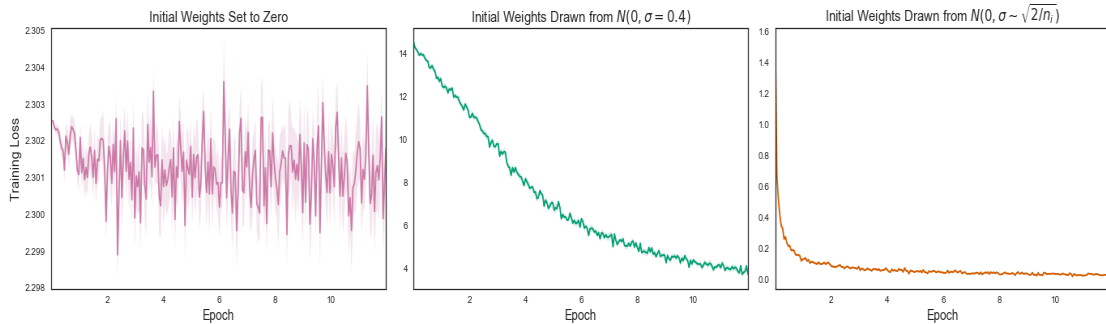


FIGURE 2.23: Training loss of CNN over 12 epochs with different weight initialisation schemes [87].

10. **Learning/Training:** the above explained components - data, connectivity, loss and activation functions and layers - are the static parts of the CNN. To consider dynamic part of the CNNs, the process of parameters learning and finding an optimal objective function are introduced. The CNNs are typically learned by minimising a loss function with respect to the parameters (weights) of the network based on a batch of training images using a certain optimisation strategy. The training of deep CNNs is almost the same learning as the traditional neural network (ANN) (i.e. gradient-descent method combined with a backpropagation algorithm). However, due to some special mechanisms in CNNs such as weight sharing, local receptive field, and pooling, slight changes in the gradient calculation in backpropagation are needed [88]. In backpropagation of CNNs, the matrix multiplications are replaced by convolutions and the pooling layer computes the error which is acquired by a single value (winning value). This value is either the maximum value within a window if it resulted from applying Max-pooling or the average value if it resulted from applying Average-pooling. Therefore, the backward

propagations in CNNs differ depending on what layer is being propagated through.

Most relevant optimisation strategies for CNNs are based on Stochastic Gradient Descent (SGD), which updates the parameters according to the rule:

$$W_{ji}^{(t+1)} \leftarrow W_{ji}^{(t)} + \Delta W_{ji}^{(t+1)} \quad \text{with} \quad W_{ji}^{(t+1)} = -\eta \frac{\partial E_B}{\partial W_{ji}} \quad (2.15)$$

Where $\eta \in (0, 1)$ is called the learning rate.

Optimisation algorithms based on SGD such as Momentum [89], AdaGrad [90], AdaDelta [91], RMSProp [92] and Adam [93] are proposed and developed to reduce the convergence time by using a certain adjustment of learning rate. For example, the Momentum scheme can be represented as follows:

$$W_{ji}^{(t+1)} \leftarrow W_{ji}^{(t)} + \Delta W_{ji}^{(t+1)} \quad \text{with} \quad W_{ji}^{(t+1)} = -\eta \frac{\partial E_B}{\partial W_{ji}} + \alpha \Delta W_{ji}^{(t)}, \alpha \in [0, 1] \quad (2.16)$$

11. **Data Augmentation:** it increases the size of data artificially (the CNN requires a huge data size to be trained) in order to decrease the possible overfitting problem. In every epoch (single pass of all training data through the network) during training, the image is randomly augmented. There are many ways to perform data augmentation, such as random rotation, random horizontal and vertical flipping, random translations, random zooming and random shearing. It has been noticed that applying a variety of data augmentation strategies can boost the network performance [94].
12. **Transfer Learning:** this can be defined as the transfer of knowledge from one learned task to a new task. Deep convolutional neural networks typically demand a huge data size in order to be trained but sometimes a provided dataset may not be adequate (too small) to train a full network. While data

augmentation is found to be effective in doubling the size of data artificially, transfer learning has also proven to be a valuable option in order to deal with the lack of enough annotated images. Transfer learning strategy can be achieved by fine-tuning the pre-trained CNN which was trained on a huge data size through removing top layers (the fully connected layers) and re-training the CNN by initialising it from the trained weights rather than from scratch. Another way to exploit this strategy can be achieved by considering the convolutional layers of the pre-trained CNN as a fixed feature extractor and then passing these features into a linear classifier to train it [95].

13. **Software/Hardware Designing:** many libraries have been introduced to design and setup the operations and components of the deep neural networks by allowing a high-level implementation rather than worrying about low-level efficient implementations. These libraries have alleviated the need for implementation from scratch, which can be time consuming. The open source frameworks such as Caffe [96], Tensorflow [97], Theano [98] and Torch [99] which are written in different programming languages like Python, C/C++ and Java, provide efficient GPU implementations (parallel computing) for the important components of the CNNs. Also, there are many packages written based on one or more of these libraries such as Lasagne (<https://github.com/Lasagne/Lasagne>) and Keras (<https://keras.io/>).

For hardware, the NVIDIA Company has launched many GPUs and suites of tools such as CUDA parallel programming platform to help in running deep learning software. Figure 2.24 shows one of the GPUs invented by NVIDIA [100].

All of the experiments in this study were conducted on an HP Z440 running Linux Mint with 16GB RAM, an Intel Xeon E5 3.50GHz processor and NVIDIA GTX TITAN X 12GB GPU card with 3072 CUDA parallel-processing cores. Python deep learning library; Lasagne, built on the top



FIGURE 2.24: NVIDIA GPU which was used to run our experiments [100].

of Theano [98] was used to implement and train our convolution neural networks. Lasagne has efficient implementations of each of the CNN layers, a diversity of activation functions, many optimisation methods, and transparently supports training networks on GPUs.

2.4.4 Deep Convolutional Neural Network Architectures

The very first convolutional neural network was developed in 1998 by Yann LeCun and called LeNet [77]. The architecture of LeNet was very fundamental, comprising of 3 layers, convolutional filters followed by subsampling operations with nonlinearities in form of *tanh* or *sigmoid*. With very slow CPUs available during that time, it was tricky to train the network. In the years from 1998 to 2010, there was no or very slow progress in developing of deep neural network. Availability more and more data and developing the computing power represented by GPUs made deep neural network progress by tackling the difficulties that had been encountered a decade ago.

In 2010, Ciresan et al. [101] published the first implementation of GPU neural network in hand-written digit recognition application. In 2012, Alex et al. [69] released AlexNet (shown in Figure 2.25) which was a deeper and wider version of LeNet and won the difficult ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (1.2 million images and 1000 classes). AlexNet was a significant breakthrough with respect to the previous methods and following this network, a variety of CNN models have been produced but all of them are attributed to Alex's work.

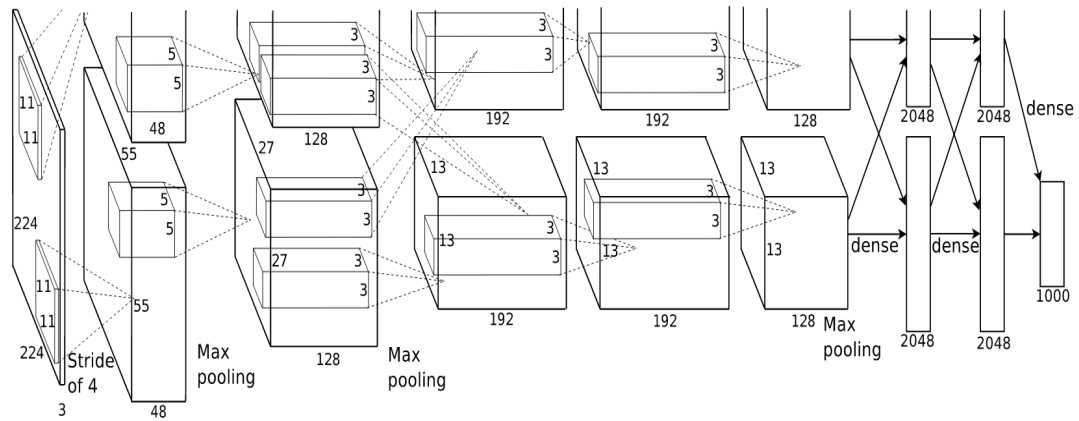


FIGURE 2.25: AlexNet architecture [69].

Recently, many convolutional neural networks (CNNs) architectures have achieved state-of-the-art performance due to their ability to learn hierarchical features from raw input data without the need of hand-crafted features such as ZF net (2013) [83], Network In Network (NIN) (2013) [102], VGG Net (2014) (shown in Figure 2.26) [103], GoogleNet (Inception) (2015) [104], Spatial transformer net (2015) [105], Spectral net (2015) [106], Fully CNN (2015) [107], U-net (2015) [108], ENet (2016) [109], ResNet (2015) [110], and DenseNet (2017) [111].

Figure 2.27 shows different CNN models comparing them in terms of the number of operations required for a single forward pass and number of parameters versus the accuracy in the ImageNet [113] challenge. These deep convolutional neural net (convnet) architectures have succeed in improving many computer vision applications such as image classification [69], object recognition [114], key-point localisation [115] and many other applications [116], [117].

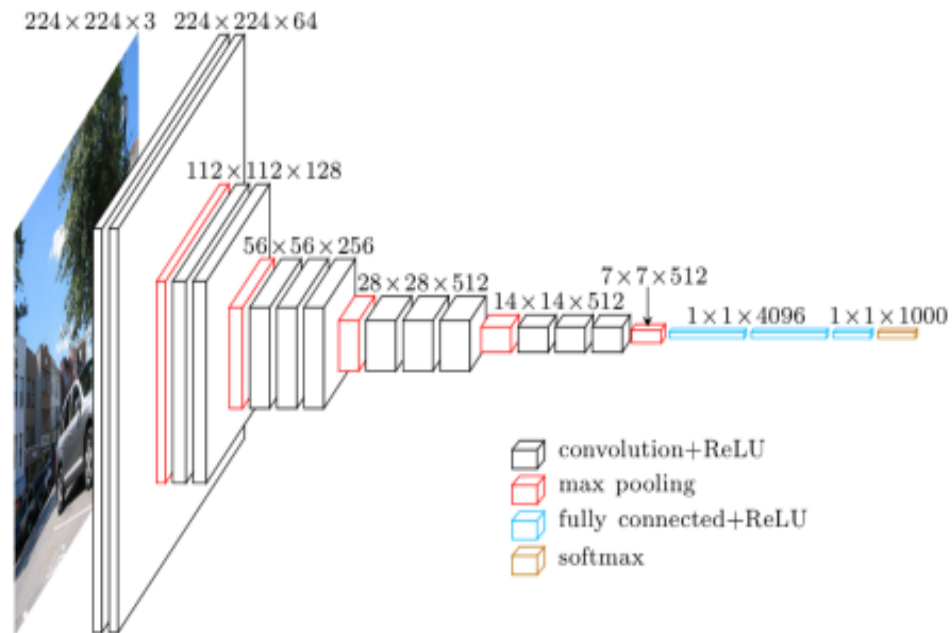


FIGURE 2.26: VGG architecture [103].

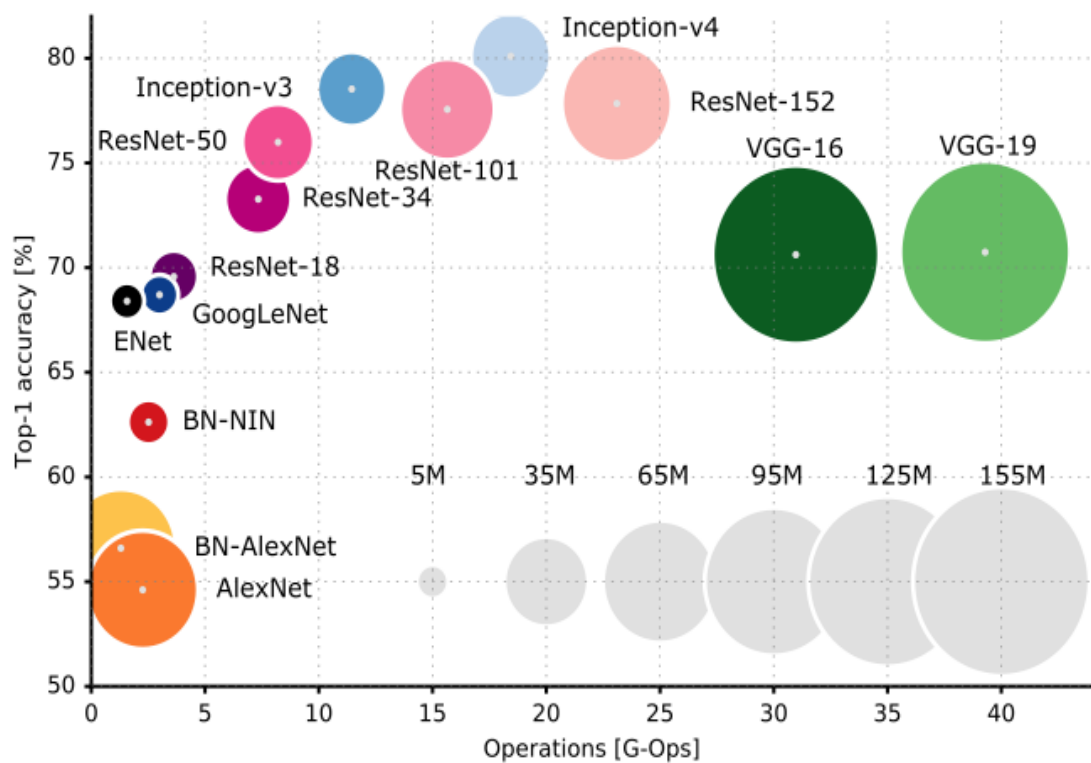


FIGURE 2.27: Performance of different CNN models in a classification task [112].

While the main layers, functions and learning process demanded in the CNN design were described in the previous section, more complex components will be discussed in this section (i.e. CNN blocks). CNN blocks are the base of the design in the recent CNN architectures; they act in a similar way to a layer, but they are themselves composed of layers. The blocks are described as follows.

1. **Inception Block:** unlike traditional CNNs (i.e. AlexNet and VGG), the CNN operations are happening in parallel in the inception block, which was introduced in GoogleNet [104] by Google. Whereas in typical CNNs only one operation at a time (either pooling or convolution) can be considered, the inception block allows all these operations to be performed in parallel. Moreover, the inception module suggests the use of more than one convolutional filter size in one block such as 1×1 , 3×3 , and 5×5 , as shown in Figure 2.28). The output of all the filters is concatenated and passed on as input to the next layer.

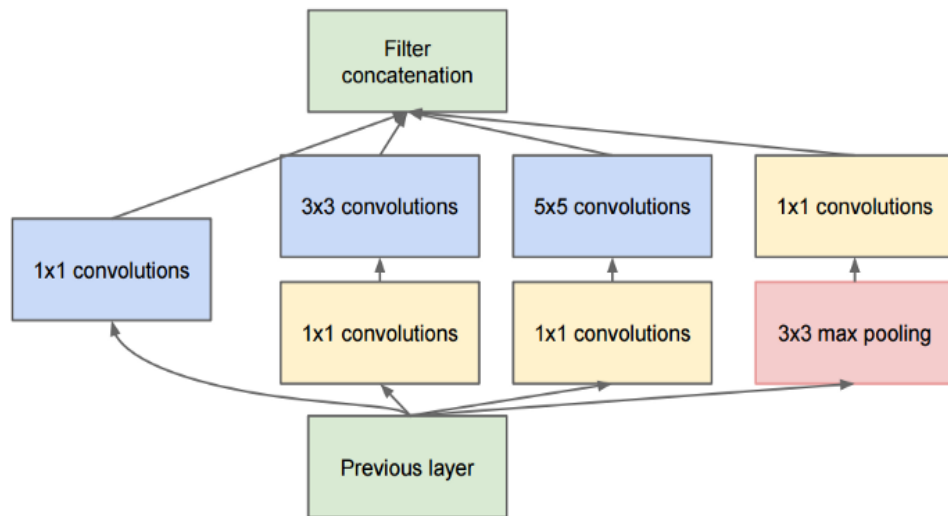


FIGURE 2.28: Inception block [104].

2. **Residual Block:** residual blocks were first introduced in the Residual Network (ResNets) [110] by Microsoft. The idea behind residual blocks is to add an identity connection which skips two layers, as shown in Figure 2.29. Let

us have an input feature, X , that goes through two convolutional filters; the output will be $F(X)$ after passing on the activation function, which is alternated with the convolution layer. That result is then added to the original input feature X to produce $H(X) = F(X) + X$. In traditional CNNs, $H(X)$ is equal to $F(X)$ where there is no more consideration for the input feature X and it does not keep any information about it. In a residual network, the identity connection adds the input feature map onto the other feature maps.

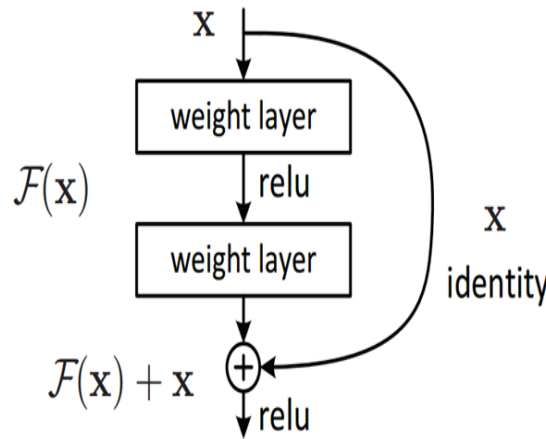


FIGURE 2.29: Residual block [110].

3. **Dense Block:** this is introduced in the Dense Network (DenseNet) [111]. Contrary to ReseNet, DenseNet suggests concatenating outputs from the previous layers instead of using the summation by connecting each convolutional layer to subsequent convolutional layers. In traditional CNNs, there are only L connections between layers but the dense block has $\frac{L(L+1)}{2}$ connections. The authors [111] claim that their proposed network prevents features from being relearned and allows many fewer filters per convolutional layer sequentially. Transition layers, which consist of pooling and 1x1 convolutions, are used to connect the dense blocks. A dense block is shown in Figure 2.30, where k represents the number of filters added per layer (growth rate).

The mentioned blocks are the basic units required to implement the most recent convlutional networks shown in Figure 2.31.

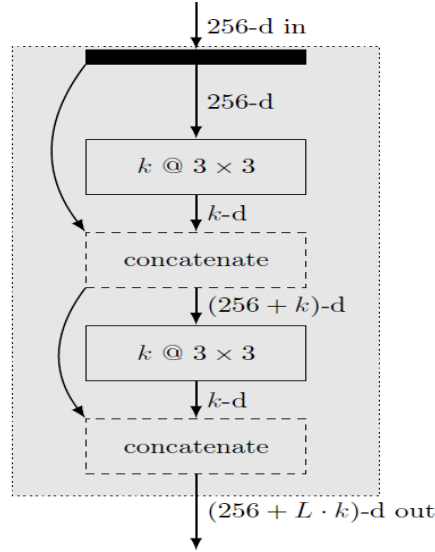


FIGURE 2.30: Dense block [118].

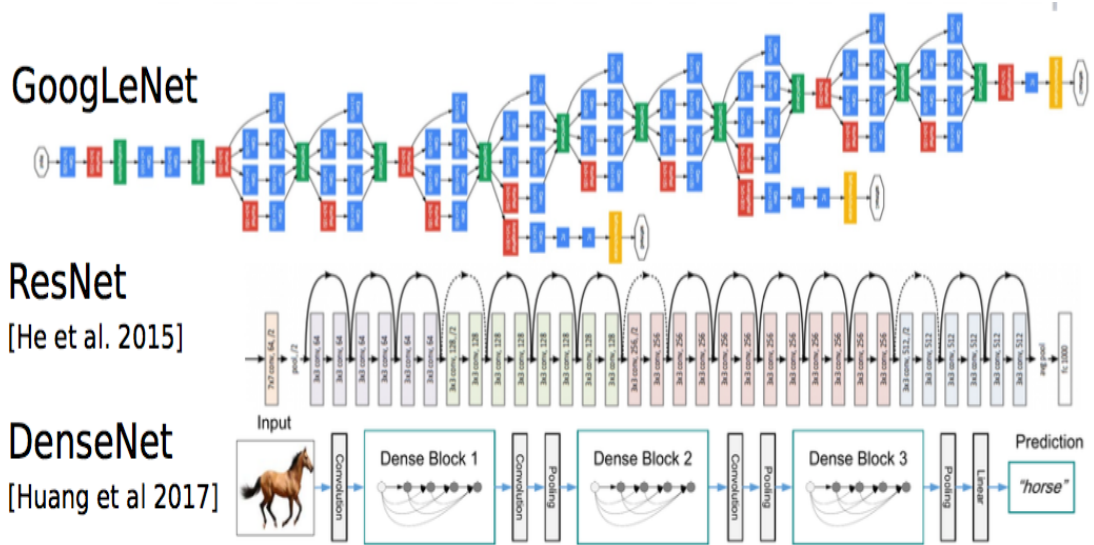


FIGURE 2.31: Advanced (block-based) deep convolutional neural network architectures.

2.5 Applications of Deep CNNs in Medical Image Analysis

Medical image analysing tasks requires abnormalities identification, quantify measurement and changes overtime. Decades ago, medical image analysis was performed using either low-level pixel processing strategies such as region growing and edge detector filters or mathematical modelling such as line fittings in order

to solve a particular task. As the size of healthcare images has been increasing due to the rapid growth and the huge advancements in image capturing devices, the analysis of medical images has become more challenging.

However, the deep learning techniques have overcome many obstacles, supported by the most successful deep learning form for medical image analysis in a variety of applications, which is deep convolutional neural networks (CNNs) [119]. The contribution of deep neural networks in medical images can be categorised according to the canonical image analysis tasks, detection, classification, segmentation, and many other tasks such as registration, content image retrieval, image generation and enhancement.

1. **Detection/Localisation:** the detection task in medical images involves anatomical/landmark localisation. Dou et al. [120] exploited a 3D fully convolutional network (FCN) model to retrieve the candidates of cerebral microbleed (CMBs), and then applied a pre-trained 3D CNN model to discriminate CMBs from hard mimics. The author of [120] evaluated their method on magnetic resonance (MR) images and achieved sensitivity of 93.16%. Furthermore, Ghesu et al. [121] merged a reinforcement learning model with hierarchical features extracted through a deep neural network to detect the anatomical landmarks in many image types, 2D magnetic resonance, 2D ultrasound and 3D computed tomography (CT) images, achieving mean detection errors of 1-2 pixels. In [122], Lu et al. presented a framework exploiting CNN to capture global context and FCN to capture the local context aiming for 3D organ localisation in CT images and achieved a mean Euclidean distance error of 3.9 pixels. Automated kidney detection was reported in [123] by Ravishankar and Sudhakar through fine-tuning a pre-trained CNN which achieved 20% higher performance than some other methods.

2. **Segmentation:** the objective of medical image segmentation is to identify the pixels/voxels that form the boundary or the area of a specific organ or lesion. One of the earliest segmentation tasks performed based on CNN was reported by Ciresan et al. [124] in 2012. They segmented neuronal membrane in electron microscopy (EM) images and achieved a pixel-error rate of 0.06. Following Ciresan's paper, many authors started exploiting the CNN in different segmentation issues. Cernazanu-Glavan and Holban [125] segmented the bone structure in X-ray images by using a CNN-based approach and obtained a pixel error rate of 0.204. Zhang et al. [126] proposed a method to segment infant brain tissue in MR images by employing CNNs and reported dice coefficients of 0.8323, 0.8531 and 0.8798 for cerebrospinal fluid (CSF), grey matter (GM) and white matter (WM) tissues, respectively. Moreover, Fu et al. [127] combined a CNN with a Conditional Random Field (CRF) to model long-range pixel interaction in order to segment the vessels of the retina. They used fundus images from different datasets and obtained accuracy of 0.9523% in an average time of 1.3s. Most recently, Zilly et al. [128] presented a method to segment the optic disc and optic cup in the retinal fundus images using ensemble learning based on CNNs. They evaluated their method on a public dataset and reported an F-score of 97.3% for optic disc segmentation and 87.1% for optic cup.
3. **Classification:** Medical image classification involves investigating the presence of a certain lesion or disease and grading the severity level of that disease in a given image. In 2013, Malon and Cosatto [129] combined shape-based features with CNN to identify the mitotic figures and achieved precision of 0.747% and F-score of 0.659% in histology images. Furthermore, Payan and Montana [130] used a 3D CNN pre-trained with sparse auto-encoders to predict the presence of Alzheimer's disease in magnetic resonance images (MRI)

and reported a classification accuracy of 89.47%. In addition, Anthimopoulos et al. [131] proposed a CNN-based method to classify interstitial lung disease into the disease stages in CT images, achieving a classification accuracy of 85.5%. Moreover, Antony et al. [132] quantified the severity of knee osteoarthritis (OA) by fine-tuning a pre-trained CNN model on OA X-ray images and achieved 94.4% correctly classified images. In [133], Wang et al. developed an automated system based on CNNs to identify the breast arterial calcifications (BACs) in mammograms and evaluated their system by linear regression analysis, yielding a coefficient of determination of 96.24%.

2.6 Summary

This chapter has presented an overview of the anatomical structure of the eye, retinal diseases, retinal imaging modalities, and the theoretical concept of deep learning including CNNs. From the works reported in this chapter it can be noticed that deep neural networks have demonstrated an interesting performance and provided state-of-the-art results in many applications on variety types of medical images. As result of that, we were encouraged to exploit deep CNNs to solve issues related to retinal images considering different tasks such as localisation in Chapter 3, segmentation in Chapter 4 and classification in Chapter 5. In Chapter 3, the work on detecting the location of the most important retinal landmarks will be presented. Next, in Chapter 4, the segmentation of choroid boundaries as well as the optic disc/cup boundaries will be taken into consideration. Finally, the classification of retinal diseases such as glaucoma and DM and the identification of the presence of retinal fluids will be considered in Chapter 5. More details about the methodologies used to achieve these tasks are explained and discussed in each individual chapter.

Chapter 3

Detection of Optic Disc and Fovea

In this chapter, the problem of locating the centres of both fovea and optic disc (OD) in the retinal fundus images is considered. In §3.1, an introduction of both of these retinal landmarks and their definition in retinal fundus images are given. §3.2 provides a brief review of the previous work related to the detection of the OD and the fovea followed by presenting the novelty of this work in §3.3. In §3.4, the proposed methodology for detecting the OD and fovea locations is presented. Next, the experimental results are described in §3.5. Finally, this work is discussed in §3.6 and the chapter is concluded in §3.7.

3.1 Introduction

The knowledge of the OD and fovea (centre of the macula) locations in the retina is considered essential for the diagnosis and screening of many retinal diseases, such as glaucoma, diabetic maculopathy (DM) and age-related macular degeneration (AMD). The additional significance of detecting the fovea is that the closer a lesion is to it, the more likely the lesion is to cause visual impairment or blindness. On the other hand, the OD centre is often regarded as a reference point for locating other retinal structures. For example, it can be used as the starting point for tracking retinal vessels in blood vessel tracking algorithms [134]. In addition, the

OD diameter (D) is usually used as the reference to measure the size and location of other anatomical and pathological structures in the retina.

The OD appears as a bright yellowish oval region within colour fundus images through which the blood vessels enter the eye. The macula is the centre of the retina which is responsible for our central vision. The fovea is a small depression in the centre of the macula. It has a darker appearance compared to the surrounding retinal tissue due to the high concentration of macular pigment. On average the vertical OD diameter is about $1800\ \mu m$. The location of the fovea centre is about $2.5D$ from the optic disc centre. The foveal radius is between $1/3$ and $1/4$ of the macula radius which is roughly equal to one optic disc diameter (D) [135],[136].

3.2 Related Work

In the literature, there has been a number of studies conducted to determine the locations of the fovea and OD. Many of these studies only locate either the OD or fovea and not both. Below is a brief review of the major algorithms published in the literature for detecting the OD, followed by fovea detection methods.

Many of the reported methods use geometric information of the vascular tree to detect the OD [137], [7]. Hoover and Goldbaum [137] exploited the spatial relationship between the OD and retinal blood vessels and proposed a fuzzy convergence algorithm to locate the origination point of the blood vessel network. This origination point was considered as the OD centre in the retinal fundus image.

Foracchia et al. [138] proposed a geometrical model to calculate the general direction of retinal blood vessels at any given location in an image using the coordinates of the OD centre as the two model parameters. Simulated annealing optimisation technique was used to identify these two parameters.

Furthermore, Fleming et al. [139] presented a method based on the elliptical form of retinal blood vessels to obtain the approximate locations of the OD and

fovea. The circular edge of the OD and the darker appearance of the fovea were exploited to refine these approximated locations.

In addition, Tobin et al. [140] used vasculature segmentation results for optic disc detection by determining density, average thickness, and average orientation of the blood vessels in relation to the position of the OD. Youssif et al. [7] described a method that can detect the optimal OD centre point by measuring the difference between the matched filter output and the vessels directions.

Niemeijer et al. [3] formulated the problem of detecting the OD and foveal centres as a regression problem. They utilised a k-nearest neighbours regressor to measure the distance in an image to the object of interest at any given location using a set of features extracted at that location.

Furthermore, a method based on Sobel operators and the Hough transform for the detection of the OD in retinal fundus images was formulated by Zhu et al. [141]. They determined the centre and radius of the OD by approximating the margin of the optic nerve head into a circle using the Hough transform.

Moreover, Lu et al. [142] designed a technique based on the circular transformation to locate the circular shape of the optic disc and colour variation across the OD boundary. The centre and the boundary of the optic disc were located by exploiting the pixels with the maximum variation along radial line segments.

Yu et al. [143] presented a method for detecting the optic disc location using template matching techniques. The OD location was determined using the characteristics of the vessels on the OD.

In [144], Dehghani et al. proposed a histogram based method which uses four images from the DRIVE dataset as a template to locate the centre of the OD where each histogram represents one colour from the RGB colour image components (red, blue, and green). The template was constructed by calculating the average of these histograms.

Harangi et al. [145] adapted the most recent OD detectors and organised them into an ensemble and complex framework in order to merge their strengths and

maximise the accuracy of OD detection. To determine the final OD position, a maximum-weighted clique was founded.

Many of the fovea localisation approaches presented in the literature have exploited the vasculature and other contextual information. Li and Chutatape [146] presented a model-based approach by combining the information provided by the main vessel arcades and the low intensity pixels in the fovea region. A parabola fitting method was used to detect the fovea and the fovea centre was identified using a thresholding scheme in the region of interest.

Niemeijer et al. [147] formulated a method based on a cost function that is based on both global and local cues to find the fovea. In addition, mathematical morphology and anatomical knowledge based methods were used to estimate the location of the fovea by Welfer et al. [148]. In their proposed system, extracting the region of interest containing the fovea was achieved initially by calculating the centre and diameter of the OD. After that, a set of fovea candidates was obtained using a morphological operation. To detect the centre of the fovea, it was selected as the centroid of the darkest candidate.

Qureshi et al. [149] proposed a method based on a combination of several algorithms for detecting the fovea and OD. They proved that ensemble algorithms can achieve better performance than a single algorithm for detecting these centres.

Moreover, a fast radial symmetry transform was used by Giachetti et al. [150] for the detection of the fovea and OD centres. The centres of symmetry of dark and bright regions were detected by applying the transform on coarsened and vessel-inpainted images and the results were combined with a vascular density estimator.

Gegundez-Arias et al. [151] detected the location of the fovea centre by means of prior known anatomical features. These features were used to localise a ROI fovea-containing sub-image. A multi-thresholding scheme using gray-level value criteria was applied and a contour map was created to calculate the fovea centre. In [8], Aquino et al. formulated a method based on combining the visual and

anatomical features of the macula and the OD for detecting the fovea centre. Table 3.1 summarises the methods reported in the literature. Detection accuracy, computational time, evaluation criterion and the dataset used are presented in this table for previous work where they are available in the original paper.

TABLE 3.1: Existing optic disc (OD) and fovea (F) detection methods in the literature.

Authors	Approach	Detected land- marks	R	Success rate	Run time	Dataset(Name ,size,[images size])
Hoover [137]	Relationship between OD and blood vessels, fuzzy convergence algorithm	OD	60	Acc.: 89%	4 min.	(STARE [152], 81, [605×700])
Foracchia [138]	Geometrical model, blood vessels direction	OD	60	Acc.: 97.53%	2 min.	(STARE, 81, [605×700])
Li [146]	Parabola fitting	F	NA	Sensitivity: 100%	NA	(Local, 35, [512×512])
Fleming [139]	Visual characteristics of blood vessels, fovea and OD	OD, F	119	Acc.: 98.4%, 96.5%	2 min.	(Local, 1056, [2160×1440])
Tobin [140]	Characteristics of blood vessels in relation to OD position	OD, F	65	Acc.: 90.4%, 92.5%	NA	(Local, 345, [1024×1152])

Niemeijer [147]	Cost func- tion and a point dis- tribution model	OD,F	50	Acc.: 98.4%, 94.4%	10 min.	(Local, 500, [768×576])
			50	Acc.: 94% , 92%	10 min.	(Local, 100, [2048×1536])
Aliaa Youssif [7]	2D Gaussian matched fil- ter	OD	60	Acc.: 98.77%	3.5 min.	(STARE, 81, [605×700])
			NA	Acc.: 100%	3.5 min.	(DRIVE [153], 40, [565×584])
Niemeijer [3]	k-NN regres- sor	OD, F	50	Acc.: 99.4%, 96.8%	7.6s	Local, 500, 768×576
			50	Acc.: 93%, 89%	7.6s	Local, 100, 2048×1536
Zhu [141]	Sobel oper- ator, Hough transform	OD	40	Acc.: 90%	NA	(DRIVE, 40, [565×584])
Lu [142]	Circular transforma- tion	OD	60	Acc.: 99.75%	5s	(STARE, 81, [605×700])
			60	Acc.: 97.5%	5s	(ARIA [154], 120, [576×768])
			NA	Acc.: 98.77%	5s	(MESSIDOR, 1200,[1440×960, 2240×1488, 2304×1536])
Welfer [148]	Selection of ROI and morphology	F	34	Acc.: 100%	NA	(DRIVE, 40, [565×584])

			34	Acc.: 92.13%	NA	(DIARETDB1 [155], 89, [640×480])
Yu [143]	Template matching technique	OD	70, 100, 110	Acc.: 99%	4.7s	(MESSIDOR, 1200, [1440×960, 2240×1488, 2304×1536])
Qureshi [149]	Combining the pre- diction of multiple algorithms	OD, F	NA NA NA	Acc.: 97.64% 96.79% Acc.: 97.79%, 98.74% Acc.: 100%, 91.73%	NA NA NA	(DIARETDB0 [156], 130, [1500×1152]) (DIARETDB1, 89, [1500×1152]) (DRIVE, 40, [565×584])
Dehghani [144]	Template implemented from three histograms	OD	NA NA NA	Acc.: 100% Acc.: 91.36% Acc.: 98.9%	27.6s 27.6s 27.6s	(DRIVE, 40, [565×584]) (STARE, 81, [605×700]) (Local, 273, [720×576])
Giachetti [150]	Fast radial symmetry transform	OD, F	70, 100, 110	Acc.: 99.66%, 99.1%	5s	(MESSIDOR, 1200, [1440×960, 2240×1488, 2304×1536])

Gegundez-Arias [151]	Priori known anatomical features and thresholding	F	68, 103, 109	Acc.: 96.92%	0.94s	(MESSIDOR, 1200, [1440×960, 2240×1488, 2304×1536])
Aquino [8]	Visual and anatomical macula and OD feature-based method	F	68, 103, 109	Acc.: 98.24%	10.88s	(MESSIDOR, 1136, [1440×960, 2240×1488, 2304×1536])
			82	Acc.: 94.38%	10.88s	(DIARETDB1, 89, [1500×1152])
Harangi [145]	Ensemble-based framework (combining probability models)	OD	NA	Precision: 98.46%	0.25s	(DIARETDB0, 130, 1500×1152)
			NA	Precision: 98.88%	0.25s	(DIARETDB1, 130, 1500×1152)
			NA	Precision: 100%	0.25s	(DRIVE, 40, [565×584])
			NA	Precision: 98.33%	0.25s	(MESSIDOR, 1200, [1440×960, 2240×1488, 2304×1536])

From the above review, it can be noticed that most of the previous studies have exploited the visual appearance or anatomical features for the detection of the OD

and fovea in order to identify their positions [137], [139], [140], [142], [148], [150] and [8]. These methods will suffer when these features are very weak or invisible due to pathologies. Some other methods rely on machine learning algorithms and feature extraction to localise and detect anatomical structures [3], [145], and [149], but the accuracy of these methods largely depends on the type and quality of the feature sets which are hand-crafted. Inspired by our observations, it is proposed to introduce new deep learning techniques to address this.

3.3 Novelty

In this chapter, a multiscale sequential deep learning technique is proposed which is aimed at detecting the centres of the OD and the fovea. The main contributions and advantages of this work are summarised as follows:

1. The application of deep convolutional neural networks to the detection of retinal landmarks is novel and promising. A suitable convolutional neural network developed to detect specifically the optic disc and fovea centres.
 - (a) **Speed and automation:** This results in a fast method requiring no user input.
 - (b) **Independence:** The method is independent on other techniques succeeding such as segmentation or detecting other landmarks.
 - (c) **No hand-crafted features:** Since features do not need to be manually defined, it has been avoided the difficulty encountered by conventional machine learning algorithms in identifying the best feature set that represents the data. This also removes the requirement of a skilled technician to identify such features manually which takes a considerable amount of time and can produce subjective results, particularly with a large dataset.

- (d) **Accurate simultaneous detection:** It has been detected more than one position simultaneously, retaining high accuracy for each.
 - (e) **Robustness:** The method is robust in the sense that it continues to work well even on poor quality images.
2. A multiscale approach to convolutional neural networks is developed to focus on the region of interest.
 - (a) **Improved Accuracy:** This approach allows the method to focus on the region of interest, removing redundant background data from consideration and facilitating refinement of the localisation. This results in significantly increased accuracy in the cases of the fovea and the optic disc.
 3. Inter-dataset training and evaluation using multiple datasets.
 - (a) **Generalisation:** This demonstrates generalisation of the method to new data, from separate datasets and graders, and captured from different devices.
 4. Variable optic disc radius (R) is incorporated into evaluation criteria.
 - (a) **Evaluation accuracy:** Incorporating this variable measure into our testing allows more accurate evaluation while others use fixed R value for evaluation.

3.4 Materials and Methods

3.4.1 Materials

The MESSIDOR [157] and Kaggle [158] datasets have been used in this work. The MESSIDOR database comprises 1200 images captured using a colour 3CCD camera on a Topcon TRC NW6 with 45 degree field of view. The MESSIDOR

images were captured using 8 bits per colour plane at a size of 1440×960 , 2240×1488 or 2304×1536 pixels. Moreover, 10,000 image from Kaggle dataset are used for training and testing. Kaggle [158] was provided by EyePACS, a free platform for retinopathy screening. The images of Kaggle dataset used in Kaggle diabetic retinopathy detection competition come from different models, resolutions and types of cameras and feature very mixed quality. Clinician have rated the presence of diabetic retinopathy in each image on a scale of 0 to 4, according to International Clinical Diabetic Retinopathy severity scale (ICDR): 0 (No DR), 1 (Mild DR), 2 (Moderate DR), 3 (Severe DR) and 4 (Proliferative DR).

The optic disc and foveal centre point coordinates were not provided in the original dataset for both datasets, for this work they were obtained from annotations from a combination of two expert graders from the Liverpool Reading Centre. An in-house program developed in Matlab (version 2016a, Mathworks Inc, Natick, MA) was used by the grader. This software program was developed to support annotations of anatomical or pathological features required by clinical trials, and allows the grader to visualise the image, selecting the location by mouse click and make correction on the selection. These annotated locations together with the images were used to train and evaluate the performance of the implemented networks.

3.4.2 Pre-processing

It is worth noting that detecting the centres of the fovea and OD is a regression task. It seems unnecessary to use colour information because the colours may just add extra complexity. For this reason, all of the images were converted to grey scale for use. For the purpose of this study, the images were resized to 256×256 pixels and the annotated centre point coordinates of both the OD and fovea were scaled accordingly. The contrast of the resized images was enhanced by applying the contrast-limited adaptive histogram equalisation technique [43] so as to reduce

uneven illumination in the images as shown in Figure 3.1. The pixel values of the enhanced images were scaled between $[0, 1]$ and the coordinates of the centre points were scaled between $[-1, 1]$.

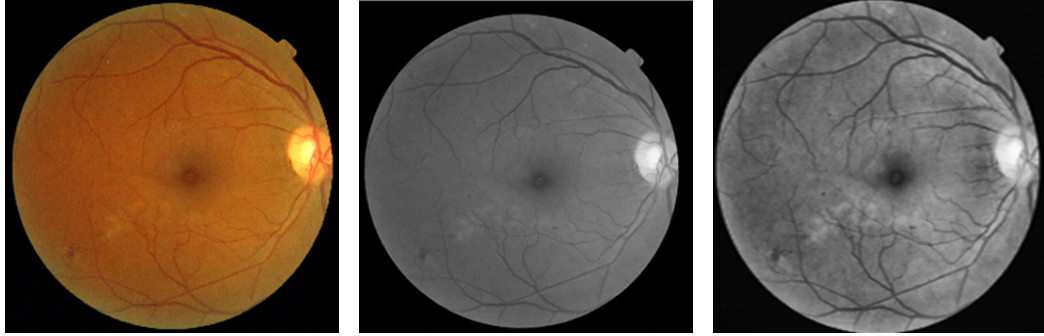


FIGURE 3.1: Image pre-processing stages. From left to right: Original image, Grayscale image, and Enhanced image.

3.4.3 Landmarks Detection by CNN

The proposed system consists of two stages, in the first stage the whole resized images along with the scaled centres are fed to the implemented CNN. The output of the first stage is the centres of both the OD and fovea (F).

In the second stage, the detected centres from the first CNN are used to obtain the refined regions of interest of both the OD and F by cropping the region around these centres by $2R$ radius value (R represents the OD radius). These resized ROI for both the F and OD along with the scaled ground truth centres are used to train the CNNs in the second stage.

Therefore, the first stage is used to obtain the ROIs for both F and OD while the second stage is aimed to detect the centres by classifying the features extracted automatically by the convolutional filters. As we go deeper through the convolutional neural network, the convolutional layers are able to describe more and more complex features. The block diagram of the proposed deep multiscale sequential convolutional neural network is presented in Figure 3.2.

Different network architectures and data augmentation strategies were evaluated in comparison to conventional (standard) neural networks. To train the

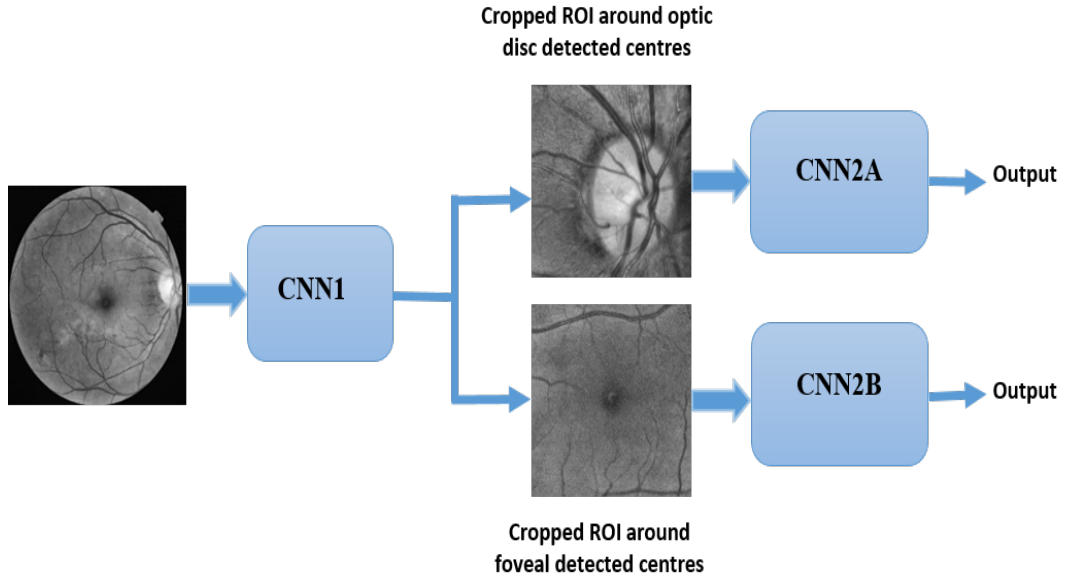


FIGURE 3.2: Block diagram of the proposed system.

networks by updating the weights, SGD with a momentum optimisation algorithm having an adaptive learning rate (start=0.03, stop=0.0001) and adaptive momentum parameter (start=0.9, stop=0.999) is used.

The weights of the kernels for the implemented convolutional layers are initialised from a uniform distribution within chosen intervals. These intervals are configured depending on the weight initialisation technique proposed in [85]. Furthermore, the objective function to be minimised is mean squared error (MSE) since we are dealing with a regression problem:

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

Where n is the number of training examples, y_i is the real value and \hat{y}_i is the predicted value given by the proposed system. In order to reduce the overfitting problem, the size of the training data is increased artificially by applying data augmentation. More specifically, the training data is augmented by flipping images left to right while the annotated OD and foveal centres were flipped accordingly. As a result of this, the size of the training data has been doubled.

The deep network was trained with 1000 epochs. An early stop strategy is used so the training will stop when there is no improvement in learning or performance on the validation set starts to worsen. The early stop value was set to 100 epochs where the learning stops after 100 epochs and the best weighting values are retained if the validation error stops improving early. The architecture of CNN with the best performance is described and shown in Table 3.2 and Figure 3.3 where the CNNs in the first and second stage of the proposed system have the same architecture.

In Table 3.2 the last column shows the size of the filters, the window size used for max-pooling, and the probability of dropping a node (Bernoulli (p)) in each layer. No zero padding and a stride of 1 pixel were used for each convolutional layer while non-overlapped pooling (stride= pool size) was used in each max-pooling layer. The probability of dropping are 0.1, 0.3, and 0.5 in the first, second, last dropout layer, respectively.

TABLE 3.2: Architecture of deep neural network with the best detection performance

Layer Name	Size	outputs	filters	Size of filter, max pooling, pro.
'inputimage'	$1 \times 256 \times 256$	65536	-	-
'conv1'	$8 \times 254 \times 254$	516128	8	filter size = (3,3)
'conv2'	$8 \times 252 \times 252$	508032	8	filter size = (3,3)
'conv3'	$8 \times 250 \times 250$	500000	8	filter size = (3,3)
'conv4'	$8 \times 248 \times 248$	492032	8	filter size = (3,3)
'dropout1'	-	-	-	dropout1-p=0.1
'conv5'	$16 \times 246 \times 246$	968256	16	filter size = (3,3)
'conv6'	$16 \times 244 \times 244$	952576	16	filter size = (3,3)
'conv7'	$16 \times 242 \times 242$	937024	16	filter size = (3,3)
'pool1'	$16 \times 121 \times 121$	234256	-	maxpool1 size = (2,2)
'dropout2'	-	-	-	dropout2-p=0.3
'conv8'	$32 \times 120 \times 120$	460800	32	filter size = (2,2)
'conv9'	$32 \times 119 \times 119$	453152	32	filter size = (2,2)
'conv10'	$16 \times 118 \times 118$	445568	32	filter size = (2,2)
'pool2'	$16 \times 59 \times 59$	111392	-	maxpool2 size = (2,2)
'FC'	350	350	-	-
'dropout3'	-	-	-	dropout3-p=0.5
'FC'	350	350	-	-
'output'	4	4	-	-

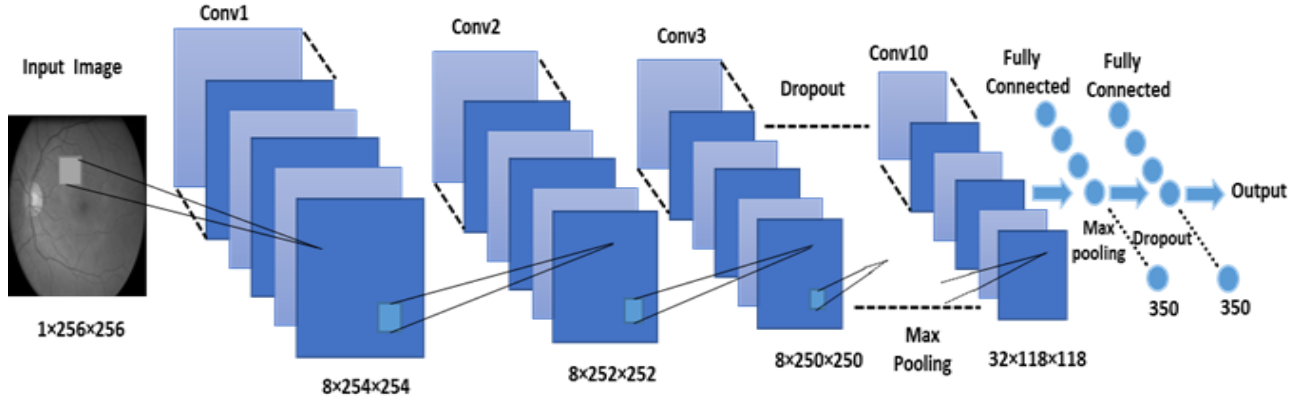


FIGURE 3.3: Block diagram of convolutional neural network.

3.4.4 Performance Evaluation

In the literature, the $1R$ criterion (where R refers to the OD radius) is the most common criterion used to evaluate the performance of retinal landmark detection methodologies. The distance between the ground truth and the obtained location of the structure of interest (i.e. the OD or foveal centre for this application) is compared with the R value in each image to determine the validity of the location determined by the automated detection methods. In this work, both the optic disc and foveal centre positions were known from expert annotations. Moreover, the location of the fovea centre is about $2.5D$ from the OD centre. The patient specific optic disc diameter (D_i) and consequently the OD radius (R_i) can be calculated for each eye i using Equation 3.2 .

$$D_i = \sqrt{\frac{(X_{OD(i)} - X_{F(i)})^2 + (Y_{OD(i)} - Y_{F(i)})^2}{2.5}} \quad (3.2)$$

Then, $R_i = 0.5D_i$ where X_{OD} , Y_{OD} , X_F , Y_F are the horizontal and vertical coordinates of the OD and fovea centres respectively marked by expert graders.

3.5 Results

For the sake of comparison, a conventional neural network with three layers (input, hidden, output) is implemented to evaluate the effect of deep learning method. This network is trained with 250 epochs and 200 neurons are used in the hidden layer. The size of the input layer is equal to the size of the input image and the size of the output layer is four neurons (x and y coordinates of the OD and fovea centres respectively).

Learning performance of the implemented networks is monitored during the training by plotting the learning curves for both training and validation sets by determining the root mean squared error (RMSE). Figure 3.4 and Figure 3.5 show the difference in terms of performance between the conventional neural network (ANN) model and the deep model during the training stage. Clearly, it can be observed that the deep neural network has improved performance with much lower error than the conventional neural network model.

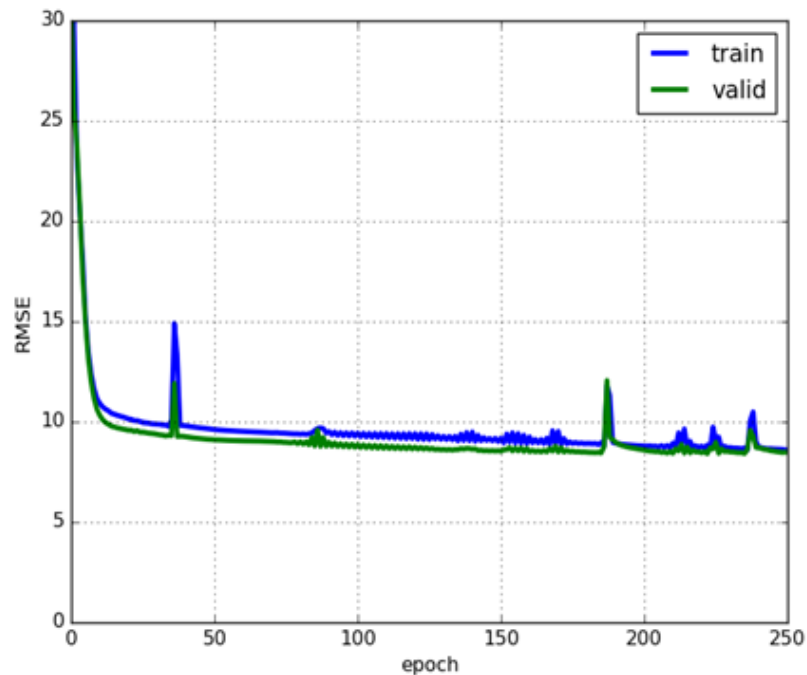


FIGURE 3.4: Performance of the conventional neural network (ANN) during training. It shows that simple model suffers from an under-fitting problem where the complexity of the network is not sufficient to capture the import features of the landmarks.

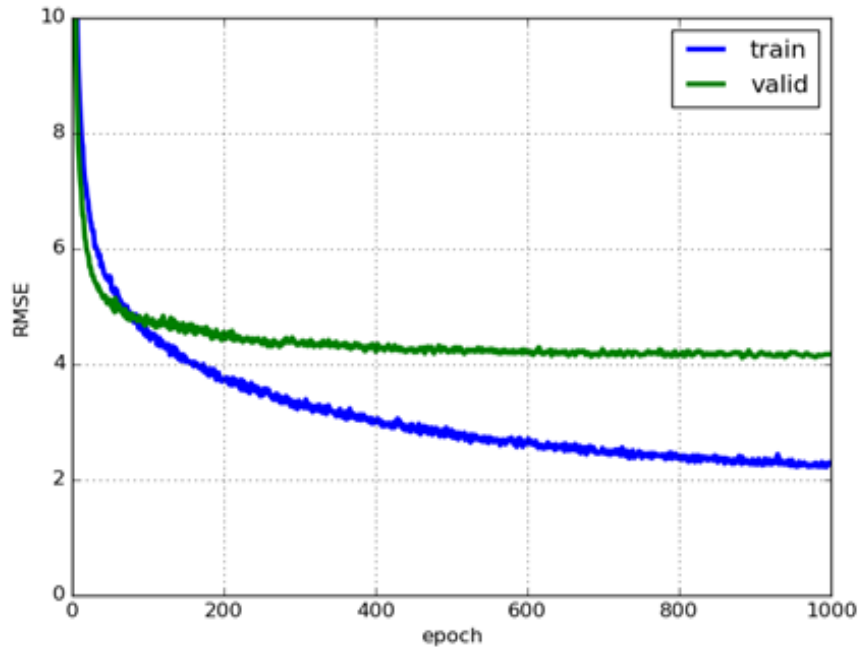


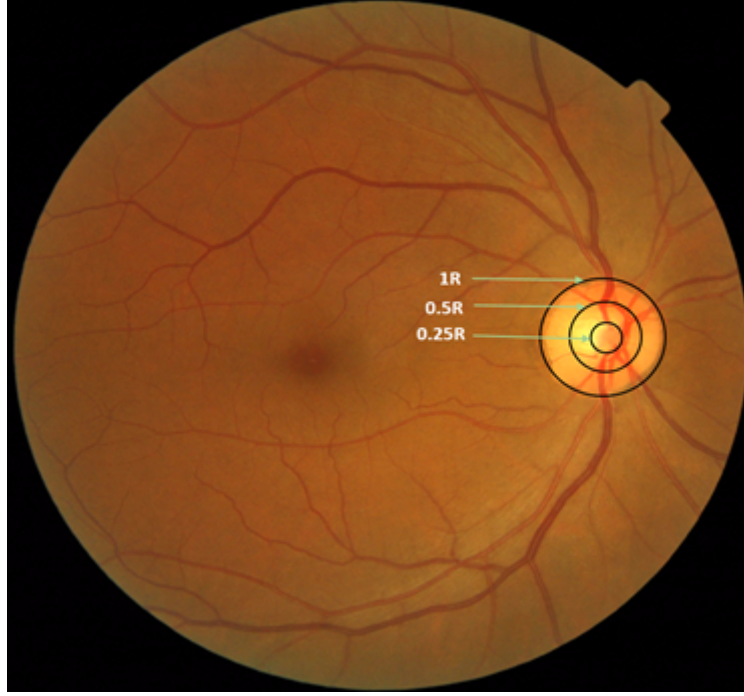
FIGURE 3.5: Performance of the deep neural network during training. It shows that the RMSE for both training and validation data is lower than the conventional neural network with slightly overfitting and thus better landmark detection performance.

For the purpose of performance analysis of the proposed system for detecting the OD and fovea, the detection accuracy was computed as the ratio between the number of testing images with detected centres satisfying the $1R$, $0.5R$ and $0.25R$ conditions (Figure 3.6 explains these criteria) and the total number of testing images.

In addition to the accuracy measure, the mean error (also called normalised localisation error) and standard deviation are also calculated. The normalised localisation error is calculated by dividing the Euclidean distance between the actual and computed OD (or foveal) centres with the D in each testing image.

The detection performance of the neural network and deep neural network is shown in Table 3.3. The effect of image enhancement is also reported in Table 3.3 for information. In this table, the MESSIDOR dataset has been randomly divided into 70% for training and validation and the remaining 30% for testing.

In Table 3.4 and Table 3.5, the proposed system was evaluated using the MESSIDOR and Kaggle datasets, where 7000 Kaggle images were used for training

FIGURE 3.6: Example shows $1R$, $0.5R$, and $0.25R$ of OD.TABLE 3.3: Performance of different networks: The networks are trained and tested on Messidor dataset ($1R$ criterion). These reported results are obtained from the first stage of the proposed system.

Model Name	Optic Disc			Fovea		
	Acc	μ	α	Acc	μ	α
Simple model (NN)	59.5	0.568	0.568	86.2	0.276	0.218
Deep model without enhancement	96.0	0.169	0.253	96.0	0.132	0.133
Deep model with enhancement	96.89	0.160	0.237	97.78	0.133	0.127

μ : Mean Error, α : Standard Deviation.

and validation (20%) of training data is used as validation data and the remaining 3000 Kaggle images and 1200 MESSIDOR images are used for testing in the first stage of the proposed system. In the second stage, the test Kaggle images from the first stage are used to train and test the second CNN where these images are divided randomly again into 80% for training and validation and 20% testing before feeding them into second stage.

Table 3.4 shows the performance of the MESSIDOR dataset in terms of the $1R$, $0.5R$ and $0.25R$ criteria for the two stages of the proposed system where on

row one, the results of CCN1 for the test set of 1200 images (TS1M) is presented. Row two shows the results of CCN1 restricted to the images that are correctly detected within the 1R criterion (TS2M). Row three shows the TS2M set which is tested with CNN2 for comparison with row two. It can be seen that the results for these images are considerably improved by CNN2. Finally, on row four, this test set is expanded to include incorrectly detected images (TS3M) from CCN1 demonstrating that, including these, the results remain strong and improved over the original idea of using CNN1 alone.

Moreover, Table 3.5 presents the accuracy of the Kaggle dataset using the same criteria where on row one, the results of CCN1 for the test set of 3000 images (TS1M) is presented. Row two shows the results of CCN1 restricted to the images that are correctly detected within the 1R criterion (TS2M). Row three shows the TS2M set tested on CNN2 for comparison with row two. It can be seen that the results for these images are considerably improved by CNN2. Finally, on row four, this test set is expanded to include incorrectly detected images (TS3M) from CCN1.

TABLE 3.4: Performance (in terms of accuracy) of the network trained on Kaggle and tested on Messidor. CNN1 refers to the first stage and CNN2 refers to the second stage.

Model Name	Optic Disc			Fovea		
	1R	0.5R	0.25R	1R	0.5R	0.25R
CNN1+TS1M	97	86.3	47.5	96.6	76	35.3
CNN1+TS2M	100	88.9	49	100	78.8	36.5
CNN2+TS2M	100	97.9	86.2	100	94.6	69.2
CNN2+TS3M	97	95	83.6	96.6	91.4	66.8

The experimental results in Table 3.4 and Table 3.5 demonstrate that the proposed method can achieve accuracies in terms of the 1R criterion of 97% and 96.6% for detection of the OD and foveal centres respectively in MESSIDOR and 96.7% and 95.6% for the detection of the OD and foveal centres respectively in the Kaggle test set.

TABLE 3.5: Performance (in terms of accuracy) of the network trained and tested on kaggle. CNN1 refers to the first stage and CNN2 refers to the second stage.

Model Name	Optic Disc			Fovea		
	1R	0.5R	0.25R	1R	0.5R	0.25R
CNN1+TS1K	96.7	87.4	51.9	95.6	83.4	51
CNN1+TS2K	100	90.1	55.6	100	87.9	54.3
CNN2+TS2K	100	99.1	93.4	100	94.9	73.3
CNN2+TS3K	96.7	95.8	90.3	95.6	90.7	70.1

On average, it only takes approximately 0.007 seconds to process a test image with two stages which is the fastest among all of the methods. Furthermore, the results show good performance when considering the $0.5R$ and $0.25R$ criteria. On the Kaggle test set, the obtained accuracies were 95.8% and 90.3% for OD detection for $0.5R$ and $0.25R$ respectively, while 90.7% and 70.1% were achieved for fovea detection in terms of these two criteria. On MESSIDOR, the detection accuracies were 95% and 83.6% for $0.5R$ and $0.25R$ for localising the OD while the obtained accuracy results for the foveal centre detection were 91.4% and 66.8% for the $0.5R$ and $0.25R$ criteria.

Figures 3.7 and 3.8 and Figure 3.9 show some example detection results on the testing dataset. In Figures 3.7 and 3.8, examples with accurate detections of the OD and fovea centres are presented while Figure 3.9 shows images with incorrect detections. Figure 3.10 shows how the second stage CNN improves the detection performance over the first stage CNN.

3.6 Discussion

A new deep neural network approach has been proposed for the detection of the OD and foveal centres in colour fundus images. Our proposed approach has produced promising results.

It is worth mentioning that many different criteria were used by others in the literature to evaluate performance in detecting the OD and foveal centres when

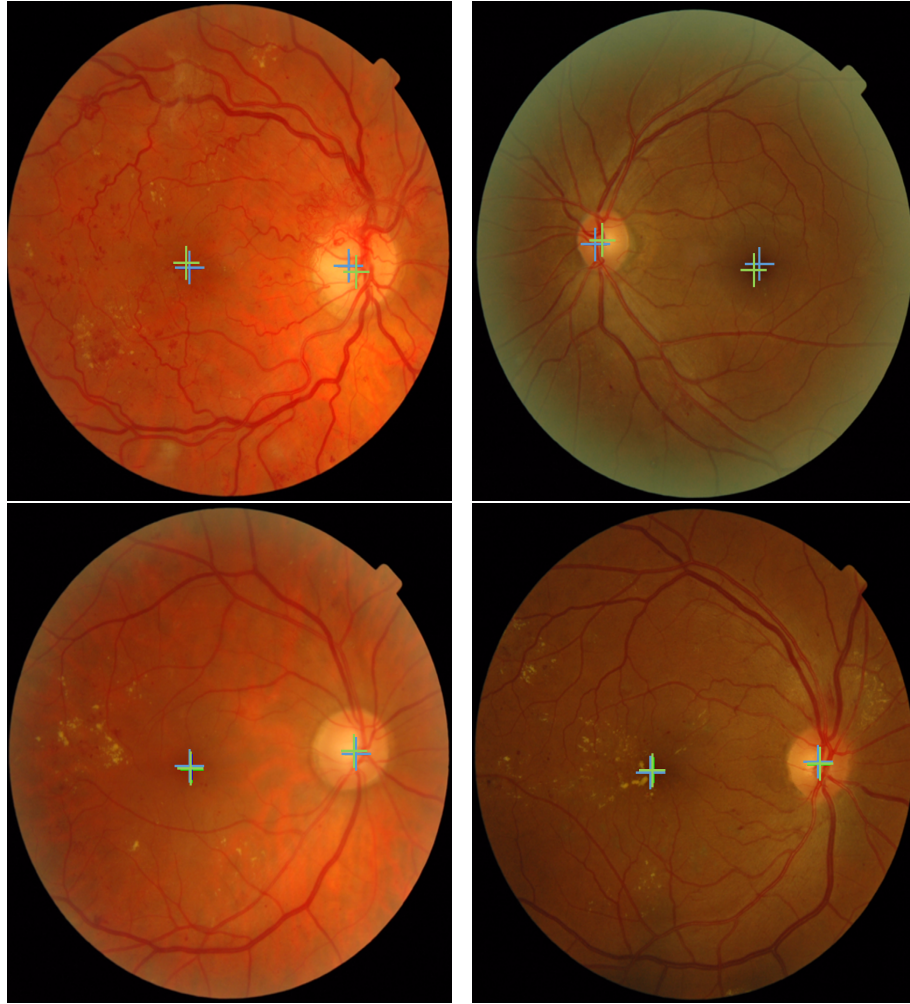


FIGURE 3.7: Examples of correct joint OD-Fovea detection results from MESSIDOR. The green plus signs refer to the locations annotated by ophthalmologists while the blue ones indicate the results of our proposed method.

compared with the ground truth. The Euclidean distance between the obtained OD and fovea centre locations and their actual locations were often used as the evaluation measure. For example, many studies [143], [150]- [8] have established that the obtained detection of the OD (or foveal) centre is correct if their Euclidean distances to the actual centres is within half the OD diameter (or one OD radius). This is the widely accepted $1R$ rule.

There is a problem in using the $1R$ rule for evaluation when the OD radius is not available. In order to alleviate this problem, Yu et al. [143] estimated the OD radius based on the field of view (FOV) of the retina and the image size. Three radii of 70, 100 and 110 pixels were used in correspondence to the three different

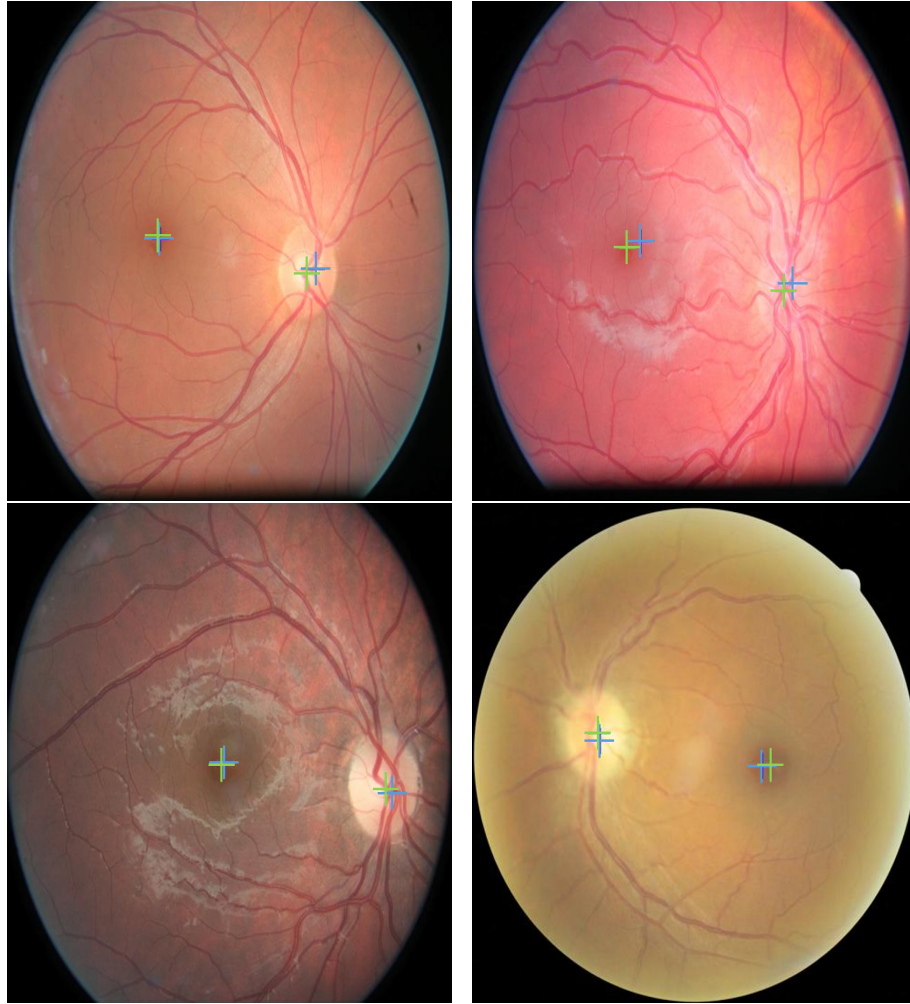


FIGURE 3.8: Examples of correct joint OD-Fovea detection results from Kaggle. The green plus signs refer to the locations annotated by ophthalmologists while the blue ones indicate the results of our proposed method.

sizes of the MESSIDOR images. Using this criterion, the authors detected the location of the OD correctly in 1189 out of the 1200 images in the MESSIDOR dataset. Following Yu's approach to estimate the OD radius, Giachetti et al. [150] reported an accuracy of 99.66% for OD detection and 99.1% for fovea detection and used the fast radial symmetry transform to achieve that. However, for the same MESSIDOR dataset Gegundez-Arias et al. [151] and Aquino et al. [8] used different OD radii in their study where the OD radii were fixed to 68, 103 and 109 pixels. Aquino et al. [8] reported an accuracy of 98.24% for the detection of the fovea.

For this study, the $1R$ rule has been followed but the OD radius was defined

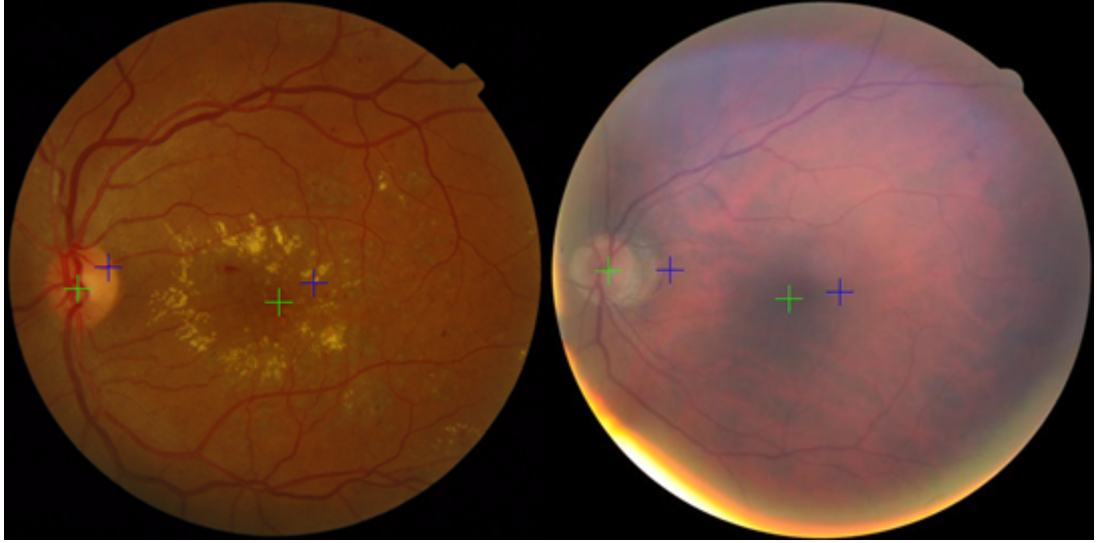


FIGURE 3.9: Examples of incorrect OD and fovea detection results. From left to right (1) Incorrect detection from MESSIDOR; (2) Incorrect detection from Kaggle. The green plus signs refer to the locations annotated by ophthalmologists while the blue ones indicate the results of our proposed method

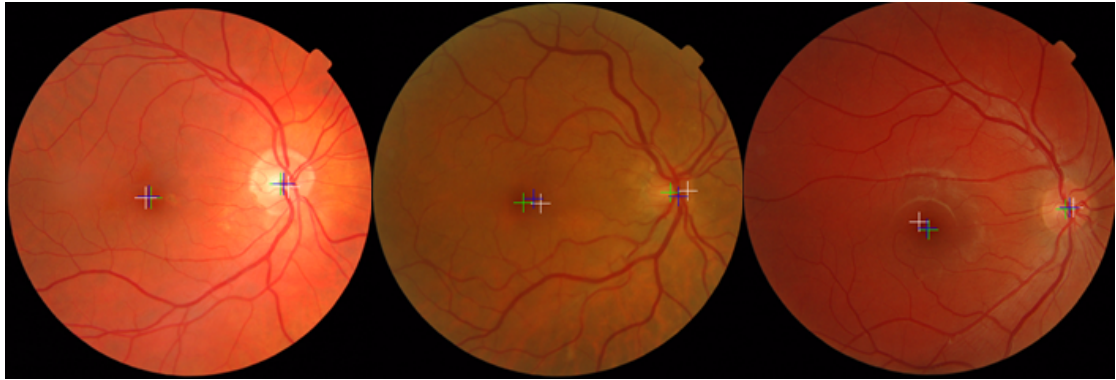


FIGURE 3.10: Examples of fundus images show the original centers (green plus), centres obtained from CNN1 (white plus), and centres from CNN2 (blue plus). It is clear that CNN2 improves the location accuracy.

by annotation results from experienced graders. As such, our rule should be more accurate. This has highlighted the issue that it is difficult to accurately compare detection performance between different methods as the criterion may be different.

The other issue for comparing results from different studies is that the number of images used were different. Even when studies used the same dataset, the way in which they used the dataset was not entirely clear. For instance, although Yu et al. [143] reported results on 1200 MESSIDOR images, they may have used the whole dataset in tailoring their detection method. This implies they have

used the data to train their method and tested on the same dataset, which means their method may have overfit the data. Our study has split the Kaggle dataset into training and testing portions. Testing images of Kaggle have not been used until the network was trained using the separate training set. This suggests that our method should have better generalisation ability. Furthermore, a completely unseen test set (MESSIDOR) is used to prove this generalisation ability.

Moreover, unlike most of the previous methods in the literature where only the $1R$ value has been reported, the accuracy based on the $0.5R$ and $0.25R$ criteria is reported in addition to the $1R$ criterion. From the $0.5R$ and $0.25R$ reported accuracies, it can be noticed that the performance has significantly improved by exploiting and analysing the ROI for both OD and F in the second stage of the proposed system.

Although our network has provided competitive results, the network architecture may not be the optimal one as training the CNNs involve many hyperparameter settings such as regularisation strength, the initial learning rate, and schedule of learning rate decay. Performing hyperparameter searches is considered a tricky and critical task [159]. Also, the number of convolution and pooling layers and the number and size of filters in each layer in CNNs are usually chosen empirically. As a result, the optimal network architecture and proper settings of these hyperparameters in the training stage are decided from experience and they are hard to find by non-expert humans [160].

In spite of these hyperparameter setting challenges in the training stage, once the network is trained, no expert is required to detect the landmarks in the test stage. Although data augmentation is considered to be useful in improving the performance of the CNNs, it is not clear what the best strategy is to achieve the best results. From our work, it is noted flipping horizontally is beneficial.

3.7 Summary

In this chapter, it has been demonstrated that our proposed method is capable of achieving excellent results in the detection of the optic disc and fovea in fundus images. One of the most important advantages of the proposed method is that it is less sensitive to preprocessing. It can be noticed that applying contrast enhancement as a preprocessing step improves the performance of the network, but not by very much. The current results were achieved without optimising the parameters of the contrast enhancement method. Another advantage of our approach is that it does not necessitate the need of vessel segmentation or border localisation in order to detect the OD and foveal centres. This will be useful when processing images of poor quality demonstrating the robustness of the proposed method. It has been proved that the ability to learn hierarchies of concepts, implementing multiple layers of abstraction in deep learning can be used for the detection landmarks in challenging medical applications. Likewise, the results of the proposed method suggest that deep learning can be used to address similar problems in other clinical applications such as screening and the diagnosis of DR, AMD, and glaucoma.

Chapter 4

Choroid and Optic Disc/Cup Segmentation

Image segmentation is one of the most important parts of medical image processing and analysis. Two segmentation approaches for medical image are presented in this chapter. In §4.1, a framework for choroid segmentation in enhanced depth imaging optic coherence tomography images (EDI-OCT) is presented. This framework comprises three main stages including patch generation by clustering, patch labelling, and then refinement where each stage is exploited to achieve a part of the final segmentation goal defined by choroid boundaries. In §4.2, another medical image segmentation task is considered for the optic disc and cup segmentation from fundus images. The details of the proposed method which is based on fully convolutional DenseNet is explained in this section along with the experimental results and discussion. Finally, the work presented in this chapter including both segmentation methods are summarised in §4.3.

4.1 Choroid Segmentation in OCT Images

4.1.1 Introduction

The choroid is a vascularised layer located between the retina and the sclera. Its inner boundary is formed by the retinal pigment epithelium layer and with sclera as the outer boundary. The choroid has many functions within the eye including providing metabolic support to the retinal pigment epithelium (RPE) of the retina. It is also involved in conditions which affect the retina and optic nerve, playing a significant role in the pathophysiology of various diseases which can result in vision loss.

Since the thickness of choroid has a strong relation to eye pathologies such as age-related choroidal atrophy, the age-related macular degeneration, and the central serous retinopathy, many scientists and clinicians have been interested in measuring it. These studies have been limited since the pigment in the RPE and choroid impedes visualisation by ophthalmoscopy, fundus photography, fluorescein angiography and optical coherence tomography (OCT), making it difficult to resolve. Scattering by dense vascular structure and absorption by the RPE are the main factors to restricting the ability of standard OCT to image the choroid clearly. Recently, enhanced depth imaging EDI-OCT [161] has emerged as a technique capable of imaging the choroid by placing the zero delay line on the choroid to obtain high resolution images and provide better information of the choroid so that it might be visualised and thus investigated. An example of a retina imaged by EDI-OCT and conventional OCT is shown in Figure 4.1 for comparison.

In order to measure the choroidal structure and thickness, manual segmentation may be undertaken but this can be subjective, particularly at the choroidal-sclera interface which remains challenging to resolve and comes at a considerable time cost, particularly for three dimensional data. Automatic segmentation of the choroidal boundaries is therefore a significant task.

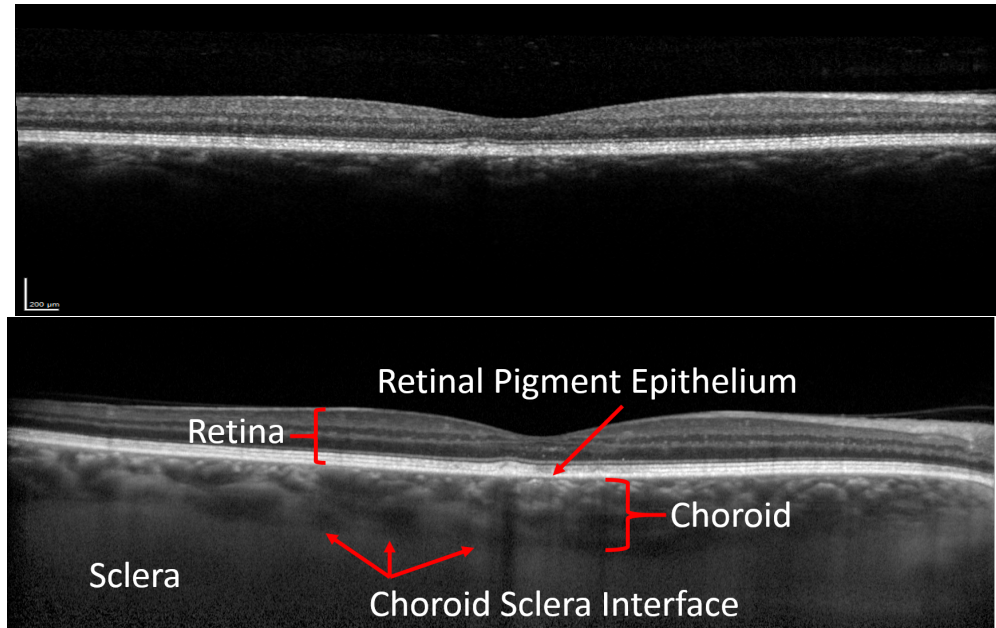


FIGURE 4.1: Example of conventional OCT image shown in the first row is compared with EDI-OCT image shown in the second row. It shows how EDI-OCT provides better information and visualisation for the choroid region.

4.1.2 Related Work

In the literature, there has been a significant amount of research in developing automated techniques for the choroid layer segmentation. Zhang et al. [162] presented a 3D approach capable of segmenting of the choroidal vessels and measuring choroidal vasculature thickness where their method was evaluated on 24 3D SD-OCT from normal subjects. A texture and shape based method was proposed in [163] by implementing a two stage statistical model to segment the choroid region in OCT images. However, their model needs extensive training and half minute to analyse each image with a mean error of 13%. In [164] and [165], the authors proposed algorithms based on the phase information extracted from polarisation sensitive optical coherence tomography (PS-OCT) to detect the edge between the choroid and sclera. The imaging modality used in their work is not available to clinicians commercially.

The authors of [166] developed a graph-search based method to detect the choroid boundaries on 45 EDI-OCT test images by searching the pixel with the

biggest gradient value above RPE and delineating the choroidal-scleral interface by finding the shortest path of the graph formed by valley pixels. In [167], a technique based on a two-stage fast active contour to segment the choroid boundary was applied on 30 EDI-OCT images. Also, EDI-OCT images were used for evaluation in [168] where a graph-searching method was proposed to automatically segment the inner and outer choroid boundaries by applying two different techniques to determine the graph weight maps in order to calculate the choroid thickness.

Moreover, a texture and wavelet-based features were extracted and Gaussian Mixture Model (GMM) was designed in [169] to segment the choroidal boundary. Inspiring by foreground/background detection in video processing and proposing a noise-estimation assisted compensating algorithm, Liu et al. [170] presented a method to segment choroidal stroma on EDI-OCT images. In [171], a machine learning based approach was presented where the low-level texture features are extracted and then passed to a SVM classifier to classify these features in order to segment the choroid region in EDI-OCT images. In another work, Zhang et al. [172] proposed fully automated three-dimensional (3D) method capable of segmenting the choroid surface from swept-source OCT (SS-OCT) and spectral-domain OCT (SD-OCT) images by utilising a combined graph-cut-graph-search method.

Furthermore, based on generating a gradual intensity distance in High Definition HD-OCT images to analyse the characteristics of the choroid-sclera interface and using an improved graph search method with curve smooth constraints, an approach was proposed to obtain the boundary of the choroid-sclera interface [173]. In another paper, the author of generating a gradual intensity distance in HD-OCT images [173] also implemented the choroid-sclera junction cost based on generated choroidal vessel image then used a graph-search method to detect the choroid-sclera interface boundary [174]. Furthermore, a 3D graph-based approach was developed, in [175], to segment the choroid boundary by exploiting the optic disc prior information in 20 spectral-domain optical coherence tomography (SD-OCT)

images of optic nerve head (ONH). In [176], a 4D graph-based method was presented for choroid surface segmentation by considering multiple 3D scans over time where 3219 OCT images from 149 patient were used for the validation. Another automated method based on graph theory, dynamic programming, and wavelet-based texture analysis was given by the authors in [177] to segment the choroid in SD-OCT images for 30 subjects. Moreover, segmenting the choroid region from $1 \mu 32$ wide-view swept source SS-OCT image volumes was presented in [178]. The authors suggested a 3-D multiresolution graph search with gradient-based cost to initially segmenting Bruch's membrane (BM) and CSI. Then, refining CSI contour was achieved by adding a regional cost, calculated from the wavelet-based gradual intensity distance.

More recently, a level set and Markov Random Field-based framework to segment the choroid on 30 3D OCT was proposed in [179] where the edge constraint and distance regularisation terms are embedded into the level set method and the region term is modelled into the framework by Markov Random Field method. Besides, an open source algorithm was developed for segmenting and quantifying the choroidal layer from 3-D OCT reconstructions [180].

Although convolutional neural networks (CNNs) was developed largely for classification, approaches for medical image segmentation using it have recently been reported. Ciresan et al. presented a sliding window approach for segmenting electron microscopy images [124]. Each pixel was classified as either membrane or non-membrane using a CNN trained on square windows centred on pixels in the raw input data. While this method shows promising results, a significant drawback is that it requires the processing of a great number of overlapping windows. This increases processing time and can be computationally redundant. More recent approaches, such as those proposed by Long et al. for computer vision applications [107] and Ronneberger et al. for microscopy image analysis [108], depend on implementing fully convolutional neural networks. These networks consider input of arbitrary size and are capable of producing correspondingly-sized results with

efficient learning and inference. However, the fully convolutional network proposed in [107] can give rather coarse resolution deeper in the network [107]. In [108], the authors developed U-net CNN architecture which extended the fully connected convolutional network by considering contracting and symmetric expanding paths, the first of which is used to capture context while the second enables precise localisation. In that paper [108], it has been reported a good performance when trained end-to-end from only a few images in three different datasets.

Most recently, CNN has been exploited (merged with other methods or individually) to segment the choroid region in [181] and [182]. In [181], a graph search based model was proposed to detect the choroid boundary where the optimal graph-edge weight values are learned from raw pixels by CNNs instead of hand-designing. While in [182], a convolution neural network was trained to directly identify the boundaries of the choroid in 62 EDI-OCT images from patients with AMD.

However, the segmentation of the choroid boundaries remains a challenge and the aforementioned methods still have some limitations for the following reasons (i) the boundary between the sclera and choroid is weak or even sometimes invisible (ii) the histogram of sclera and choroid are not distinguished or separable (iii) the inhomogeneity of choroid texture due to the presence of the vascular structure.

In this work, a new framework for image segmentation which is targeted towards EDI-OCT images of the choroid has been developed. The RPE-Choroid boundary is first segmented and image information above this contour is removed, allowing more focus on the region of interest and acceleration by reducing the amount of required computation. Then, a new energy function for partitioning the image into unlabelled clusters which should not cross the boundaries of the choroid is designed, achieving improved performance over previous work. In order to label the clusters and thus obtain the segmentation, a CNN is developed using not only image intensities but also automatically enhanced intensity information. Spatial information is further incorporated using the distance from Bruch's

membrane which can be calculated automatically given the found RPE-Choroid boundary. It has been demonstrated that this results in more accurate classification than using image intensity information alone. The segmentation result is then refined with a post-processing step which encourages connectedness of the choroid region and produces the boundary contour.

4.1.3 Materials and Method

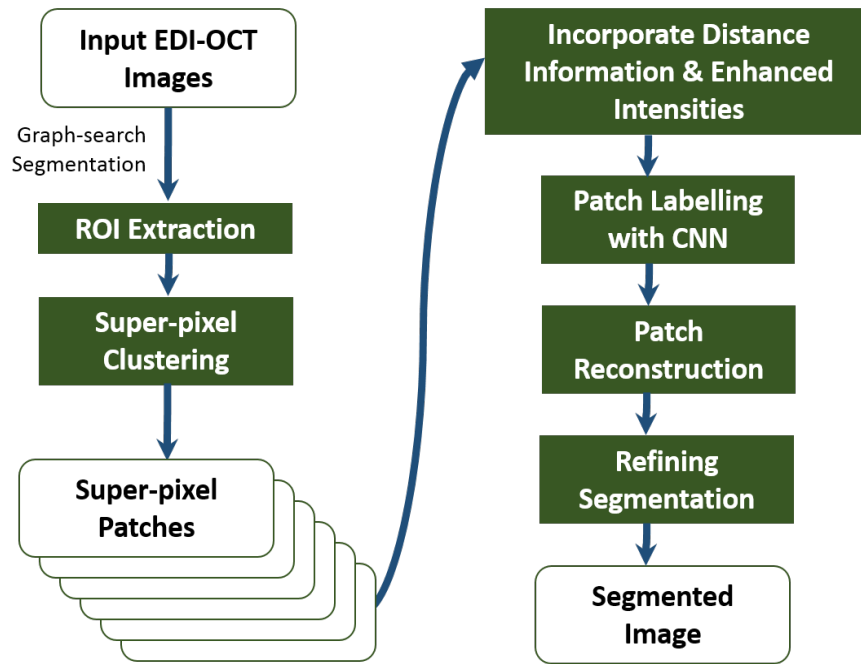


FIGURE 4.2: Block diagram shows the main stages of the proposed framework.

For all images, the inner choroidal boundary with Bruch's membrane which is located under RPE is first segmented and the image at the uppermost point of the found contour is cropped leaving primarily the sclera and then choroid. The pixels are then grouped into clusters with a super-pixel method using intensity, region and enhanced intensity information. Patches are then formed around the centres of the superpixels and those of the training data is used to train a convolutional neural network to identify choroidal and non-choroidal regions. This network is then used to classify the patches from the test set and provide the provisional segmentation. Refinement on this is achieved by encouraging connectedness and a

smooth contour in a post-processing step, producing the final segmentation result. The stages of framework are shown in Figures 4.2 and 4.3.

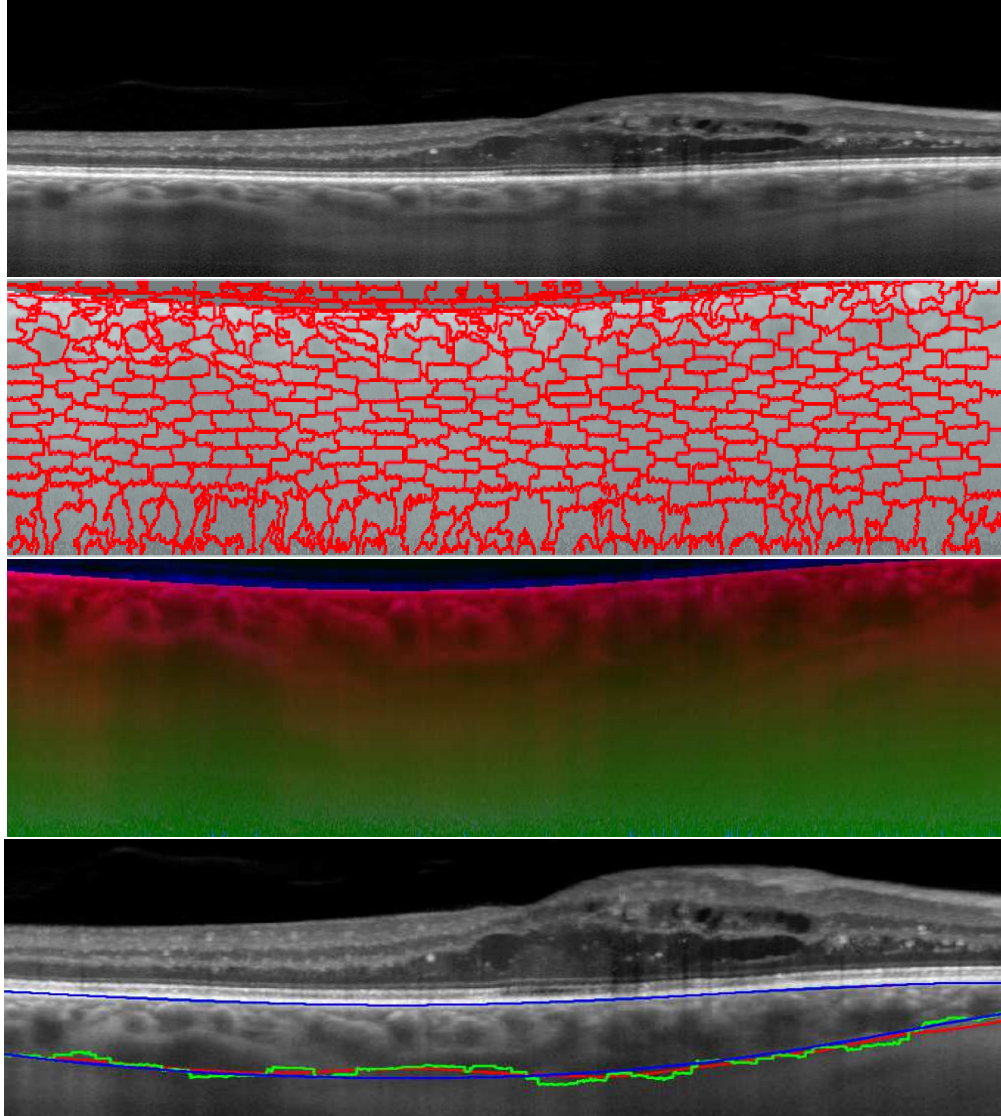


FIGURE 4.3: An EDI-OCT image of the choroid (first row) is segmented to extract the sub-RPE region which is partitioned with a superpixel technique (second row). Distance information and enhanced intensities are then calculated and combined with the image (third row) for the CNN classification of each superpixel. The CNN result is refined (forth row) [green contour] and smoothed to give the final result [red contour] which is close to the annotated segmentation [blue contour].

Dataset

Two datasets were used to evaluate the proposed framework. First dataset comprises 169 single B-scan EDI-OCT images; one image per patient; from 25 healthy

subjects and 144 patients at various stages of retinopathy and maculopathy. These images were acquired with appropriate approval at St Paul's Eye Unit, Royal Liverpool University Hospital using Heidelberg Spectralis (Heidelberg Engineering, Heidelberg, Germany). The size of images are 1024×596 pixels where the pixel size is $5.71 \times 3.85 \mu m$. To compare the performance of the proposed system to the manual segmentation and validate its performance, the inter and intra-observer agreements were determined. The manual annotations of the inner boundary (RPE-choroid) and the outer choroidal-sclera junction were provided by two expert clinical graders. The first grader provided two annotations and the second provided one annotation to serve as the ground truth in our experiments. The second dataset contains 20 EDI-OCT volumes (3D images) where each volume has 25 B-scans. This dataset was not provided with ground truth so that it has been used only for visualising the predicted boundaries. First dataset is split into training (75%) and testing (25%) sets. The CNN is trained on 75% of set one (selected randomly) and tested (without further training or modification) on the remaining 25% of set one and the whole set two.

ROI Extraction

In order to decrease the computation time, the region of interest (the sclera and choroid) in each OCT image is extracted by segmenting the inner choroidal boundary with Bruch's membrane, cropping the image out at the uppermost point of the inner boundary and removing it by using the technique introduced in [183]. In this algorithm, the pixels of an image are represented as a graph where each pixel corresponds to a node connected to neighbouring pixels by a link (edge) with certain numerical values (weights). The energy function that can be minimised by graph cuts is defined to determine the cost of travelling from one node to the other in the constructed graph.

Superpixel Segmentation

Superpixel clustering is a method of grouping image pixels based on intensity similarity and spatial difference from the superpixel centre, also known as the compactness of the superpixels. A technique similar to k -means clustering is adapted but with a reduced search space which reduces the number of calculations considerably [184]. The method is initialised with a set of κ_1 initial clusters $\mathcal{C}_i \subset \Omega$, with centres $c \in \mathcal{C}$ spaced approximately equally apart at distances of $h_x = \lfloor m/\sqrt{\kappa_1} \rfloor$ and $h_y = \lfloor n/\sqrt{\kappa_1} \rfloor$ in the lateral and depth directions respectively and then shifted to the lowest local gradient. It has been then attempt to minimise an energy function $\mathcal{F}(\mathbf{x})$ across all superpixels by iteratively considering shifts of each superpixel centre within a $2h_x \times 2h_y$ region until the residual error is sufficiently low. Previously, the Euclidean distance of pixel centres has been used along with intensity difference and a trade-off parameter α to define the energy function.

$$\mathcal{J}(\mathbf{x}, \mathbf{c}) = \sqrt{\sum_{i \in \{l, a, b\}} (z^i(\mathbf{x}) - z^i(\mathbf{c}))^2 + \alpha \sum_{i=1}^2 (x_i - c_i)^2} \quad (4.1)$$

where z^l , z^a and z^b denote the lightness and colour components of the image z , $\mathbf{x} = (x_1, x_2)^\top \in \Omega$ and $\mathbf{c} = (c_1, c_2)^\top \in \mathcal{C}$ denote the superpixel centres.

In the case of EDI-OCT choroid images, there is single-channel intensity data without colour information which allows us to simplify the energy function. However, the choroidal-sclera interface is difficult to distinguish which makes the intensity difference function less effective. This can be improved by increasing the contrast in this region, multiplying the intensity values given closer to the sclera with the cumulative sum. Building this idea into our energy function along with a non-linear intensity transform to further improve contrast, the image function $z(\mathbf{x})$ in the energy function is replaced with

$$s(\mathbf{x}) = \frac{s_1 + H_{\varepsilon_1}(s_1(\mathbf{x}) - \varepsilon_2)(\varepsilon_2 - s_1(\mathbf{x}))}{\varepsilon_2 \max(s_1(\mathbf{x}))}, \quad (4.2)$$

where

$$s_1(\mathbf{x}) = \frac{z^2(\mathbf{x})}{2 \sum_{j=x_1}^m z^2(j, x_2)},$$

and $H_{\varepsilon_1}(x)$ denotes the approximation $(1 + \exp(-2x/\varepsilon_1))^{-1}$ to the Heaviside function which may be tweaked by the parameter ε_1 which trades-off similarity to the Heaviside and smoothness, and ε_2 controls the degree of enhancement.

Furthermore, given the layered presentation of the choroid in the image and the presence of Poisson noise which is typical of OCT images, the Euclidean distance used in the energy function is replaced with the L^1 -norm of the difference and consider different weights for the lateral and depth directions. Our energy function is thus presented as

$$\mathcal{F}(\mathbf{x}, \mathbf{c}) = \sqrt{|\beta(\mathbf{x} - \mathbf{c})|_{L^1(\Omega)}^2 + \beta_3 (s(\mathbf{x}) - s(\mathbf{c}))^2}, \quad (4.3)$$

where

$$\beta = (\beta_1, \beta_2)^\top \in \mathbb{R}_{>0}^2,$$

$$|\beta(\mathbf{x} - \mathbf{c})|_{L^1(\Omega)}^2 = \sum_{i=1}^2 |\beta_i(x_i - c_i)|,$$

$\beta_1, \beta_2, \beta_3 \in \mathbb{R}_{>0}^2$ are non-negative trade-off parameters. This allows for an improved initial superpixel clustering whose boundaries better follow the choroidal-sclera boundary and gives a fast initial segmentation of the image. It remains to label the superpixels as belonging to the choroid or not. To do this, patches of the image which are centred on the superpixel centres are extracted and a CNN to label the superpixels is trained.

CNN for Patch Labelling

As mentioned, the choroid layer lies bounded between the inner boundary (retinal pigment epithelium (RPE)) and the outer boundary (sclera). So while the distance increases from RPE by going away from it towards sclera interface and, the probability of classification as being choroid region decreases. In order to obtain improved detection results for the choroid region, additional information is added to the CNN. As the inner choroidal boundary was obtained by applying graph search algorithm to segmenting it, distance information is incorporated into the network using the distance function from the inner choroidal boundary \mathcal{S}^1 given by graph search segmentation defined as follows.

$$\mathcal{D}(\mathbf{x}) = \gamma \|\mathbf{x} - C^1\|_{L^2} \quad (4.4)$$

where γ is a trade-off parameter measuring the influence of the distance function in the network, $C^1 = (x_1, x_2)^\top$ and $x_2 = \mathcal{S}^1(x_1)$. The enhanced image information $s(\mathbf{x})$ defined in equation (4.2) is also incorporated. Using this, a trichannel image is built as shown in third row in Figure 4.3.

An eight-layer CNN ($Conv_1 - 32$, $MaxP_1$, $Conv_2 - 64$, $MaxP_2$, $Drop_1(0.5)$, $Dense_1(512)$, $Drop_2(0.5)$, $Dense_2(2)$) is designed and implemented for our experiment consisting of two convolutional layers with filter size (3,3) each followed by a non-overlapped down-sampling layer with maxpooling size (2,2) and rectified linear unit as activation function, two fully connected layers where the final dense layer with softmax activation function, and two dropout layers with dropping probability 0.5. The size of input data is $33 \times 33 \times 3$. Augmentation was introduced by randomly flipping the patches during the training to reduce the likelihood of overfitting issue. The output layer consists of two neurons to classify the patches as choroid or non-choroid.

The network was trained with 300 epochs by updating the weights that were initialised using Glorot (He) weight initialisation [86]. The cross-entropy loss function

TABLE 4.1: CNN Architecture

Layer Name	Parameters
'inputlayer'	$33 \times 33 \times 3$ input image
'conv1'	32 $31 \times 31 \times 3$ with filter size: (3,3)
'maxpool1'	32 $16 \times 16 \times 3$ with maxpool size: (2,2)
'conv2'	64 $14 \times 14 \times 3$ with filter size: (3,3)
'maxpool2'	64 $7 \times 7 \times 3$ with maxpool size: (2,2)
'dropout1'	p=0.5
'dense'	512
'dropout2'	p=0.5
'output'	2

was minimised by stochastic gradient descent (SGD) algorithm with a constant learning rate 0.001. Once the network is trained, it can be used to label the test patches without further tweaking to the parameters.

Refining

After the CNN step, the segmented image regions resulted from the proposed framework is enhanced to produce smooth and accurate segmentation results. These results from the final classification layer in CNN is refined by penalising contour length and disconnected patches. Occasionally, the results of region-based segmentation approaches can be further refined using boundary-based algorithms [185] like snake active contour model [186], Chan Vese model [187], and intelligent scissors-based approach [188]. For intuitively and efficiently revising and smoothing the segmentation on the test images by modifying curved contour in less number of iterations, the following refining scheme is considered.

Let $\mathcal{L}(\mathbf{x})$ denotes the piecewise constant labelling function determined by the superpixel technique whose values are determined by the superpixel index and let $\phi^{(0)}(\mathbf{x})$ denotes a function corresponding to the segmentation achieved in the previous section. That is $\phi^{(0)}(\mathbf{x}) = 1$ if \mathbf{x} is contained in a superpixel which is classified as the choroid and $\phi^{(0)}(\mathbf{x}) = 0$ otherwise. Then the patches are relabelled

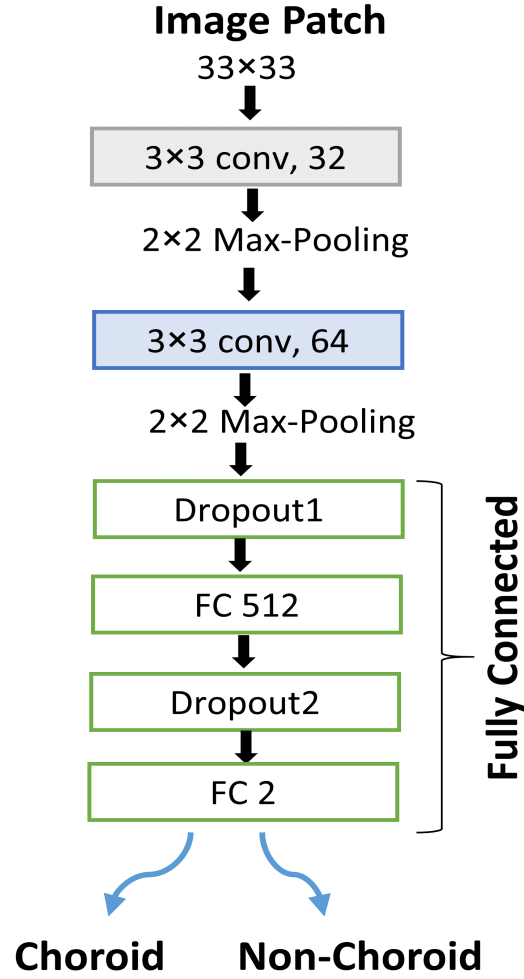


FIGURE 4.4: CNN Architecture

iteratively as

$$\phi^{(i)}(\mathbf{x}) = \phi^{(i-1)}(\mathbf{x}) + H \left(\frac{\int_{\Omega} (H([\kappa * \mathcal{L}](\mathbf{x}) - \varepsilon) - \mathcal{L}(\mathbf{x})) \phi^{(i-1)}(\mathbf{x}) d\mathbf{x}}{\int_{\Omega} H([\kappa * \mathcal{L}](\mathbf{x}) - \varepsilon) - \mathcal{L}(\mathbf{x}) d\mathbf{x}} - \frac{1}{2} \right), \quad (4.5)$$

where $*$ denotes the operation of convolution, H denotes the Heaviside step function, $\kappa(\mathbf{x})$ is a small circular filter of diameter three and ε is a very small positive parameter, set to 10^{-15} for this work. After n_i iterations, the final contour is given by $\nabla H(\phi^{(n_i)}(\mathbf{x}) - \varepsilon)$. The iterative procedure is carried out until there is no further change which typically occurred after only one or two iterations. As a final step, the outer choroidal boundary contour is smoothed by a linear average

one dimensional filter.

4.1.4 Results and Discussion

It has been demonstrated that the superpixel segmentation technique is improved using the new energy function. The suitability of the superpixel segmentation technique is checked as follows. After clustering, each superpixel is assigned a label depending on the value of the manual annotation at the superpixel's calculated centre. In this way, a parameter search was carried out to find those which give the best result for each energy function. A parameter search is performed to determine the appropriate parameters $(\alpha, \beta_1, \beta_2, \beta_3) = (10, 5.17, 48.57, 0.96)$. The super-pixel parameters were optimised only on the training images, which were then fixed for all testing and not tweaked for individual images. It demonstrates that with these parameters, the proposed model produces consistently accurate results for the previously unseen, randomly selected testing images.

To evaluate the performance of the proposed system; four different measurements including accuracy (Acc), Dice's Coefficient (DC), Tanimoto Coefficient and f_1 score have been calculated. These metrics are defined as follows:

$$Acc(\%) = \frac{tp + tn}{tp + tn + fp + fn} \times 100 \quad (4.6)$$

$$Specificity(\%) = \frac{tn}{tn + fp} \times 100 \quad (4.7)$$

$$Sensitivity(\%) = \frac{tp}{tp + fn} \times 100 \quad (4.8)$$

$$f_1(\%) = \frac{2(Precision \times Recall)}{Precision + Recall} \quad (4.9)$$

where $Precision = \frac{tp}{tp+fp}$, $Recall = Sensitivity$, tp : True Positive, tn : True Negative, fp : False Positive, fn : False Negative

$$DC = \frac{2|A.B|}{|A| + |B|} \quad (4.10)$$

$$TC = \frac{A.B}{|A|^2 + |B|^2 - A.B} \quad (4.11)$$

where A is the ground truth map of segmentation and B is the resulted segmentation map from the proposed system.

For choosing the deep learning input data, distance information and image enhancement are combined into a tri-channel image. Each 65×65 patch is resized to 33×33 by downsampling for training and testing. Using only image intensities in a single-channel image using super-pixel clustering method proposed in [184], it is achieve a mean accuracy of 0.9429 across a test set of 15698 patches of 37 images which is encouraging. It is shown from Table 4.2 that incorporating the distance measure (SLIC+In+D) with a scaling parameter of 1.3 chosen by a parameter selection scheme achieves accuracy of 0.9613 and extending this to include the enhanced image (SLIC+In+D+Enh) further improves the results to give an accuracy 0.9662 and precision and recall of 0.97 and 0.97 respectively corresponding to an F1 score of 0.97. While it may be simpler to exclude the intensity image and rely on the enhanced image only (SLIC+D+Enh), it was found that this reduces accuracy to 0.9629. Also, from Table 4.2 it can be noticed that resizing the patch not only provides acceleration for training and testing, but it demonstrates that there is no or only slightly decrease in resulted accuracy and other metrics.

Table 4.3 presents the results of our baseline method using only image information for CNN and the superpixel energy function reported in [184] but with post-processing (refining stage after patches construction). Our baseline achieves improved results over many comparable techniques as shown in Table 4.3. Our overall method considering enhanced, distance information, and refining presented in Table 4.3 and Figure 4.5, obtains improved results. It has been evaluated U-net

TABLE 4.2: The performance of patch classification in the CNN considering the patches size, the type and number of channels used in the image passed to CNN.

Method	Image Size	Acc	Precision	Recall	F ₁
SLIC+In+D	33	95.53	96	96	96
	65	96.17	96	96	96
	65to33	96.13	96	96	96
SLIC+In+D+Enh	33	95.62	96	96	96
	65	96.31	96	96	96
	65to33	96.62	97	97	97
SLIC+D+Enh	33	95.25	95	95	95
	65	95.83	96	96	96
	65to33	96.29	96	96	96

Acc: accuracy, SLIC+IN+D: SLIC algorithm with image intensities and distance information, SLIC+IN+D+Enh: SLIC algorithm with image intensities, distance information, and enhanced intensities. SLIC+D+Enh: SLIC algorithm with distance information and enhanced image intensities.

proposed in [108] on our test set and obtained accuracy, DC, TC, F1 of 97.2, 88.1, 79.2, 88.1, respectively, which are less than our obtained results. Our proposed method achieves the best results comparing to other existing methods and very close results to Shi et al. [178] although they assessed their system on a different dataset. Alonso-Caneiro et al. [168] proposed a graph-search based approach and evaluated their method on two datasets, achieving DC of 97.3 and 96.7. They reported DC better than the DC obtained by the proposed system but their algorithm takes 45 seconds per B-scan for segmentation while ours achieves that in only 2 seconds as well as the data used to evaluate their system differs from the one used for evaluating our proposed system.

Moreover, Table 4.4 presents the inter and intra-observer agreements. It shows that the proposed system is capable to get results are very close to human annotations achieving accuracies of 98.58%, 98.21%, and 98.21% versus the references: the first grader with annotation one, the first grader with annotation two, and the second grader, respectively. Furthermore, some examples of the results obtained by the proposed method versus the first grader with both two annotations and the second grader's annotations on some test images are presented in Figure 4.6

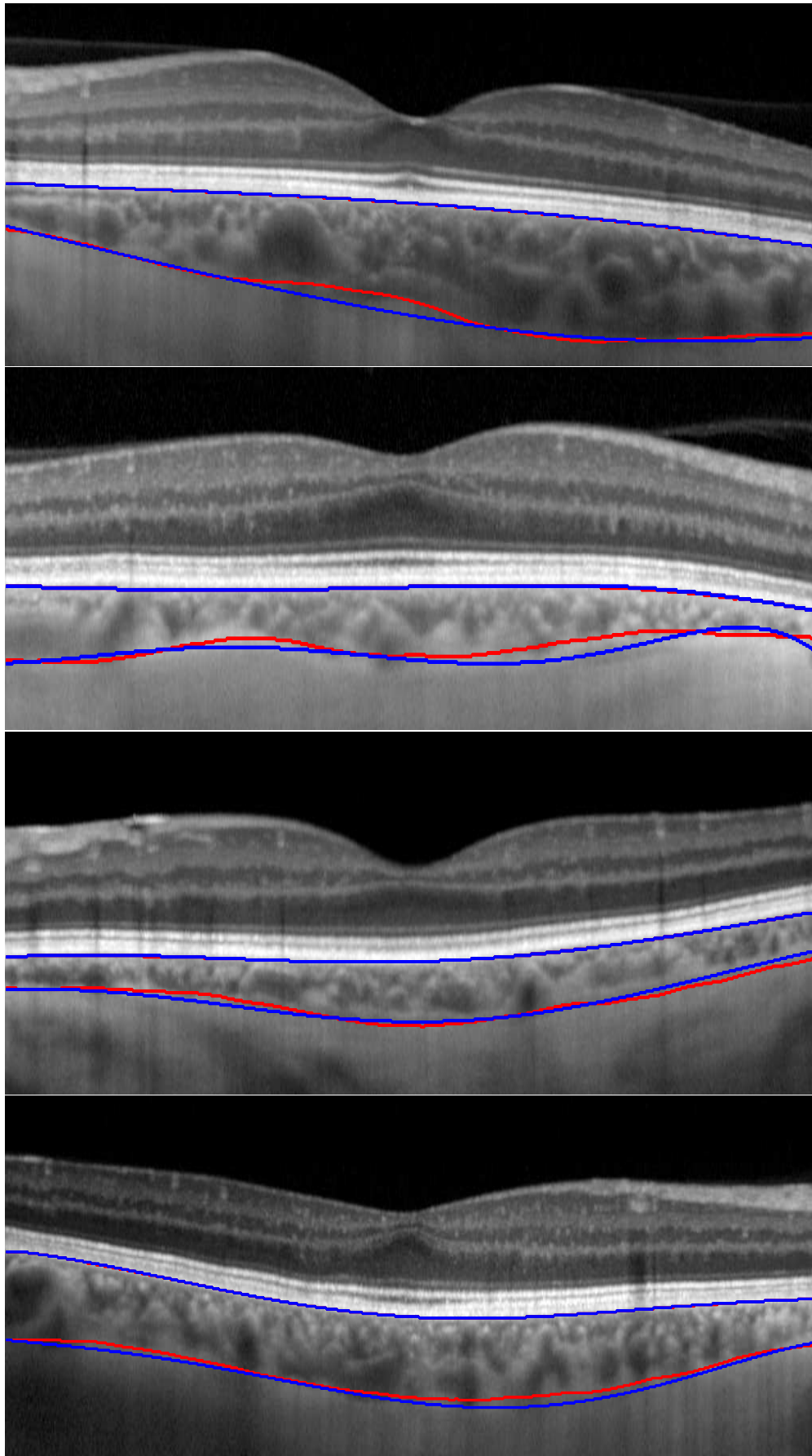


FIGURE 4.5: Results of our model for four test images comparing our result (red curve) with the expert grading (Grader1Round1)(blue curve).

TABLE 4.3: Results of our method and competing methods

Method	Acc	DC	TC	F ₁	Data
Zhang et al. [162]	-	78	-	-	24 3D SD-OCT
Tian et al. [166]	-	90.5	-	-	45 EDI-OCT
Lu et al. [167]	-	92.7	-	-	30 EDI-OCT
Alonso-Caneiro et al. [168]	-	97.3, 96.7	-	-	1083, 90 OCT
González-López et al. [171]	96.88	-	-	-	63 EDI-OCT
Chen et al. [173]	-	85.04	-	-	212 HD-OCT
Chen et al. [174]	-	86.29	-	-	319 EDI-OCT
Shi et al. [178]	-	93.17	-	-	32 3-D OCT
Wang et al. [179]	-	90	-	-	30 3-D OCT
Min Chen et al. [182]	-	83	-	-	62 EDI-OCT
U-net [108]	97.22	88.14	79.20	88.14	Our test set
Our Baseline	95.89	92.59	86.44	90.20	Our test set
Our Method	98.63	92.90	86.92	91.64	Our test set

Acc: Accuracy, DC: Dice coefficient, TC: Tanimoto coefficient.

and Table 4.5. Table 4.5 shows that in 14 cases (41.2%), our results are closer to G1A1 than either G1A2 or G2, in three of those (8.8%), our results are closer than both. In 4 (11.8%) cases, our results are closer to G1A2 than G1A1 or G2, being closer than both in one of those (2.9%). In 6 cases (17.6%), our results are closer to G2 than G1A1 or G1A2, again being closer than both in one case (2.9%). To demonstrate the performance of the framework on 3D data, Figure 4.7 shows an example of segmentation of a three dimensional EDI-OCT image.

TABLE 4.4: Mean accuracy of our results compared to two different annotations from one expert grader (Grad 1) and an annotation from a second expert grader (Grad 2). From the calculations, each of the graders gave similar annotations and our results are very close.

	Our Method	Grad 1 Ann 1	Grad 1 Ann 2
Grad 1 Ann 1	98.58	-	-
Grad 1 Ann 2	98.21	99.05	-
Grad 2	98.21	99.11	99.00

TABLE 4.5: Accuracies of our results (OM) compared to two different annotations from one expert grader (G1A1 and G1A2) and an annotation from a second expert grader (G2). The entries in bold indicate our results achieving closer segmentation than an expert grading.

Example	OM-G1A1	OM-G1A2	OM-G2	G1A1-G1A2	G1A1-G2	G1A2-G2
1	98.71	97.68	98.00	98.80	98.44	98.76
2	98.60	97.69	98.31	99.07	99.45	99.01
3	98.86	98.80	98.66	99.45	99.47	99.70
4	98.70	98.74	98.44	99.25	98.82	99.22
5	99.29	98.73	98.80	99.36	99.28	99.37
6	99.00	98.67	99.09	98.88	99.21	99.23
7	98.68	98.14	98.92	99.13	99.23	98.82
8	98.95	97.43	98.37	98.35	99.25	98.78
9	98.91	98.75	98.81	99.39	99.57	99.56
10	99.57	99.40	99.28	99.66	99.54	99.45
11	99.19	98.69	98.29	98.79	98.67	99.05
12	99.03	98.69	98.31	99.22	98.96	98.98
13	97.89	97.49	97.40	99.41	99.09	99.27
14	98.98	98.52	99.02	99.33	99.58	99.41
15	98.75	98.94	99.04	98.73	99.30	99.25
16	98.44	98.21	97.92	98.69	98.43	98.70
17	98.51	98.51	97.29	99.18	98.78	98.43
18	95.55	95.87	94.83	99.36	99.00	98.66
19	97.17	96.41	96.40	99.04	99.21	99.13
20	97.17	97.32	97.23	99.57	99.59	99.66
21	98.74	98.40	98.59	99.37	99.48	99.38
22	98.79	98.40	98.71	99.35	99.27	99.34
23	97.67	97.49	95.30	96.03	97.63	93.95
24	98.45	98.69	98.60	99.22	98.60	98.51
25	98.82	98.06	98.12	98.99	99.14	99.34
26	99.08	98.42	98.47	99.12	99.06	99.72
27	99.34	98.57	99.37	99.05	99.70	99.15
28	98.75	98.86	99.17	99.09	98.96	99.24
29	97.97	98.26	97.20	99.36	99.00	98.66
30	99.34	98.69	99.23	99.08	99.57	98.98
31	98.25	97.99	98.26	99.31	99.53	99.38
32	99.15	97.85	98.22	98.09	98.37	99.26
33	98.96	98.73	99.10	99.77	99.35	99.33
34	98.37	98.18	98.53	99.06	99.16	99.32

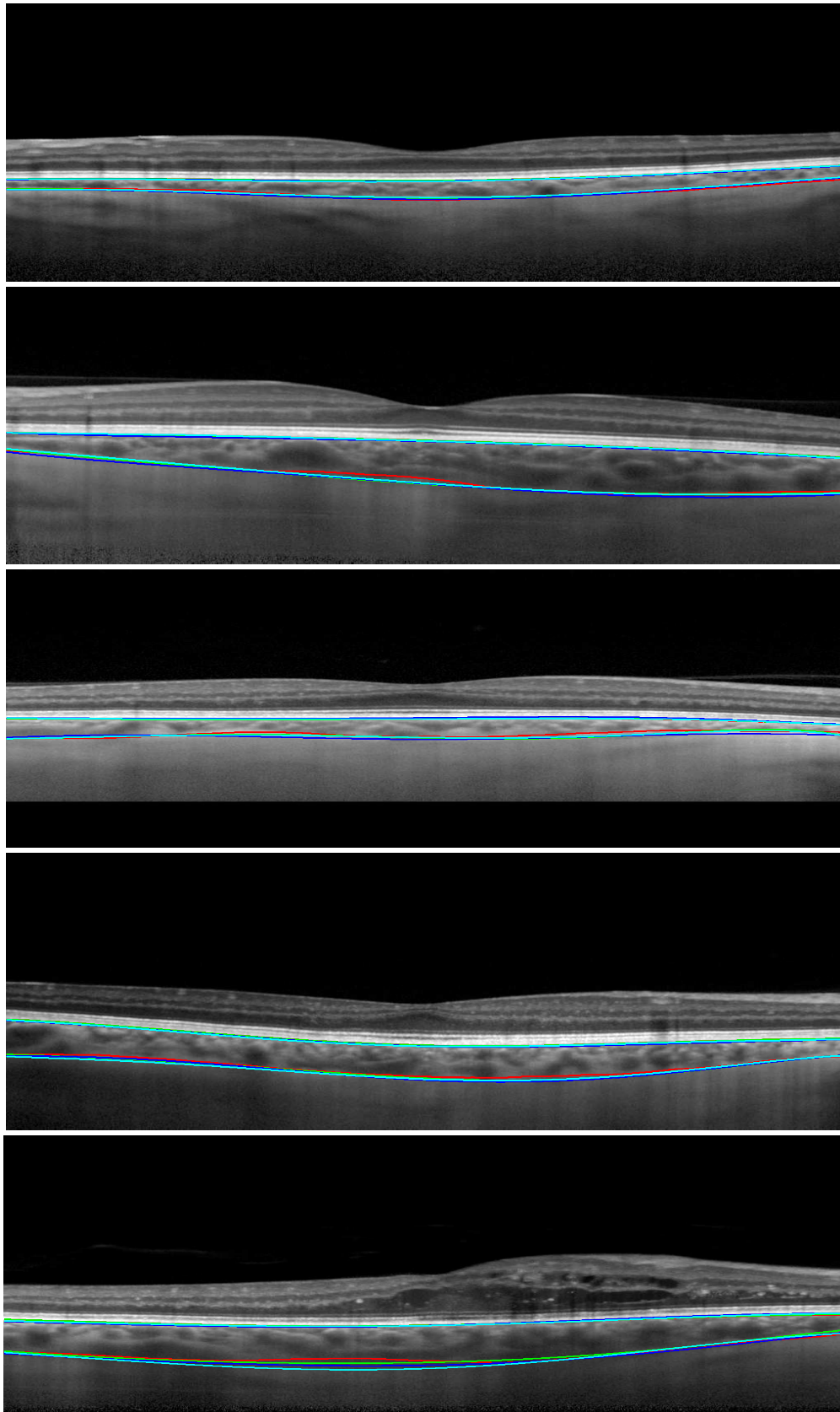


FIGURE 4.6: Examples of our results (red contour) with grader one's first annotation (green), grader one's second annotation (blue) and grader two's annotation (cyan).

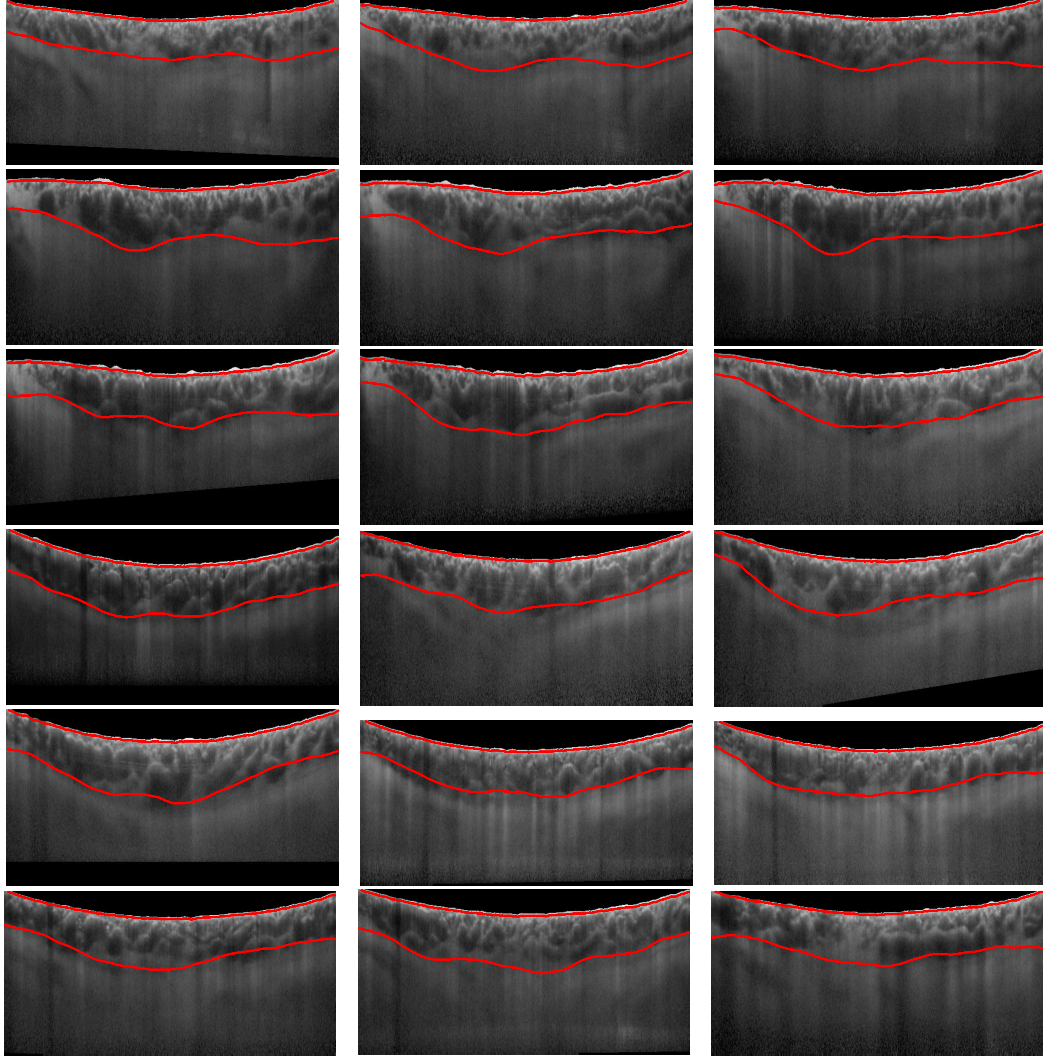


FIGURE 4.7: Example of segmentation of a three dimensional EDI-OCT choroid dataset.

4.1.5 Conclusion

An automated method for segmenting the choroid region in EDI-OCT images by combining super-pixel clustering with convolutional neural networks has been presented. It has been shown that the proposed system based on CNN and clustering has a promising ability for the automatic segmentation of the choroid boundaries. It has been obtained further improvements by defining a new super-pixel energy function involving image intensity, region and enhanced image information. The CNN classification is improved by building enhanced information as well as a distance function which produces excellent agreement with human expert's manual

segmentation. A refinement technique for obtaining an improved segmentation is later defined. It has been demonstrated that the proposed technique outperforms competing methods in terms of accuracy and overlapping given by the Tanimoto Coefficient.

4.2 Optic Disc/Cup Segmentation in Fundus Images

4.2.1 Introduction

Glaucoma is the collective name of a group of eye conditions that results in damage to the optic nerves at the back of the eye, which can cause vision loss. Glaucoma is one of the commonest causes of blindness and is estimated to affect around 80 million people worldwide by 2020 [21]. Glaucoma is known as the silent thief of vision since, in the early phases of the disease, patients do not have any noticeable pain or symptoms of vision loss. It is only when the disease progresses to a significant loss of peripheral vision that the symptoms potentially leading to total blindness may be noticed. Early detection and timely management of glaucoma is key to helping prevent patients from suffering vision loss. There are many risk factors associated with glaucoma amongst which hypertensive intra ocular pressure (IOP) is the most accepted. It is believed that IOP can cause irreversible damage to the optic nerve head, or optic disc (OD). Since the cornea is transparent, the optic disc can be imaged by several optical imaging techniques, including colour fundus photography. In two dimensional (2D) colour fundus images, the OD can be divided into two regions as shown in Figure 4.8: a peripheral zone called the neuroretinal rim and a central white region called the optic cup (OC). The ratio of the size (e.g. vertical height) of the OC to the OD, known as CDR, is often used as an indicator for the diagnosis of glaucoma [23]. Accurate segmentation of the OD and OC is essential for useful CDR measurement. However, manual delineation of the OD and OC boundaries in fundus images by human experts is a highly subjective and time consuming process, which is impractical for use in busy clinics. On the other hand, automated segmentation approaches using computers are attractive as they can be more objective and much faster than a human grader.

4.2.2 Related Work

Many approaches to segmenting of the OD and/or OC in fundus images have been proposed in the literature. The existing methods for automated OD and OC segmentation in fundus images can be broadly classified into three main categories: shape-based template matching [189–195], active contours and deformable based models [24, 143, 196–202], and more recently, machine and deep learning methods [128, 203–218]. A brief overview of the existing methods is given below.

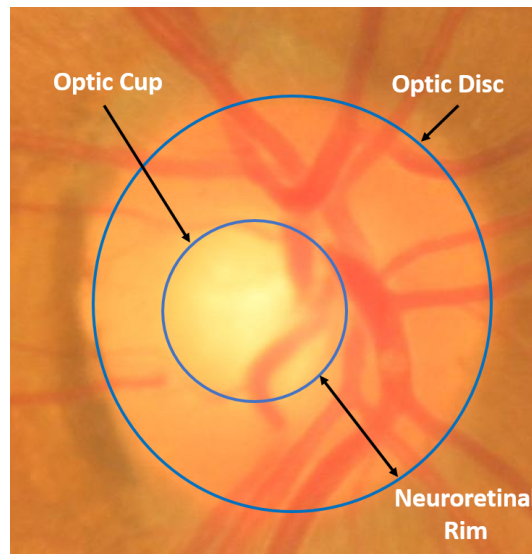


FIGURE 4.8: An example fundus image showing the optic disc and cup with their boundary contours shown in blue.

a) Shape-based and template matching models: These methods model the OD as a circular or elliptical object and try to fit a circle using the Hough transform [190, 191, 194, 195], an ellipse [189, 192] or a rounded curve using a sliding band filter [193]. These approaches typically feature in the earlier work in optic disc and cup segmentation. In general, these shape-based modelling approaches to OD and OC segmentation are not robust enough due to intensity inhomogeneity, varying image colour, changes in disc shape by lesions such as exudates present in abnormal images, and the presence of blood vessels inside and around the OD region.

b) Active contours and deformable based models: These methods have been widely applied for the segmentation of the OD and OC [24, 143, 196–202]. Active contours approaches are deformable models which convert the segmentation problem into an energy minimisation problem where different energies are derived to reflect features in the image such as intensity, texture and boundary smoothness. Lowell et al. detected the location of the OD by applying a template matching approach and then segmenting the OD region with a constrained deformable contour model [196]. Xu et al. defined a model combining smoothness, gradient, and depth information for segmenting both the OD and OC [197]. Hussain et al. presented an approach combining active contour models with genetic algorithms (GAs) for OD segmentation [198]. In [199], Joshi et al. proposed a region-based active contour approach by integrating the local image information around each point of interest and anatomical evidence such as vessel bends at the cup boundary to segment the OD and OC. Furthermore, Yu et al. [143] presented an OD boundary segmentation technique relying on directional matched filtering and combining region and local gradient information in a level set model. Moreover, Zheng et al. [200] introduced an energy function incorporating a shape prior, the location of the disc and cup, the geometric interaction of the optic disc and cup, and the rim thickness for the OD and OC segmentation. Mary et al. [201] designed an OD segmentation approach to compare the performance of the gradient vector flow (GVF) model with nine active contour model algorithms by supplying them with the initial OD contour provided by the circular Hough transform and demonstrated that the GVF model outclassed the rest. A method based on the implicit region active contour model was presented by Mittapalli et al. [24] to segment the OD by incorporating image information from multiple image channels. They also exploited the structural and gray level properties of the cup to segment the OC region. Recently, a model-based method that uses Ant Colony Optimization (ACO) meta-heuristic scheme was presented by Arnay et al. [202]. Their proposed method used the heuristic information which merges the intensity

gradient in the OD area and the curvature associated with the vessels to obtain the OC segmentation in fundus images. Active contour models are often formulated as non-linear non-convex minimisation problems, thus may not achieve the global minima due to the presence of noise and anomalies. In order to achieve good results in a short time, they require a good initialisation of the OD and OC contour provided either manually or automatically, which suggests their performance is dependent on the initialisation.

c) Machine- and deep-learning methods: Machine learning, and in particular more recent deep learning, based methods have shown promising results for OD and OC segmentation [128, 203–218]. Abramoff et al. [203] proposed a pixel feature classification algorithm to segment the OD into cup, rim, and background in colour stereo images by extracting and classifying twelve features using k-nearest neighbour (kNN) classifier. Wong et al. proposed to use support vector machine (SVM) classifier for the OD segmentation [204]. Unlike methods [203] and [204], which are pixel classification based methods, Chen et al. [205] introduced a super-pixel classification based approach to segmenting both the OD and OC in retinal fundus images. They used histograms and centre surround statistics, the location information from the OD and OC to classify each super-pixel using support vector machine (SVM) classifier. In [206], another super-pixel classification based method but with an unsupervised labelling approach was proposed by Xu et al. to segment the OC in retinal fundus images. They formulated the super-pixel classification task as a low-rank representation (LRR) problem for clustering, which can be efficiently solved in closed form. To locate the boundary of the OC, Tan et al. [207] used the existing super-pixel based approach with addressing the problems related to the classification performance variations by integrating and unifying multiple super-pixel resolutions for better cup boundary adherence. Roychowdhury et al. [208] designed a supervised method that uses six region-based features with a Gaussian mixture model classifier to provide the OD region segmentation. Moreover, Akyol et al. [209] presented an approach comprised of

five main steps which are image processing, key-point extraction, texture analysis, visual dictionary, and the random forest classifier to detect the OD contour. Furthermore, Girard et al. [210] proposed local K-means clustering combined with a regularisation step to segment the OD. In addition to that, Sedai et al. [211] presented a coupled shape regression (CSR) framework which consists of a parameter regressor to estimate the CDR and an ensemble shape regressor to predict the OD-OC boundary using the estimated CDR for OD and OC segmentation. The aforementioned machine learning based approaches highly depend on the type of extracted features which might be representative to a particular dataset but not to others. Also, extracting the features manually by hand is a tedious task and takes a considerable amount of time.

Nowadays, deep learning approaches represented by convolutional neural networks (CNNs) are an active research topic [128, 212–218]. Lim et al. [212] applied CNNs to feature-exaggerated inputs emphasizing disc pallor without blood vessel hindering to segment both the OD and OC. In [213], Maninis et al. used fully-convolutional neural network [107] based on VGG-16 net [103] for the optic disc segmentation task. For optic cup segmentation, Guo et al. [214] used large pixel patch based CNNs where the segmentation was achieved by classification of each pixel patch and post-processing. In [215], a modified version of the U-Net convolutional network [108] was presented by Sevastopolsky for automatic optic disc and cup segmentation. Furthermore, Shankaranarayana et al. [216] proposed a joint optic disc and cup segmentation scheme using fully convolutional and adversarial networks. Moreover, a framework consisting of ensemble learning based CNNs as well as entropy sampling was presented in [128] by Zilly et al. for optic cup and disc segmentation. In addition to that, Hong Tan et al. [217] proposed a single CNN with seven layers to segment the OD by classifying every pixel in the image. Most recently, Fu et al. [218] used a polar transformation with the multi-label deep learning concept by proposing a deep learning architecture, named M-Net, to segment the OD and OC simultaneously. In general, these recent deep learning

methods performed well on the basis that they were trained and tested on the same dataset. They might be incapable of achieving robustness and accuracy enough for evaluating the optic disc and cup in clinical practice as there are different type of variations such as population, camera, operators, disease, and image. These concerns of their generalisation ability should be studied thoroughly.

Given the inherent and unsolved challenges encountered in the segmentation of the OD and OC in the aforementioned methods, a new deep learning based method is proposed to segment the OD and OC. The proposed method utilises DenseNet incorporated with fully convolutional network (FCN). The FC-DenseNet, which was originally developed for semantic segmentation [219], is adapted and used for the automatic segmentation of the OD and OC. The significance of the work is two-fold: First, the proposed segmentation method has outperformed the previous methods for the simultaneous segmentation of the OD and OC; Second, generalisation performance of the proposed method is assessed on five different sets of images captured from different imaging devices and settings.

4.2.3 Materials and Methods

Image Datasets

In our experiments, five publicly available datasets of colour retinal fundus images: ORIGA [220], DRIONS-DB [221], Drishti-GS [222], ONHSD [196], and RIM-ONE [223] are used. The ORIGA dataset [220] comprises 650 fundus images with resolution of 3072×2048 pixels including 482 normal eyes and 168 glaucomatous eyes. The DRIONS-DB dataset [221] consists of 110 fundus images with resolution of 600×400 pixels. The Drishti-GS dataset [222] contains 101 fundus images centred on the OD with a Field-Of-View (FOV) of 30-degrees and resolution of 2896×1944 pixels. The ONHSD dataset [196] comprises of 99 fundus images captured using a Canon CR6 45MNf fundus camera from 50 patients. The images have a FOV of 45-degrees and resolution of 640×480 pixels. The RIM-ONE dataset [223]

comprises 169 fundus images taken using a Nidek AFC-210 fundus camera with a body of a Canon EOS 5D Mark II of 21.1 megapixels with resolution of 2144×1424 pixels. The ORIGA, Drishti-GS, and RIM-ONE datasets are provided with the OD and OC ground truth while DRIONS-DB and ONHSD are only provided with the OD ground truth.

Methods

For the OD and OC segmentation task, the proposed deep learning based approach shown in Figure 4.9 comprises three main steps: (i) Pre-processing: the image data are prepared for training with different pre-processing schemes considering the green channel only from colour (red-green-blue [RGB]) images as well as extracting and cropping the region of interest (ROI) represented by the OD region, (ii) Designing and learning: FC-DenseNet architecture [219] is adapted and used to fulfil the pixel-wise classification of images, and finally (iii) Refinement: to obtain the final segmentations by correcting the misclassified pixels located outside the OD and OC areas.

1. **Pre-processing:** first, RGB images without considering any pre-processing scheme (referred to as 'Without' through the text) are used. Pre-processing is applied on Origa data so that the network will generalise better to other datasets which are not used for training and never seen by the network during the learning. One of the consideration for colour information is achieved by training and testing the network using only green channel ('G'). Further, the region of interest represented by the OD area within 2 optic disc diameter (2ODD), has been cropped from green channel ('G+C') and used for the network training.
2. **Designing and Learning:** a FC-DenseNet network adapts the classification network DenseNet [111] to a fully-convolutional neural network (FCN)

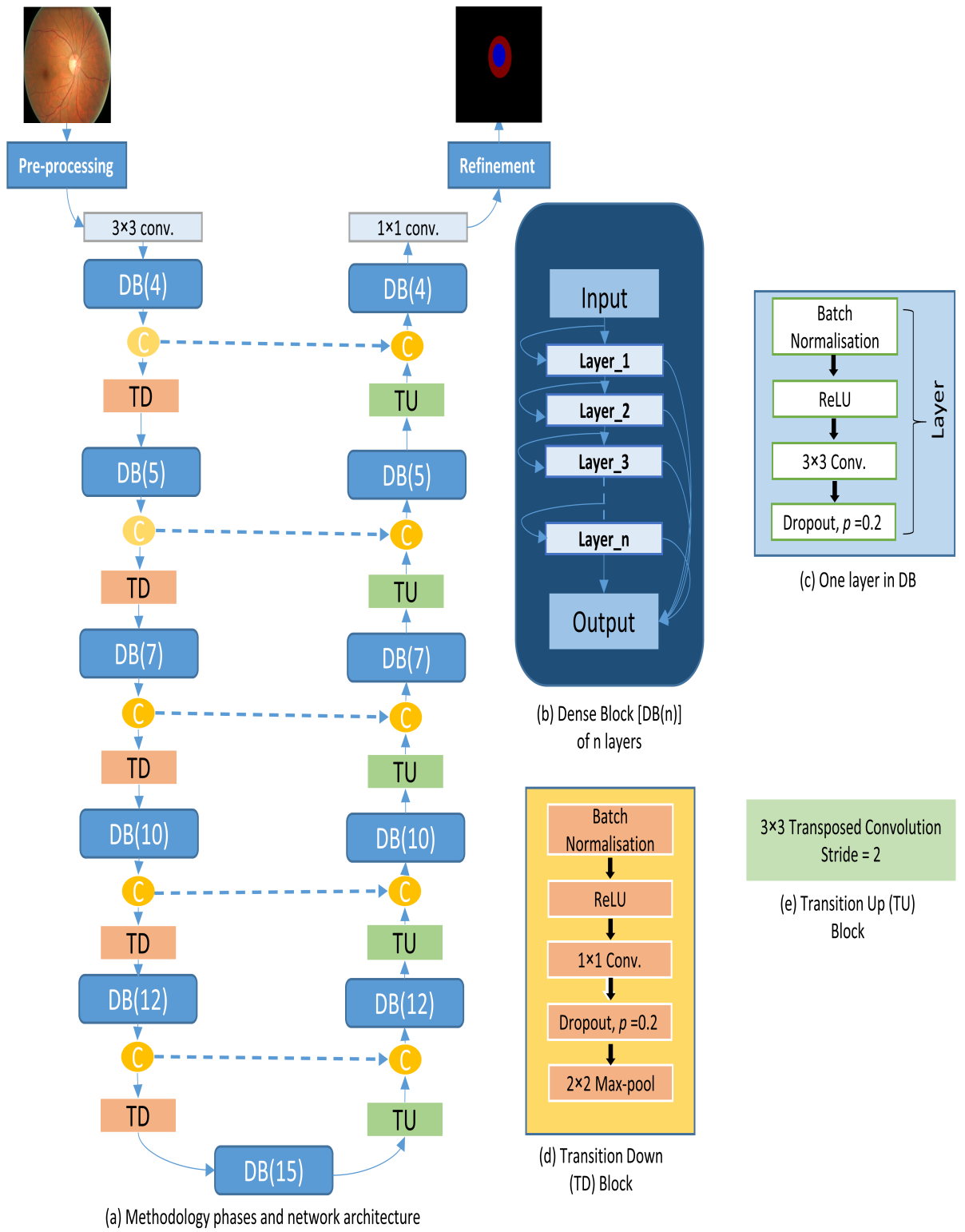


FIGURE 4.9: Block diagram of the proposed optic disc and cup segmentation system. (a) Methodology and fully convolutional DenseNet architecture, (b) Dense Blocks (DB), (c) One layer in DB, (d) Transition Down block (TD), (e) Transition Up block (TU). The circle (C) refers to concatenation process. Note, red and blue represent the cropped rim and OC respectively in the segmentation.

[107] for segmentation. A fully convolutional network is an end-to-end learnable network where the decision-making layers of the network are convolutional filters instead of fully connected layers. This adaptation on the top layers reduces the loss of spatial information caused by fully connected layers as a result of the connectivity of the output neurons of the fully connected layers to all input neurons. The key feature of DenseNet is its ability to further exploit the extracted features reuse and strengthening feature propagation by making a direct connection between each layer to every other layer. In the original paper of DenseNet [111], the authors have demonstrated that CNNs can be substantially deeper, more accurate, and efficient to train if they contain shorter connections between layers close to the input and those close to the output. This makes DenseNet the best choice over other CNN architectures proving its ability to alleviating the vanishing-gradient problem and substantially reducing the number of required parameters. The FC-DenseNet network is composed of three main blocks: dense, transition down, and transition up. Dense block (DB) consists of a batch normalisation layer, followed by rectified linear unit as an activation function, a 3×3 convolution layer, and dropout layer with a dropping rate of 0.2. A transition down (TD) block is composed of batch normalisation layer, followed by rectified linear unit as an activation function, 3×3 convolution layer, dropout layer with a dropping rate of 0.2, and 2×2 Max pooling layer. A transition Up (TU) block contains 3×3 transposed convolution layer.

The architecture of the network used in our experiments (shown in Figure 4.9) is built from one 3×3 convolution layer on the input, followed by five dense blocks each consisting of 4, 5, 7, 10, and 12 layers respectively where each dense block followed by transition down block, one dense block with 15 layers in the last layer of the down-sampling path (bottleneck), five transition up blocks each followed by dense block consisting of 12, 10, 7,

5, and 4 layers respectively, and a 1×1 convolution followed by a non-linearity represented by Softmax function. RMSprop [92], an optimisation algorithm based on stochastic gradient descent, is used for network training with a learning rate of 10^{-3} within 120 epochs with early-stop condition of 30 epochs. To increase the number of images artificially, the images are augmented with vertical flips and random crops. The weights of the network have been initialised using HeUniform [86] and cross-entropy is used as a loss function. Once the network is trained, test stage can be achieved using the trained model to segment the images in the test set

3. **Refinement:** To convert the real values resulted from the final layers of fully convolutional DenseNet into a vector of probabilities (i.e generating the probability maps for the image pixels), the Softmax function is used by squashing the outputs to be between 0 and 1. Here, the OD and OC segmentation problem are formulated as a three class classification task: class 0 as background, class 1 as OC, and class 2 as OD. Thus, the predicted class label of image pixels can be further refined by correcting the misclassified pixels in the background. This can be achieved by finding the area of all connected objects in the predicted images. The object of maximum area is retained by considering it as the OD/OC region and classifying any other small objects as background class label ('G+C+PP').

4.2.4 Results and Discussion

Origa dataset is split into 70% for training (10% of training data is randomly utilised for validation) and 30% for independent test set. The resolution of images is resized into 256×256 . The performance of the proposed method for segmenting the OD and OC when compared with the ground truth was evaluated using

many evaluation metrics such as Dice coefficient (F-Measurement), Jaccard (overlapping), accuracy (Acc), sensitivity (SEN), and specificity (SPC). Dice's and Jaccard's coefficients can be defined in terms of tp , fp , tn and fn as follows:

$$Dice(DC) = \frac{2 \times tp}{2 \times tp + fp + fn} \quad (4.12)$$

$$Jaccard(Jc) = \frac{tp}{tp + fp + fn} \quad (4.13)$$

Most comprehensive study involving five publically available datasets are carried out. This allows for evaluation with images from many different devices and conditions, and from patients of different ethnicities in comparison with previous work demonstrating the robustness of the proposed method. To assess the performance of proposed system, two evaluation scenarios are considered. First, study the performance of the system by training and testing the model on the same dataset (Origa). Second, study the performance of the system by training the model on a dataset (Origa) and testing it on other four independent datasets including DRIONS-DB, Drishti-GS, ONHSD, and RIM-ONE. Tables 4.6, 4.7, 4.8, and 4.9 show the performance of the model trained and tested on the Origa for the OD, OC, joint OD-OC segmentation, respectively. It achieves Dice score (F-measurement), Jaccard score (overlap), accuracy, sensitivity, and specificity of 0.8723, 0.7788, 0.9986, 0.8768, and 0.9994, respectively for the OC segmentation and 0.964, 0.9311, 0.9989, 0.9696, and 0.9994 for the OD segmentation. The performance of segmenting rim area located between the OD and OC contours is also calculated. It achieves Dice score (F-measurement), Jaccard score (overlap), accuracy, sensitivity, and specificity of 0.8764, 0.7849, 0.9975, 0.9028, and 0.9985 on the Origa.

TABLE 4.6: Comparison with the existing methods in the literature for only OD segmentation on different datasets.

Author	Method	Optic Disc					Dataset
		DC(F)	JC(O)	Acc	SEN	SPC	
Wong et al. [204]	Support vector machine based classification mechanism	-	0.9398	0.99	-	-	SiMES
Yu et al. [143]	Directional matched filtering and level sets	-	0.844	-	-	-	Messidor
Mookiah et al. [224]	Attanassov intuitionistic fuzzy histon (A-IFSH) based method	0.92	-	0.934	0.91	-	Private
Giachetti et al. [192]	Iteratively refined model based on contour search constrained by vessel density	-	0.861	-	-	-	MESSIDOR
Dashtbozorg et al. [193]	Sliding band filter	-	0.8900, 0.8500	-	-	-	MESSIDOR, INSPIRE-AVR
Basit and Fraz [225]	Morphological operations, smoothing filters, and the marker controlled watershed transform	-	0.7096, 0.4561, 0.5469, 0.6188	-	-	-	Shifa, CHASE-DB1, DIARETDB1, DRIVE
Wang et al. [226]	Level set method	-	0.8817, 0.8816, 0.8906	-	0.9258, 0.9324, 0.9465	0.9926, 0.9894, 0.9889	DRIVE, DIARETDB1, DIARETDB0
Hamednejad et al. [227]	DBSCAN clustering algorithm	-	-	0.7818	0.74	0.84	DRIVE

Roychowdhury et al. [208]	Region-based features and supervised classification	-	0.8067,	0.991,	0.878,	-	DRIVE, DI-ARETDB1, DIARETDB0, CHASE-DB1, MESSIDOR, STARE
			0.8022,	0.9963,	0.8815,		
			0.7761,	0.9956,	0.8660,		
			0.8082,	0.9914,	0.8962,		
			0.8373,	0.9956,	0.9043,		
			0.7286	0.9854	0.8380		
Girard et al. [210]	Local K-means clustering	-	0.9	-	-	-	MESSIDOR
Akyol et al. [209]	Keypoint detection, texture analysis, and visual dictionary	-	-	0.9438,	-	-	DIARETDB1, DRIVE, ROC
				0.9500,			
				0.9000			
Abdullah et al. [228]	Circular Hough transform and grow-cut algorithm	-	0.7860,	-	-	-	DRIVE, DI-ARETDB1, CHASE-DB1, MESSIDOR, Private
			0.8512,				
			0.8323,				
			0.8793,				
			0.8610				
Hong Tan et al. [217]	7-Layer CNN	-	-	-	0.8790	0.9927	DRIVE
Zahoor et al. [229]	Polar transform	-	0.8740,	-	-	-	DIARETDB1, MESSIDOR, DRIVE
			0.8440,				
			0.7560				
Sigut et al. [195]	Contrast based circular approximation	-	0.8900	-	-	-	MESSIDOR
Proposed	Fully convolutional DenseNet	0.9653	0.9334	0.9989	0.9609	0.9995	ORIGA

Tables 4.10 and 4.12 present the results of proposed system which is trained on the Origa dataset and assessed on the DRIONS-DB and ONHSD datasets, respectively. In these two datasets, only the optic disc segmentation performance are reported because the ground truth of the OC is not provided. The best results

TABLE 4.7: Comparison with the existing methods in the literature for only OC segmentation on different datasets.

Author	Method	Optic Cup					Dataset
		DC(F)	JC(O)	Acc	SEN	SPC	
Hatanaka et al. [230]	Detection of blood vessel bends and features determined from the density gradient	-	-	-	0.6250	1	Private
Almazroa et al. [194]	Thresholding using type-II Fuzzy method	-	-	0.7610, 0.7240, 0.8150	-	-	Bin Rushed, Magrabi, MES-SIDOR
Proposed	Fully convolutional DenseNet	0.8659	0.7688	0.9985	0.9195	0.9991	ORIGA

DC: Dice coefficient, F: F score, JC: Jaccard coefficient, O: Overlapping, Acc; Accuracy, SEN: Sensitivity, SPC: Specificity .

TABLE 4.8: Comparison with the existing methods in the literature for joint OC and OD segmentation on different datasets.

Author	Method	Optic Cup					Optic Disc					Dataset
		DC(F)	JC(O)	Acc	SEN	SPC	DC(F)	JC(O)	Acc	SEN	SPC	
Noor et al. [231]	Colour multi-thresholding segmentation	0.51	-	0.6725	0.3455	0.9995	0.59	-	0.7090	0.4200	1	DRIVE
Khalid et al. [232]	Fuzzy c-Means (FCM) and morphological operations	-	-	0.9026	0.8063	0.9989	-	-	0.937	0.8764	0.9975	DRIVE
Proposed	Fully convolutional DenseNet	0.8659	0.7688	0.9985	0.9195	0.9991	0.9653	0.9334	0.9989	0.9609	0.9995	ORIGA

have been obtained by considering the cropped green channel along with refinement ('G+C+PP') achieving Dice score (F-measurement), Jaccard score (overlap), accuracy, sensitivity, and specificity of 0.9415, 0.8912, 0.9966, 0.9232, and 0.999, respectively, on the DRIONS-DB dataset and 0.9556, 0.9155, 0.999, 0.9376, and, 0.9997 respectively on the ONHSD dataset. Further, the network trained on the Origa is tested on the Drishti-GS and RIM-ONE datasets achieved the results

TABLE 4.9: Results of the proposed method for OD and OC segmentation on **Origa** dataset compared with the existing methods in the literature.

Method	Optic Cup		Optic Disc	
	DC(F)	JC(O)	DC(F)	JC(O)
Yin et al. [233]	0.83	-	-	-
Yin et al. [234]	0.81	-	-	0.92
Xu et al. [206]	-	0.744	-	-
Tan et al. [207]	-	0.752	-	-
Fu et al. [218]	-	0.77	-	0.929
Proposed	0.8659	0.7688	0.9653	0.9334

reported in Tables 4.11 and 4.13, respectively. Also, the best obtained results on these datasets are achieved using the cropped green channel images ('G+C+PP'). Figure 4.10 shows examples of the OD and OC segmentation results on fundus image from the five datasets.

TABLE 4.10: The optic disc segmentation performance on the **DRIONS-DB** dataset considering different data processing schemes. The network is trained on Origa dataset only.

Model	Optic Disc				
	DC(F)	JC(O)	Acc	SEN	SPC
Without	0.62855	0.47715	0.98415	0.4843	0.99955
G	0.8131	0.69825	0.99055	0.73355	0.99845
G+C	0.9091	0.8403	0.99425	0.9232	0.9965
G+C+PP	0.9415	0.8912	0.9966	0.9232	0.999

For Glaucoma diagnosis, CDR is typically calculated along the vertical line passing through the optic cup centre (superior-inferior) and then a suitable ratio threshold may be defined. Varying the thresholds and comparing with the expert's glaucoma diagnosis, it has been achieved an area under the receiving operator curve (AUROC) of 0.7443 based on our segmentations which is very close to the 0.786 achieved using the ground truth segmentations. Since this limits us to considering only a few points on the optic disc, this is extended to incorporate the horizontal CDR (nasal-temporal). That is, it has been taken the average CDR vertically and horizontally and considered thresholds. It is thus achieved an AUROC of 0.7776 which is considerably higher than using only the vertical CDR

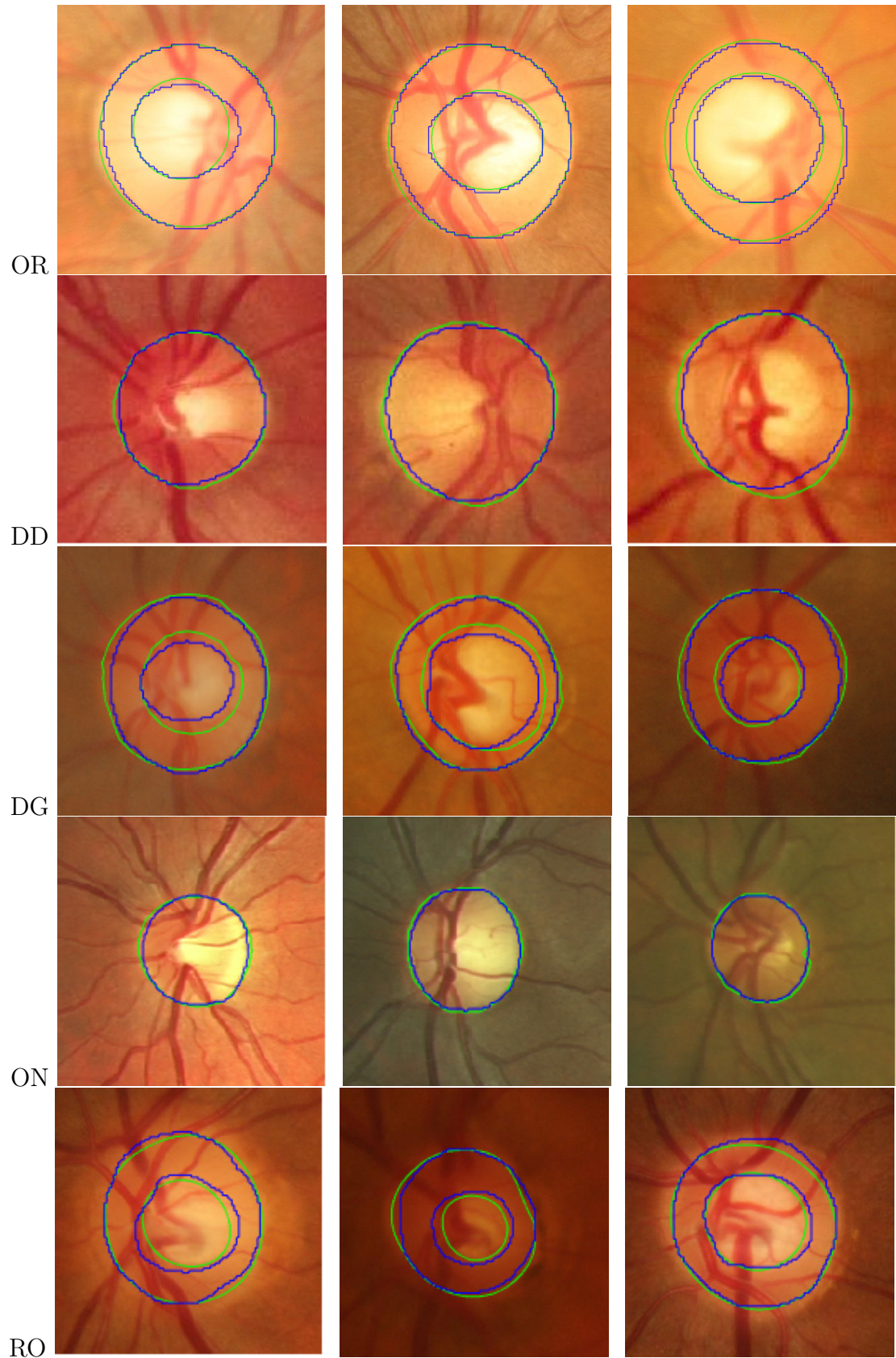


FIGURE 4.10: Examples of joint OD-OC segmentation results. From the first row to the fifth row, the examples are from the Origina (OR), DRIONS-DB (DD), Drishti-GS (DG), ONHSD (ON), and RIM-ONE (RO) respectively. The green contour refers to the ground truth provided with the images while the blue one indicates the results of our proposed method. The DRIONS-DB and ONHSD show the contour of OD only because the ground truth for OC is not provided.

TABLE 4.11: The optic disc, cup, and rim segmentation performance on the **Drishti-GS** dataset considering different data processing schemes. The network is trained on the Origa dataset only.

Model	Optic Cup				
	DC(F)	JC(O)	Acc	SEN	SPC
Without	0.6765	0.5338	0.991	0.5887	0.9989
G	0.7646	0.6259	0.9933	0.6676	0.9994
G+C	0.8045	0.6793	0.9939	0.7413	0.9986
G+C+PP	0.8282	0.7113	0.9948	0.7413	0.9995
Optic disc					
Without	0.719	0.577	0.986	0.5818	0.9997
G	0.851	0.7487	0.9916	0.7695	0.999
G+C	0.9291	0.871	0.9954	0.9268	0.9976
G+C+PP	0.949	0.9042	0.9969	0.9268	0.9992
Rim					
Without	0.3583	0.2309	0.9864	0.2841	0.9965
G	0.509	0.3557	0.987	0.5095	0.9939
G+C	0.7033	0.5601	0.9912	0.7996	0.9938
G+C+PP	0.7156	0.5743	0.9918	0.7996	0.9945

TABLE 4.12: The optic disc segmentation performance on the **ONHSD** dataset considering different data processing schemes. The network is trained on the Origa dataset only.

Model	Optic Disc				
	DC(F)	JC(O)	Acc	SEN	SPC
Without	0.6671	0.5204	0.9935	0.5646	0.9988
G	0.878	0.7924	0.9969	0.9428	0.9975
G+C	0.9392	0.8877	0.9986	0.9376	0.9993
G+C+PP	0.9556	0.9155	0.999	0.9376	0.9997

and closer to the AUROC of 0.7717 achieved by using the experts annotation. Figure 4.11 shows ROC curves of both the expert's glaucoma diagnosis and our proposed system detection, respectively.

In terms of comparing our proposed method to the existing methods in the literature, Tables 4.6, 4.7, 4.8, 4.9, 4.14, 4.15, 4.16, and 4.17 and present the comparison in terms of Dice score (F-measurement), Jaccard score (overlap), accuracy, sensitivity, and specificity measurements. Table 4.6 presents the comparison of the model trained and tested on the Origa with the existing methods proposed for the OD segmentation. The comparison with 15 methods shows that our method outperforms almost all of them. Wong et al. [204] reported segmentation overlap of 0.9398 which is slightly better than 0.9334 obtained by our proposed

TABLE 4.13: The optic disc, cup, and rim segmentation performance on the **RIM-ONE** dataset considering different data processing schemes. The network is trained on the Origa dataset only.

Model	Optic Cup				
	DC(F)	JC(O)	Acc	SEN	SPC
Without	0.2584	0.1657	0.98	0.2688	0.9886
G	0.5011	0.3627	0.9872	0.6635	0.9911
G+C	0.6096	0.4709	0.9888	0.9052	0.9904
G+C+PP	0.6903	0.5567	0.9928	0.9052	0.9944
	Optic Disc				
	DC(F)	JC(O)	Acc	SEN	SPC
Without	0.4204	0.2833	0.9629	0.2984	0.9924
G	0.6799	0.5364	0.978	0.5979	0.9946
G+C	0.8455	0.7423	0.9864	0.874	0.9915
G+C+PP	0.9036	0.8289	0.9922	0.8737	0.9976
	Rim				
	DC(F)	JC(O)	Acc	SEN	SPC
Without	0.1869	0.1088	0.9715	0.1166	0.9983
G	0.3969	0.2578	0.9741	0.3041	0.9951
G+C	0.7108	0.5666	0.9844	0.6591	0.9946
G+C+PP	0.7341	0.5942	0.9863	0.6585	0.9966

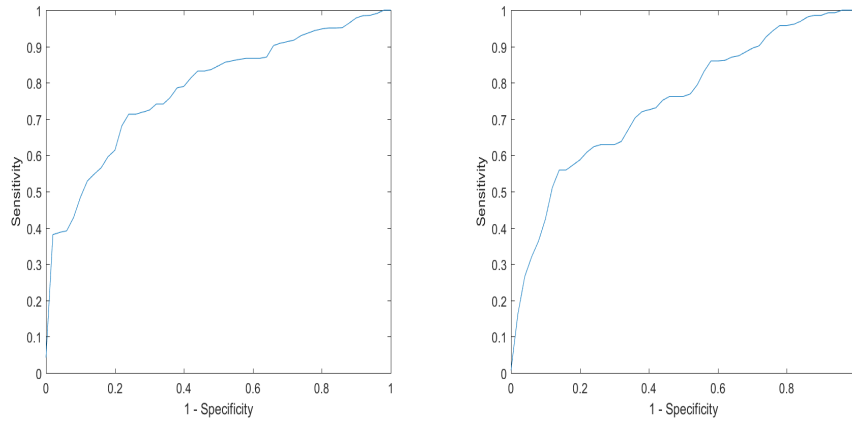


FIGURE 4.11: From left to right: ROC curves of the expert's glaucoma diagnosis and the proposed system.

system. However, their method only segments the OD region and they used features extracted manually which might be applicable to the dataset they have used but not to other datasets. For the OC region, our proposed method achieves the best results comparing to other existing methods as shown in Table 4.7. For joint OD and OC segmentation results shown in Table 4.8, our method also outperforms the proposed methods in the literature.

Tables 4.14 and 4.15 present the comparison of our system trained on Origa and tested on the Drishti-GS and RIM-ONE datasets respectively with the methods

were trained and tested on those datasets. The results of our method on Drishti-GS outperform the results reported by Nawaldgi [235] and Oktoeberza et al. [236]. Sedai et al. [211] and Zilly et al. [128] report Dice and overlap scores slightly better than ours in segmenting the OC and OD regions. However, they used the same dataset (Drishti-GS) for training and testing their system while our system is trained on the Origa images only and tested on the Drishti-GS which make it more generalisable. Furthermore, Guo et al. [214] and Sevastopolsky [215] used the same dataset (Drishti-GS) for training and testing, and segmented the OD region only. For the RIM-ONE dataset, our method is compared with three methods as shown in Tables 4.15. Similarly, these methods were tested on the same dataset used in the learning process which makes the efficacy of their system performance doubtful on other datasets. Tables 4.9, 4.16, and 4.17 show that our system trained only on Origa gives the best results compared to others on Origa, DRIONS-DB, and ONHSD datasets, respectively.

For the rim region segmentation, our system achieved an overlap of 0.7708 and balanced accuracy; which can be obtained by calculating the mean of achieved sensitivity and specificity; of 0.93 on Origa dataset. The most recent published paper for the OD and OC segmentation [218] reported rim segmentation overlap of 0.767 and balanced accuracy of 0.941 on the Origa. Their reported results are very close to ours although they have used a different scheme of data splitting for training and testing. Other existing methods in the literature have not reported rim region segmentation performance.

Furthermore, AUROC curve performance shows excellent agreement between grading done by ophthalmologist and the proposed system for glaucoma diagnosis. Combining the vertical cup to disc ratio with horizontal cup to disc ratio significantly improves the automated grading results and suggests that these diagnosis results could be further improved by using complete profile of the OD.

TABLE 4.14: Results of the proposed method for OD and OC segmentation on **Drishti-GS** dataset compared with the existing methods in the literature.

Method	Optic Cup			Optic Disc		
	DC(F)	JC(O)	Acc	DC(F)	JC(O)	Acc
Sedai et al. [211]	0.86	-	-	0.95	-	-
Sevastopolsky [215]	0.85	0.75	-	-	-	-
Guo et al. [214]	0.9373	0.8775	-	-	-	-
Nawaldgi et al. [235]	-	-	0.97	-	-	0.99
Zilly et al. [128]	0.871	0.85	-	0.973	0.914	-
Oktoeberza et al. [236]	-	-	-	-	-	0.9454
Proposed	0.8282	0.7113	0.9948	0.949	0.9042	0.9969

TABLE 4.15: Results of the proposed method for OD and OC segmentation on **RIM-ONE** dataset compared with the existing methods in the literature.

Method	Optic Cup		Optic Disc	
	DC(F)	JC(O)	DC(F)	JC(O)
Sevastopolsky [215]	0.82	0.69	0.94	0.89
Shankaranarayana et al. [216]	0.94	0.768	0.977	0.897
Arnay et al. [202]	-	0.757	-	-
Proposed	0.6903	0.5567	0.9036	0.8289

TABLE 4.16: Results of the proposed method for OD segmentation on **DRIONS-DB** dataset compared with the existing methods in the literature.

Method	Optic Disc	
	DC(F)	JC(O)
Sevastopolsky [215]	0.94	0.89
Abdullah et al. [228]	-	0.851
Zahoor et al. [229]	-	0.886
Proposed	0.9415	0.8912

TABLE 4.17: Results of the proposed method for OD segmentation on **ONHSD** dataset compared with the existing methods in the literature.

Method	Optic Disc		
	DC(F)	JC(O)	Acc
Dashtbozorg et al. [193]	0.9173	0.8341	0.9968
Girard et al. [210]	-	0.84	-
Abdullah et al. [228]	-	0.801	-
Sigut et al. [195]	-	0.865	-
Proposed	0.9556	0.9155	0.999

4.2.5 Conclusion

A novel approach based on a fully convolutional Dense network has been proposed for the joint simultaneous segmentation of the OD and OC in colour fundus images. The proposed method achieves the segmentation by extracting complex data representations from retinal images without the need of human intervention. It has been demonstrated that the performance of the proposed generalised system can outperform or achieve comparable results with competing approaches. These findings also reflect the efficiency and usefulness of FC-DenseNet for the OD and OC segmentation.

4.3 Summary

This chapter has presented two segmentation methodologies to segment the boundaries of the choroid and OD/OC in EDI-OCT and fundus images, respectively. The segmentation of the aforementioned anatomical structures is greatly required to extract certain features that help in diagnosis and subsequently treatment of the diseases affecting the retina and the choroid. Thus, the developed automated segmentation approaches has a promising potential to be an effective step towards implementing an accurate and reliable computer aided-diagnosis system. However, carrying out of an extensive validation to the choroid segmentation method would allow evaluating with images from many different devices and coexisting conditions (such as PED and SRF fluids), and from patients of different ages, genders and ethnicities. Moreover, further investigation supported by wide clinical studies would help validating the obtained findings in the clinical setting.

Chapter 5

Retinal Disease and Lesion Classification

Disease and lesion classification is very important in the development of precise computer-aided diagnosis systems (CADs). In this chapter, retinal diseases classification and lesions (fluids) identification based on deep CNN are considered in two types of images: digital colour fundus and OCT. Glaucoma detection using fundus images is considered in §5.1 based on features extracted by a CNN to train a support vector machine (SVM) classifier. Next, a deep learning multi-stage CNN model is presented to diagnosis the disease of diabetic maculopathy in §5.2. In §5.3, the concepts of fine-tuning a pre-trained CNN and ensemble learning are exploited to classify retinal fluids into three types IRF, SRF, and PED in OCT volumes. Finally, §5.4 summarises the work in this chapter.

5.1 Automated Glaucoma Grading

5.1.1 Introduction

In Section 4.2, a method of segmenting the OD and OC regions and subsequently obtaining glaucoma diagnosis based on CDR value were presented. The main aim

of the work presented in the previous chapter was to design an accurate system for OD and OC segmentation that could lead to better glaucoma diagnosis. However, in clinical settings, more features than the thinning of neuroretinal rim (NRR) measured by CDR are considered for glaucoma diagnosis. Signs and features used for glaucoma referral divisions include the presence of disc haemorrhages, NRR thickness not following the ISNT (Inferior-Superior-Nasal-Temporal) rule, alpha and beta peripapillary atrophy, and notches in the NRR [220]. In this section, the problem of glaucoma detection is considered where a small data size with pre-defined OD region are provided. Based on the transfer learning concept, it has been aimed to achieve diagnosis by extracting the features that characterise glaucoma from the region of interest.

5.1.2 Related Work

To detect glaucoma in retinal fundus images, the existing methods in the literature relied on different handcrafted features extracted from images manually. In [237], a method based on an artificial neural network (ANN) classifier and morphological features of the optic nerve was proposed to detect glaucoma. Furthermore, the authors in [238] presented a glaucoma Risk Index (GRI) approach to detect glaucoma by extracting features then feeding them to a principle component analysis (PCA) algorithm for dimensionality reduction and a SVM model to classify the images into either normal or pathological.

A method based on texture and higher order spectral (HOS) features with a random forest classifier was proposed in [4] to diagnose glaucoma in fundus images. Moreover, in [239], wavelet features with a feature selection technique were applied to detect glaucoma in the images using a sequential minimal optimisation (SMO) classifier. Recently, an approach based on Gabor features along with SVM was proposed to detect glaucoma in the images [5]. More recently, a method based on capturing the discriminative features using contextualising deep learning from the

OD after applying segmentation on the region of interest to detect glaucoma was proposed in [240].

In this work and based on a fixed feature extractor concept originating from transfer learning, an automated system to detect glaucoma in retinal fundus images is proposed. Unlike aforementioned methods, the features that discriminate the OD region from optic cup region are extracted automatically without the need for human intervention. These features are used to train a SVM model in order to detect the presence of glaucoma in a given image.

5.1.3 Material and Method

Data

A publicly available database called RIM-ONE [223], which comprises 455 high-resolution glaucoma and non-glaucoma images along with their OD and OC segmentations, is used to evaluate and test the proposed method. The images in this dataset are classified into 255 normal and 200 glaucomatous images.

Pre-processing

The images are resized to 227×227 pixels. The images are applied to the network without any enhancement or further pre-processing step except image resizing to decrease the computational time. The data is randomly divided into 70% for training and 30% for validation and evaluation of the implemented system.

Feature Extraction using CNN

Figure 5.1 shows the stages of glaucoma grading for the proposed system. In the first stage, a CNN is formed using linked layers of neurons like other neural networks with more complexity in the hidden layers. A pre-trained CNN model (AlexNet) which comprises 23 layers including convolution layers, max pooling layers, fully connected layers, softmax layer and output layer is used in the training

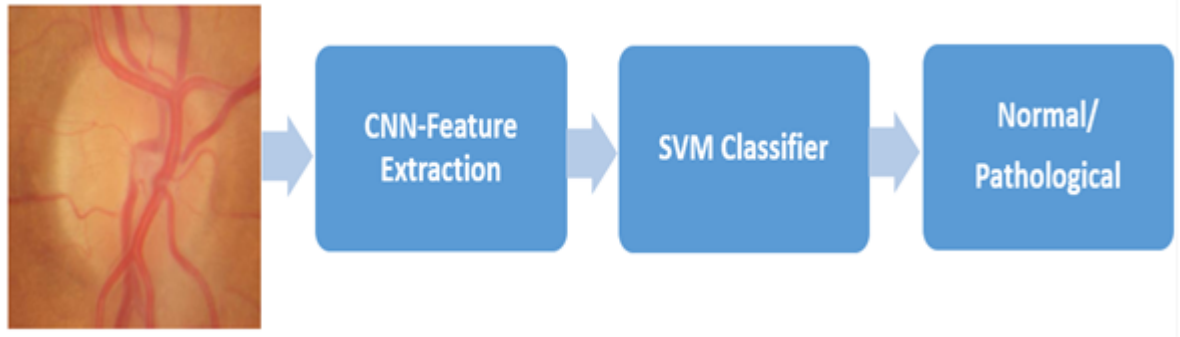


FIGURE 5.1: Proposed system stages of glaucoma grading.

stage as a feature extractor as shown in Table 5.1. Training the entire network from scratch should be avoided since the size of dataset is small. Similarly, fine-tuning the entire CNN can also be avoided due to overfitting problems. Instead, the learned AlexNet's weights which are trained on more than a million images on the ImageNet database [241] to classify images into 1000 classes are transferred. Thus, the full AlexNet (after removing the fully connected layers) is treated as a fixed-feature extractor for dataset under study. The training data is passed through the CNN model to extract the features where the images' higher-level features are available in the deeper layers of the CNN. These training features are extracted from the layer right before the classification layer (fully connected layer) 'fc7' and then fed into the SVM classifier to train it. A GPU is used in this stage rather than CPU to speed up the computation as CNNs are highly computationally intensive.

Training SVM using CNN Features and Predication

Support vector machines (SVM) are among the most widely used supervised classification methods in the field of machine learning. Hard margin classifier are considered the simplest kind of SVM, which solve an optimisation problem to find the linear classification rule with maximal geometric margin. Thus, in the linearly separable case, the hard margin SVM builds the hyperplane that classifies all data correctly and maximises the distance to the nearest training data points.

TABLE 5.1: Architecture of AlexNet.

Layer Name	Parameters Description
‘inputimage’	$227 \times 227 \times 3$ images with normalisation
‘conv1’	96 $11 \times 11 \times 3$ convolutions with stride: [4 4] and padding: [0 0]
‘relu1’	ReLU
‘norm1’	Channel normalisation
‘pool1’	3×3 max-pooling with stride: [2 2] and padding: [0 0]
‘conv2’	256 $5 \times 5 \times 48$ convolutions with stride: [1 1] and padding: [2 2]
‘relu2’	ReLU
‘norm2’	Channel normalisation
‘pool2’	3×3 max-pooling with stride: [2 2] and padding: [0 0]
‘conv3’	384 $3 \times 3 \times 256$ convolutions with stride: [1 1] and padding: [1 1]
‘relu3’	ReLU
‘conv4’	384 $3 \times 3 \times 192$ convolutions with stride: [1 1] and padding: [1 1]
‘relu4’	ReLU
‘conv5’	256 $3 \times 3 \times 192$ convolutions with stride: [1 1] and padding: [1 1]
‘relu5’	ReLU
‘pool5’	3×3 max-pooling with stride: [2 2] and padding: [0 0]
‘fc6’	4096 fully connected layer
‘relu6’	ReLU
‘fc7’	4096 fully connected layer
‘relu7’	ReLU
‘fc8’	2 fully connected layer
‘prob’	softmax
‘classificationLayer’	cross-entropy

In practice, datasets are usually not linearly separable, and therefore, the SVM optimisation problem must be modified. This modification is necessary in order to achieve a trade-off between minimising classification error on the training data points and maximising geometric margin by applying the soft margin idea. Soft margin builds a hyperplane that allows misclassification of difficult or noisy examples while maximising the distance to the nearest cleanly separated data examples [242, 243].

In this stage, the features extracted from the previous step are fed to SVM classifier for training. The parameters of SVM are tuned and set to $C = 10$ (cost parameter defines the weight of how much samples inside the margin contribute

to the overall error) and polynomial kernel with degree of 3 to achieve best performance results. Once the SVM model is learned, the test images are used to evaluate the performance of the proposed system by classifying the test data into either normal or pathological. A SVM training algorithm with a binary class predicts the labels of points in a test dataset by building a model for the training dataset. Given a set of binary-labelled training vectors, SVMs learn a decision boundary to discriminate between the two classes. The resulting classification rule can be used to classify new test examples.

5.1.4 Results and Discussion

The confusion matrix that shows the prediction performance is shown in Figure 5.2. The proposed system was evaluated on 30% of 455 images in terms of accuracy, sensitivity and specificity.

		0	1	
Output Class	0	69 50.7%	9 6.6%	88.5% 11.5%
	1	7 5.1%	51 37.5%	87.9% 12.1%
		90.8% 9.2%	85.0% 15.0%	88.2% 11.8%
		Target Class		

FIGURE 5.2: Confusion matrix shows prediction performance evaluation (0: normal; 1: glaucomas).

The network achieved 88.2%, 85%, 90.8% accuracy, sensitivity and specificity, respectively. The obtained results seem very promising where the proposed system

is implemented without applying any data augmentation or oversampling. Furthermore, the proposed method does not require prior knowledge of retinal image features like blood vessel structure.

For the sake of comparison and to study the effect of adopting pre-trained weights for a small dataset, the CNN is trained from scratch. This network achieves accuracy of 76.64%, sensitivity of 72.73%, and specificity of 81.67% which are less than the results obtained from the pre-trained network. Therefore, there is a strong motivation to learn transferable feature representations for biomedical images.

this method is also computationally simple such that it does not require segmentation of the optic disc in order to get the features for the region of interest. Moreover, all the features that represent the region of interest are extracted automatically from the data itself by the CNN without the need of manual feature extraction.

5.1.5 Conclusion

In this section, an automated approach to detect glaucoma in retinal fundus images has been presented. The proposed method based on CNN, demonstrated promising performance in diagnosing glaucoma with considerably lower computational cost compared to existing equivalent methods. The raw images were directly applied to the CNN without any enhancement or pre-processing except for image resizing to reduce the computation cost. Key features of the disease are automatically extracted from stack layers of the filters convolved along the raw image and passed to the SVM model for classification. The transfer learning used in this section has proved to be an effective strategy in situations where insufficient training data prevents a deep learning model use.

5.2 Automated Diabetic Macular Oedema Grading

5.2.1 Introduction

Diabetic Macular Oedema (DMO) or Diabetic Maculopathy (DM) is a condition characterised by the appearance of exudate close to the macula. Consequently, the central vision of the patient is affected. DMO usually develops at any time during the progression of Diabetic Retinopathy (DR). DR is associated with high blood glucose levels that cause damage to the vessels supplying blood to the retina.

According to the Early Treatment Diabetic Retinopathy Study (ETDRS) [25], the severity of DMO is mainly divided into two categories: non-clinically significant macular oedema (non-CSMO) and clinically significant macular oedema (CSMO). It is believed that the early detection and treatment of DMO may improve visual acuity. Different imaging techniques have been used for the diagnosis of DMO such as retinal thickness analyser (RTA), colour fundus photographs, fluorescein angiography (FA) and optical coherence tomography (OCT).

5.2.2 Related Work

Recently, many automated and computerised systems for DMO grading have been introduced along with associated image processing techniques for exudate, fovea detection and segmentation using retinal fundus images [15], [244]. In order to detect and grade the severity of DMO, existing methods in the literature have relied on either detection of the location and segmentation of exudate and the macula [244],[245] or the extraction of texture and image based features [246].

In [247], Tariq et al. proposed a method based on extracting morphological features and the location of exudate after segmenting the exudate using a Gabor filter bank and mathematical morphology. Finally, the distance between the exudate and the centre of the macula was calculated in order to grade the severity of

DMO in each image. Furthermore, Zaidi et al. [245] developed a grading method using Gabor filtering, mathematical morphology and Otsu thresholding with a Bayesian classifier to detect the location of exudate and positional constraints to grade the severity of DMO.

Moreover, Giancardo et al. [6] proposed an automated grading system based on an exudate probability map and wavelet decomposition. They used the Kirsch edge operator and a region growing algorithm to locate the hard exudate and the fovea region. After that, the features were extracted using wavelet analysis and fed into a SVM model for classification. In [246], the authors developed a method based on motion pattern analysis and the Radon transform to extract the features. The features were then fed into two classifiers to detect the presence of DMO in the images. Baby et al. [248] used Gaussian data description (GDD) to extract features from the wavelet sub-bands that are obtained by a dual tree complex wavelet transform (DT-CWT). Furthermore, a method based on higher order spectra features was proposed by Mookiah et al. [15].

Based on the aforementioned proposed methods, the performance of the grading systems relies on the exudate segmentation, anatomical structure localisation and feature extraction strategies. However, the detection of anatomical structures (i.e. fovea and macula) and exudate segmentation are challenging. Furthermore, the features are extracted manually and depend highly on the dataset used to evaluate the proposed methodology. Moreover, finding and engineering a feature set that is appropriate for different datasets is still a challenge since features that are representative of or descriptive for one dataset are often not representative of or descriptive for other datasets.

Automated feature learning algorithms depending on deep learning have recently emerged as a feasible approach and have proven to be effective in some computer vision applications. However, its effectiveness in DMO grading is not yet thoroughly explored in the literature. Also, as one of the transfer learning strategies was used in the previous section to detect the presence of glaucoma

in fundus images, in this section a deep learning multi-stage CNN approach is introduced to diagnose DMO. The proposed approach addresses the problem of relying on hand-crafted features as well as the time consumed in the segmentation of retinal landmarks such as the fovea and optic disc and lesions.

In traditional automated grading systems [15], [244]-[248], in order to grade DMO, the contrast of images is enhanced as a pre-processing stage and then the blood vessels are removed using matched filtering or mathematical morphology. Further to this, the location of the macula is detected and exudate segmentation is applied. After that, different texture, morphological, and image-based features are extracted from the segmented exudates. In the last stage, the DMO grading is calculated depending either on the distance of exudate from the macula or using machine-learning algorithms. The bottom diagram in Figure 5.3 shows a simplified block diagram for traditional DMO grading systems [15].

As the performance of the automated grading system is highly dependent on the extracted features, the performance of the existing methods may vary from one dataset to another because the extracted features are not always representative of different datasets. Also, these features are hand-crafted which incurs a time cost and considerable effort. So, the need to adopt a generalised and automatic feature extraction method is the best solution to counter these issues. Furthermore, traditional approaches have relied on the position of exudate around and near the macula to grade the severity of DMO. The extraction of exudate mainly depends on the efficiency of the segmentation algorithm while detection of the location of the macula and fovea highly depends on the accuracy of data mining and texture imaging techniques. Therefore, implementing a method of grading DMO which is independent of the segmentation algorithms is a crucial task.

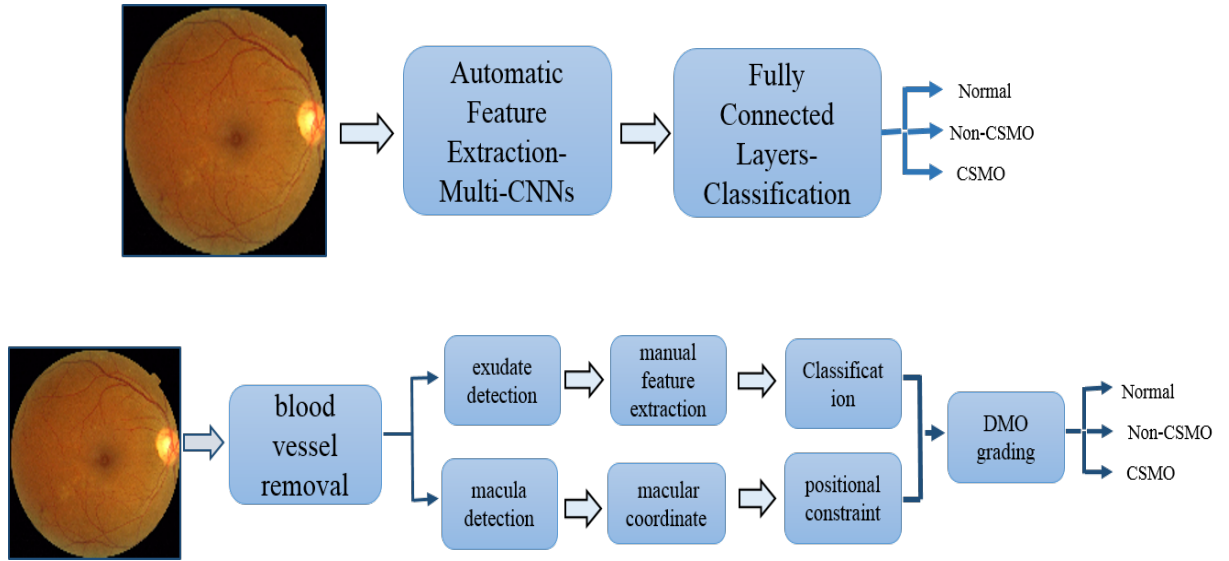


FIGURE 5.3: A block diagram showing the comparison between the traditional systems and our proposed system. The first row shows the proposed system and the second row is the traditional system. The proposed system does not depend on any kind of segmentation, hand-crafted feature and sums up many stages of the traditional system in only two automatic stages. CSMO: Clinically Significant Macular Oedema.

5.2.3 Material and Method

Data

An existing dataset (called MESSIDOR) [157] is used in this study for training and evaluation purposes because it is a fairly large dataset and it is labelled. It comprises 1200 images classified as either normal (no DMO), Non-CSMO or CSMO. These images were captured by using a color 3CCD camera on a Topcon TRC NW6 with 45-degree field of view (FOV) with resolutions of either 1440×960 , 2240×1488 or 2304×1536 pixels.

Pre-processing

The proposed feature learning and grading approach comprises three main stages: pre-processing, network design and training, and an evaluation stage. In this stage, the smallest rectangular region containing the entire FOV is automatically determined and used to crop each image. After that, the cropped images are resized into three different sizes 128×128 , 256×256 , and 512×512 pixels to

obtain acceleration while keeping the images sufficiently large to identify features such as exudates. Finally, the red, green and blue (RGB) channels in the image are scaled to have zero mean and unit variance.

Network Design and Training

The top diagram in Figure 5.3 shows a diagram of the proposed DMO grading system. In the feature extraction stage, the features are automatically learnt by multi-stage CNN and fed into a classifier for classification. The proposed system represents a promising solution to address the aforementioned concerns of traditional systems: the proposed system is automatic and does not depend on any kind of segmentation or hand-crafted feature.

The structure of the proposed CNN is shown in Figure 5.4. Three CNN architectures are implemented and trained with each of three different image sizes 128×128 , 256×256 and 512×512 pixels. The main objective of training three different networks is to obtain fast computation time. The weights of the networks trained on the smaller images are used to initialise the networks trained on the larger images. This helps to speed up the process of training without resorting to resizing images below a level where key features may not be detectable for the final classification.

The structure of the first network includes the layers in the first block along with the fully connected layers (shown in Figure 5.4). This network is trained from scratch using images of size 128×128 pixels. The architecture of the second network comprises the first and second blocks along with the fully connected layers. The weights are initialised from the first network and trained by using images with 256×256 pixels. Finally, the third network comprises the first, second and third block along with the fully connected layers and are trained on images with 512×512 pixels with the weights being initialised from the second network. The final network architecture comprises 13 convolution layers with filter of sizes 5×5 and 3×3 . Each convolutional layer is followed by a Leaky (0.01) rectified linear

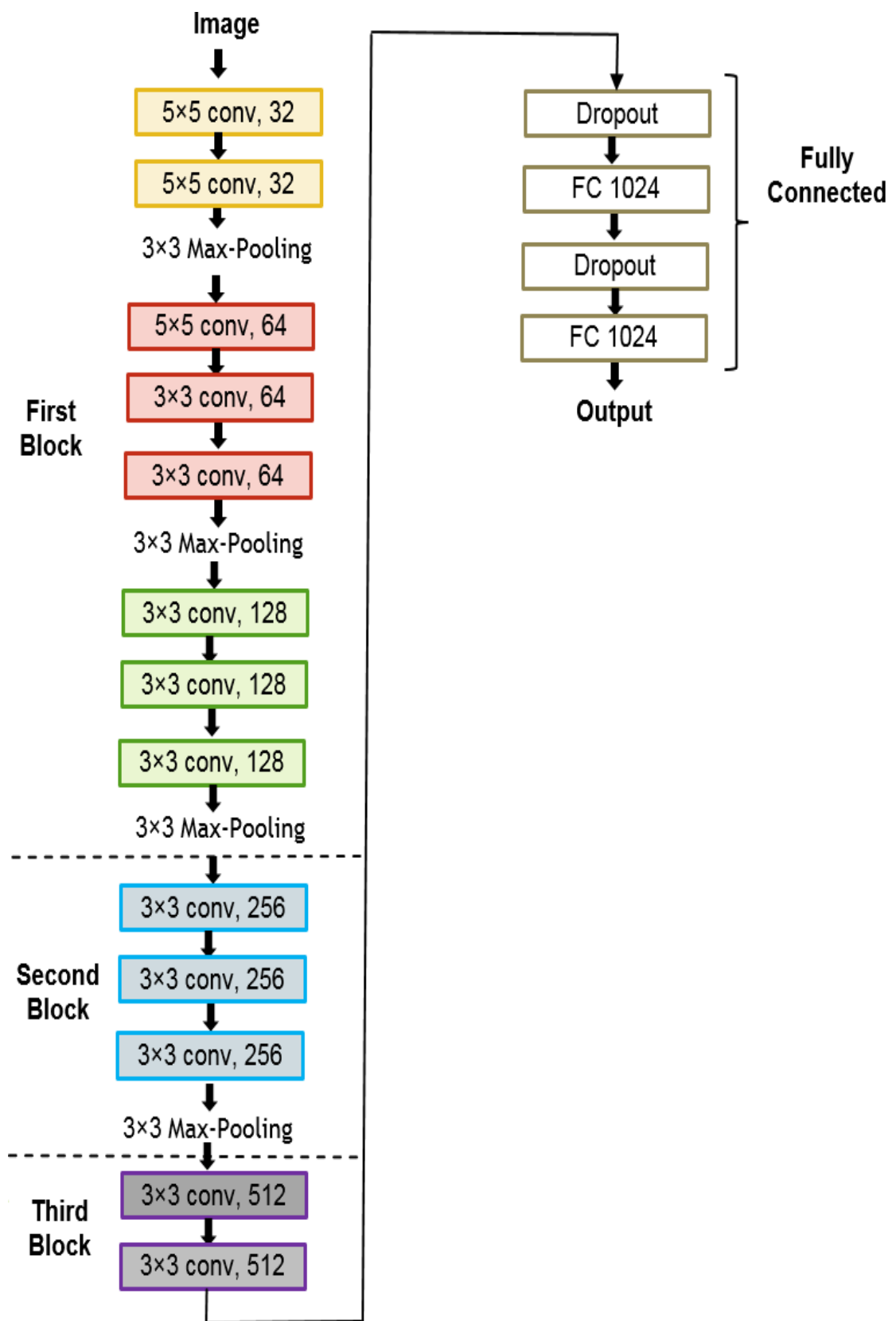


FIGURE 5.4: Proposed multi-stage convolutional neural network architecture.

unit (ReLU) step. Four max-pooling layers with window size of 3×3 , 2 fully connected layers with 1024 neurons each and two dropout layers between the fully connected layers are used.

The dataset is randomly divided into 70% for training and validation (10% of this data is used for validation), and the remaining 30% for testing. To increase the size of data artificially in order to decrease the possible overfitting problem, the data are augmented. In every epoch during training, each image is randomly augmented with: random rotation between 0-360 degrees, random horizontal and vertical flipping, random translations of between -40 and 40 pixels, random zooming and random shearing. Furthermore, to counter the impact of unevenly distributed data, oversampling is applied on the imbalanced training set in order to get more uniform distribution of classes and increase detection performance on the rare classes.

To train the networks, stochastic gradient descent SGD with the Nesterov momentum optimisation algorithm [89] is used with adaptive learning rate (start=0.003, stop=0.00003) and momentum parameter 0.9. The orthogonal weight initialisation method proposed in [84] is considered to initialise the weights of filters in the first implemented network. The first and second networks are trained with 200 epochs while the third network is trained with 250 epochs. The loss function used for optimisation is the mean squared error (MSE) with a thresholding value in order to predict and obtain the three classes (Normal (0), non-CSMO (1), CSMO (2)). Moreover, L2 regularisation with weight decay factor 0.005 is used in the convolutional layers and a dropout rate of 0.5 is used between fully connected layers. These are implemented as regularisation approaches to decrease overfitting in the network during training.

Testing

Once the network is trained, the test images are used to evaluate the performance of the implemented network by predicting the classification of previously unseen

data. The performance of the implemented system are evaluated using three measurements; sensitivity, specificity and accuracy. In this study, sensitivity is defined as the percentage of images which are correctly classified as having DMO out of the true total number of images with DMO. Specificity is defined as the percentage of images that are correctly classified as not having DMO out of the true total number of images without DMO. Accuracy is the percentage of images that classified correctly.

5.2.4 Results and Discussion

The confusion matrix shown in Figure 5.5 gives the prediction ratio per class. The proposed system was evaluated on 30% of 1200 images (360 images) where the trained CNN achieved 88.8%, 74.7% and 96.5% in terms of accuracy, sensitivity and specificity, respectively.

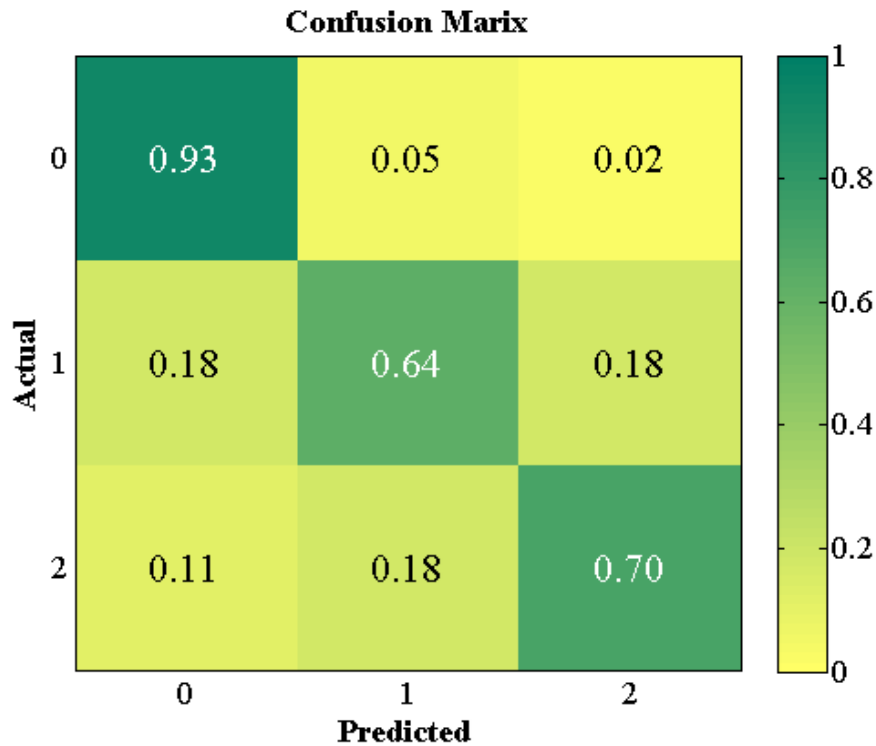


FIGURE 5.5: Prediction ratio per class in confusion matrix, (0) refers to normal images, (1) non-CSMO and (2) CSMO. It shows good prediction results despite the rarity of classes 1 and 2.

From the confusion matrix in Figure 5.5, it can be noticed that the prediction ratios for each of the classes are very encouraging. An important consideration is the imbalanced distribution of the images in the dataset, 154 images (13%) are grade 2 (CSMO), 75 images (6%) are grade 1 (non-CSMO) while 971 images (81%) are grade 0 (Normal). Consequently, the non-CSMO and CSMO classes constitute only 19% of the data while the majority of the images are class 0 (normal). This imbalanced distribution may cause the CNN to overfit to the majority classes and result in the differences observed in prediction ratios between classes.

An oversampling strategy has been used in order to balance the distribution in the case of rare or less common classes. The oversampling can be done by adding copies of images from the under-represented classes (CSMO and non-CSMO). Our proposed system demonstrates comparable results with other methods in the literature; higher accuracy than [249] who achieved 85.2% is obtained and accuracy close to machine learning methods such as [245] and [15] which had 94.1% and 95.56% respectively. Although our accuracy is close to these two models, this achieved without the need of prior feature extraction, exudate and macula segmentation or the removal of retinal blood vessels. The sensitivity obtained by the proposed system is promising and comparable to the sensitivities achieved by automated grading methods in the literature. However, the performance needs for more improvements to be applied and used in a clinical diagnostic setting.

5.2.5 Conclusion

To conclude this section, an automated method for grading the severity of DMO has been presented. It has been shown that the proposed system based on multi-stage CNN has convincing ability for automated feature extraction from retinal fundus images and grading of DMO. This technique will be valuable for future automated DR grading system.

5.3 Retinal Lesion Classification

5.3.1 Introduction

Age-related macular degeneration (AMD) is known as the main cause of the degradation or loss visual acuity in developed countries [250]. The main signs of AMD are the presence of the retinal fluids/lesions (sub-retinal fluid (SRF) and intra-retinal fluid (IRF)) and sub-retinal pigment epithelium (sub-RPE) fluid (pigment epithelium detachment (PED)). These lesions/fluids act as a biomarker for the early diagnosis of AMD and are thus helpful in analysing progress of the disease and advising a treatment.

IRF appears as contiguous fluid-filled spaces containing columns of tissue and is located between the internal limiting membrane (ILM) and the inner/outer segment (IS/OS) junction. SRF is associated with the existence of exudate in the subretinal space that is located between the IS/OS junction and retinal pigment epithelium. PED appears as three types: serous, fibrovascular, or drusenoid in which the retinal pigment epithelium (RPE) is separated from the underlying Bruch's membrane (BM) as a result of fluid accumulation [251]. These are illustrated in Figure 2.6 in Chapter 2.

Antivascular endothelial growth factor (anti-VEGF) therapy is an effective treatment for AMD. It demands a regular examination to check the changes of the retinal fluids during the anti-VEGF medication [252], [253]. A well-known monitoring and non-invasive cross sectional imaging approach is optical coherence tomography (OCT) that can help to visualise anatomic changes and various type of fluids in the retina.

As the number of acquired SD-OCT volumes and their resolutions increase, manual identification of the retinal fluids becomes challenging task. Hence, there has been a crucial need for automated techniques which give accurate identification of the abnormalities.

5.3.2 Related Work

In the literature, a few machine learning based approaches to detect and segment the retinal fluids in OCT images have been proposed. In [254], an automated method for detection of the subretinal and subretinal pigment epithelium fluids was proposed. They used a graph cut and Split Bregman-based segmentation method to detect the dark regions located between the layers. Then, potential fluids in that regions are considered to extract their features and classify them using random forest. Furthermore, a method to segment the intraretinal and subretinal fluids in 3D OCT images was proposed by combining a machine learning approach represented by an artificial neural network with segmentation method based on geodesic graph cut [255]. Moreover, a graph search and machine learning based approach to segment pigment epithelium detachment fluid (PED) in SD-OCT volumes was presented in [256].

The authors of [257] implemented a framework consisting of four stages; feature extraction, pre-segmentation, dimension reduction and supervised learning approach; to segment subretinal and intraretinal fluids. In [258], a fully automated approach to segment subretinal and intraretinal fluids in 3D OCT images was presented. The fluid segmentation was achieved by exploiting a machine learning method that combines unsupervised feature representation and heterogeneous spatial context with a graph theory segmentation approach. A non-machine learning approach was introduced by considering fuzzy level set method to detect IRF and SRF on 3D OCT images [259] in which the boundaries of detected fluids were combined to create a comprehensive volumetric segmentation.

The aforementioned approaches towards fluid identification are not end-to-end paradigms. Often, hand-crafting and heuristics are employed in choosing the graph weights. Furthermore, the testing phase of these approaches are relatively slow due to graph optimisation which is often computationally intensive. Moreover, machine learning algorithms used in these approaches are traditional methods in

which features are extracted manually which cost time and effort. To address the issue of hand-crafted features, much research has been carried out relying on deep CNN to identify the three fluid types IRF, SRF, and PED in OCT volumes [260–268].

In [260], a deep learning based approach where U-net convolutional neural network architecture was applied on OCT images to detect intraretinal fluid (IRF). Yadav et al. [261] proposed Generalised Motion Pattern (GMP) based segmentation approach using a cascade of fully convolutional networks for detection and segmentation of retinal fluids (IRF, SRF, PED) from OCT scans. Moreover, Kang et al. [262] proposed a two stage deep neural network; the first network was implemented to detect and segment fluids, while the second network was designed to enhance the robustness of the first network. In addition to that, a method based on faster R-CNN and 3D region growing was suggested in [263] to segment the fluids. In [264], the authors proposed a fully-convolutional neural network (FCNN) which merges dilated residual blocks in an asymmetric U-shape configuration to classify and segment the fluids. Tennakoon et al. [265] proposed an approach to identify and segment the fluid based on an adapted version of U-Net trained using a combined loss function including an adversarial loss term. A method based on graph-cut and fully convolutional networks was suggested by Lu et al. [266] to label the fluid pixels using random forest classification to detect the fluid regions. Morley et al. [267] proposed a framework using both ResNet and Encoder-Decoder neural network architectures for fluid identification and segmentation. Furthermore, Rashno et al. [268] presented a method based on graph shortest path algorithms and CNN for fluid detection and segmentation.

However, despite the recent efforts, retinal fluid identification in OCT images remains a challenging problem due to the strong presence of speckle noise in OCT images, the shape and location of fluids, and fluid variability size.

In this work, detecting the presence of fluids has been handled as a multi-label classification problem where all classes (fluids) coexist in a given image. A

framework has been developed to calculate the probability that IRF, SRF and PED are present in OCT volumes. The novel framework comprises three main stages: (i) preprocessing; to improve the quality of the data and make the model more effective for various devices, (ii) a deep and multi-label CNN is adapted to do the classification, and (iii) the integration of weighted ensemble learning and parameter selection is explored.

5.3.3 Materials and Methods

Data

The Retinal OCT Fluid challenge (Retouch) dataset (used in the context of the MICCAI'17 Challenge) [19] has been used in the training and evaluation of the proposed system. The dataset includes 70 volumes (4752 B-scans) accompanying reference annotations. There are 24 volumes acquired with each of the two OCT imaging devices: Cirrus (Zeiss) and Spectralis (Heidelberg, Germany), and 22 volumes acquired with T-1000 and T-2000 (Topcon). For each volume from these three imaging devices, the numbers of B-scans were 128, 49 and 128, with resolution 1024×512 , 496×512 , and 885×512 , respectively. Fluids, which are IRF, SRF and/or PED, accumulate in some B-scans due to AMD. Although not all B-scans contain fluid, there is at least one type of fluid in each volume.

Pre-processing

OCT volumes contains a lot of speckle noise because OCT imaging modalities use coherent beam reflection to capture images. This noise often degrades the quality of the image and causes problems in functioning and effectiveness of image processing algorithms. Classical denoising techniques like adaptive filtering and median filtering based approaches cannot preserve boundary information. To decrease the amount of the noise, improve the quality and enhance the images, images are pre-processed before training stage as follows:

1. **Denoising:** to reduce the effect of the noise that degrades the quality of the OCT scans, the fractional total variation based approach presented in [47, 48] was used. This method relies on the principle that signals with excessive details have high total variation (i.e the integral of the absolute gradient of the signal is high). The method proposed in [47] reduces unwanted details (the noise) while preserving important details such as edges producing smooth and piecewise constant images.
2. **Enhancement:** as the images were captured from three different imaging devices and in order to find an appropriate representation for the histogram of these images by standardising them, exact histogram specification algorithm [269] was used. Exact histogram specification converts the intensity image values so that the resulting image has a particular histogram which has a desired shape. In this work, the idea of image histogram which can be arbitrarily specified was applied by using three different histograms depending on the device utilised to acquire the images (Cirrus, Spectralis, Topcon). Furthermore, the contrast of images were enhanced by applying the contrast-limited adaptive histogram equalisation technique [43].

Figures 5.6, 5.7, and 5.8 present OCT images showing the three fluids IRF, SRF, and PED before and after processing in Cirrus, Spectralis, and Topcon, respectively.

Multi-label Classification with CNN

In this work, GoogleNet (Inception v3) CNN architecture [270] is adopted and adapted to detect the presence of three fluids. Training this network from scratch may takes several weeks where its training is a computationally intensive task. To overcome this problem, the transfer learning strategy which is proven to be effective in several cases is considered. The Inception model, which is pre-trained on ImageNet [113], is exploited by considering the pre-trained weights and then

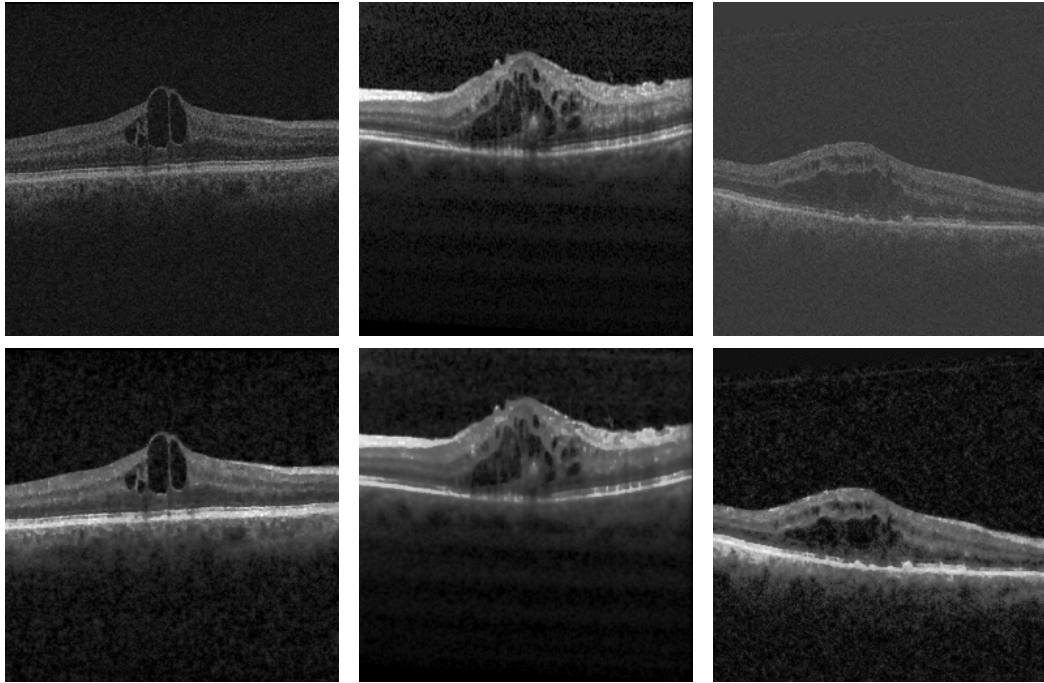


FIGURE 5.6: OCT images with IRF fluid. The images on the first row are the original images while the images on the second row are the denoised and enhanced images. The desired histogram of the enhanced images is specified by the histogram of Spectralis for the devices Cirrus, Spectralis, and Topocon, respectively.

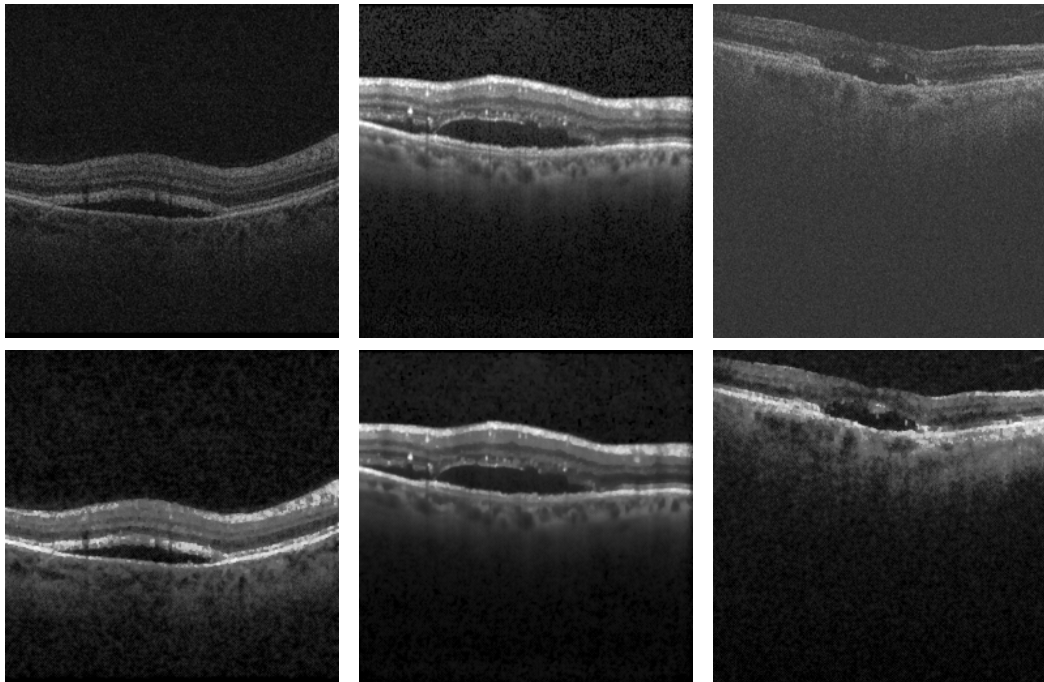


FIGURE 5.7: OCT images with SRF fluid. The images on the first row are the original images while the images on the second row are the denoised and enhanced images. The desired histogram of the enhanced images is specified by the histogram of Spectralis for the devices Cirrus, Spectralis, and Topocon, respectively.

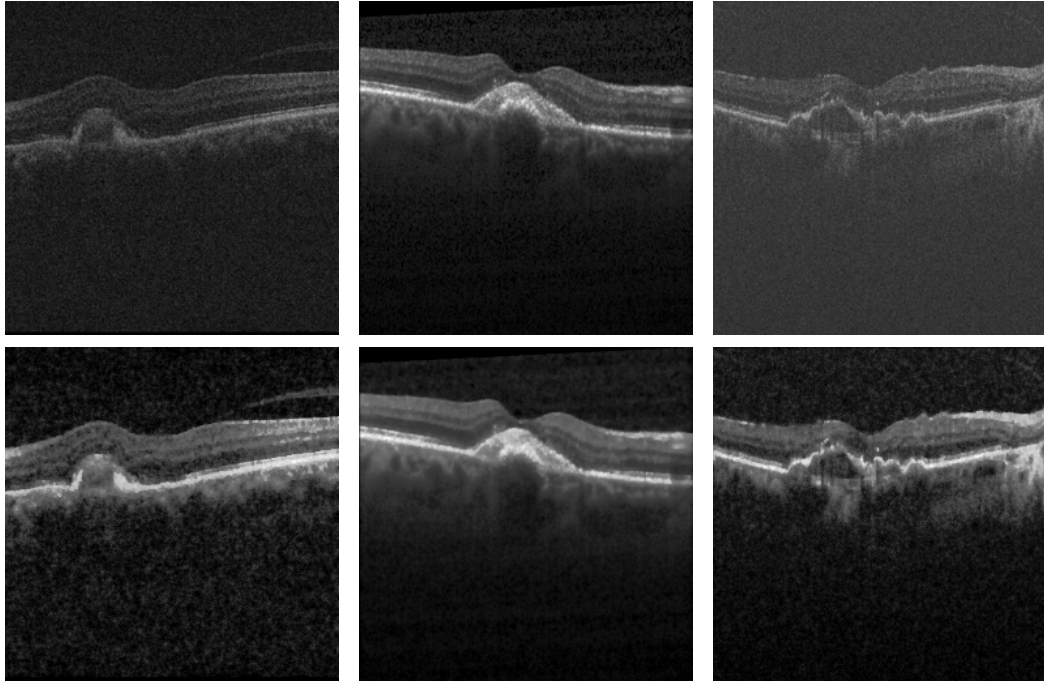


FIGURE 5.8: OCT images with PED fluid. The images on the first row are the original images while the images on the second row are the denoised and enhanced images. The desired histogram of the enhanced images is specified by the histogram of Spectralis for the devices Cirrus, Spectralis, and Topocon, respectively.

fine-tuning them. As Inception was originally trained on 1000 class (i.e 1000 output neurons), the last layer (fully connected layer) is truncated and replaced with a fully connected layer with 3 output neurons (three types of fluids).

The Inception model comprises five convolutional layers alternated with max-pooling operations, successive stacks of 11 Inception modules which are basically mini-models inside the bigger model, and softmax output layer and auxiliary softmax as an intermediate output as shown in Figure 5.9. Stacking several Inception modules makes the network architecture complicated and deep. In addition to the depth of the network, Inception modules are also wide which are designed to detect features at multiple length scales. The size of convolutional filters in each block are either 1×1 , 3×3 , 5×5 , 1×7 or 7×1 .

The idea of using 1×1 convolutional filters is to reduce the computation cost by stacking 1×1 convolutional filters in front of the 3×3 , 5×5 filters. Also, the filter size of 1×7 and 7×1 are used based on the idea of factorising convolutions

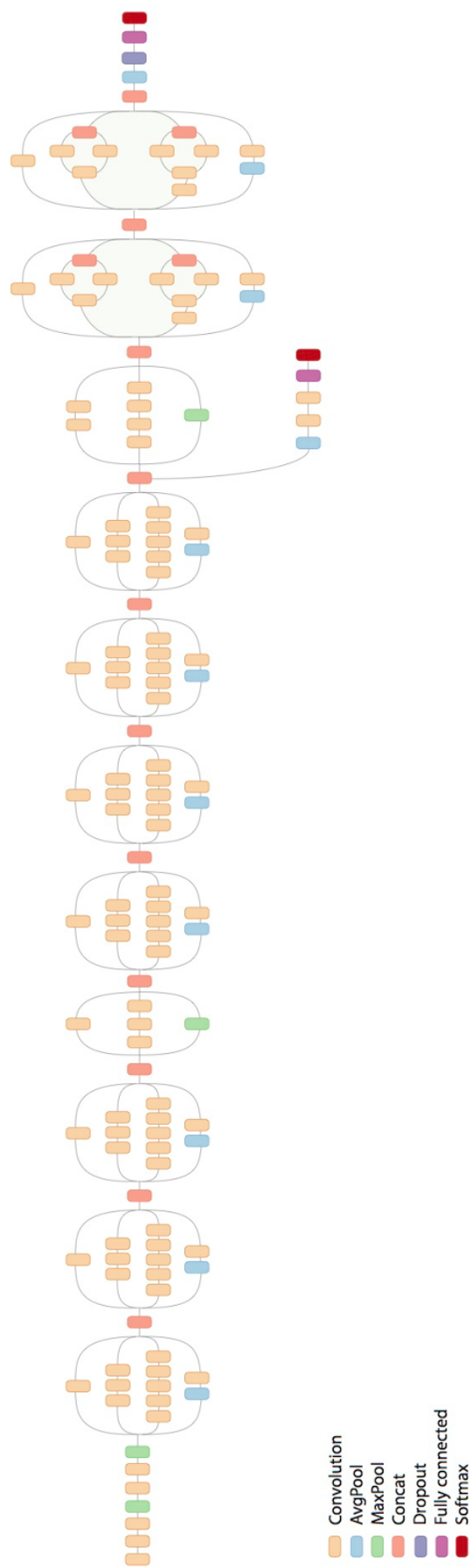


FIGURE 5.9: Schematic diagram of Inception v3 network architecture [270].

with large filter size. It is noticed that $n \times n$ convolutional filter can be replaced by a $1 \times n$ convolution followed by a $n \times 1$ convolution which results in increasing of the computational cost saving as n grows [270]. The output of these filters are then stacked before being fed into the next layer in the network. The size of max-pooling and average-pooling windows are 3×3 , except the last average pooling window (the top of Figure 5.9) is 8×8 .

The fluid identification problem we deal with is a multi-label classification task. In multi-label classification, multiple target labels can be assigned to each image instead of only one. So that, more than one type of fluid can exist in one OCT image. Inception is originally designed and used for a single-label image classification task. Therefore, modifications are made on inception CNN architecture to make it works and trains on the multi-label classification problem. These changes include replacing the cross-entropy loss function with the sigmoid cross-entropy loss function and the softmax function into logistic sigmoid function which is defined by:

$$f(x) = \frac{1}{1 + \exp^{-x}} \quad (5.1)$$

In the machine learning context, sigmoid function refers to a class of functions with S-shaped curves where their gradients are simple to calculate. These adaptations for inception CNN architecture are necessary to calculate the confidence score of each fluid in the images rather than calculating the overall probability of classes (fluids). Softmax function squashes all values into a range of $[0, 1]$ summing together to 1 which is only suitable for a single-label classification problem. With our multi-label classification problem, logistic sigmoid function is attached to each neuron in the final layer which is fully connected layer with three nodes (one for each fluid).

Initially, 58 volumes (80% of the data) are randomly chosen for training and the remaining 12 volumes (20%) were used for independent testing. Five convolutional

neural networks which have the same architecture (Inception v3) are trained with 4000 epochs on enhanced and resized into 299×299 pixels B-scan images. During the network learning, Root Mean Square Propagation (RMSPROP) SGD with momentum of value 0.9 [92] is used as an optimisation algorithm which can be given as follows:

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1 - \gamma)g_t^2 \quad (5.2)$$

Where g_t is the gradient, γ is decay term with value of 0.9, and $E[g^2]$ is the RMSprop running average of the past squared gradients. Thus, the network parameters are updated as:

$$W_{t+1} = W_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}}g_t \quad (5.3)$$

Where η is the learning rate with value 0.1 and ϵ is a small positive number used to prevent division by 0 which has value 1 in this experiment. RMSPROP works by dividing the learning rate for a weight by a running average of the magnitudes of recent gradients for that weight. Moreover, rectified linear unit (ReLU) is used as activation function in the network and the size of training batches is 32.

Each CNN is trained on training data which is preprocessed in a certain way. Three different strategies for the preprocessing of training image data are considered: images without any preprocessing (WithoutPre), images preprocessed by denoising (Den), images preprocessed by denoising and followed by enhancement using histogram specification algorithm where the target histograms are either Spectralis (DenH1), Cirrus (DenH2), or Topcon (DenH3).

Once the training of the networks is finished, the test images can be tested and the confidence score of each fluid (class) in each B-scan is determined. As the resulting confidence score of each fluid from the sigmoid function is a value in the range of $[0, 1]$, a threshold value is considered to convert this value into 0s and 1s for evaluation. The performance of the trained network on the test

data can be evaluated by comparing the predicted confidence score of the classes resulting from the classifier and converted into 0's and 1's against the real values. Threshold value of 0.5 is considered in our experiment and the evaluation of the network performance in this stage is done by considering B-scan OCT images regardless of the volume that the images belong to.

TABLE 5.2 shows the accuracy detection for each fluid in five networks where the name network refers to the type of preprocessing method applied on the training data. Although histogram specification does not help to improve the performance in IRF and SRF, it helps in PED. To overcome this issue, ensemble learning is considered by allocating a specific weight for each network.

TABLE 5.2: Accuracy detection in terms of B-scans of three fluids in each network. DenH1: Denoising and Spectralis image histogram, DenH2: Denoising and Cirrus image histogram, DenH3: Denoising and Topcon image histogram, IRF: Intraretinal fluid, SRF: Subretinal fluid, PED, Pigment epithelium detachment.

Network	IRF	SRF	PED
DenH1	0.8147	0.9107	0.9264
DenH2	0.7742	0.9152	0.9325
DenH3	0.7877	0.9137	0.9415
Den	0.8335	0.9152	0.9264
WithoutPre	0.8282	0.8912	0.9325

Ensemble Learning and Parameter Selection

From the results shown in TABLE 5.2, it can be noticed that is no single network which can give the best accuracy for all fluids. In order to optimise the results and inspired by the ensemble learning concept in machine learning, the output of these five networks is combined by taking into consideration a specific weight for each network.

To choose the best weight for each network to each fluid type, brute force parameter selection method has been used to tune the five weight parameters in order to choose the best values that can be used for test data. The range of

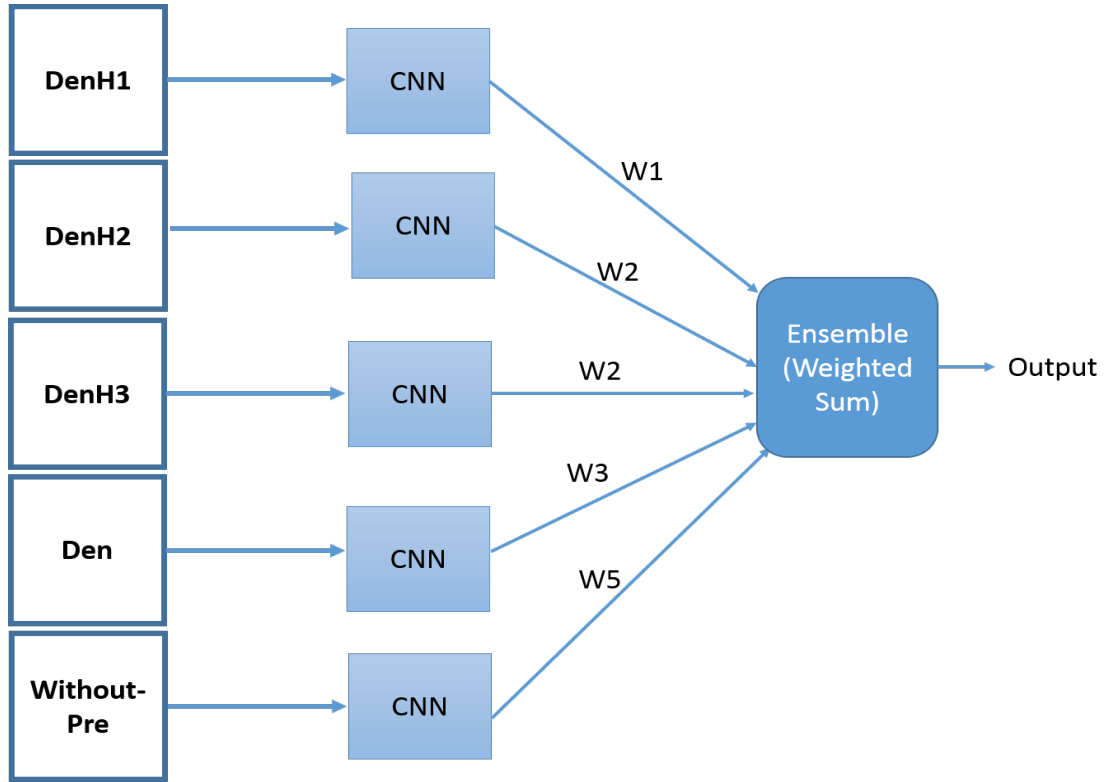


FIGURE 5.10: Ensemble of CNN combined by weighted sum. DenH1: Denoising and Spectralis image histogram, DenH2: Denoising and Cirrus image histogram, DenH3: Denoising and Topcon image histogram, W: Weight.

weight values in each network is $[0, 1]$ with a step 0.1 where a mesh grid of size $(11, 11, 11, 11, 11)$ is constructed for this parameter selection problem. 110 values for the threshold have been considered ranging in $[0, 1]$ with a step of 0.01.

5.3.4 Results and Discussion

In order to evaluate the performance of the combined weighted networks, Area Under Curve of Receiver Operating Curve (ROC) for each fluid is calculated as shown in TABLE 5.3 and Figure 5.11 .

The first row in TABLE 5.3 shows the AUR for the three fluids in terms of OCT B-Scan. To consider the volume to which each B-Scan belongs in the AUR calculations, three different approaches are suggested; mean, maximum, and the maximum mean of 3 consecutive B-scans in a volume. In the mean method, the mean of the probabilities of fluid presence in all B-Scans that belong to the same

TABLE 5.3: Area under curve of each fluid in each method considering the volumes. B-scan scheme represents the AUC calculation without taking the volumes into consideration

Method	Fluid	W1	W2	W3	W4	W5	AUC
B-Scan	IRF	0	0	0	0.3	1	0.8660
	SRF	0.9	0.1	0	0	0	0.9522
	PED	0	0	0.8	0.2	0.9	0.9794
Mean	IRF	0.5	0	0	0	0.2	0.7708
	SRF	0.1	0	0	0	0	0.9271
	PED	0.1	0	0	0	0	1
Max	IRF	0	0	0.1	0	0	0.8333
	SRF	0.1	0	0	0	0	0.8750
	PED	0.1	0	0	0	0	1
Max the mean of 3 consecutive B scans	IRF	0	0	0	0.1	0.2	0.8542
	SRF	0.1	0	0	0	0	0.8750
	PED	0.1	0	0	0	0	1

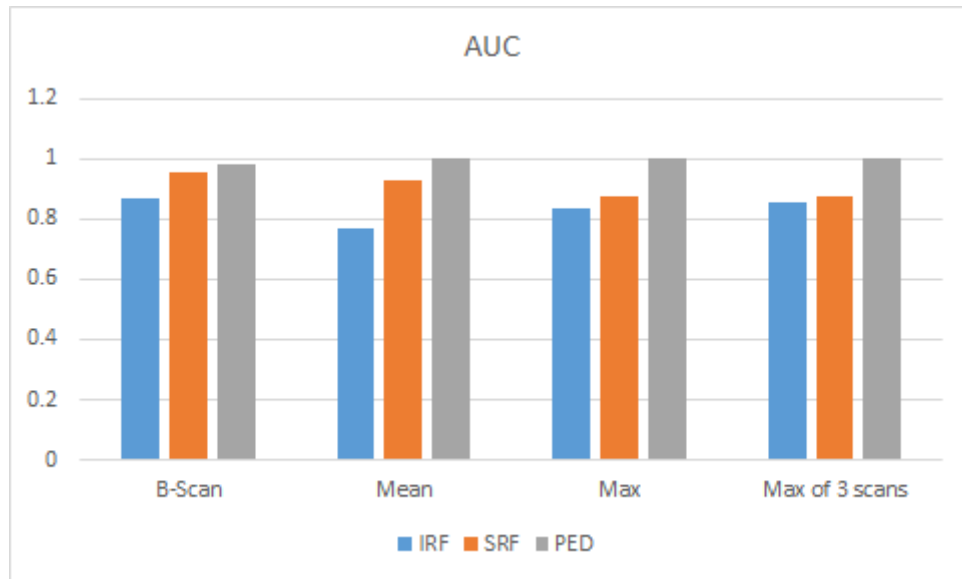


FIGURE 5.11: Area under curve for retinal fluid detection.

volume is calculated to represent the probability of the fluid detection in that volume. In the maximum method, the maximum of the probabilities of a fluid presence in all B-Scans that belongs to a certain volume is found to represent the probability of the fluid detection in that volume. Moreover, the maximum mean of 3 consecutive B-scans in a volume works by calculating the mean probability of each 3 consecutive B-scans that belong to a certain volume and then finding the

maximum mean probability value of each B-Scan that belongs to that volume.

Overall, the best results of fluid identification considering the volume can be achieved by using maximum the mean of 3 consecutive B-scans scheme. In this scheme, the AUC of fluids detection are 0.8542, 0.875, and 1 for IRF, SRF, and PED, respectively. For the sake of comparison and using the same dataset, Yadav et al. [261] achieved AUC 0.85, 0.84, and 0.87 for IRF, SRF, and PED identification, respectively which are less than our obtained results. Although Lu et al. [266] reported results higher than ours, they used a Random forest classifier with hand-crafted features to identify the fluids. Unlike our method, the manually extracted features are not necessary to have a good generalisation ability on other datasets. Furthermore, the promising fluid identification results were obtained by evaluating the proposed method on OCT volume images captured by three different devices with high presence of noise and pathologies which demonstrates the robustness of the fluids' identification framework.

5.3.5 Conclusion

This section proposed a deep learning method for fluids classification from OCT volumes. Both concept of the transfer learning and ensemble learning have been exploited to identify the fluids IRF, SRF, and PED in 3D OCT images. The pre-trained Inception CNN has been modified to fit the multi-label classification problem. The ensemble learning concept has been used to combine many CNNs with a specific weight for each one, where the weight has been chosen depending on a parameter selection scheme. The obtained results appear promising where the proposed framework has been successful in identifying the presence of retinal fluids with a good prediction ability.

5.4 Summary

Exploiting CNNs is still facing numerous challenges; of these, the size of training datasets can be considered the most challenging. CNNs typically demand a huge data size in order to be trained. Although data augmentation is considered a valuable option to increase the size of the dataset artificially, transfer learning has also proven to be effective when there is no adequate data to train the CNN. This chapter aimed to exploit the transfer learning idea into building deep CNNs to learn invariant and more transferable feature representations from fundus and OCT images. It has been demonstrated that transferable learned features are used effectively in glaucoma, DMO and retinal fluid identification by designing models which are more stable and accurate.

Chapter 6

Conclusions and Future Work

In this thesis, efficient and accurate deep learning-based methodologies have been developed in order to produce effective automated techniques for the interpretation of retinal images and the diagnosis of retinal diseases. In this final chapter, we conclude this work and present some ideas for future research arising from the work carried out in this thesis.

6.1 Summary

Various retinal image analysis tasks including localisation, segmentation and classification have been investigated and explored through the designing and developing of several automated analysis methods. These presented methods have examined the development of deep learning approaches based on Convolutional Neural Networks (CNNs) alongside with the use of colour fundus and OCT images and image processing techniques.

Automated detection of the optic disc and fovea centres has the potential for great impact in the automated diagnosis of glaucoma and diabetic maculopathy. The developed optic disc and fovea detection approach based on multiscale CNNs allows to focus on the region of interest, removing redundant background data from consideration and facilitating refinement of the localisation. This results in

significantly increased accuracy in the cases of simultaneous fovea and optic disc detection in fundus images.

Important components of the eye's anatomical structure, including the choroid, and the optic disc and cup were segmented by developing new segmentation approaches. Our choroid segmentation framework achieved high quality automatic segmentation of the choroids boundaries in Enhanced Depth Imaging (EDI)-OCT images by making use of deep learning and clustering techniques. Improvements have been obtained by proposing a new super-pixel energy function combining image intensity, region and enhanced image information. It has been demonstrated that this allows us to obtain significantly improved results over alternative methods in the literature. Furthermore, by leveraging the combination of a fully convolutional network and DenseNet, high quality simultaneous OD and OC segmentation in retinal fundus images has been achieved. This proposed deep network allows improved segmentation, particularly of the OC. The calculation of the optic disc diameter (ODD) is used to crop the images to $2ODD$ and rescale, reducing the image to the region of interest, which reduces computation time without requiring excessive reduction of the image resolution.

In another set of experiments and by exploiting the transfer learning concept in different ways, three automated grading disease/lesion identification systems were proposed and implemented successfully. The adoption of the fixed feature extractor concept originating from transfer learning to automatically detect the presence of glaucoma in retinal fundus images demonstrated that the use of transfer learning is an effective strategy to detect glaucoma in situations where the data size is small for training. The automated method for grading the severity of DMO in fundus images based on developing multi-stage CNNs achieved promising results without the need of prior feature extraction, exudate and macula segmentation or the removal of retinal blood vessels. The developed retinal fluid identification framework including SRF, IRF and PED fluid identification from OCT volumes

exploited the concepts of transfer learning and ensemble learning and achieved promising identification results with an acceptable prediction ability.

6.2 Discussion

One of the most important advantages of the proposed OD and fovea detection method is that it is less sensitive to preprocessing. This is particularly useful in processing images of poor quality. Another important advantages of this proposed approach is that it does not require vessel segmentation or border localisation in order to detect the OD and foveal centres. Moreover, in proposed choroid segmentation method, incorporating distance information and enhanced intensity information with image intensities improved the CNN classification which produced excellent agreement with manual segmentation. Furthermore, the OD and cup segmentation approach achieved state-of-the-art results on a large dataset, outperforming the previous methods and we demonstrated the effectiveness of the method on other datasets without the need of re-training the model using images from those datasets. These findings also suggest feasible implementation of the two developed segmentation approaches in a wide range of medical image processing applications. The developed automated disease grading / lesion identification methods achieved promising results; however, they require further refinement and verification to be implemented in a clinical setting.

The solutions provided in the developed methods address the main weaknesses of traditional computer-aided systems where handcrafted features are considered key requirements. In contrast to conventional machine learning classifiers (shallow architectures) including neural networks and support vector machines, for which a feature extraction step is essential, hierarchies of significant features are learnt by deep CNNs directly from the raw input data. The obtained results demonstrated that the proposed feature learning methods by implementing multiple layers of

learning filters are superior to current approaches in extracting the features automatically from the data, and thus solved issues related to the workload in extracting the features manually from various types of datasets. Furthermore, the proposed methods have proven to be more robust, less time consuming, more generalisable, and most often more accurate. The significance of the obtained findings was therefore achieved by exploiting hierarchical feature representations learned solely from data, instead of potentially subjective hand-crafted features mostly designed based on domain-specific knowledge.

Although deep learning algorithms -especially CNNs- have dominated computer vision over the last few years, achieving top scores on many image analysis tasks, these algorithms still face many challenges. These algorithms are trained to learn progressively using data. Large datasets are needed to make sure that the machine delivers the desired results. Providing large sets of annotated and labelled medical data by human expert graders is a challenging task. Deep learning algorithms are also a quite resource-demanding approaches. They require more powerful GPUs, high-performance graphics processing units, and a large amount of storage to train the models. Furthermore, they need more time to train in comparison with traditional machine learning methods.

6.3 Future Work

There are many different directions that can be arising from the work presented in this thesis as future work.

1. **Further development of proposed systems:** The proposed fovea and optic disc detection methodology involves hyper-parameter settings such as regularisation strength, the initial learning rate, and schedule of learning rate decay which are set manually for training. Therefore, the proposed method could be further developed by introducing an automated hyper-parameter search method. Furthermore, a possible future direction for the proposed

fluid identification method is segmenting the identified fluids. The aim of segmentation after identification is to visualise these regions of fluid and obtain quantitative measurements of size. This can improve the understanding, diagnosis and prediction of retinal diseases.

2. **Evaluation/Validation:** Conducting an extensive validation to the proposed methods which would allow evaluating with images from many different devices and conditions, and from patients of different ethnicities, genders and ages. This would help to explore whether camera type and patients age, gender, and ethnicity influence the developed automated retinal image analysis systems output and performance. Furthermore, perform a clinical evaluation of the developed approaches would assess its feasibility in clinical ophthalmology.
3. **Extension the developed methods to other conditions/diseases:** It has been demonstrated that neuroretinal alterations are early in diabetes, preceding microvascular damage and it has been shown that a reduction of the ganglion cell-inner plexiform layer (GC-IPL) and retinal nerve fiber layer (RNFL) may occur in diabetic patients without any signs of DR. A possible future direction from this thesis is to extend choroid segmentation method to segment these layers automatically and subsequently measuring their thickness from OCT images acquired from different imaging equipment manufacturers.
4. **Development of new deep learning algorithms to identify other diseases:** The advent of SD-OCT has provided an understanding and recognition of the anatomical abnormalities seen in acute macular neuroretinopathy (AMNR). SD-OCT shows that the condition affects the outer retina and is associated with disruption of the inner/outer segment (IS-OS) junction. Another possible direction for future work can be taken into consideration by

designing an automated system based on deep learning to extract the disease's features from the SD-OCT images in order to detect the presence of the disease automatically.

5. **Deep learning development:** Addressing the limitations of deep learning algorithms including the need for larger labelled training data size and powerful computation/storage resources (GPUs/Memory). This will help to further improve the performance of the developed systems in terms of accuracy of detection/identification and reduce the execution time.

6.4 Conclusion

All in all, this work has successfully combined skills and expertise from different disciplines including ophthalmology, artificial intelligence and image processing to developing retinal image analysis and disease diagnosis methods. The developed retinal image analysis methods have great potential to not only provide less subjective diagnosis but also speed up the diagnosis process for a variety of retinal diseases including age-related macular degeneration, diabetic maculopathy and glaucoma and consequently for implementation in a clinical setting. This will hopefully help reduce the required number of primary care appointments, unnecessary referrals and patient waiting time, thus improving the patients quality of life and reducing cost in clinical ophthalmology. The presented deep learning-based approaches will also be valuable tools to be used for other challenging medical problems.

Bibliography

- [1] L. M. Jampol, J. Shankle, R. Schroeder, P. Tornambe, R. F. Spaide, and M. R. Hee. Diagnostic and therapeutic challenges. *Retina*, 26(9):1072–1076, 2006.
- [2] A. Tufail, C. Rudisill, C. Egan, V. V. Kapetanakis, S. Salas-Vega, C. G. Owen, A. Lee, V. Louw, J. Anderson, G. Liew, et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology*, 124(3):343–351, 2017.
- [3] M. Niemeijer, M. D. Abràmoff, and B. Van Ginneken. Fast detection of the optic disc and fovea in color fundus photographs. *Medical Image Analysis*, 13(6):859–870, 2009.
- [4] U. R. Acharya, S. Dua, X. Du, C. K. Chua, et al. Automated diagnosis of glaucoma using texture and higher order spectra features. *IEEE Transactions on Information Technology in Biomedicine*, 15(3):449–455, 2011.
- [5] U. R. Acharya, E. Ng, L. W. J. Eugene, K. P. Noronha, L. C. Min, K. P. Nayak, and S. V. Bhandary. Decision support system for the glaucoma using gabor transformation. *Biomedical Signal Processing and Control*, 15:18–26, 2015.
- [6] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, and E. Chaum. Exudate-based diabetic macular edema detection in fundus

- images using publicly available datasets. *Medical Image Analysis*, 16(1): 216–226, 2012.
- [7] A. A.-H. A.-R. Youssif, A. Z. Ghalwash, and A. A. S. A.-R. Ghoneim. Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter. *IEEE Transactions on Medical imaging*, 27(1):11–18, 2008.
- [8] A. Aquino. Establishing the macular grading grid by means of fovea centre detection using anatomical-based and visual-based features. *Computers in Biology and Medicine*, 55:61–73, 2014.
- [9] E. Soto-Pedre, A. Navea, S. Millan, M. C. Hernaez-Ortega, J. Morales, M. C. Desco, and P. Pérez. Evaluation of automated image analysis software for the detection of diabetic retinopathy to reduce the ophthalmologists' workload. *Acta Ophthalmologica*, 93(1), 2015.
- [10] D. Veiga, C. Pereira, M. Ferreira, L. Gonçalves, and J. Monteiro. Quality evaluation of digital fundus images through combined measures. *Journal of Medical Imaging*, 1(1):014001–014001, 2014.
- [11] U. Akram. Retinal image preprocessing: background and noise segmentation. *Indonesian Journal of Electrical Engineering and Computer Science*, 10(3):537–544, 2012.
- [12] Eye anatomy: parts of the eye. <http://www.allaboutvision.com/resources/anatomy.htm>. Accessed: 2017-09-22.
- [13] B. Cassin and S. Solomon. *Dictionary of Eye Terminology*. Wiley Online Library, 1990.
- [14] C. W. Oyster. *The human eye: structure and function*. Sinauer Associates, 1999.

- [15] M. R. K. Mookiah, U. R. Acharya, H. Fujita, J. H. Tan, C. K. Chua, S. V. Bhandary, A. Laude, and L. Tong. Application of different imaging modalities for diagnosis of diabetic macular edema: a review. *Computers in Biology and Medicine*, 66:295–315, 2015.
- [16] Age-related macular degeneration (AMD). <https://www.nhs.uk/conditions/age-related-macular-degeneration-amd/>. Accessed: 2017-11-03.
- [17] Diabetes. <http://www.eyedocsottawa.com/services-special-vision/special-vision-concerns/diabetes/>. Accessed: 2017-11-03.
- [18] G. S. Hageman, K. Gehrs, L. V. Johnson, and D. Anderson. *Age-related macular degeneration (AMD)*. University of Utah Health Sciences Centre, 2008.
- [19] Retinal fluids. <https://retouch.grand-challenge.org/background/>. Accessed: 2017-09-22.
- [20] F. C. DeCroos, C. A. Toth, S. S. Stinnett, C. S. Heydary, R. Burns, G. J. Jaffe, C. R. Group, et al. Optical coherence tomography grading reproducibility during the comparison of age-related macular degeneration treatments trials. *Ophthalmology*, 119(12):2549–2557, 2012.
- [21] H. A. Quigley and A. T. Broman. The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology*, 90(3):262–267, 2006.
- [22] B. D. Tay Xue Chao. Glaucoma. <https://www.fastbleep.com/biology-notes/20/288>. Accessed: 2017-11-06.
- [23] T. Damms and F. Dannheim. Sensitivity and specificity of optic disc parameters in chronic glaucoma. *Investigative Ophthalmology & Visual Science*, 34(7):2246–2250, 1993.

- [24] P. S. Mittapalli and G. B. Kande. Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma. *Biomedical Signal Processing and Control*, 24:34–46, 2016.
- [25] E. T. D. R. S. R. Group et al. Treatment techniques and clinical guidelines for photocoagulation of diabetic macular edema: Early treatment diabetic retinopathy study report number 2. *Ophthalmology*, 94(7):761–774, 1987.
- [26] F. Ghanchi, D. R. G. W. Group, et al. The royal college of ophthalmologists’ clinical guidelines for diabetic retinopathy: a summary. *Eye*, 27(2):285, 2013.
- [27] R. Simó and C. Hernández. Neurodegeneration in the diabetic eye: new insights and therapeutic perspectives. *Trends in Endocrinology & Metabolism*, 25(1):23–33, 2014.
- [28] A. J. Barber. Diabetic retinopathy: recent advances towards understanding neurodegeneration and vision loss. *Science China Life Sciences*, 58(6):541–549, 2015.
- [29] A. R. Santos, L. Ribeiro, F. Bandello, R. Lattanzio, C. Egan, U. Frydkjaer-Olsen, J. García-Arumí, J. Gibson, J. Grauslund, S. P. Harding, et al. Functional and structural findings of neurodegeneration in early stages of diabetic retinopathy: cross-sectional analyses of baseline data of the eurocondor project. *Diabetes*, 66(9):2503–2510, 2017.
- [30] S. K. Lynch and M. D. Abràmoff. Diabetic retinopathy is a neurodegenerative disorder. *Vision research*, 139:101–107, 2017.
- [31] I. Klaassen, C. J. Van Noorden, and R. O. Schlingemann. Molecular basis of the inner blood-retinal barrier and its breakdown in diabetic macular edema and other pathological conditions. *Progress in Retinal and Eye Research*, 34:19–48, 2013.

- [32] N. Bhagat, R. A. Grigorian, A. Tutela, and M. A. Zarbin. Diabetic macular edema: pathogenesis and treatment. *Survey of Ophthalmology*, 54(1):1–32, 2009.
- [33] P. J. Saine and M. E. Tyler. *Ophthalmic photography: retinal photography, angiography, and electronic imaging*, volume 132. Butterworth-Heinemann Boston, 2002.
- [34] F. C. Delori, E. S. Gragoudas, R. Francisco, and R. C. Pruett. Monochromatic ophthalmoscopy and fundus photography: the normal fundus. *Archives of Ophthalmology*, 95(5):861–868, 1977.
- [35] W. Drexler and J. G. Fujimoto. State-of-the-art retinal optical coherence tomography. *Progress in Retinal and Eye Research*, 27(1):45–88, 2008.
- [36] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, et al. Optical coherence tomography. *Science (New York, NY)*, 254(5035):1178, 1991.
- [37] C. K. Hitzenberger, P. Trost, P.-W. Lo, and Q. Zhou. Three-dimensional imaging of the human retina by high-speed optical coherence tomography. *Optics Express*, 11(21):2753–2761, 2003.
- [38] J. Yang, L. Liu, T. Jiang, and Y. Fan. A modified gabor filter design method for fingerprint image enhancement. *Pattern Recognition Letters*, 24(12):1805–1817, 2003.
- [39] F. Hossain and M. R. Alsharif. Image enhancement based on logarithmic transform coefficient and adaptive histogram equalization. In *Convergence Information Technology, 2007. International Conference on*, pages 1439–1444. IEEE, 2007.

- [40] S. Thapar and S. Garg. Study and implementation of various morphology based image contrast enhancement techniques. *International Journal of Computing Business Research*, 2012.
- [41] S. Vimal and P. Thiruvikraman. Automated image enhancement using power law transformations. *Sadhana*, 37(6):739–745, 2012.
- [42] S. Kim, W. Kang, E. Lee, and J. Paik. Wavelet-domain color image enhancement using filtered directional bases and frequency-adaptive shrinkage. *IEEE Transactions on Consumer Electronics*, 56(2), 2010.
- [43] K. Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc., 1994.
- [44] P. Patidar, M. Gupta, S. Srivastava, and A. K. Nagawat. Image de-noising by various filters for different noise. *International journal of computer applications*, 9(4), 2010.
- [45] J. C. Church, Y. Chen, and S. V. Rice. A spatial median filter for noise removal in digital images. In *Southeastcon, 2008. IEEE*, pages 618–623. IEEE, 2008.
- [46] S. Zhong and V. Cherkassky. Image denoising using wavelet thresholding and model selection. In *International Conference on Image Processing*,, volume 3, pages 262–265. IEEE, 2000.
- [47] B. M. Williams, J. Zhang, and K. Chen. A new image deconvolution method with fractional regularisation. *Journal of Algorithms & Computational Technology*, 10(4):265–276, 2016.
- [48] J. Zhang and K. Chen. A total fractional-order variation model for image restoration with nonhomogeneous boundary conditions and its numerical solution. *SIAM Journal on Imaging Sciences*, 8(4):2487–2518, 2015.

- [49] B. R. Masters, R. C. Gonzalez, and R. Woods. Digital image processing. *Journal of Biomedical Optics*, 14(2):029901, 2009.
- [50] M.-O. Baradez, C. P. McGuckin, N. Forraz, R. Pettengell, and A. Hoppe. Robust and automated unimodal histogram thresholding and potential applications. *Pattern Recognition*, 37(6):1131–1148, 2004.
- [51] P. Ganesan, V. Rajini, and R. I. Rajkumar. Segmentation and edge detection of color images using CIELAB color space and edge detectors. In *Emerging Trends in Robotics and Communication Technologies (INTERACT), 2010 International Conference on*, pages 393–397. IEEE, 2010.
- [52] H. G. Kaganami and Z. Bei. Region-based segmentation versus edge detection. In *Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP'09. Fifth International Conference on*, pages 1217–1221. IEEE, 2009.
- [53] Y. W. Lim and S. U. Lee. On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. *Pattern Recognition*, 23(9):935–952, 1990.
- [54] O. Rotem, H. Greenspan, and J. Goldberger. Combining region and edge cues for image segmentation in a probabilistic gaussian mixture framework. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.
- [55] A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in Neural Information Processing Systems*, pages 841–848, 2002.
- [56] L. Deng. Three classes of deep learning architectures and their applications: a tutorial survey. *APSIPA Transactions on Signal and Information Processing*, 2012.

- [57] Y. Bengio et al. Learning deep architectures for AI. *Foundations and Trends® in Machine Learning*, 2(1):1–127, 2009.
- [58] L. Deng, D. Yu, et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4):197–387, 2014.
- [59] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [60] M. T. Jones. Deep learning architectures, the rise of artificial intelligence. Technical report, IBM, 09 2017. 2017-09-08.
- [61] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh. *Feature extraction: foundations and applications*, volume 207. Springer, 2008.
- [62] Jagreet. Log analytics with deep learning and machine learning. <https://www.xenonstack.com/blog/log-analytics-with-deep-learning-and-machine-learning>. Accessed: 2017-11-1.
- [63] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.
- [64] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006.
- [65] R. Salakhutdinov and G. Hinton. Deep boltzmann machines. In *Artificial Intelligence and Statistics*, pages 448–455, 2009.
- [66] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

- [67] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.
- [68] H. Sak, A. Senior, and F. Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.
- [69] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.
- [70] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(10):1533–1545, 2014.
- [71] S. Lai, L. Xu, K. Liu, and J. Zhao. Recurrent convolutional neural networks for text classification. In *AAAI*, volume 333, pages 2267–2273, 2015.
- [72] J. Gauthier. Conditional generative adversarial nets for convolutional face generation. *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester*, 2014(5):2, 2014.
- [73] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2625–2634, 2015.
- [74] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2016.

- [75] A. Dosovitskiy and T. Brox. Generating images with perceptual similarity metrics based on deep networks. In *Advances in Neural Information Processing Systems*, pages 658–666, 2016.
- [76] W. Liu, R. Ji, and S. Li. Towards 3D object detection with bimodal deep boltzmann machines over RGBD imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3013–3021, 2015.
- [77] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [78] P. Velickovic. Deep learning for complete beginners: convolutional neural networks with keras. <https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html>. Accessed: 2017-11-1.
- [79] F. J. Huang, Y.-L. Boureau, Y. LeCun, et al. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.
- [80] X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.
- [81] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015.

- [82] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [83] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*, pages 818–833. Springer, 2014.
- [84] A. M. Saxe, J. L. McClelland, and S. Ganguli. Exact solutions to the non-linear dynamics of learning in deep linear neural networks. *arXiv preprint arXiv:1312.6120*, 2013.
- [85] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.
- [86] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1026–1034, 2015.
- [87] A. Perunovic. Understanding neural network weight initialization. <https://intoli.com/blog/neural-network-initialization/>, 2017. Accessed: 2017-10-26.
- [88] G. Gwardys. Convolutional neural networks backpropagation: from intuition to derivation. <https://grzegorzwardys.wordpress.com/2016/04/22/8/>, 2016. Accessed: 2017-10-26.
- [89] D. E. Rumelhart, G. E. Hinton, R. J. Williams, et al. Learning representations by back-propagating errors. *Cognitive Modeling*, 5(3):1, 1988.
- [90] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011.

- [91] M. D. Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.
- [92] G. Hinton, N. Srivastava, and K. Swersky. Lecture 6a overview of mini-batch gradient descent. *Coursera Lecture slides <https://class.coursera.org/neuralnets-2012-001/lecture>*, Online, 2012.
- [93] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [94] J. Wang and L. Perez. The effectiveness of data augmentation in image classification using deep learning. 2017.
- [95] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, pages 3320–3328, 2014.
- [96] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.
- [97] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- [98] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley, and Y. Bengio. Theano: new features and speed improvements. *arXiv preprint arXiv:1211.5590*, 2012.
- [99] R. Collobert, K. Kavukcuoglu, and C. Farabet. Torch7: A matlab-like environment for machine learning. In *BigLearn, NIPS Workshop*, number EPFL-CONF-192376, 2011.

- [100] Bring the power of deep learning to your data. <http://www.nvidia.co.uk/object/gpu-computing-uk.html>. Accessed: 2017-10-19.
- [101] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber. Deep big simple neural nets excel on handwritten digit recognition. 80, 2010. <http://arxiv.org/abs/1003.0358>.
- [102] M. Lin, Q. Chen, and S. Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.
- [103] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. <https://arxiv.org/abs/1409.1556>.
- [104] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.
- [105] M. Jaderberg, K. Simonyan, A. Zisserman, et al. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2015.
- [106] O. Rippel, J. Snoek, and R. P. Adams. Spectral representations for convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 2449–2457, 2015.
- [107] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

- [108] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- [109] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147*, 2016.
- [110] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [111] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [112] A. Canziani, A. Paszke, and E. Culurciello. An analysis of deep neural network models for practical applications. *arXiv preprint arXiv:1605.07678*, 2016. <https://arxiv.org/abs/1605.07678>.
- [113] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3): 211–252, 2015.
- [114] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014.

- [115] J. L. Long, N. Zhang, and T. Darrell. Do convnets learn correspondence? In *Advances in Neural Information Processing Systems*, pages 1601–1609, 2014.
- [116] S. Li and A. B. Chan. 3D human pose estimation from monocular images with deep convolutional neural network. In *Asian Conference on Computer Vision*, pages 332–347. Springer, 2014.
- [117] G. Levi and T. Hassner. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 34–42, 2015.
- [118] M. Thoma. Analysis and optimization of convolutional neural network architectures. Masters thesis, Karlsruhe Institute of Technology, Karlsruhe, Germany, June 2017. URL <https://martin-thoma.com/msthesis/>.
- [119] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis. *arXiv preprint arXiv:1702.05747*, 2017.
- [120] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng. Automatic detection of cerebral microbleeds from mr images via 3D convolutional neural networks. *IEEE Transactions on Medical Imaging*, 35(5):1182–1195, 2016.
- [121] F. C. Ghesu, B. Georgescu, T. Mansi, D. Neumann, J. Hornegger, and D. Comaniciu. An artificial agent for anatomical landmark detection in medical images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 229–237. Springer, 2016.

- [122] X. Lu, D. Xu, and D. Liu. Robust 3D organ localization with dual learning architectures and fusion. In *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 12–20. Springer, 2016.
- [123] H. Ravishankar, P. Sudhakar, R. Venkataramani, S. Thiruvenkadam, P. Anangi, N. Babu, and V. Vaidya. Understanding the mechanisms of deep transfer learning for medical images. In *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 188–196. Springer, 2016.
- [124] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in Neural Information Processing Systems*, pages 2843–2851, 2012.
- [125] C. Cernazanu-Glavan and S. Holban. Segmentation of bone structure in x-ray images using convolutional neural network. *Advances in Electrical and Computer Engineering*, 13(1):87–94, 2013.
- [126] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, 108:214–224, 2015.
- [127] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 132–139. Springer, 2016.
- [128] J. Zilly, J. M. Buhmann, and D. Mahapatra. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Computerized Medical Imaging and Graphics*, 55:28–41, 2017.

- [129] C. D. Malon and E. Cosatto. Classification of mitotic figures with convolutional neural networks and seeded blob features. *Journal of Pathology Informatics*, 4, 2013.
- [130] A. Payan and G. Montana. Predicting alzheimer’s disease: a neuroimaging study with 3D convolutional neural networks. *arXiv preprint arXiv:1502.02506*, 2015.
- [131] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Transactions on Medical Imaging*, 35(5):1207–1216, 2016.
- [132] J. Antony, K. McGuinness, N. E. O’Connor, and K. Moran. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*, pages 1195–1200. IEEE, 2016.
- [133] J. Wang, H. Ding, F. Azamian, B. Zhou, C. Iribarren, S. Molloy, and P. Baldi. Detecting cardiovascular disease from mammograms with deep learning. *IEEE Transactions on Medical Imaging*, 2017.
- [134] L. Gagnon, M. Lalonde, M. Beaulieu, and M.-C. Boucher. Procedure to detect anatomical structures in optical fundus images. In *Proceedings of SPIE*, volume 4322, pages 1218–1225, 2001.
- [135] H. W. Larsen. *The ocular fundus: a color atlas*. WB Saunders Company, 1976.
- [136] J. Schwiegerling. Field guide to visual and ophthalmic optics. Spie, 2004.
- [137] A. Hoover and M. Goldbaum. Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Transactions on Medical Imaging*, 22(8):951–958, 2003.

- [138] M. Foracchia, E. Grisan, and A. Ruggeri. Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *IEEE Transactions on Medical Imaging*, 23(10):1189–1195, 2004.
- [139] A. D. Fleming, K. A. Goatman, S. Philip, J. A. Olson, and P. F. Sharp. Automatic detection of retinal anatomy to assist diabetic retinopathy screening. *Physics in Medicine and Biology*, 52(2):331, 2006.
- [140] K. W. Tobin, E. Chaum, V. P. Govindasamy, and T. P. Karnowski. Detection of anatomic structures in human retinal imagery. *IEEE Transactions on Medical Imaging*, 26(12):1729–1739, 2007.
- [141] X. Zhu, R. M. Rangayyan, and A. L. Ells. Detection of the optic nerve head in fundus images of the retina using the hough transform for circles. *Journal of Digital Imaging*, 23(3):332–341, 2010.
- [142] S. Lu. Accurate and efficient optic disc detection and segmentation by a circular transformation. *IEEE Transactions on Medical Imaging*, 30(12):2126–2133, 2011.
- [143] H. Yu, E. S. Barriga, C. Agurto, S. Echegaray, M. S. Pattichis, W. Bauman, and P. Soliz. Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets. *IEEE Transactions on Information Technology in Biomedicine*, 16(4):644–657, 2012.
- [144] A. Dehghani, H. A. Moghaddam, and M.-S. Moin. Optic disc localization in retinal images using histogram matching. *EURASIP Journal on Image and Video Processing*, 2012(1):19, 2012.
- [145] B. Harangi and A. Hajdu. Detection of the optic disc in fundus images by combining probability models. *Computers in Biology and Medicine*, 65:10–24, 2015.

- [146] H. Li and O. Chutatape. Automated feature extraction in color retinal images by a model based approach. *IEEE Transactions on Biomedical Engineering*, 51(2):246–254, 2004.
- [147] M. Niemeijer, M. D. Abramoff, and B. Van Ginneken. Segmentation of the optic disc, macula and vascular arch in fundus photographs. *IEEE Transactions on Medical Imaging*, 26(1):116–127, 2007.
- [148] D. Welfer, J. Scharcanski, and D. R. Marinho. Fovea center detection based on the retina anatomy and mathematical morphology. *Computer Methods and Programs in Biomedicine*, 104(3):397–409, 2011.
- [149] R. J. Qureshi, L. Kovacs, B. Harangi, B. Nagy, T. Peto, and A. Hajdu. Combining algorithms for automatic detection of optic disc and macula in fundus images. *Computer Vision and Image Understanding*, 116(1):138–145, 2012.
- [150] A. Giachetti, L. Ballerini, E. Trucco, and P. J. Wilson. The use of radial symmetry to localize retinal landmarks. *Computerized Medical Imaging and Graphics*, 37(5):369–376, 2013.
- [151] M. E. Gegundez-Arias, D. Marin, J. M. Bravo, and A. Suero. Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques. *Computerized Medical Imaging and Graphics*, 37(5):386–393, 2013.
- [152] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000.
- [153] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.

- [154] Y. Zheng, M. H. A. Hijazi, and F. Coenen. Automated disease/no disease grading of age-related macular degeneration by an image mining approach. *Investigative Ophthalmology & Visual Science*, 53(13):8310–8318, 2012.
- [155] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Ranninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä. The diaretdb1 diabetic retinopathy database and evaluation protocol. In *BMVC*, pages 1–10, 2007.
- [156] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, H. Uusitalo, H. Kälviäinen, and J. Pietilä. Diaretdb0: Evaluation database and methodology for diabetic retinopathy algorithms. *Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology, Finland*, 73, 2006.
- [157] E. Decencière, X. Zhang, G. Cazuguel, B. Laÿ, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, et al. Feedback on a publicly distributed image database: the messidor database. *Image Analysis & Stereology*, 33(3):231–234, 2014.
- [158] Diabetic retinopathy detection. <https://www.kaggle.com/c/diabetic-retinopathy-detection>. Accessed: 2016-10-08.
- [159] J. Snoek, O. Rippel, K. Swersky, R. Kiros, N. Satish, N. Sundaram, M. Patwary, M. Prabhat, and R. Adams. Scalable bayesian optimization using deep neural networks. In *International Conference on Machine Learning*, pages 2171–2180, 2015.
- [160] G. Montavon, G. B. Orr, and K.-R. Müller. Tricks of the trade. 2012.
- [161] R. F. Spaide, H. Koizumi, and M. C. Pozonni. Enhanced depth imaging spectral-domain optical coherence tomography. *American journal of Ophthalmology*, 146(4):496–500, 2008.

- [162] L. Zhang, K. Lee, M. Niemeijer, R. F. Mullins, M. Sonka, and M. D. Abramoff. Automated segmentation of the choroid from clinical SD-OCT automated segmentation of choroid from SD-OCT. *Investigative Ophthalmology and Visual Science*, 53(12):7510–7519, 2012.
- [163] V. Kajić, M. Esmaeelpour, B. Považay, D. Marshall, P. L. Rosin, and W. Drexler. Automated choroidal segmentation of 1060 nm OCT in healthy and pathologic eyes using a statistical model. *Biomedical Optics Express*, 3(1):86–103, 2012.
- [164] T. Torzicky, M. Pircher, S. Zotter, M. Bonesi, E. Götzinger, and C. K. Hitzenberger. Automated measurement of choroidal thickness in the human eye by polarization sensitive optical coherence tomography. *Optics Express*, 20(7):7564–7574, 2012.
- [165] L. Duan, M. Yamanari, and Y. Yasuno. Automated phase retardation oriented segmentation of chorio-scleral interface by polarization sensitive optical coherence tomography. *Optics Express*, 20(3):3353–3366, 2012.
- [166] J. Tian, P. Marziliano, M. Baskaran, T. A. Tun, and T. Aung. Automatic segmentation of the choroid in enhanced depth imaging optical coherence tomography images. *Biomedical Optics Express*, 4(3):397–411, 2013.
- [167] H. Lu, N. Boonarpa, M. T. Kwong, and Y. Zheng. Automated segmentation of the choroid in retinal optical coherence tomography images. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 5869–5872. IEEE, 2013.
- [168] D. Alonso-Caneiro, S. A. Read, and M. J. Collins. Automatic segmentation of choroidal thickness in optical coherence tomography. *Biomedical Optics Express*, 4(12):2795–2812, 2013.

- [169] H. Danesh, R. Kafieh, H. Rabbani, and F. Hajizadeh. Segmentation of choroidal boundary in enhanced depth imaging OCTs using a multiresolution texture based modeling in graph cuts. *Computational and Mathematical Methods in Medicine*, 2014, 2014.
- [170] X. Liu, X. Jiang, Y. Sun, A. Xu, and H. Lu. Inter-frame information based choroid segmentation in enhanced optical coherence tomography images. In *Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference on*, pages 211–216. IEEE, 2014.
- [171] A. González-López, B. Remeseiro, M. Ortega, M. G. Penedo, and P. Charlón. A texture-based method for choroid segmentation in retinal EDI-OCT images. In *International Conference on Computer Aided Systems Theory*, pages 487–493. Springer, 2015.
- [172] L. Zhang, G. H. Buitendijk, K. Lee, M. Sonka, H. Springelkamp, A. Hofman, J. R. Vingerling, R. F. Mullins, C. C. Klaver, and M. D. Abramoff. Validity of automated choroidal segmentation in SS-OCT and SD-OCT. *Investigative Ophthalmology and Visual Science*, 56(5):3202–3211, 2015.
- [173] Q. Chen, W. Fan, S. Niu, J. Shi, H. Shen, and S. Yuan. Automated choroid segmentation based on gradual intensity distance in HD-OCT images. *Optics Express*, 23(7):8974–8994, 2015.
- [174] Q. Chen, S. Niu, S. Yuan, W. Fan, and Q. Liu. Choroidal vasculature characteristics based choroid segmentation for enhanced depth imaging optical coherence tomography images. *Medical Physics*, 43(4):1649–1661, 2016.
- [175] Z. Hu, C. A. Girkin, A. Hariri, and S. R. Sadda. Three-dimensional choroidal segmentation in spectral OCT volumes using optic disc prior information. In *Proceeding of SPIE Vol*, volume 9697, pages 96971S–1, 2016.

- [176] I. Oguz, M. D. Abramoff, L. Zhang, K. Lee, E. Z. Zhang, and M. Sonka. 4D graph-based segmentation for reproducible and sensitive choroid quantification from longitudinal oct scans. *Investigative Ophthalmology and Visual Science*, 57(9):OCT621–OCT630, 2016.
- [177] M. D. Twa, K. L. Schulle, S. J. Chiu, S. Farsiu, and D. A. Berntsen. Validation of macular choroidal thickness measurements from automated SD-OCT image segmentation. *Optometry and Vision Science*, 93(11):1387–1398, 2016.
- [178] F. Shi, B. Tian, W. Zhu, D. Xiang, L. Zhou, H. Xu, and X. Chen. Automated choroid segmentation in three-dimensional 1- μ m wide-view OCT images with gradient and regional costs. *Journal of Biomedical Optics*, 21(12):126017–126017, 2016.
- [179] C. Wang, Y. Li, and Y. X. Wang. Automatic choroidal layer segmentation using markov random field and level set method. *IEEE Journal of Biomedical and Health Informatics*, 2017.
- [180] J. Mazzaferri, L. Beaton, G. Hounye, D. N. Sayah, and S. Costantino. Open-source algorithm for automatic choroid segmentation of OCT volume reconstructions. *Scientific Reports*, 7, 2017.
- [181] X. Sui, Y. Zheng, B. Wei, H. Bi, J. Wu, X. Pan, Y. Yin, and S. Zhang. Choroid segmentation from optical coherence tomography with graph-edge weights learned from deep convolutional neural networks. *Neurocomputing*, 237:332–341, 2017.
- [182] M. Chen, J. Wang, I. Oguz, B. L. VanderBeek, and J. C. Gee. Automated segmentation of the choroid in EDI-OCT images with retinal pathology using convolution neural networks. In *Proceedings of the Fetal, Infant and Ophthalmic Medical Image Analysis*, pages 177–184, 2017.

- [183] H. Lin, Y. Dong, D. Markl, B. M. Williams, Y. Zheng, Y. Shen, and J. A. Zeitler. Measurement of the intertablet coating uniformity of a pharmaceutical pan coating process with combined terahertz and optical coherence tomography in-line sensing. *Journal of Pharmaceutical Sciences*, 106(4): 1075–1084, 2017.
- [184] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.
- [185] M. Deriche, A. Amin, and M. Qureshi. Color image segmentation by combining the convex active contour and the chan vese model. *Pattern Analysis and Applications*, pages 1–15.
- [186] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- [187] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001. doi: 10.1109/83.902291.
- [188] N. A. M. Isa, S. A. Salamah, and U. K. Ngah. Adaptive fuzzy moving k-means clustering algorithm for image segmentation. *IEEE Transactions on Consumer Electronics*, 55(4), 2009.
- [189] P. Pallawala, W. Hsu, M. L. Lee, and K.-G. A. Eong. Automated optic disc localization and contour detection using ellipse fitting and wavelet transform. In *European Conference on Computer Vision*, pages 139–151. Springer, 2004.
- [190] X. Zhu and R. M. Rangayyan. Detection of the optic disc in images of the retina using the hough transform. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 3546–3549. IEEE, 2008.

- [191] A. Aquino, M. E. Gegúndez-Arias, and D. Marín. Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE Transactions on Medical Imaging*, 29(11):1860–1869, 2010.
- [192] A. Giachetti, L. Ballerini, and E. Trucco. Accurate and reliable segmentation of the optic disc in digital fundus images. *Journal of Medical Imaging*, 1(2):024001–024001, 2014.
- [193] B. Dashtbozorg, A. M. Mendonça, and A. Campilho. Optic disc segmentation using the sliding band filter. *Computers in Biology and Medicine*, 56:1–12, 2015.
- [194] A. Almazroa, S. Alodhayb, K. Raahemifar, and V. Lakshminarayanan. Optic cup segmentation: type-ii fuzzy thresholding approach and blood vessel extraction. *Clinical Ophthalmology*, 11:841, 2017.
- [195] J. Sigut, O. Nunez, F. Fumero, M. Gonzalez, and R. Arnay. Contrast based circular approximation for accurate and robust optic disc segmentation in retinal images. *PeerJ*, 5:e3763, 2017.
- [196] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy. Optic nerve head segmentation. *IEEE Transactions on Medical Imaging*, 23(2):256–264, 2004.
- [197] J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. C. T. Kuan. Optic disk feature extraction via modified deformable model technique for glaucoma analysis. *Pattern Recognition*, 40(7):2063–2076, 2007.
- [198] A. R. Hussain. Optic nerve head segmentation using genetic active contours. In *Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on*, pages 783–787. IEEE, 2008.

- [199] G. D. Joshi, J. Sivaswamy, and S. Krishnadas. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE Transactions on Medical Imaging*, 30(6):1192–1205, 2011.
- [200] Y. Zheng, D. Stambolian, J. OBrien, and J. C. Gee. Optic disc and cup segmentation from color fundus photograph using graph cut with priors. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 75–82. Springer, 2013.
- [201] M. C. V. S. Mary, E. B. Rajsingh, J. K. K. Jacob, D. Anandhi, U. Amato, and S. E. Selvan. An empirical study on optic disc segmentation using an active contour model. *Biomedical Signal Processing and Control*, 18:19–29, 2015.
- [202] R. Arnay, F. Fumero, and J. Sigut. Ant colony optimization-based method for optic cup segmentation in retinal images. *Applied Soft Computing*, 52:409–417, 2017.
- [203] M. D. Abramoff, W. L. Alward, E. C. Greenlee, L. Shuba, C. Y. Kim, J. H. Fingert, and Y. H. Kwon. Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. *Investigative Ophthalmology & Visual Science*, 48(4):1665–1673, 2007.
- [204] D. W. K. Wong, J. Liu, N. M. Tan, F. Yin, B.-H. Lee, and T. Y. Wong. Learning-based approach for the automatic detection of the optic disc in digital retinal fundus photographs. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 5355–5358. IEEE, 2010.
- [205] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, and T. Y. Wong. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Transactions on Medical Imaging*, 32(6):1019–1032, 2013.

- [206] Y. Xu, L. Duan, S. Lin, X. Chen, D. W. K. Wong, T. Y. Wong, and J. Liu. Optic cup segmentation for glaucoma detection using low-rank superpixel representation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 788–795. Springer, 2014.
- [207] N.-M. Tan, Y. Xu, W. B. Goh, and J. Liu. Robust multi-scale superpixel classification for optic cup localization. *Computerized Medical Imaging and Graphics*, 40:182–193, 2015.
- [208] S. Roychowdhury, D. D. Koozekanani, S. N. Kuchinka, and K. K. Parhi. Optic disc boundary and vessel origin segmentation of fundus images. *IEEE Journal of Biomedical and Health Informatics*, 20(6):1562–1574, 2016.
- [209] K. Akyol, B. Şen, and Ş. Bayır. Automatic detection of optic disc in retinal image by using keypoint detection, texture analysis, and visual dictionary techniques. *Computational and Mathematical Methods in Medicine*, 2016, 2016.
- [210] F. Girard, C. Kavalec, S. Grenier, H. B. Tahar, and F. Cheriet. Simultaneous macula detection and optic disc boundary segmentation in retinal fundus images. In *Medical Imaging: Image Processing*, page 97841F, 2016.
- [211] S. Sedai, P. K. Roy, D. Mahapatra, and R. Garnavi. Segmentation of optic disc and optic cup in retinal fundus images using shape regression. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 3260–3264. IEEE, 2016.
- [212] G. Lim, Y. Cheng, W. Hsu, and M. L. Lee. Integrated optic disc and cup segmentation with deep learning. In *Tools with Artificial Intelligence (IC-TAI), 2015 IEEE 27th International Conference on*, pages 162–169. IEEE, 2015.

- [213] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool. Deep retinal image understanding. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 140–148. Springer, 2016.
- [214] Y. Guo, B. Zou, Z. Chen, Q. He, Q. Liu, and R. Zhao. Optic cup segmentation using large pixel patch based CNNs. In *Proceedings of the Ophthalmic Image Analysis Workshop*, pages 129–136. University of Iowa, 2016.
- [215] A. Sevastopolsky. Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. *arXiv preprint arXiv:1704.00979*, 2017.
- [216] S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam. Joint optic disc and cup segmentation using fully convolutional and adversarial networks. In *Fetal, Infant and Ophthalmic Medical Image Analysis*, pages 168–176. Springer, 2017.
- [217] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, and S. Sivaprasad. Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. *Journal of Computational Science*, 20:70–79, 2017.
- [218] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *arXiv preprint arXiv:1801.00926*, 2018.
- [219] S. Jégou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1175–1183. IEEE, 2017.
- [220] Z. Zhang, F. S. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, J. Cheng, and T. Y. Wong. Origa-light: An online retinal fundus image database for glaucoma analysis and research. In *Engineering in Medicine and Biology*

- Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 3065–3068. IEEE, 2010.
- [221] E. J. Carmona, M. Rincón, J. García-Feijoó, and J. M. Martínez-de-la Casa. Identification of the optic nerve head with genetic algorithms. *Artificial Intelligence in Medicine*, 43(3):243–259, 2008.
- [222] J. Sivaswamy, S. Krishnadas, G. D. Joshi, M. Jain, and A. U. S. Tabish. Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. In *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, pages 53–56. IEEE, 2014.
- [223] F. Fumero, S. Alayón, J. Sanchez, J. Sigut, and M. Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. In *Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on*, pages 1–6. IEEE, 2011.
- [224] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, L. C. Min, E. Y.-K. Ng, M. M. Mushrif, and A. Laude. Automated detection of optic disk in retinal fundus images using intuitionistic fuzzy histon segmentation. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 227(1):37–49, 2013.
- [225] A. Basit and M. M. Fraz. Optic disc detection and boundary extraction in retinal images. *Applied optics*, 54(11):3440–3447, 2015.
- [226] C. Wang and D. Kaba. Level set segmentation of optic discs from retinal images. *Journal of Medical and Bioengineering*, 4(3):213–220, 2015.
- [227] G. Hamednejad and H. Pourghassem. Retinal optic disk segmentation and analysis in fundus images using dbscan clustering algorithm. In *Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical*

- Engineering (ICBME), 2016 23rd Iranian Conference on*, pages 122–127. IEEE, 2016.
- [228] M. Abdullah, M. M. Fraz, and S. A. Barman. Localization and segmentation of optic disc in retinal images using circular hough transform and grow-cut algorithm. *PeerJ*, 4:e2003, 2016.
- [229] M. N. Zahoor and M. M. Fraz. Fast optic disc segmentation in retina using polar transform. *IEEE Access*, 5:12293–12300, 2017.
- [230] Y. Hatanaka, Y. Nagahata, C. Muramatsu, S. Okumura, K. Ogohara, A. Sawada, K. Ishida, T. Yamamoto, and H. Fujita. Improved automated optic cup segmentation based on detection of blood vessel bends in retinal fundus images. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 126–129. IEEE, 2014.
- [231] N. Noor, N. Khalid, and N. Ariff. Optic cup and disc color channel multi-thresholding segmentation. In *Control System, Computing and Engineering (ICCSC), 2013 IEEE International Conference on*, pages 530–534. IEEE, 2013.
- [232] N. E. A. Khalid, N. M. Noor, and N. M. Ariff. Fuzzy c-means (FCM) for optic cup and disc segmentation with morphological operation. *Procedia Computer Science*, 42:255–262, 2014.
- [233] F. Yin, J. Liu, D. W. Wong, N. M. Tan, J. Cheng, C.-Y. Cheng, Y. C. Tham, and T. Y. Wong. Sector-based optic cup segmentation with intensity and blood vessel priors. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 1454–1457. IEEE, 2012.

- [234] F. Yin, J. Liu, D. W. K. Wong, N. M. Tan, C. Cheung, M. Baskaran, T. Aung, and T. Y. Wong. Automated segmentation of optic disc and optic cup in fundus images for glaucoma diagnosis. In *Computer-based Medical Systems (CBMS), 2012 25th International Symposium on*, pages 1–6. IEEE, 2012.
- [235] S. Nawaldgi and Y. Lalitha. A novel combined color channel and isnt rule based automatic glaucoma detection from color fundus images. *Indian Journal of Science and Technology*, 10(13), 2017.
- [236] K. W. Oktoeberza, H. A. Nugroho, and T. B. Adji. Optic disc segmentation based on red channel retinal fundus images. In *International Conference on Soft Computing, Intelligence Systems, and Information Technology*, pages 348–359. Springer, 2015.
- [237] J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T.-C. Lim. Automated diagnosis of glaucoma using digital fundus images. *Journal of Medical Systems*, 33(5):337, 2009.
- [238] R. Bock, J. Meier, L. G. Nyúl, J. Hornegger, and G. Michelson. Glaucoma risk index: automated glaucoma detection from color fundus images. *Medical Image Analysis*, 14(3):471–481, 2010.
- [239] S. Dua, U. R. Acharya, P. Chowriappa, and S. V. Sree. Wavelet-based energy features for glaucomatous image classification. *IEEE Transactions on Information Technology in Biomedicine*, 16(1):80–87, 2012.
- [240] X. Chen, Y. Xu, S. Yan, D. W. K. Wong, T. Y. Wong, and J. Liu. Automatic feature learning for glaucoma detection based on deep learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 669–677. Springer, 2015.

- [241] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.
- [242] V. N. Vapnik and V. Vapnik. *Statistical learning theory*, volume 1. Wiley New York, 1998.
- [243] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [244] J. Nayak, P. S. Bhat, and U. Acharya. Automatic identification of diabetic maculopathy stages using fundus images. *Journal of Medical Engineering & Technology*, 33(2):119–129, 2009.
- [245] Z. Y. Zaidi, M. U. Akram, and A. Tariq. Retinal image analysis for diagnosis of macular edema using digital fundus images. In *Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on*, pages 1–5. IEEE, 2013.
- [246] K. S. Deepak and J. Sivaswamy. Automatic assessment of macular edema from color retinal images. *IEEE Transactions on Medical Imaging*, 31(3):766–776, 2012.
- [247] A. Tariq, M. U. Akram, A. Shaukat, and S. A. Khan. Automated detection and grading of diabetic maculopathy in digital retinal images. *Journal of Digital Imaging*, 26(4):803–812, 2013.
- [248] C. G. Baby and D. A. Chandy. Content-based retinal image retrieval using dual-tree complex wavelet transform. In *Signal Processing Image Processing & Pattern Recognition (ICSIPR), 2013 International Conference on*, pages 195–199. IEEE, 2013.

- [249] S. T. Lim, W. M. D. W. Zaki, A. Hussain, S. Lim, and S. Kusalavan. Automatic classification of diabetic macular edema in digital fundus images. In *Humanities, Science and Engineering (CHUSER), 2011 IEEE Colloquium on*, pages 265–269. IEEE, 2011.
- [250] N. M. Bressler. Age-related macular degeneration is the leading cause of blindness... *Jama*, 291(15):1900–1901, 2004.
- [251] S. M. Waldstein, A.-M. Philip, R. Leitner, C. Simader, G. Langs, B. S. Gerendas, and U. Schmidt-Erfurth. Correlation of 3-dimensionally quantified intraretinal and subretinal fluid with visual acuity in neovascular age-related macular degeneration. *JAMA ophthalmology*, 134(2):182–190, 2016.
- [252] P. J. Rosenfeld, D. M. Brown, J. S. Heier, D. S. Boyer, P. K. Kaiser, C. Y. Chung, and R. Y. Kim. Ranibizumab for neovascular age-related macular degeneration. *New England Journal of Medicine*, 355(14):1419–1431, 2006.
- [253] D. M. Brown, P. K. Kaiser, M. Michels, G. Soubrane, J. S. Heier, R. Y. Kim, J. P. Sy, and S. Schneider. Ranibizumab versus verteporfin for neovascular age-related macular degeneration. *New England Journal of Medicine*, 355(14):1432–1444, 2006.
- [254] W. Ding, M. Young, S. Bourgault, S. Lee, D. A. Albiani, A. W. Kirker, F. Forooghian, M. V. Sarunic, A. B. Merkur, and M. F. Beg. Automatic detection of subretinal fluid and sub-retinal pigment epithelium fluid in optical coherence tomography images. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 7388–7391. IEEE, 2013.
- [255] H. Bogunović, M. D. Abràmoff, and M. Sonka. Geodesic graph cut based retinal fluid segmentation in optical coherence tomography. In *Proceedings of the Ophthalmic Medical Image Analysis International Workshop*, pages 49–56. University of Iowa, 2015.

- [256] Z. Sun, H. Chen, F. Shi, L. Wang, W. Zhu, D. Xiang, C. Yan, L. Li, and X. Chen. An automated framework for 3D serous pigment epithelium detachment segmentation in SD-OCT images. *Scientific Reports*, 6, 2016.
- [257] A. Breger, M. Ehler, H. Bogunovic, S. M. Waldstein, A. M. Philip, U. Schmidt-Erfurth, and B. S. Gerendas. Supervised learning and dimension reduction techniques for quantification of retinal fluid in optical coherence tomography images. *Eye*, 31(8):1212, 2017.
- [258] A. Montuoro, S. M. Waldstein, B. S. Gerendas, U. Schmidt-Erfurth, and H. Bogunović. Joint retinal layer and fluid segmentation in OCT scans of eyes with severe macular edema using unsupervised representation and auto-context. *Biomedical Optics Express*, 8(3):1874–1888, 2017.
- [259] J. Wang, M. Zhang, A. D. Pechauer, L. Liu, T. S. Hwang, D. J. Wilson, D. Li, and Y. Jia. Automated volumetric segmentation of retinal fluid on optical coherence tomography. *Biomedical Optics Express*, 7(4):1577–1589, 2016.
- [260] C. S. Lee, A. J. Tying, N. P. Deruyter, Y. Wu, A. Rokem, and A. Y. Lee. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomedical optics express*, 8(7):3440–3448, 2017.
- [261] S. Yadav, K. Gopinath, and J. Sivaswamy. A generalized motion pattern and FCN based approach for retinal fluid detection and segmentation. In *Retouch Challenge in Conjunction with the 4th MICCAI '17 Workshop on Ophthalmic Medical Image Analysis (OMIA)*, 2017.
- [262] S. H. Kang, H. S. Park, J. Jang, and K. Jeon. Deep neural networks for the detection and segmentation of the retinal fluid in OCT images. In *Retouch Challenge in Conjunction with the 4th MICCAI '17 Workshop on Ophthalmic Medical Image Analysis (OMIA)*, 2017.

- [263] Q. Chen, Z. Ji, T. Wang, Y. Tang, C. Yu, O. I. Paul, and L. B. Sappa. Automatic segmentation of fluid-associated abnormalities and pigment epithelial detachment in retinal SD-OCT images. In *Retouch challenge in conjunction with the 4th MICCAI '17 Workshop on Ophthalmic Medical Image Analysis (OMIA)*, 2017.
- [264] S. Apostolopoulos, C. Ciller, R. Sznitman, and S. De Zanut. Simultaneous classification and segmentation of cysts in retinal OCT. In *Retouch challenge in conjunction with the 4th MICCAI '17 Workshop on Ophthalmic Medical Image Analysis (OMIA)*, 2017.
- [265] R. Tennakoon, A. K. Gostar, R. Hoseinnezhad, and A. Bab-Hadiashar. Retinal fluid segmentation and classification in OCT images using adversarial loss based CNN. In *Retouch challenge in conjunction with the 4th MICCAI '17 Workshop on Ophthalmic Medical Image Analysis (OMIA)*, 2017.
- [266] D. Lu, M. Heisler, S. Lee, G. Ding, M. V. Sarunic, and M. F. Beg. Retinal fluid segmentation and detection in optical coherence tomography images using fully convolutional neural network. In *Retouch challenge in conjunction with the 4th MICCAI '17 Workshop on Ophthalmic Medical Image Analysis (OMIA)*, 2017.
- [267] D. Morley, H. Foroosh, S. Shaikh, and U. Bagci. Simultaneous detection and quantification of retinal fluid with deep learning. In *Retouch challenge in conjunction with the 4th MICCAI '17 Workshop on Ophthalmic Medical Image Analysis (OMIA)*, 2017.
- [268] A. Rashno, D. D. Koozekanani, and K. K. Parhi. Detection and segmentation of various types of fluids with graph shortest path and deep learning approaches. In *Retouch challenge in conjunction with the 4th MICCAI '17 Workshop on Ophthalmic Medical Image Analysis (OMIA)*, 2017.

-
- [269] D. Coltuc, P. Bolon, and J.-M. Chassery. Exact histogram specification. *IEEE Transactions on Image Processing*, 15(5):1143–1152, 2006.
- [270] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.