Anatomical Structure Segmentation in Retinal Images with Some Applications in Disease Detection

Thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
in
Computer Science and Engineering

by

Arunava Chakravarty
201199525
arunava.chakravarty@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
November 2019

International Institute of Information Technology

Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Anatomical Structure Segmentation in Retinal Images with Some Applications in Disease Detection" by Arunava Chakravarty, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date	Adviser: Prof. Jayanthi Sivaswamy

To my parents

Acknowledgments

I take this opportunity to express my deep gratitude towards Prof. Jayanthi Sivaswamy for her guidance as my Ph.d supervisor. She taught me how to do research, write and present research papers, and provided a lot of independence in my research during the later years of my Ph.d.

I would like to thank Prof. C.V. Jawahar, Prof. P.J. Narayanan and the administrative staff at the Centre for Visual Information Technology (CVIT) research centre for providing access to the GPU servers and other computational resources in the lab twenty four hours a day and seven days a week without which this thesis wouldnot have been possible.

I sincerely thank my friends (Vijay, Pritish, Aniket, Aditya, Rajvi, Praveen, Saurabh, Devendra and Viresh) and seniors (Anand Mishra, Gopal Joshi and Natraj) at CVIT for constantly boosting my morale and providing invaluable suggestions, comments and feedback on my research which played a key role in the successful completion of this thesis.

I am grateful to the Tata Consultancy Services for the financial support under their Ph.d. research scholarship program.

Finally, I would like to thank my parents for believing in my capabilities at all times and encouraging me to put my best in the research work.

Abstract

Color Fundus (CF) imaging and Optical Coherence Tomography (OCT) are widely used by ophthalmologists to visualize the retinal surface and the intra-retinal tissue layers respectively. An accurate segmentation of the anatomical structures in these images is necessary to visualize and quantify the structural deformations that characterize retinal diseases such as Glaucoma, Diabetic Macular Edema (DME) and Age-related Macular Degeneration (AMD). In this thesis, we propose different frameworks for the automatic extraction of the boundaries of relevant anatomical structures in CF and OCT images.

First, we address the problem of the segmentation of Optic Disc (OD) and Optic Cup (OC) in CF images to aid in the detection of Glaucoma. We propose a novel boundary-based Conditional Random Field (CRF) framework to jointly extract both the OD and OC boundaries in a single optimization step. Although OC is characterized by the relative drop in depth from the OD boundary, the 2D CF images lack explicit depth information. The proposed method estimates depth from CF images in a supervised manner using a coupled, sparse dictionary trained on a set of image-depth map (derived from OCT) pairs. Since our method requires a single CF image per eye during testing it can be employed in the large-scale screening of glaucoma where expensive 3D imaging is unavailable.

Next, we consider the task of the intra-retinal tissue layer segmentation in cross-sectional OCT images which is essential to quantify the morphological changes in specific tissue layers caused by AMD and DME. We propose a supervised CRF framework to jointly extract the eight layer boundaries in a single optimization step. In contrast to the existing energy minimization based segmentation methods that employ handcrafted energy cost terms, we linearly parameterize the total CRF energy to allow the appearance features for each layer and the relative weights of the shape priors to be learned in a joint, end-to-end manner by employing the Structural Support Vector Machine formulation. The proposed method can aid the oph-

thalmologists in the quantitative analysis of structural changes in the retinal tissue layers for clinical practice and large-scale clinical studies.

Next, we explore the Level Set based Deformable Models (LDM) which is a popular energy minimization framework for medical image segmentation. We model the LDM as a novel Recurrent Neural Network (RNN) architecture called the Recurrent Active Contour Evolution Network (RACE-net). In contrast to the existing LDMs, RACE-net allows the curve evolution velocities to be learned in an end-to-end manner while minimizing the number of network parameters, computation time and memory requirements. Consistent performance of RACE-net on a diverse set of segmentation tasks such as the extraction of OD and OC in CF images, cell nuclei in histopathological images and left atrium in cardiac MRI volumes demonstrates its utility as a generic, off-the-shelf architecture for biomedical segmentation.

Segmentation has many clinical applications especially in the area of computer aided diagnostics. We close this dissertation with some illustrative applications of the segmentation information. We consider the case of disease detection in CF and OCT images. We explore and benchmark two classification strategies for the detection of glaucoma from CF images based on deep learning and handcrafted features respectively. Both the methods use a combination of appearance features directly derived from the CF image and structural features derived from the OD and OC segmentation. We also construct a Normative Atlas for the macular OCT volumes to aid in the detection of AMD. The irregularities in the Bruch's membrane caused by the deposit of drusen are modeled as deviations from the normal anatomy represented by the Atlas Mean Template.

Contents

Chapter Page

1 Introduction . . . 1
1.1 Clinical Background . . . 3
1.1.1 Color Fundus Imaging . . . 4
1.1.2 Optical Coherence Tomography . . . 6
1.2 Scope and Contributions . . . 8
1.3 Outline . . . 10

2 Optic Disc and Cup Boundary Extraction from Monocular Fundus Images . . . 11
2.1 Background . . . 12
2.2 Method . . . 14
2.2.1 Region of Interest (ROI) extraction . . . 15
2.2.2 Supervised Depth Estimation . . . 16
2.2.2.1 Depth estimation from Luminance . . . 17
2.2.2.2 Depth estimation from Chrominance . . . 17
2.2.2.3 Patch-level Feature Extraction . . . 17
2.2.2.4 Coupled Sparse Dictionary . . . 19
2.2.3 Joint OD-OC Segmentation . . . 20
2.3 Materials . . . 24
2.4 Results . . . 24
2.4.1 Evaluation of depth estimation . . . 25
2.4.2 Evaluation of OD-OC Segmentation . . . 27
2.4.2.1 Evaluation Metrics . . . 27
2.4.2.2 Benchmarking against state of the art . . . 28
2.4.2.3 Cup-to-Disc Diameter Ratio (CDR) analysis . . . 31
2.4.2.4 Analysis of Rim thickness . . . 32
2.4.3 Glaucoma Screening . . . 32
2.5 Discussion . . . 33
2.6 Conclusions . . . 36

3 Joint Multi-layer Segmentation for Retinal Optical Coherence Tomography Images . . . 38
3.1 Background . . . 40
3.2 Methods . . . 44
3.2.1 Image Preprocessing . . . 44
3.2.2 Modelling Joint Multi-Layer Segmentation as a CRF . . . 45

3.2.3 Linear Parameterization of CRF Energy . . . 47
3.2.3.1 Unary Boundary Cost . . . 47
3.2.3.2 Pairwise Intra-Layer Cost . . . 48
3.2.3.3 Pairwise Inter-Layer Cost . . . 49
3.2.4 The Structural Support Vector Machine Formulation . . . 50
3.3 Materials . . . 52
3.4 Results . . . 53
3.4.1 Performance Metrics . . . 54
3.4.2 Performance on the NORMAL-1 dataset . . . 56
3.4.3 Cross-testing Performance on the NORMAL-2 dataset . . . 57
3.4.4 Performance in the presence of Age-Related Macular Degeneration . . . 59
3.4.5 Performance in the presence of Diabetic Macular Edema . . . 61
3.4.6 Performance on DME cases alone . . . 64
3.5 Discussions . . . 65
3.6 Conclusions . . . 70

4 RACE-net: A Recurrent Neural Network for Biomedical Image Segmentation . . . 72
4.1 Background . . . 73
4.2 Method . . . 76
4.2.1 A generalized PDE for curve evolution . . . 76
4.2.2 Single time-step of the Curve Evolution . . . 78
4.2.2.1 Network Architecture of $g(I, \phi)$ and $h(I)$ . . . 79
4.2.3 Curve Evolution as a RNN . . . 81
4.2.3.1 Loss Function . . . 83
4.3 Results . . . 84
4.3.1 Optic Disc and Cup Segmentation . . . 84
4.3.2 Cell Nuclei segmentation . . . 87
4.3.3 Left Atrium Segmentation . . . 90
4.4 Discussion . . . 94
4.4.1 Performance . . . 94
4.4.2 Failure cases . . . 95
4.4.3 Computational time and network size . . . 96
4.4.4 Boundary initialization and network hyperparameters . . . 98
4.4.5 Effect of the regularization term in the Loss Function . . . 99
4.5 Conclusion . . . 100

5 Applications in Retinal Disease Detection . . . 101
5.1 Glaucoma detection in Fundus Images . . . 101
5.1.1 Method based on Handcrafted Features . . . 104
5.1.1.1 Extraction of Segmentation based features . . . 105
5.1.1.2 Extraction of Image based features . . . 106
5.1.2 Method based on Deep Learning . . . 107
5.1.3 The CNN architecture . . . 107
5.1.4 Implementation details . . . 109
5.1.5 Post-processing . . . 110
5.1.6 Experiments . . . 111

5.1.7 Results . . . 111
5.1.8 Conclusion . . . 115
5.2 AMD detection in 3D OCT Volumes . . . 117
5.2.1 Method . . . 119
5.2.1.1 Atlas Construction . . . 119
5.2.1.2 The pairwise registration algorithm . . . 121
5.2.1.3 Initial Template Selection . . . 123
5.2.1.4 AMD Classification . . . 124
5.2.1.5 Coarse Retinal Tissue Localization . . . 126
5.2.2 Experimental Setup . . . 126
5.2.3 Results . . . 126
5.2.4 Conclusion . . . 128

6 Conclusions . . . 130
6.1 Summary . . . 130
6.2 Future Directions . . . 131

Appendix A: Conditional Random Field Inference . . . 133
A.1 Conditional Random Field . . . 133
A.2 Belief Propagation . . . 135
A.3 Tree-Reweighted Message Passing -Sequential Algorithm . . . 136

Appendix B: Structural Support Vector Machine . . . 139
B.1 The Structural Support Vector Machine formulation . . . 139
B.1.1 The dual optimization problem . . . 140
B.2 The Frank-Wolfe Optimization Algorithm . . . 141

Appendix C: Level Set based Deformable Models . . . 143
C.1 E-L for the Boundary Cost . . . 143
C.2 E-L for the Regional Cost . . . 145
C.3 Curve Evolution in Level Set Representation . . . 147

Appendix D: Public Datasets . . . 148

List of Figures

Figure Page

1.1 The population distribution of the leading ocular diseases in the world (Fig a) and India (Fig. b). Image Credits: a. International Centre for Eye Health, London School of Hygiene and Tropical Medicine; b. Glaucoma Society of India. 1

1.2 A diagrammatic representation of the Human Eye (Fig. a) and a cross-sectional representation of the intra-retinal tissue layers (Fig. b). The image has been adapted from https://webvision.med.utah.edu/book/part-i-foundations/gross-anatomy-of-the-eye/ and https://anatomychartpad.com/nervous-layer-of-the-eye/nervous-layer-of-the-eye-visual-system-sensory-system-part-1/ with modifications. . . . 3

1.3 a. Color fundus photographs provide a true color image of the retinal surface. b. A macula centric color fundus image taken at a 45^o Field of View. c. An Optic Disc centric color fundus image taken at a 30^o Field of View. d. Examples of portable fundus cameras and smartphone attachments useful in mass screening programs. The images have been adapted from https://ophthalmology.med.ubc.ca/patient-care/ophthalmic-photography/color-fundus-photography/ (Fig. a), and https://mandarinoptomedic.com/product/optomed-smartscope-pro-handheld-imaging-system/, https://www.bocinstruments.com.au/shop/item/welch-allyn-i-examiner (Fig. d). Fig. b, c are part of the MESSIDOR [5] and DRISHTI-GS1 [6] public datasets. 5

1.4 The Optic Disc (in Green) and Cup (in blue) boundaries are marked on a small Region of Interest in the Color Fundus image along with a diagrammatic illustration of the expected cross-sectional view for a Normal (Fig. a) and a Glaucomatous (Fig. b) eye. The cross-sectional diagrams have been adapted from https://www.nature.com/articles/nrdp201667/figures/1 with modifications. 6

1.5 A 3D OCT volume is composed of a series of cross-sectional slices called B-scans. Each 1D column profile within a B-scan is called an A-scan. The retinal tissue is composed of multiple layers which are depicted on a macular B-scan of the OCT volume. 7

1.6 **a.** A B-scan with AMD characterized by the irregularities (indicated by white arrows) in the BM (Green) boundary. **b.** A B-scan of a DME patient with intra-retinal (IRF) and sub-retinal (SRF) fluid-filled regions. 8

2.1 a. A color fundus image. b. Cropped Region of interest of a. with Optic disc and cup boundaries. c. Topographical representation of b., cup boundary defined by the drop in depth from disc edge. 12

2.2 Outline of the proposed system. Training(test) modules are enclosed within red(green) dotted box. 15

2.3 Overview of the proposed supervised depth estimation method. 16
2.4 The parameterization and graphical model representation of OD-OC boundaries. 21
2.5 OD regions from 3 sample images (column a). Corresponding depth estimates are shown as greysacle image and topographical maps with input image wrapped onto depth surface. Columns b,c are ground truth; Col. d,e are computed results. 27
2.6 Qualitative results on some challenging cases. Disc and cup boundaries are depicted in green and blue respectively. a. cropped region around OD; b. Ground Truth; Results of c. Proposed Method; d. Superpixel based [42]; e. Multiview [74] f. Graph Cut prior [75] g. Vessel bend [36] 30
2.7 Box plots representing the distribution of vertical CDR values estimated with the proposed method. 32
2.8 Box plots representing the distribution of the absolute error in rim thickness ratio (ratio of cup radius to disc radius) for 90 equidistant orientations, 4 degrees apart along the X-axis. 33
2.9 ROC curves of glaucoma classification on RIM-ONE v2 for five fold cross-validation.Solid and Dotted lines indicate performance obtained using 14-D feature and CDR alone respectively. 34

3.1 Retinal layer boundaries in a macular OCT B-scan. 39
3.2 a. A retinal B-scan: a local patch is enlarged to depict the speckle noise and indistinct layer boundaries; the vessel shadows are indicated by red arrows. b. ILM(Red), BM (Green) and RPE_{out} (Blue) boundaries in a B-scan with AMD; irregularities in the BM boundary are indicated by white arrows. c. The 8 layer boundaries in a B-scan with DME; the fluid-filled region is indicated by a white arrow. 40
3.3 Overview of the proposed joint multi-layer segmentation pipeline. The supervised training is indicated by dashed lines. 44
3.4 a. Raw OCT B-scan; b. Corresponding preprocessed Region of Interest. 45
3.5 The Conditional Random Field for joint multi-layer OCT Segmentation. 47
3.6 Illustration of the Structural Support Vector Machine (Structural Support Vector Machine (SSVM)) formulation. 52
3.7 Qualitative results on 3 B-scans of healthy subjects from the *NORMAL-1* dataset is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark. Region within the white dashed rectangle in the first row is magnified at the bottom for comparison. 57
3.8 Qualitative results on 3 B-scans of healthy subjects from the *NORMAL-2* dataset is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark. Region within the white dashed rectangle in the first row is magnified at the bottom. 58
3.9 Qualitative results on 3 B-scans with AMD is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark. 60
3.10 Qualitative results on 3 B-scans from the *DME-1* dataset is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark. 62

3.11 Variation in performance with respect to a) the filter size and b) the regularization weight λ. Lower values of U-BLE in pixels indicate better performance. . . . 66

4.1 **a.** The Feedforward neural network (Feedforward Neural Network (FFNN)) architecture that models the evolution of the level set function ϕ in a single time step t. **b.** The details of the customized $C(\phi)$ module used within the FFNN in Fig. a. to compute the normal and curvature of ϕ. 79

4.2 The proposed Convolutional Neural Network (CNN) architecture to model $h(I)$ and $g(I, \phi)$. 80

4.3 **a.** Curve Evolution as a RNN. **b.** The unrolling of the RNN in a. over time. The recurrent feedback connections are depicted in red. 81

4.4 **a.** depicts the curve evolution of RACE-net over the intermediate time-steps, T=0, 2, 4, 6 for the OD boundary (in green) and over T=0, 1, 3, 5 for the OC boundary (in blue) from left to right. **Fig. b** depicts four sample results for OD and OC boundaries. In each subimage **i-iv**, the cropped region of interest is depicted in column 1, the result of the RACE-net for the OD and OC boundaries (in green) are depicted in column 2 followed by a comparison with the Ground truth markings (in blue) for OD (column 3) and OC (column 4). 85

4.5 **a.** The curve evolution over time steps T=0,2,3,4,6 (from top to bottom). **Fig. b.** depicts the qualitative results of the RACE-net: 1^{st} **row:** input image, 2^{nd} **row:** Ground truth markings (in red), 3^{rd} **row:** result of RACE-net (in green), 4^{th} **row:** the result of RACE-net (in green) and the Ground truth markings (in red) are overlapped for comparison. 89

4.6 **a.** The curve evolution over time steps T=1,2,3,5 and 7 (from left to right). **b.** The 3D surface rendering of the segmented left atrium: 1^{st} row is the Ground truth and second row depicts the corresponding segmentation by the RACE-net. **c.** Qualitative results for the individual slices depicted in the 1^{st} column are provided. Only a small region marked by the red bounding box has been magnified for better visibility of the result. 2^{nd} column depicts the Ground Truth markings in blue, 3^{rd} column depicts the result of RACE-net depicted in green. The Ground truth and the result of the RACE-net are overlapped for comparison in the 4^{th} column. 92

4.7 Examples of failure cases for: **a.** Optic disc, **b.** Optic cup, **c.** cell nuclei and **d.** left atrium. The Ground truth is marked in blue in **a.**, **b**, **d** and yellow in **c**. The results of RACE-net are marked in green in each case. The results obtained using U-net (marked in red in each case) are provided for comparison. 96

4.8 Each row depicts the evolution of the Curve for the segmentation of OD over time steps T=0,2,4,6 (from left to right). The proposed method is not sensitive to the exact location of the OD with respect to the initial boundary. 99

5.1 **a.** Block Diagram of the proposed method for glaucoma classification based on handcrafted features. **b.** OD and OC boundaries in a cropped fundus image; **c.** Color based clustering of b. used for BoW computation; **d.** polar representation of the red channel of b.; **e.** LBP of Radon transform of d. for ToP compuation. . 104

5.2 Outline of the proposed DL based method for glaucoma classification. 107

5.3 The proposed Multi-task Convolutional Neural Network architecture for joint segmentation of Optic disc, cup and image level classification of glaucoma using a combination of image appearance and structural features. 108
5.4 Area under the ROC curve for the proposed method based on handcrafted features (1^{st} row) and Deep Learning (2^{nd} row) on the private test set (1^{st} column), RIM-ONE (2^{nd} column) and the REFUGE (3^{rd} column) dataset respectively. In each plot, the ROC curve for the segmentation features (green), appearance features alone (blue) and the combined features (red) are depicted. 113
5.5 OCT B-scan of an AMD case with the three relevant layer boundaries. 117
5.6 The normative OCT atlas consists of a mean intensity template (MT) obtained by the average of a set of co-registered healthy OCT Volumes and probability maps (P-maps) which give the probability of observing a particular tissue layer at a given location. 118
5.7 Atlas Construction pipeline. 120
5.8 AMD classification pipeline. 124
5.9 ROC for AMD detection. 127
5.10 (left to right) the mean intensity atlas (1^st row) and the corresponding probability maps for the seven tissue layers for the 30^{th}, 50^{th} and 60^{th} B-scans. 128

A.1 An example of a Conditional Random Field with corresponding Unary and Pairwise cost tables. 134
A.2 Ordering of message updates for the Belief Propagation Algorithm for a tree structured graph. 136
A.3 A grid structured loopy graph with nine nodes is decomposed into 6 sub-trees: $\{T_1, T_2...T_6\}$. Each node is a member of two sub-trees and each edge occurs in a single sub-tree. 137
A.4 The monotonic chain ordering of nodes during a forward and backward pass of the Sequential Tree Re-weighted Message Passing (TRW-S) algorithm. The numbering of each node indicates the order in which it is considered in the TRW-S algorithm. 138

List of Tables

Table Page

2.1 Dataset Specifications. All datasets are publicly accessible except $Dataset-1$. . 25
2.2 Performance of the Depth Estimation. (mean/standard deviation) 26
2.3 Segmentation performance on DRISHTI-GS1 29
2.4 p-values for the paired T-test between the proposed and the benchmark methods. A p-value < 0.05 indicates statistically significant improvement. 29
2.5 OD segmentation performance on DRIONS-DB. (mean/std) 31
2.6 OD segmentation performance on MESSIDOR. (mean/std) 31
2.7 CDR error (mean/std.) of the proposed method. Lower values indicate better performance. 32
2.8 Average Computation time (seconds/image) 35
2.9 Analysis of Energy Terms: λ_1,λ_2,λ_3,λ_4 are weights associated with the disc smooothness (λ_1), cup data (λ_2), cup smoothness (λ_3) and the OD-OC interaction (λ_4) terms. 35

3.1 Dataset Description. The **voxel resolution** is reported in *axial, lateral, azimuthal* directions. # **Images** reports the (total number of B-scans/ acquired from the number of OCT volumes). # **GT** reports the number of layer boundaries for which the Ground Truth markings are available. 53
3.2 Unsigned and Signed Boundary Localization Errors (mean ± standard deviation in pixels) on the NORMAL-1 dataset. The best result in each column is indicated in bold. 56
3.3 Layer Thickness Error in pixels and Dice coefficient (mean ± standard deviation) for 7 tissue regions on the NORMAL-1 dataset. The best result in each column is indicated in bold. 57
3.4 Unsigned, Signed Boundary Localization Errors (mean ± standard deviation in pixels) on the NORMAL-2 dataset. Best result among the automated methods in each column is indicated in bold. 58
3.5 Layer Thickness Error in pixels and Dice coefficient (mean ± standard deviation) for 7 tissue regions on the NORMAL-2 dataset. The best result among the automated methods is indicated in bold. 59
3.7 Layer Thickness Error in pixels and Dice coefficient (mean ± standard deviation) for 3 layer boundaries on the combined AMD and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold. 60

3.6 Boundary Localization Errors (mean ± std. deviation in pixels) on the combined *AMD* and *NORMAL-1* dataset. Best result among the automated methods in each column is indicated in bold. 61
3.8 Unsigned Boundary Localization Errors (mean ± standard deviation) on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold. 62
3.9 Signed Boundary Localization Errors (mean ± standard deviation) on the combined DME and NORMAL-1 dataset. Best result among the automated methods in each column is indicated in bold. 63
3.10 Mean Layer Thickness Error (mean ± standard deviation) in pixels for 7 tissue regions on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold. 63
3.11 Dice coefficient (mean ± standard deviation) for 7 tissue regions on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold. 64
3.12 Mean Dice coefficient and the layer thickness error (in pixels) for 7 tissue regions evaluated on the *DME-1* dataset alone. 65
3.13 p-values for the paired T-test between the U-BLE metric of the proposed method against the best performing method among CASERAL, OCTSEG, IRA. A p-value < 0.05 indicates statistically significant improvement. 67

4.1 Segmentation performance on DRISHTI-GS1 test set. (mean/standard deviation) 87
4.2 Performance of OD segmentation on the MESSIDOR dataset. (mean/standard deviation) 88
4.3 Cell nuclei Segmentation Performance on the UCSB Bio-segmentation Benchmark dataset 88
4.4 3D Left Atrium Segmentation Performance on the STACOM 2013 Challenge dataset (mean/standard deviation) 93
4.5 Statistics for the Dice and Boundary Error of the proposed RACE-net on various tasks. 95
4.6 Average time (in seconds) to segment each image using CPU. 97

5.1 Dataset Description 112
5.2 Performance of Optic Disc and Cup Segmentation. (mean Dice coefficient) 113
5.3 Individual Contribution of the Segmentation and Appearance based features. . . 114
5.4 Benchmarking Glaucoma Classification Performance against the state of the art. 116
5.5 AMD classification performance. 127
5.6 Dice coefficients (mean ± standard deviation) for the segmentation of 7 retinal tissue layers on the NORMAL-2 dataset. 128

List of Abbreviations

AIC	Akaike Information Criterion
AMD	Age-related Macular Degeneration
BCFW	Block Co-ordinate Frank-Wolfe
BLE	Boundary Localization Error
BM	Bruch's Membrane
BPTT	Back Propagation Through Time
CAD	Computer Aided Diagnosis
CCA	Canonical Correlation Analysis
CDR	Cup-to-Disc Diameter Ratio
CF	Color Fundus
CNN	Convolutional Neural Network
CRF	Conditional Random Field
CSD	Coupled Sparse Dictionary
DME	Diabetic Macular Edema
DR	Diabetic Retinopathy
FCN	Fully Convolutional Network
FFNN	Feedforward Neural Network
$GCL-IPL$	Ganglion Cell and Inner Plexicon layer
GMM	Gaussian Mixture Model
GT	Ground Truth
ILM	Inner Limiting Membrane
INL	Inner Nuclear Layer
LDM	Level Set based Deformable Models
MCCS	Minimum Cost Closed Set
NFL	Nerve Fiber Layer
OC	Optic Cup
OCT	Optical Coherence Tomography

OD	Optic Disc
$ONL - IS$	Outer Nuclear and Inner Segment
OPL	Outer Plexiform Layer
OS	Outer Segment layer
QP	Quadratic Programming
PDE	Partial Differential Equation
RACE-net	Recurrent Active Contour Evolution Network
RNN	Recurrent Neural Network
ROI	Region of Interest
RPE	Retinal Pigment Epithelium Layer
RPE_{out}	outer boundary of the Retinal Pigment Epithelium Layer
SDF	Signed Distance Function
SFS	Shape From Shading
SSVM	Structural Support Vector Machine
TRW-S	Sequential Tree Re-weighted Message Passing

Chapter 1

Introduction

Around 39 million people in the world are blind and another 217 million people suffer from moderate to severe visual impairments [1]. India has the largest population of people suffering from visual defects in the world. The population growth and increase in the life expectancy have led to a marked increase in the chronic and age-related retinal diseases. As depicted in Fig. 1.1, Glaucoma, Diabetic Retinopathy (DR) and Age-related Macular Degeneration (AMD) are among the leading causes of irreversible blindness in India and the world. However, about 80% of all vision impairments can be either prevented or cured through an early detection of the disease and timely treatment [2].

Retinal imaging modalities provide a non-invasive way to visualize the anatomical changes in the eye. They aid the ophthalmologists in the detection and tracking the progression of various retinal diseases over time, both in a primary health-care and large scale screening scenario. In a primary health-care setting, patients suffering from disease symptoms visit an

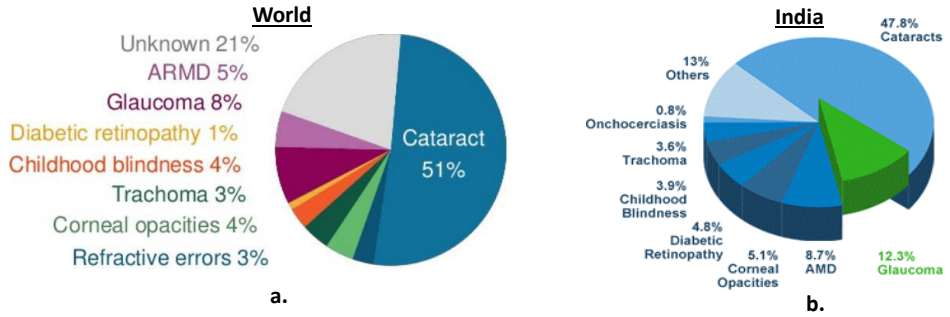

Figure 1.1: The population distribution of the leading ocular diseases in the world (Fig a) and India (Fig. b). Image Credits: a. International Centre for Eye Health, London School of Hygiene and Tropical Medicine; b. Glaucoma Society of India.

ophthalmologist for treatment. However, many retinal diseases are asymptomatic in nature implying that the patients donot feel any discomfort in the early stages. Therefore, screening programs are employed as a proactive strategy to detect the retinal diseases at an early stage. A large population of people (not necessarily suffering from disease symptoms) are examined by a trained reader often at remote rural centers as a public health initiative and suspicious cases are referred back to the experts. This has led to an increasing demand for mass screening programs which form an essential part of initiatives such as the *Sankara Netralaya Teleophthalmology* Project and the *Aravind Teleophthalmology Network* [3] in India.

Currently, there is a dearth of medical experts in India. The ophthalmologist to population ratio is 1:219,000 in rural and 1:25,000 in the urban areas respectively [4]. Out of the 19 million people who suffer from visual defects in India, 15 million reside in rural areas [3]. The availability of the ocular images in a digital format has enabled the application of image processing and machine learning techniques to develop Computer Aided Diagnosis (CAD) tools for the automated analysis of retinal diseases. Effective CAD tools can reduce the ophthalmologist's workload and aid in minimizing the inter and intra-observer subjectivity in the diagnosis.

The segmentation of anatomical structures forms an essential component of any CAD solution. Many ocular diseases such as Glaucoma, AMD and Diabetic Macular Edema are characterized by the deformations in the morphology of relevant anatomical structures in the retina. An accurate segmentation of these structures is essential in deriving quantitative clinical measurements that aid in the detection of the disease, tracking its progression and effective treatment planning. However, manual segmentation requires domain expertise and is a tedious and time consuming process. Moreover, the manual segmentations are subjective and lead to a considerable inter and intra-observer variability.

This work is aimed towards the development of robust solutions for the automated segmentation of relevant anatomical structures in retinal Color Fundus (CF) imaging and Optical Coherence Tomography (OCT) modalities. Additionally, some applications related to the detection of diseases such as Glaucoma in CF images and AMD in OCT volumes have also been explored. In some cases, our contributions towards the development of a novel segmentation algorithm have been shown to have potential applications in a diverse set of medical images such as histopathology and cardiac MRI volumes apart from retina.

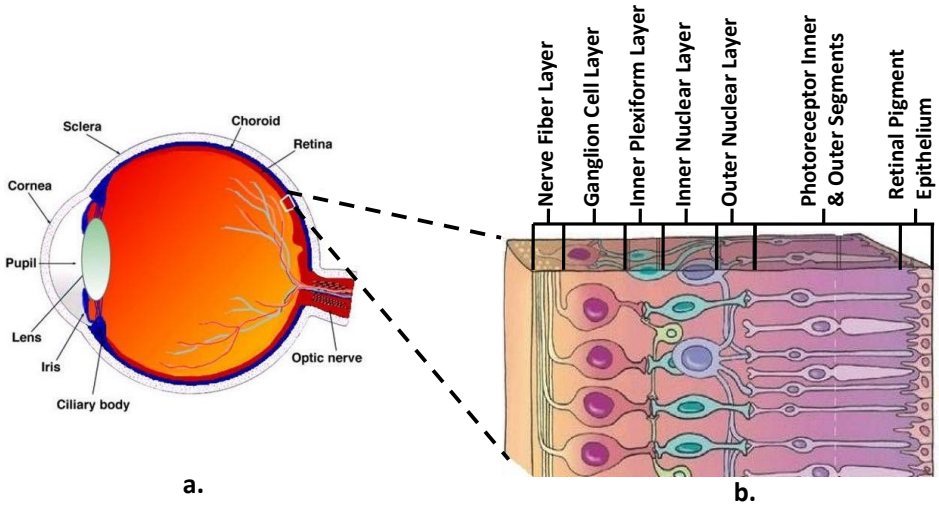

Figure 1.2: A diagrammatic representation of the Human Eye (Fig. a) and a cross-sectional representation of the intra-retinal tissue layers (Fig. b).

The image has been adapted from https://webvision.med.utah.edu/book/part-i-foundations/gross-anatomy-of-the-eye/ and https://anatomychartpad.com/nervous-layer-of-the-eye/nervous-layer-of-the-eye-visual-system-sensory-system-part-1/ with modifications.

1.1 Clinical Background

Human eye is a roughly spherical structure (see Fig. 1.2 a) where light enters through the pupil, passes through the lens and is projected onto the back of the eye called the retina. The size of the pupil (adjusted by the dilation of iris) controls the amount of light entering the eye, while the lens shape is adjusted by the ciliary muscles to focus the light onto the retinal surface. The retina is a complex multi-layered structure which converts light into electrical signals and relays it to the brain, where it is processed to produce the sensation of vision. It is composed of different layers of tissue as shown in Fig. 1.2 b and contain light-sensitive photoreceptors (the sixth layer in Fig. 1.2 b) called the rods and cones. The ocular diseases such as the AMD, DR and Glaucoma effect different layers of the retina and cause irreversible blindness.

CF (Color fundus) imaging and OCT (Optical Coherence Tomography) have emerged as the two most widely used retinal imaging modalities. While OCT provides a colorless, 3-D tomographic view of the retinal tissue, CF imaging provides a 2D true color image of the retinal surface. Since OCT provides a richer 3D information of the anatomy, it is often used in the diagnosis of suspect cases, tracking disease progression and treatment planning for severe cases of AMD and Diabetic Macular Edema (DME) in primary healthcare clinics. DME is an

advanced stage of DR which is characterized by the swelling of the retinal tissue due to the build-up of fluid (edema) in the macular region of the retina and is the most common cause of visual loss due to DR. The cost, portability and acquisition time of the OCT scanners currently inhibits their use in mass screening. In contrast, CF imaging is widely available, portable, relatively inexpensive and hence frequently used for the early detection of diseases in screening programs.

In this work, we have focussed on developing CAD solutions for these two imaging modalities. Specifically, we have analyzed methods for the assessment of glaucoma in CF images through the segmentation of the Optic Disc and Cup. We have also addressed the problem of intra-retinal layer segmentation and the detection of AMD in OCT images. Next, we briefly discuss the details of the CF and OCT imaging and the disease symptoms relevant to our work.

1.1.1 Color Fundus Imaging

CF photography is an optical imaging modality that provides a magnified, 2D true color image of the retinal surface (see Fig. 1.3 a). They are acquired using specialized fundus cameras that consist of a low powered microscope attached to a digital camera. However, they can image only a small region of the entire retinal surface at a time. A typical fundus camera has 30 to 50° Field of View with a 2.5x magnification. The CF images can be taken from different viewpoints to focus on different parts of the retina as depicted in Fig. 1.3 b, c.

Currently, CF imaging is widely used in the primary healthcare clinics due to its ease of use and wide availability. The development of low cost portable fundus cameras and specialized attachments to acquire fundus images using smartphones (see Fig. 1.3 d) have made it an appropriate imaging modality for screening programs as well [7], [8]. The dilation of the pupil by using mydriatic eye drops prior to the imaging helps in obtaining better quality images. Though this process is commonly employed in clinics, it is not feasible in a mass screening setting. Thus, the development of CAD tools for mass screening programs pose additional challenges since they must be able to handle poor quality and/or low-resolution images that are acquired using low-cost portable cameras without dilation of the pupils. Moreover, the data collected from screening also show a large class imbalance where the number of negative class (healthy subjects) is much larger than the disease cases. This is because an entire community is examined in a screening scenario where most of the people are healthy.

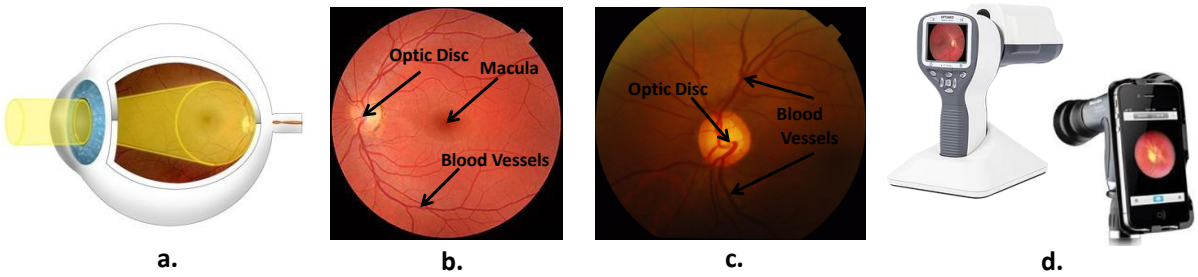

Figure 1.3: a. Color fundus photographs provide a true color image of the retinal surface. b. A macula centric color fundus image taken at a 45^o Field of View. c. An Optic Disc centric color fundus image taken at a 30^o Field of View. d. Examples of portable fundus cameras and smartphone attachments useful in mass screening programs.

The images have been adapted from https://ophthalmology.med.ubc.ca/patient-care/ophthalmic-photography/color-fundus-photography/ (Fig. a), and https://mandarinoptomedic.com/product/optomed-smartscope-pro-handheld-imaging-system/, https://www.bocinstruments.com.au/shop/item/welch-allyn-i-examiner (Fig. d). Fig. b, c are part of the MESSIDOR [5] and DRISHTI-GS1 [6] public datasets.

Optic Disc (OD), macula and the retinal blood vessels are the major anatomical structures on the retinal surface (Fig. 1.3 b, c). Macula is the region of the retina with the maximum concentration of photoreceptors and responsible for colored, sharp and detailed central vision. The clinical examination of the macular region plays an important role in determining the severity of AMD and DR. Hence, the macula-centric CF images are commonly employed for their assessment.

OD is an elliptical bright structure where the blood vessels appear to converge (see Fig. 1.3 b,c). It does not have any photoreceptors and hence also known as the blind spot. The retinal nerve fibers throughout the retinal surface collect the electrical signals from the photoreceptors and exit the eye through the OD to connect to the brain for further processing. The OD has a central depression called the Optic Cup (OC) which is devoid of nerve fibers (see Fig. 1.4). Thus, the nerve fibers bend into the OD in the annular region between the OD and OC boundaries which is known as the neuro-retinal rim.

Glaucoma is a chronic optic neuropathy that leads to a gradual but irreversible loss of retinal nerve fibers. This results in the thinning of the neuroretinal rim and the consequent enlargement of the OC. Thus, CF images with an OD-centric view are commonly used to assess the structural changes in the OD region for the early diagnosis of Glaucoma. Though the automated analysis of retinal diseases such as DR have received significant attention from the

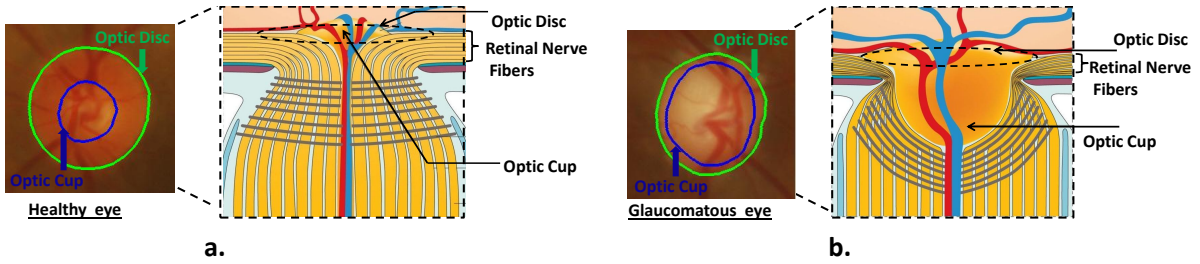

Figure 1.4: The Optic Disc (in Green) and Cup (in blue) boundaries are marked on a small Region of Interest in the Color Fundus image along with a diagrammatic illustration of the expected cross-sectional view for a Normal (Fig. a) and a Glaucomatous (Fig. b) eye.

The cross-sectional diagrams have been adapted from https://www.nature.com/articles/nrdp201667/figures/1 with modifications.

research community [9], [10], [11], the development of CAD solutions for the assessment of glaucoma from CF images still remains an open research problem. This is due to the fact that the OC is primarily defined by the 3D cross-sectional depth information within the OD which is unavailable in the 2D fundus images. In this work, we have explored different methods for the accurate segmentation of OD and OC in chapters 2 and 4. We have also explored and benchmarked different strategies for the image-level detection of glaucoma in CF images based on Deep learning and handcrafted features in chapter 5.

1.1.2 Optical Coherence Tomography

Optical Coherence Tomography (OCT) is a relatively new non-invasive imaging modality that provides a high resolution, 3D cross-sectional view of the tissues lining the retina using the properties of infra-red light reflectivity [12]. An example of an OCT volume is depicted in Fig. 1.5. The OCT volumes are composed of a series of cross-sectional 2D slices called the B-scans. Each B-scan comprises a series of 1D image columns called the *A-scans* that lies in the direction parallel to the propagation of light into the tissue. The intra-retinal tissue is a multi-layered structure which transforms light into neural signals for further use by the brain. It is commonly divided into 7 adjacent layers [13] separated by 8 boundaries. The boundaries ordered from the top to bottom as depicted in Fig. 1.5 are the : i) Inner Limiting Membrane (ILM) separating the vitreous and Nerve Fiber Layer(NFL), ii) NFL/GCL boundary separating NFL from the Ganglion Cell and Inner Plexicon layer (GCL-IPL), iii) IPL/INL separating GCL-IPL from the Inner Nuclear Layer (INL), iv) INL/OPL separating INL from the Outer Plexiform Layer

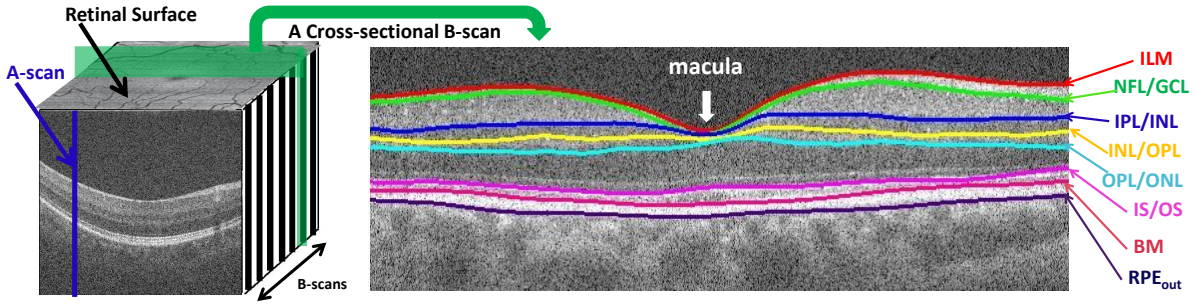

Figure 1.5: A 3D OCT volume is composed of a series of cross-sectional slices called B-scans. Each 1D column profile within a B-scan is called an A-scan. The retinal tissue is composed of multiple layers which are depicted on a macular B-scan of the OCT volume.

(OPL), v) OPL/ONL separating OPL from the Outer Nuclear and Inner Segment (ONL-IS) region, vi) IS/OS separating ONL-IS from the Outer Segment (OS) vii)The Bruch's Membrane (BM) separating OS from the Retinal Pigment Epithelium Layer (RPE) layer and finally the viii) RPE_{out} boundary separating RPE from the choroid.

OCT is a tomographic imaging technique similar to ultrasound. However, instead of sound-waves as used in ultra-sound, OCT uses broadband, low frequency infrared light waves which allows it to acquire high-resolution images in a micrometer range in the axial direction (along the A-scans). OCT imaging is performed one A-scan at a time by directing a source light beam at a specific position on the retinal surface. The beam of light is split into two arms called a sample arm and the reference arm which are directed towards the retinal tissue and a moving mirror respectively. The energy of the interference obtained by superimposing the light back-scattered by the mirror (from the reference arm) and the light back-scattered by the retinal tissue (from the sample arm) is detected by a photodetector and encoded as the intensity of the OCT image. The entire 1D column of the A-scan is obtained by varying the position of the mirror in the sample arm. This OCT construction is known as the Time Domain OCT and it's acquisition speed is limited to around 400 A-scans per second due to the mechanical movement of the mirror [12]. This limitation has been overcome by the Spectral domain OCT imaging which employs a spectrometer with a diffraction grating in place of the photodetector to acquire the fourier spectrum of the entire A-scan simultaneously without moving the mirror in the reference arm. Spectral domain OCT have significantly reduced the acquisition time to around 27,000 image B-scans per second [12]. This has improved the image quality by lower-

ing the risk of motion artefacts due to eye movements and increasing the image resolution by allowing a denser sampling of the A-scans in the OCT volumes.

OCT imaging can be used to visualize and quantify the structural changes or pathologies in specific layers of the retina that characterize the presence and progression of various ocular diseases. Early stages of AMD is characterized by the accumulation of extracellular materials called drusen in the RPE layer which lead to irregularities and undulations in the Bruch's membrane [14] as depicted in Fig. 1.6 a. DME is characterized by the presence of fluid-filled regions that appear as dark holes in the OCT images around the macula leading to the swelling of the retinal tissue [15],[16] as shown in Fig. 1.6 b. The fluid-filled regions are clinically classified into intra-retinal fluids if they occur above the INL/OPL boundary and sub-retinal fluids if they occur beneath the IS/OS boundary [17].

In this work, we have explored a method for the joint segmentation of the eight intra-retinal tissue layer boundaries in chapter 3. We have also explored the construction of a normative atlas from OCT volumes in chapter 5 with potential application in AMD detection.

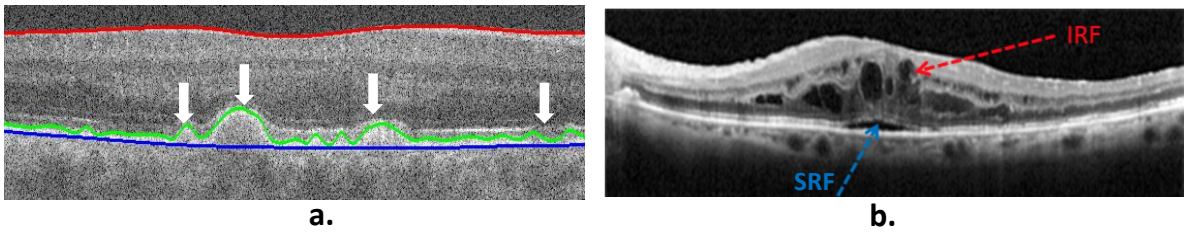

Figure 1.6: a. A B-scan with AMD characterized by the irregularities (indicated by white arrows) in the BM (Green) boundary. **b.** A B-scan of a DME patient with intra-retinal (IRF) and sub-retinal (SRF) fluid-filled regions.

1.2 Scope and Contributions

This thesis addresses the task of segmentation of anatomical structures in CF and OCT images to aid in the detection and assessment of retinal diseases. In contrast to the region based segmentation methods which treat segmentation as a pixel-labeling problem, we explore different energy minimization frameworks based on discrete Conditional Random Field (CRF) and Level Set based Deformable Models (LDM) to extract the anatomical structure boundaries. Treating segmentation as a boundary extraction problem allows us to incorporate constraints

on the boundary to preserve its overall shape and smoothness. We have explored solutions that jointly extract the boundaries of multiple but related anatomical structures in a single optimization step. This allows us to improve the computational efficiency and model the shape priors and high level dependencies between the related anatomical structures. Instead of handcrafting the energy for a specific segmentation task, we have also looked at methods based on structured prediction and deep learning to allow the energy to be learned in a supervised end-to-end manner.

In this context, we have explored CRF formulations for the joint extraction of the OD and OC boundaries in CF and the eight tissue layer boundaries in the OCT images. The structural changes in the OD and OC play a vital role in the diagnosis of glaucoma in CF images. Similarly, the variations in the thickness of the various intra-retinal tissue layers around the macula can be useful in characterizing the presence and severity of diseases such as DME and AMD. A novel Recurrent Neural Network called the Recurrent Active Contour Evolution Network (RACE-net) has also been investigated which models the Level set based deformable models within a deep learning framework.

Finally, we have also looked at the task of the image-level detection of retinal diseases in CF and OCT images. We have explored two classification strategies based on deep learning and handcrafted features respectively, to detect glaucoma in CF images and modeled the abnormalities caused by AMD in 3D OCT volumes as significant deviations from an OCT atlas. Thus, the main contributions of this thesis are:

1. We have proposed a novel CRF formulation for the joint extraction of the OD and OC boundaries. Since, OC is primarily characterized by the depth information, our method explicitly estimates depth from the single 2D fundus image itself during testing. This is achieved using a coupled, sparse dictionary that maps the image appearance to the corresponding depth values.

2. We have extended the CRF framework to jointly segment the seven intra-retinal tissue layers in OCT images. Additionally, we have eliminated the need to handcraft each term of the CRF energy by learning the appearance cost terms and the relative weights of the shape priors in a joint, end-to-end manner using a Structural Support Vector Machine formulation.

3. We have also explored a novel Recurrent Neural Network (RNN) architecture which models the level set based deformable models. The curve evolution velocities at each time step are modeled using a feed-forward architecture inspired by the multi-scale image pyramid. Apart from OD and OC segmentation, the effectiveness of this method has also been demonstrated on histopathology images and 3D cardiac MRI volumes.

4. We have explored different strategies for the image level detection of glaucoma in CF images and AMD in OCT volumes respectively. The irregularities and undulations caused by the deposition of drusen in the RPE layer has been modeled as significant deviation from the normal anatomy captured by the construction of a Normative atlas. Two classification strategies based on handcrafted features and deep learning have been explored for the task of the detection of glaucoma in CF images. Both the strategies use a combination of appearance features derived directly from the CF image and structural/shape features derived from the OD-OC segmentation to aid in glaucoma classification.

1.3 Outline

This thesis is organized as follows. In Chapter 2, we present a CRF based framework for the joint segmentation of OD and OC in CF images. In Chapter 3, the CRF framework is extended to jointly extract the eight intra-retinal layer boundaries in OCT images in a single optimization step. The CRF energy is linearly parameterized and learned in a supervised end-to-end manner by modeling it as a structured prediction problem. In Chapter 4, we propose a Deep Learning based alternative for the extraction of anatomical boundaries which is inspired by the level set based deformable models. Some applications related to the image-level detection of retinal diseases have been explored in Chapter 5. In this context we looked at the problem of the detection of glaucoma in color fundus images and AMD in OCT volumes respectively. Finally, in Chapter 6, we conclude this thesis by summarizing our contributions and discussing some possible directions for future research.

Chapter 2

Optic Disc and Cup Boundary Extraction from Monocular Fundus Images

Glaucoma, a chronic ocular disorder caused by the accelerated degeneration of the retinal optic nerve fibers, accounts for 12.3% of the total blindness world-wide [18] and projected to affect 79.86 million people by 2020 [19]. It results in a gradual loss of sight which starts with the peripheral vision and slowly progresses towards complete and irreversible vision loss. Due to its asymptomatic nature in the early stages, currently 70-90% of the glaucomatous population worldwide are reported to be unaware of their condition [20],[21]. Large scale screening can play a vital role in preventing blindness through early detection and treatment.

Early diagnosis of glaucoma is primarily based on the assessment of structural changes in the Optic Disc (OD), though factors such as functional visual field assessment and intraocular pressure are also considered. OD is characterized by a bright elliptical region in the Color Fundus (CF) images. It contains a central depression called the optic cup (Optic Cup (OC)) which is surrounded by the neuro-retinal rim consisting of the retinal nerve fibers that bend into the OC (Fig. 2.1 a,b). [1] Loss of nerve fibers leads to rim thinning and a consequent enlargement of the OC. Several clinical measures such as the vertical Cup-to-Disc Diameter Ratio (CDR) and the ISNT rule based on the sector-wise rim thickness distributions [22], are used to quantify these structural changes. Automatic OD and OC segmentation can aid the measurement of such clinical indicators for an efficient and objective glaucoma assessment.

OC is primarily characterized by the depth information (Fig. 2.1 b,c). Topcon Imagenet and Humphrey Retinal Analyzers define the cup boundary at 125 μm and 120 μm below the

[1] Also see the discussion on Glaucoma in Chapter 1, Section 1.1.1, page 5-6.

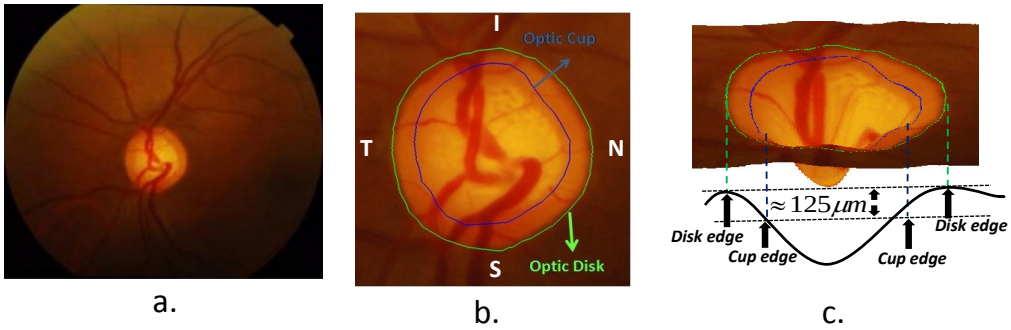

Figure 2.1: a. A color fundus image. b. Cropped Region of interest of a. with Optic disc and cup boundaries. c. Topographical representation of b., cup boundary defined by the drop in depth from disc edge.

OD edge [23] respectively. Several studies also define the cup edge at one-third or half drop in depth from the OD edge to the deepest point in optic cup [24], [25], [26]. In contrast to 3D imaging techniques such as Optical Coherence Tomography (OCT) and stereo fundus cameras that provide true depth estimation, monocular CF images are 2D projections of the retinal surface and lack explicit depth information. Though OCT and stereo imaging is widely used in hospitals, they cannot be employed in a large scale screening due to their cost and portability. In contrast, CF imaging is relatively inexpensive to acquire and widely available.

In this work, we explore a method to jointly segment both OD and OC in a single optimization step. This is achieved by a novel, boundary-based Conditional Random Field (CRF) formulation which is effective in modeling the drop in depth between the OD and OC boundaries. In a clinical setting, both OCT and CF images are available for an eye, whereas during screening, only CF images are available. Hence, we employ a supervised depth estimation strategy to relate the appearance information from a CF image to corresponding depth estimates which requires only a single CF image during testing.

2.1 Background

Majority of the existing methods segment OD first, followed by the OC in a sequential order. Since, OD segmentation is relatively easy compared to OC, more methods have been explored for the former.

OD Segmentation: Techniques based on template matching, supervised classification, deformable and active shape models have been employed for this task. Template based methods often rely on the Hough Transform [27],[28],[29] to fit a circle or ellipse to the edge maps extracted from CF images. An alternative method is explored in [30] where the OD center is characterized by the maximum response of sliding bank filters applied at multiple scales and its boundary is obtained by smoothing the pixel locations that contribute to the maximum filter response. In the template based methods, the analysis is often restricted to the brighter regions in the image to improve the accuracy and efficiency [31]. For example, in [32], a two-step thresholding operation is applied to the image after enhancing the bright regions using iterative morphological operations. These methods suffer from inaccuracies due to the vessel occlusions in the OD region and inflexible shape assumptions. In [33], blood vessel inpainting is explored to handle the vessel occlusions followed by an adaptive threshold based region-growing technique for OD segmentation.

Deformable models such as Snakes [34], level sets [35], and the modified Chan-Vese model [36] improve on the template-based methods by iteratively refining the boundaries using energy minimization. The energy terms are often based on the image gradient computed in multiple color and feature channels. Recently, in [37] an active disc based deformable model has been explored. Active disc comprises a pair of concentric inner and outer discs corresponding to the OD boundary and a local background around it respectively, which is used to define a local contrast energy. These methods are sensitive to poor initialization which can be improved by combining multiple OD detectors through majority voting and data fusion [38]. Further, the gradient information is sensitive to ill-defined boundaries and the presence of peripapillary atrophy near the OD boundary. Active Shape Models in [39],[40] incorporate a statistical shape prior. A set of landmark points on the OD is initialized using the mean shape from the training images and iteratively adapted to the test image, while being consistent with the point distribution model representing the shapes encountered during training.

Classification-based methods label each pixel [41], or superpixel [42] into OD or background classes using features such as Gaussian steerable filter responses on color opponency channels, disparity values extracted from stereo image pairs [41], color histograms and center-surround statistics [42]. Reliance on low level features make these methods susceptible to image noise and vessel occlusions. Moreover, the segmentations may contain multiple connected components.

OC segmentation: OC segmentation is restricted to the region inside OD. Since OC is largely characterized by a discontinuity in the depth of the retinal surface, proposed solutions either rely on explicit depth measurement or depth cues derived from appearance of the CF images. In the former approach, depth is obtained from OCT [43],[44] or from stereo-based disparity maps [41], [45], [46]. Recently, in [47], information of Bruch's membrane opening from OCT is combined with fundus imaging for a joint *multi-modal* OD-OC segmentation. Factors such as cost, portability, and acquisition time (of OCT or stereo CF images) inhibit the widespread usage of these solutions in a large scale screening setting.

In monocular CF images, appearance of the pallor is characterized by the region of maximum color contrast within the OD and vessel bends as they enter into the cup. Using the pallor information alone [35] [48] leads to inaccurate OC boundaries as a distinct pallor region is often absent. Moreover, while in normal cases, pallor and the OC boundary appear nearby, glaucomatous cases have a much larger cup encompassing the pallor [49]. Therefore, additional information based on vessel kinks (detected using wavelet transform or curvature information) have been employed in [36], [50], and [51] to segment OC. Since a majority of blood vessels enter OC from the inferior and superior directions, boundary estimates in nasal and temporal sectors tend to be inaccurate. Further, only a small subset of locations where the vessels bend actually lie on the OC boundary, requiring several heuristics for selection of actual vessel kinks.

Supervised classification based methods have also been explored for OC segmentation [42]. A convolution neural network based method is explored in [52] where the filters are learnt over several layers. In [53], a supervised active shape model has been proposed. The initial boundary is represented by a set of landmark points which is iteratively refined using a cascade of regression functions. In each iteration, a separate regression function is learnt to map the image appearance features derived from the current boundary estimates to shape increment vectors which is used to refine the boundary.

2.2 Method

A rough Region of Interest (ROI) is extracted from the given CF image during pre-processing and provided as input to the proposed method for joint OD-OC segmentation. The ROI extraction is discussed in Section 2.2.1. Our proposed method depicted in Fig. 2.2. consists

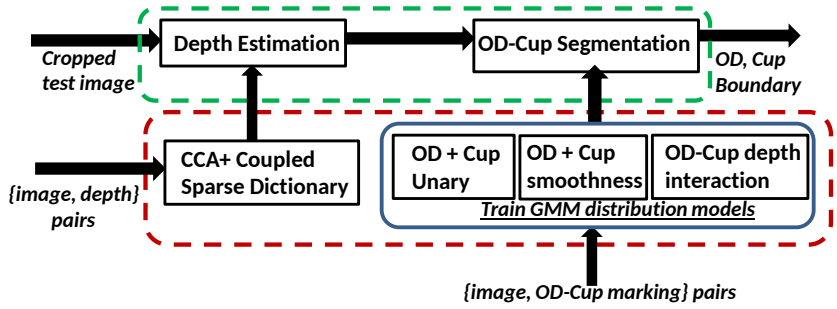

Figure 2.2: Outline of the proposed system. Training(test) modules are enclosed within red(green) dotted box.

of 2 stages : i) Supervised depth estimation from the input ROI; ii) extraction of OD-OC boundaries using the depth estimates and color gradients extracted from the ROI.

To estimate depth, Canonical Correlation Analysis (CCA) and Coupled Sparse Dictionary (CSD) basis are learnt from a set of fundus image-depth map pairs during training to relate image appearance to depth values. The learnt basis vectors are used during testing to estimate the depth from the CF images alone. The details are provided in Section 2.2.2.

The proposed CRF formulation for the joint OD-OC segmentation is presented in Section 2.2.3. Its various energy terms model the expected distribution of the color gradients and relative drop in depth between the OD and OC boundaries. The probability distributions are learnt from a separate set of training CF images images with manual OD and OC markings from experts.

2.2.1 ROI extraction

A Hough transform based algorithm based on [36] is used to localise the OD region. First, the candidate regions for OD are identified by thresholding the red channel of the CF images at 0.95 after normalizing it to [0,1]. Thereafter, the vessels are supressed within the selected candidate regions using morphological top-hat operation on the green channel to obtain a rough vessel mask followed by a diffusion based inpainting [54]. Next, an edge map is extracted by applying Canny edge detector at a very low threshold. Circular Hough transform is employed to obtain a rough estimate of the OD center and radius R. A rough alignment of the detected region is performed using a simple image processing based method by rotating the image in the range of ± 20 degrees about the detected OD center such that the distribution of the blood

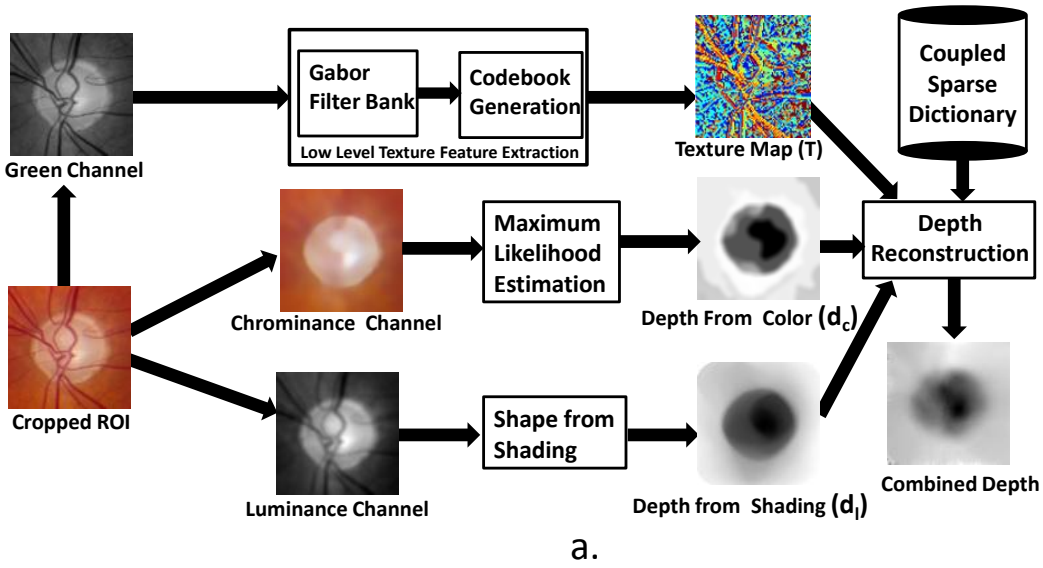

Figure 2.3: Overview of the proposed supervised depth estimation method.

vessels in the Inferior and Superior Sectors become roughly symmetric. Moreover, the images of the left eyes are detected based on the clinical knowledge that the nasal region tends to have a higher density of the thick blood vessels in comparison to the temporal side and reflected about the vertical axis to align the left and right eye images. Finally, a square region of interest (ROI) of size $3R$ (with a margin of 0.5 R on all sides) is extracted (see Fig. 2.1 a,b).

2.2.2 Supervised Depth Estimation

The proposed depth estimation pipeline is depicted in Fig. 2.3. Initially, two separate estimates of depth denoted by d_l and d_c are derived from the luminance and chrominance features respectively. While d_c is obtained by relating the color features at each pixel to probable depth values, d_l is derived from the intensity (gray-scale) image using an unsupervised Shape From Shading (SFS) algorithm. The details of extracting d_l and d_c are discussed below in Sections 2.2.2.1 and 2.2.2.2 respectively. Finally, d_l and d_c are integrated at a *patch level* along with Gabor filter-bank based texture features and mapped to ground-truth (GT) depth values using coupled sparse dictionary (CSD). The details of the patch level feature extraction in the appearance and depth feature spaces is described in Section 2.2.2.3 followed by the details of the CSD based mapping in section 2.2.2.4.

2.2.2.1 Depth estimation from Luminance

The luminance channel L is obtained as the average of the R,G and B channels of the color fundus image ROI followed by the suppression of the high color gradients using the method in [55]. The depth estimate d_l is obtained from L using the simple but fast unsupervised shape from shading algorithm in [56]. Some simplistic assumptions are made to have a tractable solution: i) retinal surface is assumed to be Lambertian. ii) The albedo (surface reflectivity) of the retinal surface is uniform. Since albedo of blood vessels is different from that of the retinal surface, vessel inpainted images obtained during ROI extraction are used. iii) The albedo and illumination direction is computed from the image itself following the method in [57] under the assumption that the surface normals are distributed evenly in the 3D space. Since, the SFS algorithm implicitly assumes bright regions to be closer to the camera, the complement of its output gives the actual depth estimate. Though an exact reconstruction of depth from a single view is not possible due to the simplified assumptions made in the SFS algorithm, results (Section. 2.4.1) indicate a strong positive correlation between d_l and the true depth values.

2.2.2.2 Depth estimation from Chrominance

The mean and variance of each of the R, G, B color channels of the vessel inpainted ROI in the fundus image is first standardized to fixed values (the average values computed from the INSPIRE dataset) and then normalised $(j/(r+g+b); j = r, g, b)$ to obtain a 3-long illumination invariant color feature C for each pixel. The colour-based depth estimate d_c is obtained from C using a supervised approach: for each depth value $d \in [0, 255]$, the conditional $P(C \mid d)$ is learnt from a training set of image-depth pairs, using a Gaussian Mixture Model (GMM) with number of Gaussians selected in the range 1-6 that maximizes the Akaike Information Criterion (AIC) [58]. During testing, the maximum a posteriori estimate for $P(d \mid C)$ is computed. The lack of one-to-one correspondence between the color and depth values and treating each pixel independent of its neighborhood leads to inaccuracies in the d_c estimates.

2.2.2.3 Patch-level Feature Extraction

To obtain a robust and accurate depth estimate, d_l and d_c are integrated at a patch level along with texture features to obtain the final depth estimate. Each 8×8 image patch at

location i is represented by a 444 dimensional *appearance* feature p_i extracted from the fundus image and a 192 dimensional *depth* feature q_i extracted from the corresponding patch in the GT depth map derived from OCT. While both image and the corresponding GT depth map d_{oct} is available during training, only fundus image is available during testing. The goal is to learn a mapping between the two feature spaces to predict the depth feature of a patch given its appearance feature.

To obtain the appearance feature, at first each pixel in the image patch is represented by a 6-D appearance feature f_a obtained by concatenating d_l,d_c and their gradients $\frac{\partial d_l}{\partial x}$, $\frac{\partial d_l}{\partial y}$, $\frac{\partial d_c}{\partial x}$, $\frac{\partial d_c}{\partial y}$. The i^{th} patch in a CF image is represented by a vector $p_i \in R^{444}$ obtained by concatenating f_a of all 64 pixels in the patch and a 60-bin histogram of a texture word map T within the patch.

T is extracted from the green channel of the fundus image. Each pixel is represented by the responses of a gabor filter bank comprising of 36 filters along with their 1^{st} and 2^{nd} order derivatives in the two directions resulting in a $(36 \times (1+2+2)$ =180-long feature which is clustered (during training) into 60 words. Each pixel is then represented by the nearest word index [59] to obtain T. The filter bank consists of gabor filters in 6 orientations ($\{0, 30, 60, 90, 120, 150\}$ degrees) at 6 scales ($\sigma = \{0.04 \times 1.6^s \mid 1 \leq s \leq 5, s \in Z\}$) with the aspect ratio and wavelength fixed at 0.5 and $\frac{\sigma}{0.56}$ respectively.

The *depth* feature is obtained by representing each pixel location by a 3-D depth feature vector f_d obtained by concatenating the GT depth d_{oct} along with it's gradients $\frac{\partial d_{oct}}{\partial x}$,$\frac{\partial d_{oct}}{\partial y}$. A patch-level aggregation of f_d results in the depth feature $q_i \in R^{192}$.

The publicly available INSPIRE dataset [60] is used to train our depth estimation method which provides the corresponding depth map d_{oct} for each fundus image as the GT. In INSPIRE, the depth maps were computed by additionally collecting the 3D OCT images corresponding to each fundus image. The depth of the retinal surface was extracted from the OCT and manually registered to the fundus image.

Thus, M image patches, are represented in the appearance and depth space by matrices P and Q whose M columns are p_i and q_i, respectively. The patch features yield a more robust representation relative to the individual pixel-level depth estimates. The dimensionality of P and Q is jointly minimized by employing CCA [61]. CCA projects both P and Q into two separate 192-D feature spaces. In contrast to other dimensionality reduction techniques such as PCA which can be applied to each feature space independently, CCA jointly computes a pair of

basis $\phi_{img} \in R^{444\times192}$ and $\phi_{depth} \in R^{192\times192}$ for the two feature spaces such that for each image patch, the correlation between its appearance and the depth feature is maximized. ϕ_{img} and ϕ_{depth} projects P and Q to $P_{cca} = \phi_{img}^T.P$ and $Q_{cca} = \phi_{depth}^T.Q$ respectively. The dimensionality reduction in the appearance features from 444 to 192 helps to reduce the risk of overfitting on the training data as well as decreases the memory and computational requirements for learning the coupled sparse dictionaries. The dimensionality of the depth features in Q is not reduced in Q_{cca} to enable an exact reconstruction of the depth feature vector using the inverse of the ϕ_{depth} basis.

2.2.2.4 Coupled Sparse Dictionary

The task of predicting a 192-D depth feature from the corresponding appearance feature is a multi-output regression problem. Therefore, the regression based methods that predict a single output are unsuitable for this purpose. CSD provides a way to model such mappings and has been shown to be effective in applications such as image super-resolution where the image patches in a low resolution are mapped to the corresponding image patches in the high resolution [62].

Majority of the sparse dictionary learning methods focus on training an over-complete dictionary in a single feature space for various signal recovery and recognition tasks such as image impainting [63], face recognition [64] and object tracking [65]. In contrast, the CSD considers two feature spaces which in our case are the appearance and the depth feature spaces respectively. The unknown mapping function between them is modelled by jointly learning a separate over-complete dictionary in each feature space such that the sparse representation of a feature in the appearance space can be used to reconstruct its paired feature in the depth space [62].

Let U and V denote the two overcomplete CSD, each with 1100 dictionary atoms, in the P_{cca} and Q_{cca} feature space respectively. The sparse code α is shared in the two representations: $P_{cca} \approx U.\alpha$ and $Q_{cca} \approx V.\alpha$ for all the training patches. Hence, estimation of U,V is posed as

$$argmin_{U,V,\alpha} \parallel P_{cca} - U.\alpha \parallel_2 + \parallel Q_{cca} - V.\alpha \parallel_2 + \lambda. \parallel \alpha \parallel_0 . \tag{2.1}$$

E.q. 2.1 can be rewritten by concatenating P and Q resulting in the standard sparse dictionary learning problem [62] which can be formulated as

$$argmin_{U,V,\alpha} \left\| \begin{bmatrix} P_{cca} \\ Q_{cca} \end{bmatrix} - \begin{bmatrix} U \\ V \end{bmatrix} .\alpha \right\|_2 + \lambda. \parallel \alpha \parallel_0 . \tag{2.2}$$

An online dictionary learning algorithm [66] was used for learning the sparse dictionary $\begin{bmatrix} U \\ V \end{bmatrix} \in R^{384\times1100}$ from the feature set $\begin{bmatrix} P_{cca} \\ Q_{cca} \end{bmatrix}$ using a batch size of 600, sparsity coefficient $\lambda = 0.6$ and max-iteration= 800. The learnt basis is split horizontally to obtain U and V.

Coupled Sparse Dictionary Testing The given test image ROI is densely sampled into overlapping patches which are represented in the patch level image space by P_{test}. The sparse code α^* is estimated by solving the LASSO optimization problem,

$$\alpha^* = argmin_\alpha ||U.\alpha - P_{test}||_2^2 \quad s.t. \quad ||\alpha||_1 \leq \lambda. \tag{2.3}$$

The corresponding estimated depth in CCA space, Q_{est} is then obtained by projecting α onto the depth basis V using

$$Q_{est} = V.\alpha^*. \tag{2.4}$$

Q_{est}, is backprojected onto the CCA basis to obtain $D_{est} = (\phi_{depth})^{-1}.Q_{est}$, which consists of the depth value d and its gradients $\frac{\partial d}{\partial x}$ and $\frac{\partial d}{\partial y}$ at each pixel. The refined depth value is taken as the average of d and the depth estimated from $\frac{\partial d}{\partial x}$ and $\frac{\partial d}{\partial y}$ using gradient inversion method in [55]. Thus, having estimated the depth map for a given CF image, next, we present details on how it is used to segment the OD and OC.

2.2.3 Joint OD-OC Segmentation

The proposed joint segmentation framework seeks to extract both the OD and OC boundaries from a given image ROI by formulating it as a CRF based energy minimization problem. The OD and OC boundaries are modelled as concentric closed curves about the ROI centre denoted by $O : (x_o, y_o)$. The two curves are parameterized by N points uniformly spaced in orientation which are represented in the polar coordinates by $(x_n^i, \theta_n), n \in \{1, 2, \ldots, N\}$. Here x_n^i represents the radial distance of the n^{th} point from O, $\theta_n = (n-1) \times \frac{360^\circ}{N}$ represents its angular orientation with respect to the horizontal and $i \in \{d, c\}$ represents points on the OD and OC boundary respectively.

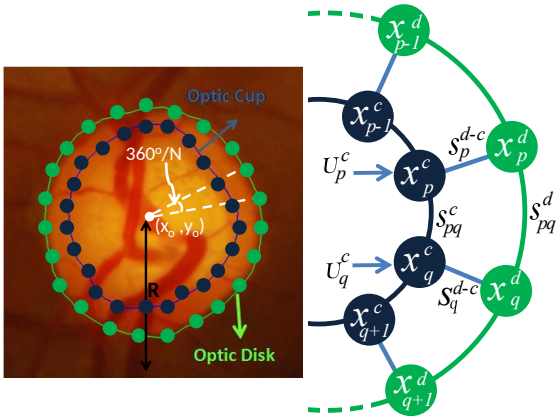

Figure 2.4: The parameterization and graphical model representation of OD-OC boundaries.

Each x^d_n and x^c_n is a discrete random variable that can take values from the label set $L = \{l | 1 \leq l \leq R, l \in \mathbb{Z}^+\}$, where R represents the radius of the largest circle that can be enclosed within the ROI and $\mathbb{Z}^+$ is the set of all positive integers. The set of random variables $X = \{x^d_n; x^c_n\}_{n=1}^N$ defines a *Random Field* and $\mathbf{x} \in L^{2N}$ denotes a feasible labeling of X obtained by assigning a label from L to each x^i_n. The graphical model corresponding to X is depicted in Fig.2.4, where each node belonging to OD and OC correspond to the random variables x^d_n, x^c_n and are colour coded green and blue, respectively.

For each random variable x^d_n, a Disc unary term $U^d_n(x^d_n = l)$ is defined to capture the probability that the OD boundary passes through the point $(l, (n-1) \times \frac{360^o}{N})$, given a set of features that capture the gradient characteristics of OD boundary at that point. An identical set of features is used to define the Cup Unary term $U^c_n(x^c_n = l)$ which captures the pallor gradient at cup boundary.

Moreover, based on the Markovian assumption, the label of each x^i_n is also considered to be dependent on its immediate neighbors in the same as well as the adjacent boundary. The smoothness of the OD and OC boundary is captured by the disc and cup smoothness energy terms denoted by $S^d_{p,q}(x^d_p, x^d_q)$ and $S^c_{p,q}(x^c_p, x^c_q)$ respectively. They are defined between each pair of adjacent nodes (p, q) lying in the OD or OC boundary. A disc-cup interaction term $S^{d-c}_n(x^d_n, x^c_n)$ is defined between the corresponding nodes across the two boundaries that lie in the same orientation to capture the relative drop in the depth between the OD and OC

boundaries. Thus, the CRF energy $E(X)$ is defined as

$$\begin{aligned} argmin_{\mathbf{x}} E(\mathbf{x}) = &\sum_{n=1}^{N} U_n^d(x_n^d) + \lambda_1 \sum_{p,q \in N_d} S_{p,q}^d(x_p^d, x_q^d) \\ &+ \lambda_2 \sum_{n=1}^{N} U_n^c(x_n^c) + \lambda_3 \sum_{p,q \in N_c} S_{p,q}^c(x_p^c, x_q^c) \\ &+ \lambda_4 \sum_{n=1}^{N} S_n^{d-c}(x_n^d, x_n^c), \end{aligned} \tag{2.5}$$

where N_d and N_c represents the set of adjacent pair of nodes in the OD and OC boundaries respectively and defined as $N_i = \{(x_k^i, x_{mod(k,N)+1}^i) | 1 \leq k \leq N\}$. The labeling $\mathbf{x}$ that minimizes $E(\mathbf{x})$ corresponds to the desired segmentation. During implementation, e.q. 2.5 is solved using the Sequential Tree Re-weighted Message Passing (TRW-S) algorithm in [67]. The details of the various energy terms are described below.

Disc and cup Unary Terms : The Unary cost terms attempt to model the edge information at OD and OC boundaries. A given image location at a radial distance of l from the center of the ROI in the direction θ_n is represented using a 5-D feature vector $\mathbf{f}_{n,l}$. The feature consists of the image gradient magnitudes *computed along the radial direction* θ_n in a 31×31 neighborhood in *R of RGB, V of HSV, and all 3 color channels in YCbCr color space* respectively. The gradient values along each radial direction are further normalized to [0,1]. The neighborhood size of 31×31 was experimentally determined and the good performance of a relatively large neighborhood size could be attributed to the fact that in many cases the OD and OC boundaries are characterized by a gradual transition in color, intensity and lack a sharp distinct edge. The exploration of different color spaces was inspired by the existing methods such as [42] which have used a combination of different color models for OD and OC Segmentation.

The probability distributions of the color gradients at the OD and OC boundatries denoted by $P_n^d(\mathbf{f}_{n,l})$ and $P_n^c(\mathbf{f}_{n,l})$ respectively, are modeled independently for each orientation θ_n using Gaussian Mixture models(GMM) in the $\mathbf{f}_{n,l}$ feature space. The GMMs are learned from the training images and their corresponding ground truth markings. Since, $E(X)$ is to be minimized, we define the unary terms $U_n^d(x_n^d = l)$ and $U_n^c(x_n^d = l)$ to be inversely related to the probability as

$$\begin{aligned} U_n^d(x_n^d = l) &= 1 - P_n^d(\mathbf{f}_{n,l}) \; and \\ U_n^c(x_n^c = l) &= 1 - P_n^c(\mathbf{f}_{n,l}). \end{aligned} \tag{2.6}$$

Smoothness Terms : The smoothness cost terms denoted by S^d and S^c for the OD and OC boundaries respectively, attempt to model their smoothness. The probability distribution of the difference between the radial distances of two adjacent boundary points are modeled using univariate GMMs for each orientation θ_n independently and learned from the ground truth markings of the training images. The Pairwise Smoothness cost terms in the CRF are defined as

$$\begin{aligned} S^d_{p,q}(x^d_p = l, x^d_q = m) &= 1 - P^{(p,q)}_d(\mid l - m \mid) \; and \\ S^c_{p,q}(x^c_p = l, x^c_q = m) &= 1 - P^{(p,q)}_c(\mid l - m \mid). \end{aligned} \tag{2.7}$$

Here, $P^{(p,q)}_d$ and $P^{(p,q)}_c$ represents the probability distribution of the spatial variation in the adjacent nodes, $(x^d_p, x^d_q) \in N_d$ and $(x^c_p, x^c_q) \in N_d$ on OD and OC boundary respectively.

Disc-Cup Interaction Term : This pairwise term captures the relationship between corresponding landmark points on the disc x^d_n and the cup x^c_n along the direction θ_n. A radial profile of length R is extracted from the estimated depth map D along θ_n and normalized in the range $[0, 1]$ such that the deepest point has a depth of 1. Rather than using an empirical ratio for the drop in depth (for e.g., one-third was used in [45]), we learn its probability distribution for each θ_n separately from the training images. Thus, the disc-cup interaction term is defined as

$$S^{d-c}_n(x^d_n, x^c_n) = \begin{cases} 1 - P^d_n(D(x^d_n) - D(x^c_n)), & \text{if } x^d_n \geq x^c_n \\ \infty, & \text{otherwise,} \end{cases} \tag{2.8}$$

where P^d_n models the drop in depth $D(x^d_n) - D(x^c_n)$ between OD and cup boundary in the n^{th} direction and modeled using a univariate GMM. Configurations in which the radial distance of cup landmark x^c_n exceeds that of the corresponding disc landmark x^d_n are infeasible because OC always lies within the disc. This is ensured by assigning ∞ (practically, a very large cost value during implementation) to such disc-cup interaction terms.

The probability distributions in eqs. 2.6, 2.7 and 2.8 are modeled using GMMs and learned from expert-marked OD and OC boundaries in the training images using Expectation-Maximization. The optimal number of Gaussians in the GMMs are selected by searching in the range between 1-5 and selecting the one which *maximizes* the AIC [58]. If k represents the number of Gaussian distributions in the GMM, and $\hat{L}$ represents the likelihood ie., the probability of obtaining the training data from the GMM then AIC is given by $2.ln(\hat{L}) - 2k$.

The AIC metric derived from information theory rewards the goodness of fit (assessed by the likelihood function) but penalizes the number of estimated parameters (k) to prevent overfitting on the training data.

2.3 Materials

The proposed method has been evaluated on six datasets: five publicly available, namely, INSPIRE [60], DRISHTI-GS1 [6], RIM-ONE version 2 [68], DRIONS-DB [69] , MESSIDOR [5] and a privately collected dataset referred to as DATASET-1. The datasets cover a range of ethnicity in population, imaging protocols, fundus cameras and image quality. Comprehensive experiments were performed aimed at testing the segmentation and its use in glaucoma diagnosis. Depending on the availability of Ground Truth (GT), different evaluation criterions were used for each dataset. A summary of each dataset and the experiments performed on them is provided in Table 2.1. GT for *MESSIDOR* is available from University of Huelva [70] which has been used to benchmark several OD segmentation algorithms [71],[29],[72]. The locally sourced *Dataset-1* contains 18 normal and 10 glaucomatous images acquired from Aravind Eye Care Hospital, Madurai, India for which both CF and OCT imaging is available. The CDR in the OCT generated report of corresponding fundus images was taken as the gold standard. Structural markings of OD-OC boundaries by an expert was also collected for comparison.

Images in all datasets are OD-centric with the exception of *MESSIDOR* that provides macula-centric images. *DRIONS-DB* contains challenging cases such as illumination artefacts, blurred or missing rim, peripapillary atrophy and strong pallor distractor [73]. The INSPIRE dataset provides normalized ground-truth depth maps for each CF image obtained using manual layer segmentation and registration of corresponding 3D-OCT images.

2.4 Results

The supervised depth estimation method is evaluated in Section 2.4.1. Various experiments to evaluate the joint OD-OC segmentation is presented in Section 2.4.2. This includes a *comparison with other state of the art* methods (5.2.2); errors in computed *CDR* (5.2.3) and the *rim thickness error across sectors* (5.2.4). The utility of the segmentation in *glaucoma detection* is evaluated in Section 2.4.3.

Dataset	# images	Protocol	Camera	Source	Ground Truth	Evaluation
INSPIRE	30	4096×4096	Nidek 3Dx digital stereo	USA	OCT based depth	Depth Estimation
DRISHTI-GS1	50 train, 51 test	2896×1944 30^o FOV	Zeiss Visucam NM/FA	India	Manual OD, OC marking	a) OD,OC segmentation b) CDR error
Dataset-1	28	2896×1944 30^o FOV	Zeiss Visucam NM/FA	India	CDR from OCT & an expert marking	CDR error
RIM-ONE v2	455	cropped ROI around OD	Nidek AFC-210	Spain	image-level glaucoma diagnosis	Glaucoma classification
DRIONS-DB	110	600×400 OD-centric,	digitized using HP-Photo Smart-S20	Spain	manual OD marking	OD segmentation
MESSIDOR	1200	1440×960, 2304×1536 , 2240×1488, 45^o FOV	Topcon TRC NW6	France	manual OD marking	OD segmentation

Table 2.1: Dataset Specifications. All datasets are publicly accessible except $Dataset-1$.

2.4.1 Evaluation of depth estimation

Since, OC boundary is primarily defined by the *relative* drop in depth values, depth maps defined up to an arbitrary scale factor contain sufficient information for cup segmentation. Moreover, the ground-truth depth maps in the INSPIRE dataset used in our evaluation are normalized to [0,255]. Hence, *Pearson product-moment correlation coefficient* ρ is used for assessment. $\rho \in [-1, 1]$ where -1 (or 1) indicates a total negative (or positive) correlation while 0 indicates no correlation. It is defined as

$$\rho(d, D) = \frac{\sum_m \sum_n (d_{m,n} - \bar{d})(D_{m,n} - \bar{D})}{\sqrt{(\sum_m \sum_n (d_{m,n} - \bar{d}))^2 (\sum_m \sum_n (D_{m,n} - \bar{D}))^2}}, \tag{2.9}$$

where d (or D) denotes the computed (or GT) depth map with mean $\bar{d}$ ($\bar{D}$) and (m, n) denote pixel locations. Due to a limited data availability, Leave-One-Out cross-validation was employed for evalutation, using 29 images to train and 1 image to test in each fold. Qualitative results are depicted in Fig. 2.5 where darker pixels in the 2^{nd} and 4^{th} columns indicate higher depth values.

	Pearson Correlation Coeff.	
	With Preprocessing	Without Preprocessing
Color Features alone (GMM regression)	0.72/0.12	0.71/0.09
Shape from Shading (unsupervised)	0.76/0.13	0.74/0.14
Combined using L2- ridge regression	0.77/0.14	—
Combined using Coupled Sparse Dictionary	0.80/0.12	0.80/0.11

Table 2.2: Performance of the Depth Estimation. (mean/standard deviation)

The first column in Table 2.2 presents the performance of the combined depth estimate as well as the individual contributions of the different depth cues employed in the proposed method. Depth estimation using the unsupervised Shape from Shading (SFS) alone performs better than mapping color intensities directly to corresponding depth values using GMM with a Pearson Correlation Coefficient (ρ) of 0.76 and 0.72 respectively. The proposed method which combines cues from SFS, GMM based depth from color and texture features using the coupled sparse dictionary(CSD) further improves the depth estimation with a ρ of 0.80. The proposed CSD based method was benchmarked against L2-normalized ridge regression that maps the appearance features to corresponding depth values at the central pixel location in each patch. The method resulted in a ρ of 0.77 in comparison to 0.80 obtained by the proposed method. The better performance of the CSD based method can be attributed to the fact that it allows a richer representation of the depth feature space consisting of depth values as well as its gradients that capture the local variation at a patch level.

The proposed method employed simple preprocessing steps such as a rough inpainting of the blood vessels using morphological operations within the region of interest and a grayworld intensity normalization to reduce the effect of uneven illumination. To assess the impact of these preprocessing steps, the performance of depth estimation was evaluated without any preprocessing and the results have been presented in the second column of Table 2.2. While there is a slight improvement in the individual depth estimates with the preprocessing, it didnot have any impact on the final combined depth estimate.

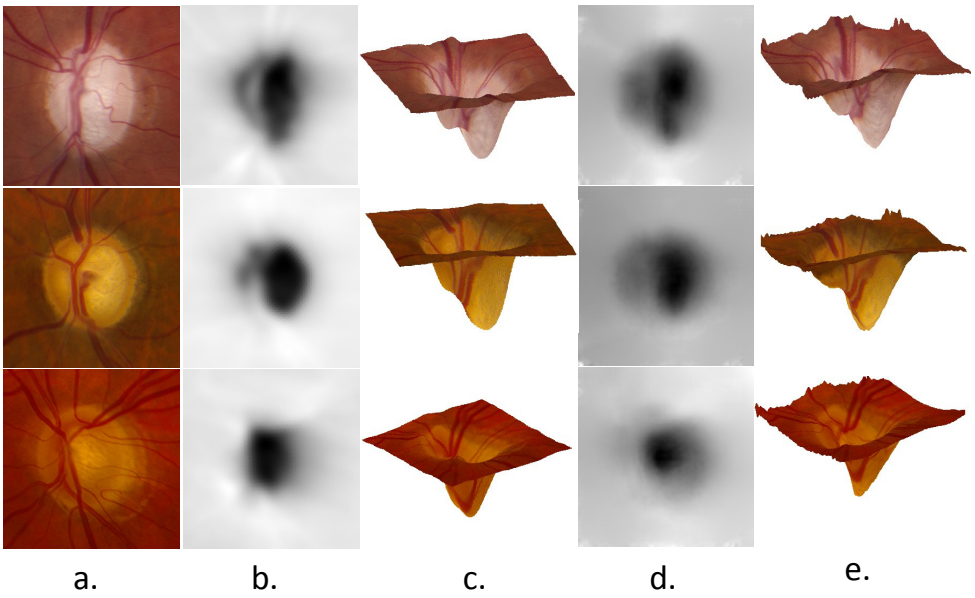

Figure 2.5: OD regions from 3 sample images (column a). Corresponding depth estimates are shown as greysacle image and topographical maps with input image wrapped onto depth surface. Columns b,c are ground truth; Col. d,e are computed results.

2.4.2 Evaluation of OD-OC Segmentation

2.4.2.1 Evaluation Metrics

Both region and boundary localization based metrics are used to evaluate the segmentation results. *Dice* similarity coefficient $D(X,Y)$ measures the extent of overlap between the set of pixels in the segmented region X and ground truth Y and can be defined as

$$D(X,Y) = \frac{2.\mid X \cap Y \mid}{\mid X \mid + \mid Y \mid}. \tag{2.10}$$

Since Dice coefficient is unsuitable to gauge the segmentation performance at a local (boundary) level, the average *Boundary Localization Error (BLE)* is also employed to measure the distance (in pixels) between the computed (C_o) and GT (C_g) boundaries. It is defined as

$$BLE(C_g, C_o) = \frac{1}{n}\sum_{\theta=1}^{\theta_n} \mid r_\theta^g - r_\theta^o \mid, \tag{2.11}$$

where r_θ^o (r_θ^g) denotes the radial euclidean distance of the estimated GT boundary point from the centroid of the GT in the direction θ; 24 equi-spaced points were considered in the evaluation. The desirable value for BLE is 0.

2.4.2.2 Benchmarking against state of the art

The segmentation performance of the proposed method was compared with 4 other methods on DRISHTI-GS1. Sample qualitative results are shown in Fig. 2.6. Quantitative results are presented in Table 2.3. The best figures for Dice and BLE are shown in bold fonts. The proposed method is seen to achieve the best figures consistently. The exception is [42] which achieves the best figures for OC segmentation on the training set, which however, degrades on the Test set. The high dimensionality (1025) feature-based superpixel method [42] tends to overfit the training data. The next best-performing method is the depth-based method in [74]. This however, requires two CF images per eye (during train and test phase) unlike our method which requires single CF image per eye during testing, thus, providing significant advantage. [75] also adopts a joint OD-OC segmentation framework but treats it as a pixel labeling problem. It requires direct edges in a 15 pixel neighborhood for each pixel and doesnot consider depth information. In contrast, the proposed method uses a *boundary based* formulation requiring labeling of only 72 landmark points and lends a natural way to model the depth-based inter-dependence between OD and OC by 36 edges (OD-cup interaction term) resulting in improvement in speed as well as accuracy as seen in the results of [75] on DRISHTI-GS1 in Fig. 2.6 and Table 2.3.

The statistical significance of the improvement in performance of the proposed method has been evaluated using a paired T-test against the segmentation results of the benchmark methods in Table 2.4. The Multi-view [74] method explored a novel method for OC segmentation alone and employed the same method as the Vessel Bend [36] based method for OD segmentation. On the task of OD segmentation, the proposed method achieved a statistically significant improvement in performance in comparison to the Superpixel [42] and Graph Cut prior [75] based method but the marginal improvement of our method (dice of 0.97 against 0.96) over the [36] was not found to be statistically significant. On the task of OC segmentation, the proposed method achieved a statistically significant improvement against all the benchmark methods.

Our method also outperforms the recent methods in [37] and [52] on OD segmentation and is comparable to [52] on the OC segmentation task. The unsupervised active disc energy based method in [37] reported a dice of 0.91 for OD on the combined training and test set of DRISHTI-GS1. The deep convolution network based method in [52] achieved a dice coefficient

Method	Training				Testing			
	OD		Cup		OD		Cup	
	Dice	BLE	Dice	BLE	Dice	BLE	Dice	BLE
Sequential approaches								
Vessel Bend [36]	0.96/0.05	8.61/8.89	0.74/0.20	33.91/25.14	0.96/0.02	8.93/2.96	0.77/0.20	30.51/24.80
Superpixel [42]	0.95/0.04	9.39/6.81	**0.84/0.12**	**16.89/8.1**	0.95/0.02	9.38/5.75	0.80/0.14	22.04/12.57
Multiview [74]	0.96/0.05	8.61/8.89	0.77/0.17	24.24/16.90	0.96/0.02	8.93/2.96	0.79/0.18	25.28/18.00
Joint approaches								
Graph cut prior [75]	0.93/0.06	12.94/11.67	0.76/0.15	29.65/ 18.01	0.94/0.06	14.74/15.66	0.77/0.16	26.70/16.67
Proposed	**0.97/0.02**	**6.42/3.36**	**0.84/0.14**	17.44/12.80	**0.97/0.02**	**6.61/3.55**	**0.83/0.15**	**18.61/13.02**

Table 2.3: Segmentation performance on DRISHTI-GS1

of 0.95 for OD and 0.83 for OC respectively using a five fold cross-validation on the training set of DRISHTI-GS1. While the performance of [52] for OC segmentation is comparable to our method on the training set, its performance on the test set is not available for comparison.

	vs. Superpixel [42]	**vs. Graph Cut prior [75]**	**vs. Multiview [74]**	**vs. Vessel Bend [36]**
Dice				
Optic Disc	8×10^{-5}	5×10^{-4}	—	0.400
Optic Cup	0.0266	8×10^{-5}	2×10^{-6}	8×10^{-5}
Boundary Error (pixels)				
Optic Disc	8×10^{-4}	4×10^{-4}	—	0.2770
Optic Cup	0.0241	2×10^{-4}	2×10^{-6}	4×10^{-5}

Table 2.4: p-values for the paired T-test between the proposed and the benchmark methods. A p-value < 0.05 indicates statistically significant improvement.

The OD segmentation results of the proposed method has also been reported on DRIONS-DB in Table 2.5 and MESSIDOR in Table 2.6. Though the proposed method jointly computes both OD and OC segments, the accuracy of OC could not be evaluated on these datasets in the absence of ground truth cup markings. The results reported in [29], [71], [72], [73], [76], [31], [77], [32] have been reproduced in Table 2.5 and 2.6 to benchmark the performance of the proposed method. While Dice coeffcient was used as the region-based metric for DRIONS-DB, Jaccard similarity coefficient J was used for MESSIDOR following the norms of the previously published results for a direct comparison. J also measures the extent of overlap between the

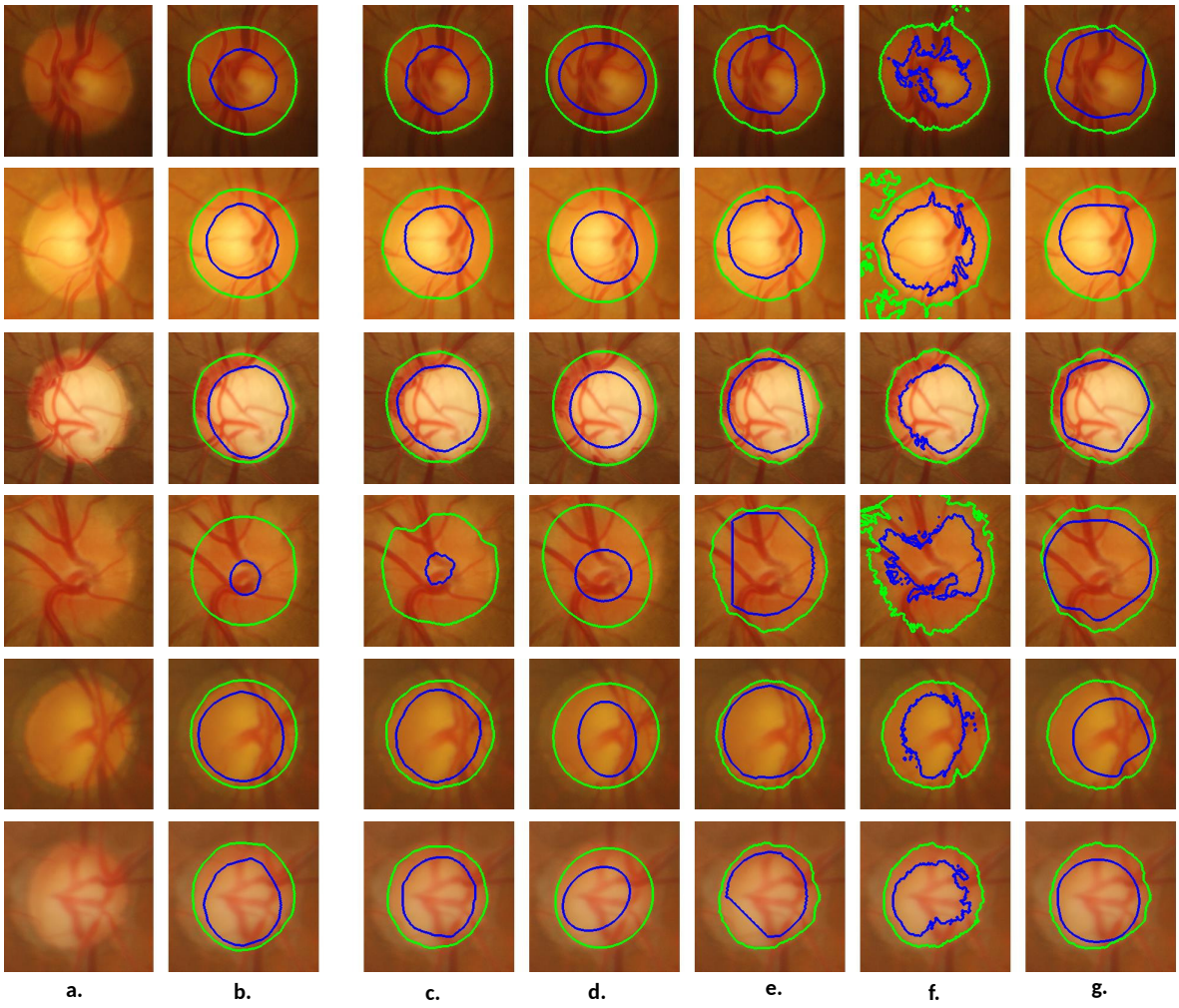

Figure 2.6: Qualitative results on some challenging cases. Disc and cup boundaries are depicted in green and blue respectively. a. cropped region around OD; b. Ground Truth; Results of c. Proposed Method; d. Superpixel based [42]; e. Multiview [74] f. Graph Cut prior [75] g. Vessel bend [36]

set of pixels in the segmented region X and ground truth Y as

$$J(X,Y) = \frac{\mid X \cap Y \mid}{\mid X \cup Y \mid}. \tag{2.12}$$

On DRIONS-DB, the proposed method outperforms the state of the art and performs comparably to human experts, while in MESSIDOR, the performance is comparable to other state of the art methods.

	Dice	BLE
Walter et al. [77]	0.68/0.39	-
Morales et al. [76]	0.88/0.17	-
Morales et al. [73]	0.91/0.10	-
Proposed Method	0.95/0.03	2.27/1.24
Expert 2	0.96/0.03	2.11/1.13

Table 2.5: OD segmentation performance on DRIONS-DB. (mean/std)

	Jaccard	BLE
Morales et al. [73]	0.82/0.14	-
Yu et al. [72]	0.84	-
Roychowdhury et. al. [31]	0.84	-
Aquino et al. [29]	0.86	-
Marin et. al. [32]	0.87	-
Giachetti et al. [71]	0.88	-
Proposed Method	0.87/0.22	4.32/4.71

Table 2.6: OD segmentation performance on MESSIDOR. (mean/std)

2.4.2.3 CDR analysis

Vertical CDR is a key metric widely used in the clinical assessment of glaucoma. The proposed method was evaluated on 2 datasets: DRISHTI-GS1 and Dataset-1. For Dataset 1, corresponding OCT reports were also available and used as Gold standard. The range of estimated CDR values is plotted in Fig. 2.7. While the median CDR values for the healthy and glaucomatous classes are well separated (0.6 versus 0.75), for both datasets, the standard deviation is higher for Dataset-1. While a clear separation between the 2 classes cannot be obtained from the estimated CDR alone, it can act as a useful feature in automated glaucoma classification. The mean and standard deviation of absolute error in estimated CDR are listed in Table 2.7. On the DRISHTI-GS 1 test set, the proposed method outperformed the Vessel Bend [36], Multiview [74] and the Superpixel [42] based methods with a mean CDR error of 0.08. However, the proposed method's CDR error increases to 0.16 when compared to the OCT based Gold standard in Dataset-1. In comparison, human experts achieved a better estimate of CDR against the Gold standard with a mean error of 0.04.

	Normal	Glaucoma	Combined
	DRISHTI-GS1 Train Set		
Proposed Method	0.11/0.11	0.07/0.05	0.08/0.08
	DRISHTI-GS1 Test Set		
Vessel Bend [36]	0.29/0.19	0.08/0.07	0.14/0.15
Multiview [74]	0.28/0.21	0.08/0.09	0.13/0.16
Superpixel [42]	0.13/0.11	0.10/0.06	0.11/0.08
Proposed Method	0.14/0.15	0.06/0.05	0.08/0.09
	Dataset-1		
Proposed Method	0.16/0.19	0.16/0.24	0.16/0.20
Expert 1	0.05/ 0.03	0.03 / 0.03	0.04/ 0.03

Table 2.7: CDR error (mean/std.) of the proposed method. Lower values indicate better performance.

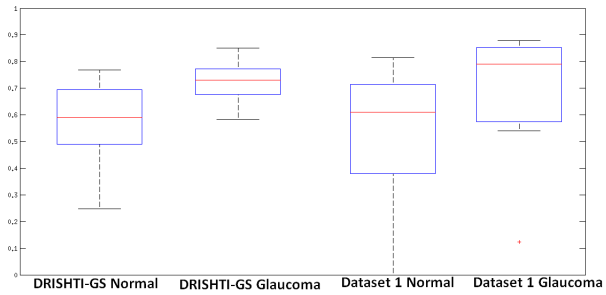

Figure 2.7: Box plots representing the distribution of vertical CDR values estimated with the proposed method.

2.4.2.4 Analysis of Rim thickness

The distribution of the overall rim thickness along all sectors of OD also plays an important role in glaucoma diagnosis. Rim thickness ratio (RTR) is defined as the ratio of cup radius to disc radius along different orientation angles about the OD center. The absolute error in RTR between the estimated and GT segmentations was evaluated. The overall mean/std. in RTR error was 0.102/0.098 for the Test set and 0.096/0.094 for Training set of DRISHTI-GS1 across all orientations. Fig. 2.8. depicts the distribution of RTR error across 360 degrees, 4 degrees apart. The smooth trend indicates a lack of regional bias in rim thickness error as opposed to vessel-kink based methods which tend to be more erroneous in the nasal and temporal regions due to absence of vessels in these sectors.

2.4.3 Glaucoma Screening

The utility of our segmentation method in the detection of glaucoma was evaluated on RIM-ONE v2. The OD-OC segmentations for the 455 images in RIM-ONE was obtained by training the proposed system on DRISHTI-GS1. Thereafter, a 14D feature was extracted from the

segmentations: a) vertical CDR and rim to disc area ratio; b) vertical OD diameter; c) ratio of horizontal to vertical CDR d) 6 features computed by $\frac{r_i - r_j}{\sigma_i + \sigma_j}$ with $i \neq j$ and $(i, j) \in \{(I, S, N, T)\}$; e) $\sigma(r_k)$with $k \in \{I_N, I_T, S_N, S_T\}$ to capture local deformations in these sectors. I,S, N and T represents the Inferior, Superior, Nasal and temporal sectors of OD (see Fig. 2.1 b), which are further sub-divided into Inferio-nasal (I_N), inferio-temporal (I_T), superio-nasal (S_N), superio-temporal (S_T) quadrants. The mean and standard deviation of the rim thickness between the OD and OC boundaries are denoted by r_i and σ_i respectively. A five fold cross-validation was performed using SVM with RBF kernel as the classifier. *The system achieved an average area under the curve (AUC) of* 0.85 ± 0.04 *and an average accuracy of* 0.77 ± 0.02. *AUC decreases to* 0.71 ± 0.03 *with only CDR, indicating that accurate segmentation in all sectors is critical to improve the classification performance.* The ROC curves of the 5 folds is shown in Fig. 2.9.

2.5 Discussion

In this work, a joint boundary based OD-OC segmentation formulation is proposed that explicitly models the drop in depth between the two boundaries. In the absence of true depth in 2D color fundus images, a supervised method is employed to estimate the depth from the image appearance. The proposed method outperformed several existing methods on the task of OC segmentation in the DRISHTI-GS1 dataset. For OD segmentation, the proposed method outperformed the existing methods on DRIONSDB and DRISHTI-GS1, while the performance was comparable to the state of the art on the MESSIDOR dataset. However, in terms of error

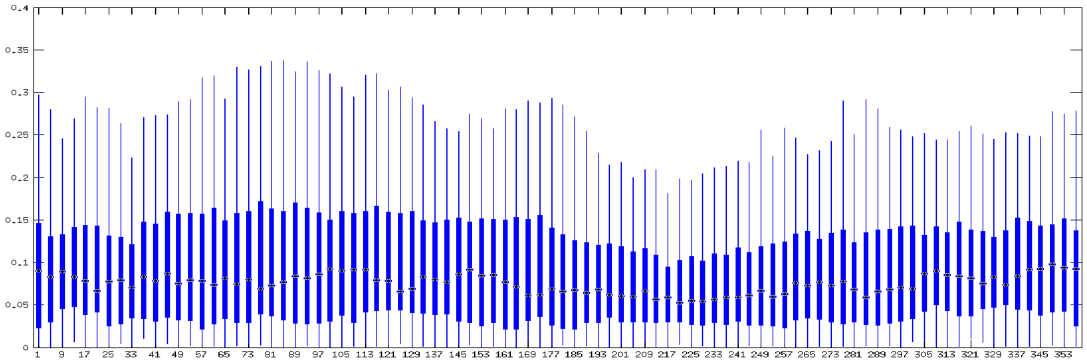

Figure 2.8: Box plots representing the distribution of the absolute error in rim thickness ratio (ratio of cup radius to disc radius) for 90 equidistant orientations, 4 degrees apart along the X-axis.

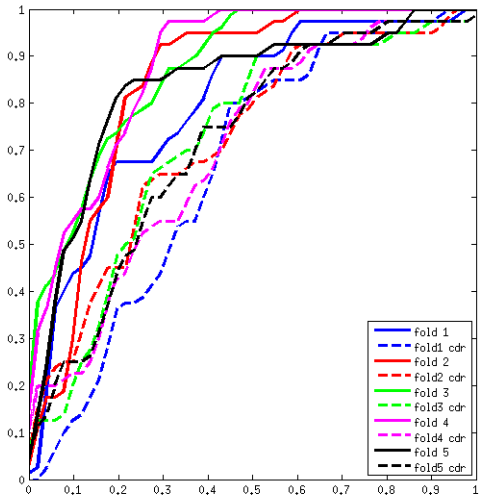

Figure 2.9: ROC curves of glaucoma classification on RIM-ONE v2 for five fold cross-validation.Solid and Dotted lines indicate performance obtained using 14-D feature and CDR alone respectively.

in CDR estimates, a human expert was found to surpass our method's performance against the OCT based Gold standard, indicating a scope for further improvement. An analysis of the rim-thickness error across all directions did not indicate any sector-wise regional bias in the segmentation accuracy. A set of features extracted from the OD-OC segmentation was able to achieve an AUC of 0.85 on the task of glaucoma classification on the RIM-ONE dataset indicating the potential use of the proposed method in glaucoma screening.

The proposed CRF formulation represents the OD and OC boundaries in polar coordinates in terms of its radial distance from the center of the extracted ROI around the OD. As a result, the proposed method assumes a reasonably accurate localizsation of the ROI and would fail if the center of the ROI is not contained within the OC region. However, empirically we found a simple image processing based method for the ROI extraction based on intensity thresholding and Hough transform based circle detection to work reasonably well on a diverse set of color fundus images.

The proposed method has been implemented in Matlab and it takes an average of 9.3 seconds per image to obtain OD and OC segmentations on a 3 Ghz, i7 processor with 8GB RAM. A comparison of the average computation time during testing in Table 2.8 shows that the proposed method is faster than some of the comparable benchmark methods.

Proposed	Multiview [74]	Superpixel [42]
9.3	12.6	22.4

Table 2.8: Average Computation time (seconds/image)

	Train		Test	
(λ_1,λ_2,λ_3,λ_4)	Dice	BLE	Dice	BLE
OD Segmentation				
A:(0, 0, 0, 0)	0.96/0.02	6.84/3.17	0.96/0.02	7.97/4.92
B:(1, 0, 0, 0)	0.97/0.02	6.41/3.37	0.97/0.02	6.64/3.55
C:(0, 0.1, 1, 0.1)	0.96/0.02	6.87/3.17	0.96/0.02	7.90/4.80
D:(1, 0.1, 1, 0.1)	0.97/0.02	6.42/3.36	0.97/0.02	6.61/3.55
Cup Segmentation				
E: (1, 0.1, 1, 0.1)	0.84/0.14	17.44/12.80	0.83/0.15	18.61/13.02
F:(1, 0.1, 1, 0)	0.77/0.23	24.64/25.52	0.70/0.27	32.32/30.27
G: (1, 0.1, 0, 0.1)	0.78/0.09	28.73/10.83	0.74/0.12	33.96/11.65
H: (1, 0, 1, 0.1)	0.82/0.14	20.13/13.87	0.82/0.17	21.49/16.25

Table 2.9: Analysis of Energy Terms: λ_1,λ_2,λ_3,λ_4 are weights associated with the disc smooothness (λ_1), cup data (λ_2), cup smoothness (λ_3) and the OD-OC interaction (λ_4) terms.

The parameters λ_1, λ_2, λ_3 and λ_4 in eq. 2.5 control the relative weights of the unary, smoothness and the disc-cup interaction terms with unit weight to the Disc unary term. Their optimal values were determined via a grid search over the Training set of DRISHTI-GS1. Tuning was done by searching values in $\{10^{-i}| -4 \leq i \leq 4, i \in Z\}$ and the parameters were finally set to $(\lambda_1 = 1, \lambda_2 = 0.1, \lambda_3 = 1, \lambda_4 = 0.1)$.

To analyze the impact of the various terms in eq 5. on OD and OC segmentation, we performed an additional set of experiments (Table 2.9) on the DRISHTI-GS1 test set by setting the weights of one or more of the energy terms to 0 while retaining the optimal weights for others. *OD segmentation* was analysed with parameter settings A through D. In setting A, the smoothness and pairwise terms are set to zero and the disc unary term alone is seen to be adequate for fairly accurate segmentation ($BLE = 7.97$ pixels on Test set). Addition of the disc smoothness term (setting B) marginally improves the BLE by 1.33 pixels on the Test set. The OD-cup interaction term has no significant impact (settings C and D) on OD segmentation.

OC segmentation was analysed in settings E through H. Inclusion of cup smoothness and OD-OC interaction terms are seen (Table 2.9) to significantly improve segmentation performance. The Dice value improves for these cases by 0.09 (E versus G) and 0.13 (E versus F), while BLE is nearly halved in both cases. The cup unary term (E versus H) only leads to a marginal improvement. The significant contribution from OD-OC interaction to OC segmentation indicates that the change in depth is more reliable in defining the OC boundary in comparison to the pallor edge information captured by the cup unary term. In contrast, the OD boundary is primarily defined by the color gradients captured by the disc unary term.

2.6 Conclusions

In this Chapter, we proposed a joint OD-OC segmentation framework that modeled depth based interaction between the OD and OC boundaries, using supervised depth estimates from the fundus image in addition to the color gradients at the OD and pallor boundaries. Despite training the supervised depth estimation method on just 30 fundus image-depth map pairs from the INSPIRE dataset, there was a good correlation between the estimated and true depth maps (mean ρ of 0.80).

The proposed joint segmentation method outperformed several state of the art methods including the existing joint OD-OC segmentation method in [75] for both OD and OC on multiple public datasets.

A comprehensive set of features derived from the OD-OC segmentation was able to distinguish between the Normal and glaucomatous images with an AUC of 0.85 on the RIM-ONE dataset. The experiments indicated that accurate segmentation in all sectors is critical for achieving good classification performance rather than the vertical direction alone.

Chapter 3

Joint Multi-layer Segmentation for Retinal Optical Coherence Tomography Images

Optical Coherence Tomography (OCT) is a non-invasive imaging modality that provides a 3D, cross-sectional view of the retinal tissue. The intra-retinal tissue is a multi-layered structure that can be anatomically divided into 7 adjacent layers [13], separated by 8 boundaries as depicted in Fig. 3.1. The boundaries ordered from the top to bottom are the : i) Inner Limiting Membrane (ILM) separating the vitreous and Nerve Fiber Layer(NFL), ii) NFL/GCL boundary separating NFL from the Ganglion Cell and Inner Plexicon layer (GCL-IPL), iii) IPL/INL separating GCL-IPL from the Inner Nuclear Layer (INL), iv) INL/OPL separating INL from the Outer Plexiform Layer (OPL), v) OPL/ONL separating OPL from the Outer Nuclear and Inner Segment (ONL-IS) region, vi) IS/OS separating ONL-IS from the Outer Segment (OS) vii)BM separating OS from the Retinal Pigment Epithelium (RPE) layer and finally the viii) RPE_{out} boundary separating RPE from the choroid. [1]

The accurate segmentation of these layers is necessary to quantify the morphological changes in the retinal tissue that characterize the presence and progression of various retinal diseases such as glaucoma [78], Diabetic Macular Edema (DME) [79] and Age-related Macular Degeneration (AMD) [14]. These morphological changes have also been correlated to neuro-degenerative diseases such as Multiple Sclerosis [80]. While the development of spectral domain OCT has led to the fast acquisition of a large number of B-scans per OCT volume in a short duration, their manual segmentation is a tedious, time-consuming and subjective task. Most commercial

[1] We refer to the discussion in Chapter 1, Section 1.1.2, pages 7-8 for further details on OCT imaging.

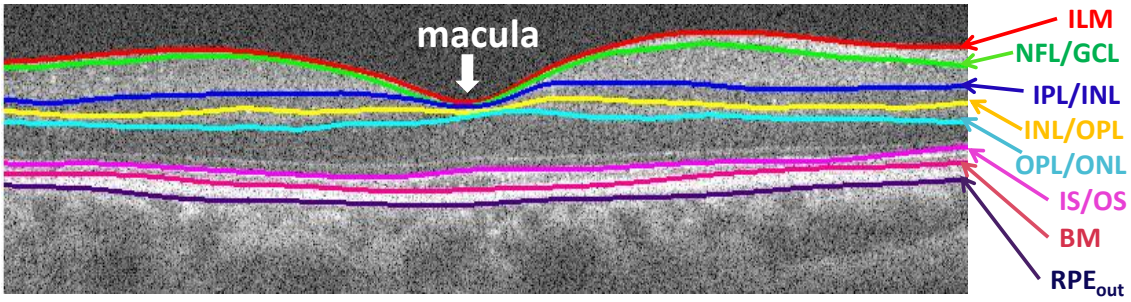

Figure 3.1: Retinal layer boundaries in a macular OCT B-scan.

OCT systems are equipped to segment only two or three layers and often fail in the presence of pathologies [81].

The main challenges in the automated layer segmentation are depicted in Fig. 3.2 and consist of the speckle noise, vessel shadows, indistinct layer boundaries, inter-scanner variations and the presence of pathologies. The vessel shadows are caused by the occurence of a retinal blood vessel at the beginning of an A-scan which absorbs the reflected light from locations beneath it [82]. Ocular diseases also lead to significant changes in the tissue morphology. In AMD, the drusen deposits in the RPE layer lead to irregularities and undulations in the BM boundary [14] as depicted in Fig. 3.2 b. DME is characterized by the presence of fluid-filled regions [15],[16] in the OPL and INL layers around the macula leading to the swelling of the retinal tissue as shown in Fig. 3.2 c.

In this chapter, we extend the work presented in Chapter 2 to explore a supervised Conditional Random Field (CRF) based framework for the joint multi-layer segmentation in OCT B-scans. The proposed CRF energy consists of multiple cost terms to capture the appearance and the shape priors for each layer. The appearance is captured by two convolutional filter banks, one to give high response at specific layer boundaries and the other to capture the appearance of each intermediate tissue region. The shape priors on the boundary smoothness and the thickness of each layer are modeled using Gaussian distributions. The CRF energy is *linearly parameterized* to allow an end-to-end training of its constituent cost terms (both the filter banks and the relative weights of the shape priors) by employing a Structural Support Vector Machine (SSVM) formulation. Being supervised in nature, our method can be adapted to different pathologies by training it on appropriate images without the need for handcrafting.

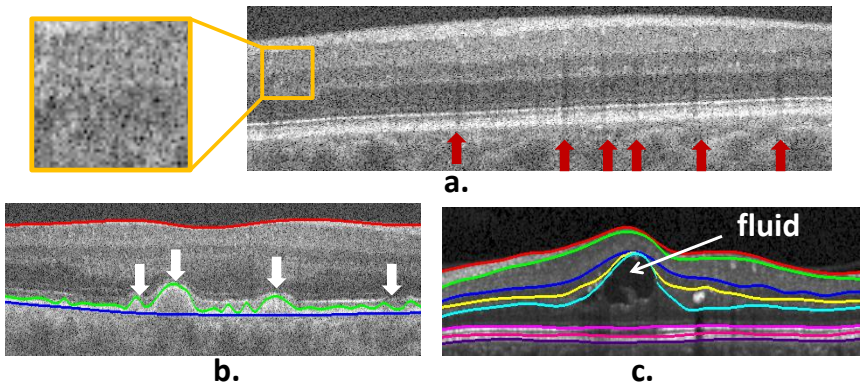

Figure 3.2: a. A retinal B-scan: a local patch is enlarged to depict the speckle noise and indistinct layer boundaries; the vessel shadows are indicated by red arrows. b. ILM(Red), BM (Green) and RPE_{out} (Blue) boundaries in a B-scan with AMD; irregularities in the BM boundary are indicated by white arrows. c. The 8 layer boundaries in a B-scan with DME; the fluid-filled region is indicated by a white arrow.

To summarize, the contributions of this work are as follows. First, we propose a joint segmentation framework based on a novel CRF formulation that extracts all retinal boundaries in a single optimization step. Second, a supervised strategy is explored to learn the CRF energy in a joint, end-to-end manner, eliminating the need to handcraft the individual cost terms or fine-tune their relative weights. Third, the robustness of the method has been demonstrated on data acquired across multiple centres with different scanners and at different resolutions. Finally, we have also evaluated the adaptibility of the proposed method to the morphological changes in the presence of pathologies related to AMD and DME.

3.1 Background

The initial attempts to segment the layers in retinal OCT images employed simple image processing techniques and focussed on the segmentation of only few (2-4) prominent layers. Each A-scan in the OCT slice was segmented individually based on peak, valley and/or signed gradient analysis of the intensity profile. This was followed by regularization across adjacent columns based on rule based heuristics [83] or iterative refinement by incorporating 3D information [84]. Their performance suffered due to the lack of a strong intensity gradient at the boundaries and overlapping intensities between the adjacent layers.

To overcome these challenges, deformable models have been proposed to incorporate shape priors in addition to the appearance of the retinal layers. The edge based active contour models were explored in [85], [86]. While constraints on the layer thickness and boundary smoothness were imposed in [85] using a coupled level set framework, an approximate parallelism constraint between the adjacent layer boundaries was incorporated in [86]. The region based Chan-Vese model was adapted in [87] to simultaneously segment multiple layers using a circular arc based shape regularization. An active appearance based statistical model for the layer shape and texture has also been explored in [88]. However, these methods require a good initial estimate of the layer boundaries failing which the deformable model can be entrapped in a local Energy minima or require a high convergence time.

Graph based optimization methods have also been explored for layer segmentation. In [13] the layers in each OCT B-scan were segmented by finding the shortest path in a specially constructed undirected graph using the Dijkstra's algorithm. The nodes in the graph represented the pixels and the edge weights were defined using the gradient intensity and the eucledian distance between the pixels in the image. It employed a sequential approach, restricting the search space for each layer based on the previously segmented layers. This method has been further refined in [89] to reduce the computation time by leveraging the spatial dependency between the adjacent B-scans of an OCT volume. Another method explored in [90], [91],[92] employs an energy minimization approach which is formulated as a Minimum Cost Closed Set (MCCS) problem on a specially constructed geometric graph. The Energy comprises multiple cost terms defined to capture the appearance and shape priors for each layer. However, each cost term is *handcrafted* manually and then combined using empirically determined relative weights.

Most of the deformable model and graph based energy minimization methods fail in the presence of pathologies. Few attempts have been made to adapt these methods by incorporating handcrafted disease specific modifications to handle a particular pathology. The Dijkstra's shortest path based method [13] has been adapted for DME in [93] by employing an explicit segmentation of the fluid-filled regions and attempts have also been made to adapt it for AMD cases in [94]. Similarly, the geometric graph based method has been adapted in [95] to handle Serous Pigment Epithelial Detachments and the coupled level set based deformable model [85] has been extended to handle DME cases in [96] by modelling the lesions as an additional space-

variant layer delineated by auxiliary interfaces. However, designing a single method that works equally well on both healthy and abnormal cases without the prior knowledge of the presence and the type of abnormalities in the OCT B-scan still remains an open problem.

Recently, deep learning based methods have also been explored for this task. In [97], a Convolutional Neural Network (CNN) based on [98] was applied to 33×33 image patches extracted from the OCT B-scans to obtain the probabilities of the central pixel in the patch for each layer boundary. Alternatively, Fully Convolutional Network (FCN) have been employed in [99], [100] and [101] to obtain the probability maps for each tissue layer. Unlike the patch based method, FCNs estimate the per-pixel class probabilities for the entire image in a single forward pass.

The U-net architecture [102] was employed in [99] and [100] which consists of a contracting path of an encoder block followed by an expansive path of a decoder block. Additional skip connections are also employed in these architectures to directly provide the output feature maps from each encoder layer as input to the corresponding decoder layer. The *ReLayNet* architecture in [99] is comprised of an encoder with four convolutional layers of 64, 7×3 filters separated by max-pooling layers. The decoder consists of a sequence of four unpooling and convolutional layers (each with 64, 7×3 filters) to successively restore the resolution of the feature maps. The *ReLayNet* was trained on sub-images obtained by slicing the B-scans width-wise into a set of non-overlapping regions, each consisting of 64 A-scans. Alternatively, in [101], the Dense-net [103] architecture was explored where, for each layer, the feature-maps of all the preceding layers were used as direct inputs using skip connections.

The main advantage of CNNs is their ability to learn hierarchical features in a data-driven end-to-end manner. However, since CNNs pose segmentation as a pixel-labeling problem, they cannot explicitly incorporate any shape priors to capture the boundary smoothness or model the dependencies between the adjacent boundaries. As a result, complex post-processing methods have been employed to refine the probability maps obtained from the CNNs. For eg., in [97] and [100], the dijsktra's shortest path based energy minimization method was modified to refine the layer boundaries obtained from the CNN, while [101] employed a Gaussian process based regression for the same. Moreover, CNN architectures in general, require a large amount of training data to prevent over-fitting.

In this work, we propose a method that combines the advantages of both the CNN and the energy minimization based methods. Similar to CNNs, the proposed supervised CRF formulation learns two sets of filter banks to capture the appearance of each tissue layer and their boundaries, thereby eliminating the need to handcraft the CRF energy. Additionally, it offers some key advantages over a CNN. In comparison to the CNN architectures in [97] and [99] which employ 85,578 and 900,170 network parameters respectively, the proposed CRF energy is linearly parameterized using only 5,438 learnable parameters. Fewer parameters reduce the risk of overfitting and allow the models to be trained using very few training samples. For example, only 87 and 55 training samples were employed to train the proposed method in the experiments reported in Sections 3.4.2 and 3.4.6 respectively.

Another key advantage of the proposed method over CNN is its ability to explicitly incorporate shape priors on the layer boundaries, thereby eliminating the need for any addititonal post-processing. The distance in height between the adjacent points in each layer boundary and the expected thickness of each tissue region are modelled as Gaussian distributions and large deviations of the segmentation result from these distributions are penalized within the proposed CRF energy. Hard constraints are also employed within the CRF framework to maintain the anatomically correct ordering of the tissue layers and prevent the intersection of the layer boundaries. The relative weights between the appearance and the shape prior based cost terms are automatically learned during the SSVM based end-to-end training. Unlike CNNs, the SSVM objective function used to train the CRF is convex and has a unique (global) optima which leads to a more efficient and robust training. As a result, our method has a low computational requirement and can be trained without a GPU.

The advantages of the proposed method over the existing graph and deformable models based energy minimization methods are as follows. Since the learned energy is optimally adapted to the segmentation problem, our method outperforms the existing energy minimization methods with similar but handcrafted energy cost terms such as [91]. Moreover, our CRF formulation efficiently incorporates both hard and soft constraints on the shape priors using a single undirected edge as opposed to the MCCS formulation that requires additional directed edges in the graph construction to incorporate soft constraints [92]. Finally, in this work all layers are jointly segmented in a single optimization step. In contrast, [91] takes a two-step approach by

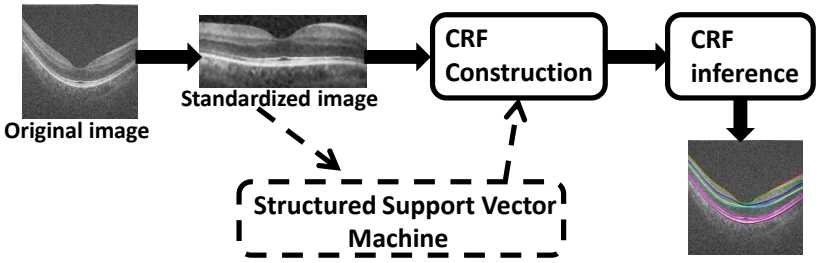

Figure 3.3: Overview of the proposed joint multi-layer segmentation pipeline. The supervised training is indicated by dashed lines.

segmenting the outer layer boundaries (ILM, BM and RPE_{out}) first, followed by the remaining inner layers, while [13], [14] segments each layer sequentially.

3.2 Methods

An overview of the proposed method is presented in Fig. 3.3. A preprocessing step described in Section 3.2.1 is applied to standardize both the training and test OCT B-scan images. The joint extraction of the multiple layer boundaries is formulated within a CRF based Energy Minimization framework in Section 3.2.2. During training, the CRF energy is linearly parameterized as detailed in Section 3.2.3 and its parameters are learned in a supervised, end-to-end manner by posing it as a SSVM optimization problem as discussed in Section 3.2.4. During testing, the optimal labeling of the CRF is inferred to extract the multiple layer boundaries in a single optimization step.

3.2.1 Image Preprocessing

The preprocessing step involves the flatenning of the retinal curvature, extraction of the Region of Interest (ROI) containing the retinal tissue, reducing the speckle noise and intensity standardization of the OCT B-scans. Retinal curvature flatenning is a crucial preprocessing step that reduces the variations in the spatial location of the retinal tissue across the A-scans in an OCT image. It aids in obtaining a tighter ROI thereby reducing the time and memory requirements and also provides a more consistent shape of the layers for segmentation. Each B-scan is flatenned using the method employed in [13] as follows. At first, a rough estimate of the RPE_{out} boundary is obtained by fitting a quadratic polynomial to a set of candidate pixels.

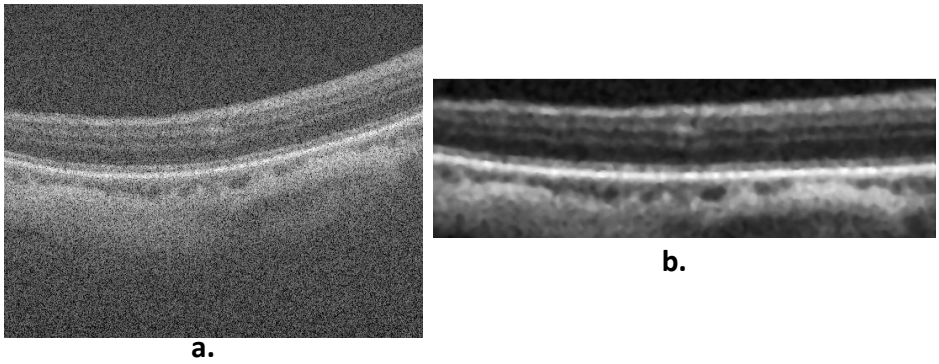

Figure 3.4: a. Raw OCT B-scan; b. Corresponding preprocessed Region of Interest.

Since, RPE_{out} appears as the brightest boundary in the retinal OCT images, the candidate pixels are obtained by detecting the brightest pixel in each A-scan and removing the outliers. Then each column is shifted (cyclically) by an offset such that the detected boundary lies on a straight line.

The retinal tissue is surrounded by a dark background at the top and bottom in each B-scan (see Fig. 3.4 a). To estimate the ROI, the input image is smoothed by a large Gaussian filter with $\sigma = 9$ to reduce the effect of the speckle noise in the background and remove the dark regions within the retinal tissue. Then the ROI is estimated as the maximum extent of the largest connected component obtained by thresholding the smoothed image at 0.3 after scaling the pixel intensities in the B-scan to [0,1].

Finally, the speckle noise in the OCT images is reduced by applying the speckle reducing anisotropic diffusion [104] for 30 iterations in time-steps of 0.1 and a histogram based intensity standardization scheme based on [105] is applied to handle the inter and intra-scanner intensity variations. The intensity of each B-scan is scaled to [0,1]. The ROI is resized to 190×600 to handle any variations in image resolution across the images.

3.2.2 Modelling Joint Multi-Layer Segmentation as a CRF

The joint multi-layer segmentation problem seeks to extract the L layer boundaries in an OCT B-scan image I of size $H \times Y$. Each boundary is labeled from $1 \leq l \leq L$ in the increasing order of its height along H as depicted in Fig. 3.5. The boundaries are uniformly sampled at N equidistant columns y_n along Y such that the l^{th} boundary intersects y_n at a height of $x_{l,n}$. Each $x_{l,n}$ can be interpreted as a discrete random variable that can take a value from the

label set $\Omega = \{1 \leq i \leq H, i \in \mathbb{Z}^+\}$, where $\mathbb{Z}^+$ represents the set of positive integers. The set of all random variables is defined as the random field $X = \{x_{l,n} | 1 \leq l \leq L, 1 \leq n \leq N\}$. A feasible *labeling* denoted by $\mathbf{x} \in \Omega^{L \times N}$ can be obtained by assigning an arbitary label from Ω to each $x_{l,n}$ in X. Our objective is to define an energy $E(\mathbf{x}, I)$ over the Random Field X for each image I such that the optimal labeling that maximizes $E(\mathbf{x}, I)$ corresponds to the desired layer boundaries.

$E(\mathbf{x}, I)$ consists of multiple cost terms. A unary boundary cost $\varepsilon^{l}_{bnd}(x_{l,n})$ is defined for each $x_{l,n}$ to capture the likelihood that the l^{th} boundary passes through the point $(x_{l,n}, y_n)$ given the local appearance of the OCT B-scan around that location. The label of each $x_{l,n}$ is also dependent on its immediate neighbors $x_{l,n+1}$ on the same boundary and $x_{l+1,n}$ on the adjacent $(l+1)^{th}$ boundary resulting in a second order CRF with a maximum clique size of 2. An undirected graphical representation of the CRF for a local *4-neighborhood* is depicted in Fig. 3.5, where each node represents a random variable. By the Markovian property, the labeling of $x_{l,n}$ given the labeling of its immediate 4 neighbors is independent of all other nodes in the graph. The Intra-layer pairwise cost $\varepsilon^{l,n}_{intra}(x_{l,n}, x_{l,n+1})$ captures the smoothness and the similarity in appearance between the adjacent points on a boundary. Additionally, the pairwise Inter-layer Energy Cost $\varepsilon^{l,n}_{inter}(x_{l,n}, x_{l+1,n})$ is defined between the adjacent l and $(l+1)^{th}$ boundaries to enforce the correct layer ordering, the expected layer thickness and capture the appearance of the tissue layer between them. All the three cost terms are dependent on the observed image I. To simplify the notation, the input arguments for the cost terms have been omitted in the rest of the chapter and simply represented by ε^{l}_{bnd}, $\varepsilon^{l,n}_{intra}$ and $\varepsilon^{l,n}_{inter}$ respectively. Thus, the CRF inference problem for I is defined as

$$\begin{aligned} \underset{\mathbf{x}}{\operatorname{argmax}}\, E(\mathbf{x}, I) &= \sum_{l=1}^{L}\sum_{n=1}^{N} \varepsilon^{l}_{bnd} + \sum_{l=1}^{L}\sum_{n=1}^{N-1} \varepsilon^{l,n}_{intra} + \sum_{l=1}^{L-1}\sum_{n=1}^{N} \varepsilon^{l,n}_{inter} \\ &= E_{bnd}(\mathbf{x}, I) + E_{intra}(\mathbf{x}, I) + E_{inter}(\mathbf{x}, I), \end{aligned} \tag{3.1}$$

where $E_{bnd}(\mathbf{x}, I)$, $E_{intra}(\mathbf{x}, I)$ and $E_{inter}(\mathbf{x}, I)$ are the sum of all the unary, intra-layer and the inter-layer cost terms in the entire CRF respectively. During implementation, the CRF inference in eq. 1 is converted into a minimization problem by taking the negative of all the unary and pairwise cost terms, and solved using the Sequential Tree Re-weighted Message Passing (TRW-S) algorithm [67]. Instead of handcrafting the individual cost terms, we parameterize $E(\mathbf{x}, I)$

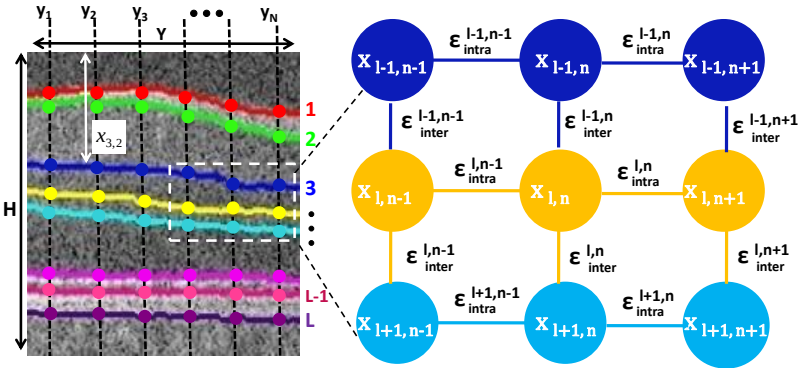

Figure 3.5: The Conditional Random Field for joint multi-layer OCT Segmentation.

by a set of parameters θ which can then be learned from a set of training images. Next, we develop an appropriate definition of $E_\theta(\mathbf{x}, I)$ for the multi-layer OCT segmentation problem.

3.2.3 Linear Parameterization of CRF Energy

Since E(x,I) is defined as a sum of the unary and pairwise Cost terms, its linear parameterization involves the decomposition of each of its cost terms into a linear function. We define the individual cost terms in eq. 3.1 as,

$$\begin{aligned} E_{bnd}(\mathbf{x}, I) &= \mathbf{w}_{bnd}^\top . F_{bnd}(\mathbf{x}, I), \\ E_{intra}(\mathbf{x}, I) &= \mathbf{w}_{intra}^\top . F_{intra}(\mathbf{x}, I) \text{ and} \\ E_{inter}(\mathbf{x}, I) &= \mathbf{w}_{inter}^\top . F_{inter}(\mathbf{x}, I). \end{aligned} \tag{3.2}$$

The details of the linear decomposition of each cost term is detailed below in Sections 3.2.3.1, 3.2.3.2 and 3.2.3.3. Thus, the net CRF energy defined in eq. 3.1 can be rewritten in the linear form by substituting the definition of the individual cost terms from eq. 3.2 as

$$\begin{aligned} E_\theta(\mathbf{x}, I) &= \mathbf{w}_{bnd}^\top . F_{bnd}(\mathbf{x}, I) + \mathbf{w}_{intra}^\top . F_{intra}(\mathbf{x}, I) + \mathbf{w}_{inter}^\top . F_{inter}(\mathbf{x}, I) \\ &= \theta^\top . F(\mathbf{x}, I), \end{aligned} \tag{3.3}$$

where $\theta^\top = [\mathbf{w}_{bnd}^\top \, \mathbf{w}_{intra}^\top \, \mathbf{w}_{inter}^\top]$ and $F(\mathbf{x}, I) = [F_{bnd}^\top \, F_{intra}^\top \, F_{inter}^\top]^\top$.

3.2.3.1 Unary Boundary Cost

To capture the image appearance at each layer boundary, we aim to learn a convolutional filter bank $\{\mathbf{u_l}\}_{l=1}^{L}$. Each $\mathbf{u_l}$ is a $p \times p$ filter which should have a high response only at the pixels

lying on the l^{th} boundary. Let $\mathbf{I_{l,n}}$ represent a $p \times p$ image patch centered at $(x_{l,n}, y_n)$. Both $\mathbf{u}_l$ and $\mathbf{I_{l,n}}$ are linearly indexed to $p^2 \times 1$ column vectors so that the response of the convolution filter at $(x_{l,n}, y_n)$ is obtained by their dot product. Thus, the boundary cost for each $x_{l,n}$ is defined as $\varepsilon^l_{bnd}(x_{l,n}) = \mathbf{u}_l^\top . \mathbf{I_{l,n}}$ and the total Boundary cost over the entire CRF is given by

$$E_{bnd}(\mathbf{x}, I) = \sum_{l=1}^{L} \sum_{n=1}^{N} \mathbf{u}_l^\top . \mathbf{I_{l,n}} = \sum_{l=1}^{L} \mathbf{u}_l^\top \left\{ \sum_{n=1}^{N} \mathbf{I_{l,n}} \right\} = \mathbf{w}_{bnd}^\top . F_{bnd}(\mathbf{x}, I), \tag{3.4}$$

where $F_{bnd}(\mathbf{x}, I) = \left[\left(\sum_{n=1}^{N} \mathbf{I_{1,n}} \right) \left(\sum_{n=1}^{N} \mathbf{I_{2,n}} \right) \ldots \left(\sum_{n=1}^{N} \mathbf{I_{L,n}} \right) \right]^\top$ and $\mathbf{w}_{bnd}^\top = [\mathbf{u}_1^\top \, \mathbf{u}_2^\top \ldots \mathbf{u}_L^\top]$.

3.2.3.2 Pairwise Intra-Layer Cost

The interaction between each pair of adjacent points $x_{l,n}$ and $x_{l,n+1}$ on the l^{th} boundary is modeled as a linear combination of a shape prior and an appearance term. The shape prior between $(x_{l,n}, x_{l,n+1})$ is a soft constraint that penalizes large deviations of the signed gradient of the height values $(x_{l,n+1} - x_{l,n})$ to preserve the local smoothness of the l^{th} boundary. The deviation is modeled by a Gaussian function $d^{l,n}_{intra}(x_{l,n}, x_{l,n+1}) = \exp\left\{ -\frac{1}{2} . (\frac{(x_{l,n+1} - x_{l,n}) - \mu^{l,n}_{intra}}{\sigma^{l,n}_{intra}})^2 \right\}$. The mean $\mu^{l,n}_{intra}$ and the standard deviation $\sigma^{l,n}_{intra}$ of the signed gradient are pre-computed for each layer l and column y_n using the ground truth layer markings of the training images.

The presence of abnormalities such as AMD and DME has an adverse effect on the boundary smoothness as depicted in Fig. 3.2 b,c. Hence, using the shape priors alone can lead to large segmentation errors in such cases. To overcome this, an additional term is introduced to favour labelings where the adjacent boundary points are also similar in appearance. The similarity term $S(x_{l,n}, x_{l,n+1})$ measures the dissimilarity between the two $p \times p$ image patches centered at the adjacent boundary points $(x_{l,n}, y_n)$ and $(x_{l,n+1}, y_{n+1})$. A histogram intersection [106] based dissimilarity measure is defined as $S(x_{l,n}, x_{l,n+1}) = 1 - min \sum_{k=1}^{255} min\{h_k(x_{l,n}, y_n), h_k(x_{l,n+1}, y_{n+1})\}$, where h_k represents the k^{th} bin of the normalized 255-bin histograms computed over the two image patches.

The pairwise intra-layer energy for each $(x_{l,n}, x_{l,n+1})$ is defined as $\varepsilon^{l,n}_{intra} = \alpha_l . d^{l,n}_{intra}(x_{l,n}, x_{l,n+1}) + \beta_l . S(x_{l,n}, x_{l,n+1})$, where α_l and β_l are the scalar relative weights defined for each layer boundary l. These weights are crucial to obtain an accurate segmentation as they not only provide the relative importance of the shape and the appearance term but also define the weightage of the entire Intra-Layer Pairwise term with respect to the Unary

Boundary and the Pairwise Inter-Layer Cost terms. The total Intra-layer pairwise energy is given by,

$$\begin{aligned} E_{intra}(\mathbf{x}, I) &= \sum_{l=1}^{L}\sum_{n=1}^{N-1}\Big\{\alpha_l.d_{intra}^{l,n}(x_{l,n}, x_{l,n+1}) + \beta_l.S(x_{l,n}, x_{l,n+1})\Big\} \\ &= \sum_{l=1}^{L}\alpha_l.\left\{\sum_{n=1}^{N-1} d_{intra}^{l,n}(x_{l,n}, x_{l,n+1})\right\} + \sum_{l=1}^{L}\beta_l.\left\{\sum_{n=1}^{N-1} S(x_{l,n}, x_{l,n+1})\right\} \\ &= \mathbf{w}_{intra}^{\top}.F_{intra}(\mathbf{x}, I). \end{aligned} \tag{3.5}$$

Here, $E_{intra}(\mathbf{x}, I)$ is linearized by taking $\mathbf{w}_{intra}^{\top} = [\,\alpha_1\,\alpha_2\,...\alpha_L\,\beta_1\,\beta_2\,...\beta_L\,]$ and $F_{intra}(\mathbf{x}) = [d^1\ d^2\,...\,d^L\,S^1\,S^2\,...\,S^L]^{\top}$, where $d^i = \sum_{n=1}^{N-1} d_{intra}^{i,n}(x_{i,n}, x_{i,n+1})$ and $S^i = \sum_{n=1}^{N-1} S(x_{i,n}, x_{i,n+1})$ respectively.

3.2.3.3 Pairwise Inter-Layer Cost

The interaction between the corresponding points $x_{l,n}$ and $x_{l+1,n}$ in the l^{th} and the $(l+1)^{th}$ layer boundaries respectively is captured by the Pairwise Inter-Layer Cost $\varepsilon_{inter}^{l,n}$ and modeled as a *linear combination* of a shape prior and a regional appearance term.

The shape prior denoted by $d_{inter}^{l,n}$ imposes the restriction that the $(l+1)^{th}$ boundary must lie below the l^{th} boundary at each y_n. It also enforces a soft constraint on the layer thickness $(x_{l+1,n} - x_{l,n})$ by penalizing its deviation from the expected value which is modeled by a Gaussian function with a mean $\mu_{inter}^{l,n}$ and a standard deviation $\sigma_{inter}^{l,n}$. The layer ordering is ensured by constraining the layer thickness to lie within a minimum T_{mn}^{l} and a maximum T_{mx}^{l} range which is defined for each layer. This is achieved by assigning $-\infty$ to the infeasible labelings that donot satisfy this criteria. $T_{mn}^{l} > 0$ ensures that the layer boundaries donot intersect. Therefore,

$$d_{inter}^{l,n} = \begin{cases} \exp\left\{-\frac{1}{2}.\left(\frac{(x_n^{l+1} - x_n^{l}) - \mu_{inter}^{l,n}}{\sigma_{inter}^{l,n}}\right)^2\right\}, & if\ T_{mn}^{l} \le (x_{l+1,n} - x_{l,n}) \le T_{mx}^{l} \\ -\infty,\ otherwise. \end{cases} \tag{3.6}$$

The parameters $\mu_{inter}^{l,n}$, $\sigma_{inter}^{l,n}$, T_{mn}^{l} and T_{mx}^{l} are precomputed for each layer l and column y_n from a set of training images.

The second term in the $\varepsilon_{inter}^{l,n}$ seeks to capture the appearance of the intermediate tissue regions lying between the adjacent boundaries. Let R_l denote the tissue region between the l and $(l+1)^{th}$ boundary. The appearance of each of the $L-1$ regions is captured by a

convolutional filter bank $\{\mathbf{v_l}\}_{l=1}^{L-1}$. Each $\mathbf{v}_l$ is a $p \times p$ filter that captures the appearance of R_l and has the maximum *average* filter response in each column y_n within R_l . The average filter bank response is obtained using the dot product $\frac{1}{|x_{l+1,n}-x_{l,n}|}(\sum_{j=x_{l,n}}^{x_{l+1,n}} \mathbf{v}_l^\top.\mathbf{I_{j,n}})$, where $\mathbf{I_{j,n}}$ represents a $p \times p$ image patch centered at $(x_{j,n}, y_n)$. Both $\mathbf{v}_l$ and $\mathbf{I_{j,n}}$ are linearly indexed to $p^2 \times 1$ column vectors.

Therefore the local Pairwise Inter-Layer Cost at each $(x_{l,n}, x_{l+1,n})$ is defined as $\varepsilon_{inter}^{l,n} = \frac{1}{|x_{l+1,n}-x_{l,n}|}(\sum_{j=x_{l,n}}^{x_{l+1,n}} \mathbf{v}_l^\top.\mathbf{I_{j,n}}) + \gamma_l.d_{inter}^{l,n}(x_{l,n}, x_{l+1,n})$. The relative weight of the regional appearance term is implicitly learned by an appropriate scaling of the weights in $\mathbf{v}_l$, while γ_l controls the relative weightage to the shape prior with respect to the regional appearance term as well as the Boundary and the pairwise intra-layer cost terms. Both $\mathbf{v}_l$ and γ_l are learned in an end-to-end manner. The total Inter-layer pairwise Energy is given by,

$$\begin{aligned} E_{inter}(\mathbf{x}, I) &= \sum_{l=1}^{L-1}\sum_{n=1}^{N}\left\{\frac{1}{\mid x_{l+1,n} - x_{l,n} \mid}\left(\sum_{j=x_{l,n}}^{x_{l+1,n}} \mathbf{v}_l^\top.\mathbf{I_{j,n}}\right) + \gamma_l.d_{inter}^{l,n}\right\} \\ &= \sum_{l=1}^{L-1}\mathbf{v}_l^\top\left\{\sum_{n=1}^{N}\frac{1}{\mid x_{l+1,n} - x_{l,n} \mid}\sum_{j=x_{l,n}}^{x_{l+1,n}}\mathbf{I_{j,n}}\right\} + \sum_{l=1}^{L-1}\gamma_l\left\{\sum_{n=1}^{N}d_{inter}^{l,n}\right\} \\ &= \mathbf{w}_{inter}^\top.F_{inter}(\mathbf{x}, I). \end{aligned} \tag{3.7}$$

Here, $E_{inter}(\mathbf{x}, I)$ is linearized by taking $\mathbf{w}_{inter}^\top = [\,\mathbf{v}_1\,\mathbf{v}_2\,...\mathbf{v}_{L-1}\,\gamma_1\,\gamma_2\,...\,\gamma_{L-1}\,]$ and $F_{inter}(\mathbf{x}) = [\,\mathbf{r^1}\,\mathbf{r^2}\,...\,\mathbf{r^{L-1}}\,t^1\,t^2\,...\,t^{L-1}]^\top$, where $t^i = \sum_{n=1}^{N} d_{inter}^{i,n}(x_{i,n}, x_{i+1,n})$ and $\mathbf{r^i} = \sum_{n=1}^{N}\frac{1}{|x_{i+1,n}-x_{i,n}|}\sum_{j=x_{i,n}}^{x_{i+1,n}}\mathbf{I_{j,n}}$ respectively.

3.2.4 The Structural Support Vector Machine Formulation

In Section 3.2.3, the proposed CRF energy for the joint multi-layer segmentation was linearly parameterized into $E_\theta(\mathbf{x}, I) = \theta^\top.F(\mathbf{x}, I)$ (eq. 3.3), where $F(\mathbf{x}, I)$ is known as the joint feature function. The model parameter θ consists of the weights of the convolutional filters that capture the appearance of each layer and their boundaries as well as the relative weights given to the shape priors to preserve the smoothness and the expected thickness of each layer. The problem of learning θ during training is posed as a structural Support Vector Machine (SSVM) [107] formulation. Let $\{I^{(k)}, \mathbf{x}^{(k)}\}_{k=1}^{K}$ denote a set of K training OCT B-scans, where $I^{(k)}$ denotes the k^{th} training image with the corresponding Ground Truth (GT) labeling $\mathbf{x}^{(k)}$. Let

$\mathcal{Y}_k = \{\mathbf{x}|\mathbf{x} \in \Omega^{L\times N}\} - \{\mathbf{x}^{(k)}\}$ denote the set of all feasible but incorrect labelings. To quantify the segmentation error of $\mathbf{x}$, we define a loss function $\Delta(\mathbf{x}^{(k)}, \mathbf{x}) = \sum_{l=1}^{L}\sum_{n=1}^{N} \mid x_{l,n}^{(k)} - x_{l,n} \mid$ as the sum of the unsigned distances between the corresponding labels in $\mathbf{x}$ and the GT $\mathbf{x}^{(k)}$.

$E_\theta(\mathbf{x}, I)$ maps each feasible labeling $\mathbf{x}$ of an image I to a scalar score value. Our objective is to learn a θ such that for each I^k, i) the GT labeling has the maximum score, ie., $\mathbf{x}^{(k)} = \underset{\mathbf{x}}{\text{argmax}}\ E_\theta(\mathbf{x}, I^{(k)})$ and ii) the higher the loss $\Delta(\mathbf{x}^{(k)}, \mathbf{x})$ of a feasible labeling $\mathbf{x} \in \mathcal{Y}_k$, the lower is its energy $E_\theta(\mathbf{x}, I^{(k)})$ with respect to that of the correct labeling. This can be posed as the following SSVM [107] formulation,

$$\begin{aligned} \underset{\theta,\xi\geq 0}{\text{argmin}} \quad & \frac{\lambda}{2} \mid\mid \theta \mid\mid^2 + \frac{1}{M}\sum_{k=1}^{M}\xi_k \\ \text{s.t.} \quad & \theta^\top.\Big\{F(\mathbf{x}^{(k)}, I^{(k)}) - F(\mathbf{x}, I^{(k)})\Big\} \geq \Delta(\mathbf{x}^{(k)}, \mathbf{x}) - \xi_k \quad \forall k,\ \forall \mathbf{x} \in \mathcal{Y}_k, \end{aligned} \tag{3.8}$$

where ξ_k are the slack variables. The $L2$ regularization on θ is employed to ensure good generalization on unseen test images and λ is the regularization weight. The constraints in eq. 3.8 ensure that the difference in E_θ between the GT given by $E_\theta(\mathbf{x}^{(k)}, I^{(k)}) = \theta^\top.F(\mathbf{x}^{(k)}, I^{(k)})$ and each incorrect labeling given by $E_\theta(\mathbf{x}, I^{(k)}) = \theta^\top.F(\mathbf{x}, I^{(k)})$ is greater than a margin which is scaled by the loss function $\Delta(\mathbf{x}^{(k)}, \mathbf{x})$. This is further illustrated in Fig. 3.6. Here, the point $F(\mathbf{x}^{(k)}, I^{(k)})$ represents the joint feature vector corresponding to the correct GT labeling of an image $I^{(k)}$ and the point $F(\mathbf{x}, I^{(k)})$ corresponds to a feasible but incorrect labeling $\mathbf{x}$ for the same image. The distance of the points $F(\mathbf{x}^{(k)}, I^{(k)})$ and $F(\mathbf{x}, I^{(k)})$ from a hyperplane parameterized by θ is given by $\theta^\top.F(\mathbf{x}^{(k)}, I^{(k)})$ and $\theta^\top.F(\mathbf{x}, I^{(k)})$ respectively. Our objective is to learn θ such that the margin between the GT and an incorrect labeling given by $\theta^\top.(\mathbf{x}^{(k)}, I^{(k)}) - \theta^\top.(\mathbf{x}, I^{(k)})$ is greater than or equal to the loss $\Delta(\mathbf{x}^{(k)}, \mathbf{x})$. We note that shifting the hyperplane by an arbitary bias term b doesnot change the margin ($\{\boldsymbol{\theta}^\top.(\boldsymbol{x}^{(k)}, I^{(k)}) + b\} - \{\boldsymbol{\theta}^\top.(\boldsymbol{x}, I^{(k)}) + b\}$) as the bias term gets cancelled out. Hence a bias term is ignored in the SSVM formulation.

By substituting $E_\theta(\mathbf{x}, I) = \theta^\top.F(\mathbf{x}, I)$, the constraints in eq. 3.8 for each I_k can be rearranged as $\xi_k \geq \Delta(\mathbf{x}^{(k)}, \mathbf{x}) - \{E_\theta(\mathbf{x}^{(k)}, I^{(k)}) - E_\theta(\mathbf{x}, I^{(k)}\}, \quad \forall \mathbf{x} \in \mathcal{Y}_k$. These $|\mathcal{Y}_k|$ constraints can be replaced by a single most violating constraint of the form $\xi_k \geq H_k$, where H_k is obtained by solving the *max oracle* optimization problem $H_k = \underset{\mathbf{x}}{\text{argmax}}\ \Delta(\mathbf{x}^{(k)}, \mathbf{x}) - \{E_\theta(\mathbf{x}^{(k)}, I^{(k)}) - E_\theta(\mathbf{x}, I^{(k)}\}$. Since, $E_\theta(\mathbf{x}^{(k)}, I^{(k)})$ is the Energy of the GT labeling and inde-

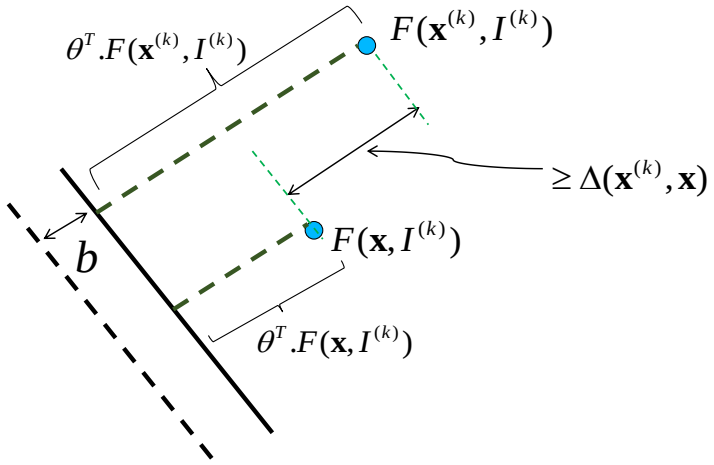

Figure 3.6: Illustration of the Structural Support Vector Machine (SSVM) formulation.

pendent of $\mathbf{x}$, the max oracle optimization problem reduces to

$$H_i = \underset{\mathbf{x}}{\text{argmax}} \quad \Delta(\mathbf{x}^{(k)}, \mathbf{x}) + E_\theta(\mathbf{x}, I^{(k)}) \tag{3.9}$$

Since in our case, $\Delta(\mathbf{x}^{(k)}, \mathbf{x})$ is separable at each $x_{l,n}$, eq. 3.9 can be solved using the TRW-S algorithm (discussed in APPENDIX A) similar to the CRF inference in eq. 3.1 with an additional term $\mid x_{l,n}^{(k)} - x_{l,n} \mid$ added to the Unary Boundary Cost for each $x_{l,n}$.

Though eq. 3.8 is a convex Quadratic Programming (QP) Problem, it cannot be solved directly due to the extremely large number of constraints. For each training sample, there are an exponentially large number of possible incorrect labelings in $\mathcal{Y}_k$ resulting in a total of $\sum_k |\mathcal{Y}_k|$) constraints. Hence an iterative Block Co-ordinate Frank Wolfe Algorithm [108] is employed to make the optimization tractable. At each iteration, a training image I_k is randomly selected. Then the label $\mathbf{x}$ corresponding to the *most violating* constraint is obtained by solving eq. 3.9 keeping θ fixed. Next, a Frank-Wolfe update step is performed on θ considering the most violating constraint alone. We refer to APPENDIX B for further details and a pseudocode of the Block Co-ordinate Frank-Wolfe (BCFW) algorithm.

3.3 Materials

The proposed method has been extensively evaluated on 4 public datasets that contain B-scans of healthy subjects as well as patients suffering from AMD and DME. The datasets are

Dataset	B-scan Size (pixels)	Voxel Resolution (μm)	Scanner	# Images	#GT
NORMAL-1	400 × 400/ 400 × 800.	3.23, 13.4, 67/ 3.23,6.7, 33.5.	Bioptigen Inc.	108/10	8
NORMAL-2	496 × 768	3.9,10-12, 120-140	Spectralis	110/10	8
AMD-1	512 × 1000	3.06-3.24, 6.50-6.60, 65-69.8	Bioptigen Inc. (4 clinics)	220/20	3
DME-1	496 × 768	3.87,10.94-11.98, 118-128	Spectralis	110/10	8

Table 3.1: Dataset Description. The **voxel resolution** is reported in *axial, lateral, azimuthal* directions. # **Images** reports the (total number of B-scans/ acquired from the number of OCT volumes). # **GT** reports the number of layer boundaries for which the Ground Truth markings are available.

summarized in Table 3.1 and cover a range of image resolution, scanners and image quality. The *NORMAL-1* [13] and *NORMAL-2* [89] datasets contain B-scans from healthy subjects. Out of the 10 volumes in the *NORMAL-1* dataset, five volumes were acquired at a resolution of 400× 400 and the other half at 400 × 800 pixels respectively. The *AMD-1* dataset [14] was acquired from 4 clinics at varying resolutions and contain B-scans of subjects suffering from intermediate AMD which is characterized by the presence of drusen and geographic atrophy. The B-scans in the *DME-1* dataset [93] contain fluid filled regions associated with DME.

For each dataset, the manual ground truth (GT) markings by a senior grader is available for all the 8 boundaries with the exception of the *AMD-1* dataset for which the markings of only 3 clinically relevant boundaries, the ILM, BM and RPE_{out} (see Fig. 3.2 b) are available. Due to the tedium involved in obtaining manual GT, only a few non-adjacent, linearly spaced B-scans from each OCT volume are provided in each dataset that encompass both the foveal and the peripheral regions. To evaluate the inter-observer variance, the manual marking from a second expert is also available for all the datasets except *NORMAL-1*, for which only a subset of 28 B-scans were marked by a second expert [13].

3.4 Results

In this Section, we present various experiments to validate our joint multi-layer OCT segmentation framework. Both boundary and region based metrics have been defined in Section

3.4.1 to evaluate the segmentation performance. In Section 3.4.2, a five-fold cross validation is performed on the *NORMAL-1* dataset to evaluate the performance on healthy OCT B-scans. This is followed by cross-testing on the *NORMAL-2* dataset in Section 3.4.3 after training the CRF model on B-scans from the *NORMAL-1* dataset alone.

The proposed method is also evaluated in the presence of pathologies associated with AMD and DME in Sections 3.4.4, 3.4.5 and 3.4.6. Ideally, a single CRF model should be able to segment both healthy and abnormal cases without any prior knowledge about the presence of pathologies in the given image. Hence, these experiments are performed by combining the *NORMAL-1* dataset to the *AMD-1* dataset in Section 3.4.4 and the *DME-1* dataset in Section 3.4.5. The performance of our method when trained and evaluated on the *DME-1* dataset alone has also been presented in Section 3.4.6.

The performance of the proposed method is compared to the results obtained from three publicly available OCT segmentation softwares, CASEREL [109], the Iowa Reference Algorithm (IRA) [110] and OCTSEG [111]. The qualitative and quantitative results of these methods were obtained by running their publicly available implementations on each dataset using their default parameters and weights. CASEREL is based on [13] and provides the segmentation of seven layer boundaries (except BM). IRA is based on [91] and segments eleven layer boundaries out of which the boundaries 1, 2, 4, 5, 6, 8, 10 and 11 correspond to our GT markings. OCTSEG is based on [112] and segments six layer boundaries with the exception of BM and INL/OPL.

The layer segmentations of the proposed method are obtained by mapping the results of the CRF inference back into the original image coordinate space by reversing the image flattening, resizing and ROI extraction operations which were performed during the image preprocessing.

3.4.1 Performance Metrics

Let $\mathbf{x}^{gt}$ denote the GT and $\mathbf{x}$ denote the corresponding estimated layer boundary markings for a given test image I with H rows and Y columns. Using a notation similar to Section 3.2.2, $\mathbf{x} = \{x_{l,y} | 1 \leq x_{l,y} \leq H, 1 \leq l \leq L,\ 1 \leq y \leq Y,\ l, y \in \mathbb{Z}^+\}$, where $\mathbb{Z}^+$ represents the set of positive integers and $x_{l,y}$ represents the height at which the l^{th} boundary passes through column y. The L boundaries divide the retinal tissue into $L-1$ adjacent layers. Let R_l denote the layer that lies between the l^{th} and the $(l+1)^{th}$ boundary.

The Unsigned Boundary Localization Error(U-BLE) is a boundary based performance metric which is defined as the average unsigned distance in pixels between the corresponding points in $\mathbf{x}$ and $\mathbf{x}^{gt}$ along each column in the image. Thus, the U-BLE for the l^{th} boundary is defined as

$$U\text{-}BLE_l(\mathbf{x}, \mathbf{x}^{gt}) = \frac{1}{Y} \sum_{y=1}^{Y} \mid x_{l,y} - x_{l,y}^{gt} \mid . \tag{3.10}$$

The signed Boundary Localization Error (S-BLE) is defined in a similar manner where signed distance between the corresponding points is computed instead of the absolute distance. Thus,

$$S\text{-}BLE_l(\mathbf{x}, \mathbf{x}^{gt}) = \frac{1}{Y} \sum_{y=1}^{Y} \left(x_{l,y} - x_{l,y}^{gt} \right). \tag{3.11}$$

While U-Boundary Localization Error (BLE) gives a true measure of the segmentation error, S-BLE measures the overall bias of the method to over-estimate or under-estimate the boundary. This is because in S-BLE, the positive and negative errors across the columns tend to cancel each other out. A positive value of the S-BLE indicates that the GT tends to lie above the estimated boundary and viceversa. The U-BLE metric is similar to the loss function Δ used during training (in Section 3.2.4) with the exception that while the loss function is computed only at the N equidistant points used to represent the boundary, U-BLE is computed across all the columns in the image. Ideally, U-BLE and the magnitude of the S-BLE should be close to 0.

The Dice coefficient and the average error in the layer thickness measurements (LTE) are used as the region based metrics. Dice measures the extent of overlap between the l^{th} layer R_l and the corresponding GT denoted by R_l^{gt} and defined as

$$Dice(R_l, R_l^{gt}) = \frac{2.|R_l \cap R_l^{gt}|}{|R_l| + |R_l^{gt}|}. \tag{3.12}$$

The thickness of various retinal tissue layers provide clinically relevant information that aids in the detection and tracking the progresssion of ocular diseases [113]. Hence, LTE is defined to measure the average absolute difference in the thickness (in pixels) between the extracted and GT tissue regions across each column in the image. Since, $(x_{l+1,y} - x_{l,y})$ represents the thickness of the tissue region between the l^{th} and the $(l+1)^{th}$ boundary at column y, LTE for the region R_l is defined as

$$LTE_l(\mathbf{x}, \mathbf{x}^{gt}) = \frac{1}{|Y|} \sum_{y=1}^{|Y|} \mid (x_{l+1,y}^{gt} - x_{l,y}^{gt}) - (x_{l+1,y} - x_{l,y}) \mid . \tag{3.13}$$

The Dice coefficient is bounded between [0,1] and should ideally be close to 1. The LTE being a measure of error should be close 0. While Dice provides a global measure of segmentation accuracy across all columns, the LTE is more sensitive to the localized estimation errors at each column. On the other hand, in contrast to Dice, LTE is not sensitive to the absolute position of the boundaries as the thickness of R_l remains constant even if the two adjacent boundaries are translated by constant values in each column.

3.4.2 Performance on the NORMAL-1 dataset

The proposed method has been evaluated on the *NORMAL-1* dataset which consists of 108 B-scans of healthy subjects from 10 OCT volumes. A five-fold cross-validation was performed by randomly dividing the dataset into five parts, each containing the B-scans from 2 OCT volumes. Three parts consist of 22 B-scans and the remaining two parts contain 21 B-scans respectively. In each fold, the proposed method was tested on one part after being trained on all the remaining 86 or 87 B-scans. The various performance metrics for each boundary are reported in Table 3.2 and 3.3 respectively. Sample qualitative results are depicted in Fig. 3.7.

The manual markings by a second expert was also available on a subset of 28 images. On these images, the U-BLE of our method on the eight boundaries ordered from $l = 1$ to 8 was found to be 1.09, 1.75, 1.64, 1.80, 2.03, 1.26, 1.54 and 1.67 pixels respectively. In comparison, the second expert marking had a U-BLE of 1.65, 1.56, 1.76, 2.59, 2.06, 1.97, 1.91, and 1.80 pixels on the eight boundaries with respect to the GT.

	ILM	NFL/GCL	IPL/INL	INL/OPL	OPL/ONL	IS/OS	*BM*	RPE_{out}
	U-BLE							
CASEREL	**0.99±0.27**	2.83±2.59	4.57±2.17	5.10±2.11	5.05±4.16	1.23±0.89	—	1.70±0.60
OCTSEG	1.87±3.19	6.56±2.93	3.73±3.67	—	3.48±2.99	1.11±1.49	—	1.59±1.69
IRA	1.55±0.72	2.56±1.17	1.67±0.77	1.79±0.57	2.25±1.25	**0.97±0.40**	1.95±1.02	**1.38±0.63**
Proposed	1.09±0.28	**1.66±0.64**	**1.51±0.47**	**1.68±0.55**	**1.95±0.81**	1.15±0.85	**1.47±0.75**	1.67±0.76
	S-BLE							
CASEREL	**0.00±0.59**	1.36±3.14	-1.24±3.96	-1.30±4.26	-1.71±5.68	**0.00±1.04**	—	0.13±1.03
OCTSEG	-0.36±2.97	0.42±5.33	1.73±4.35	—	1.54±3.36	-0.22±1.42	—	**0.01±1.78**
IRA	0.00±1.04	**0.00±2.30**	0.00±1.36	**0.00±1.38**	0.47±1.97	-0.26±0.72	-0.11±2.05	0.22±1.13
Proposed	0.00±0.73	0.17±1.23	**0.00±0.97**	0.10±1.30	**-0.06±1.63**	0.28±1.03	**0.01±1.37**	-0.21±1.52

Table 3.2: Unsigned and Signed Boundary Localization Errors (mean ± standard deviation in pixels) on the NORMAL-1 dataset. The best result in each column is indicated in bold.

	NFL	GCL-IPL	INL	OPL	ONL-IS	OS	RPE
	LTE						
CASEREL	3.11±2.60	4.76±1.96	2.30±0.65	4.98±1.53	5.25±3.96	—	—
OCTSEG	6.84±2.74	5.53±1.67	—	—	3.36±2.61	—	—
IRA	2.41±1.01	2.45±0.89	**1.10±0.64**	2.47±1.03	2.38±1.26	2.15±0.93	1.90±0.90
Proposed	**1.97±0.67**	**2.06±0.59**	1.85±0.53	**2.32±0.96**	**2.26±1.08**	**1.73±0.77**	**1.83±0.85**
	Dice						
CASEREL	0.80±0.14	0.83±0.09	0.63±0.15	0.61±0.15	0.89±0.07	—	—
OCTSEG	0.61±0.14	0.79±0.12	—	—	0.91±0.06	—	—
IRA	0.79±0.08	0.91±0.04	0.86±0.05	0.78± 0.09	0.93±0.02	0.84± 0.07	0.84±0.06
Proposed	**0.84±0.06**	**0.93±0.02**	**0.87±0.03**	**0.80.±0.07**	**0.94±0.02**	**0.86±0.07**	**0.85±0.06**

Table 3.3: Layer Thickness Error in pixels and Dice coefficient (mean ± standard deviation) for 7 tissue regions on the NORMAL-1 dataset. The best result in each column is indicated in bold.

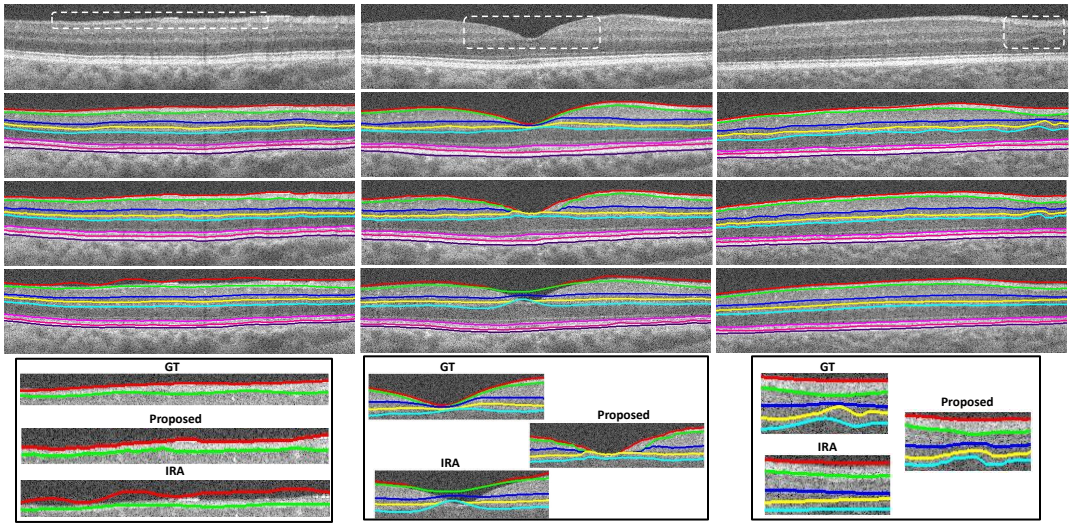

Figure 3.7: Qualitative results on 3 B-scans of healthy subjects from the *NORMAL-1* dataset is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark. Region within the white dashed rectangle in the first row is magnified at the bottom for comparison.

3.4.3 Cross-testing Performance on the NORMAL-2 dataset

A cross-testing based evaluation has been performed to test the generalizability of the proposed method on unseen test data. In this experiment, our method was trained on the 108 B-scans from the *NORMAL-1* dataset and tested on the 110 B-scans in the *NORMAL-2* dataset. The quantitative and qualitative results are depicted in Tables 3.4, 3.5 and Fig. 3.8 respectively.

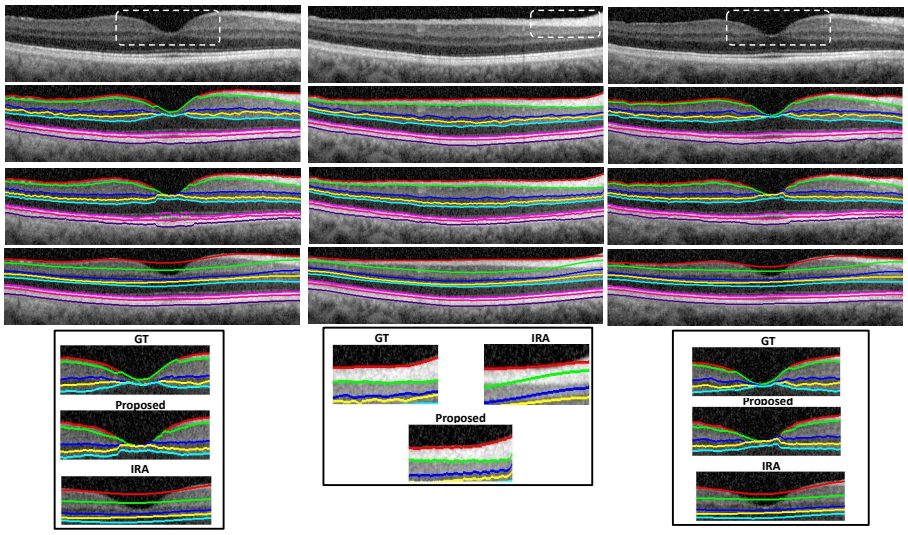

Figure 3.8: Qualitative results on 3 B-scans of healthy subjects from the *NORMAL-2* dataset is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark. Region within the white dashed rectangle in the first row is magnified at the bottom.

	ILM	NFL/GCL	IPL/INL	INL/OPL	OPL/ONL	IS/OS	*BM*	RPE_{out}
U-BLE								
CASEREL	**0.89± 0.38**	2.96±2.50	4.46±1.66	4.91±1.52	3.59±2.81	0.63± 0.22	—	**1.03±0.39**
OCTSEG	1.11±1.45	5.39±5.14	2.30±2.86	—	2.29±1.58	1.26±1.56	—	1.04±0.37
IRA	1.36±0.84	7.79±2.78	5.90±1.66	3.99±1.05	2.36±1.17	0.77±0.48	**1.08±0.54**	1.05±0.44
Proposed	0.96±0.26	**1.47± 1.06**	**1.13± 0.56**	**1.16± 0.32**	**1.11±0.31**	**0.61±0.20**	1.13±0.48	1.35±0.54
Manual Expert 2	0.96±0.26	1.29±0.53	1.40±0.37	1.30±0.32	1.38±0.45	0.74±0.20	2.38±0.10	1.10±0.35
S-BLE								
CASEREL	-0.10±0.57	1.47±3.33	-2.10±3.07	-1.71±3.21	-1.51±3.63	**0.03±0.32**	—	0.15±0.74
OCTSEG	0.45±1.45	4.20±5.66	2.07±3.19	—	1.17±1.89	-0.71±1.77	—	**0.00±0.89**
IRA	-0.12±0.52	**0.14±6.16**	-0.83±4.50	0.33±2.90	0.51±1.82	-0.11±0.70	-0.17±0.94	-0.16±0.84
Proposed	**0.00± 0.54**	-0.27± 1.57	**-0.21±0.76**	**-0.05± 0.60**	**0.02± 0.66**	-0.08± 0.36	**0.10±1.02**	-0.33±1.12
Manual Expert 2	0.52±0.45	0.63±0.84	0.86±0.67	-0.15±0.78	0.33±0.90	0.30±0.38	2.23±1.21	0.36±0.74

Table 3.4: Unsigned, Signed Boundary Localization Errors (mean ± standard deviation in pixels) on the NORMAL-2 dataset. Best result among the automated methods in each column is indicated in bold.

	NFL	GCL-IPL	INL	OPL	ONL-IS	OS	RPE
LTE							
CASEREL	3.23±2.45	4.72± 1.72	1.83±0.35	4.06± 1.10	3.66±2.76	—	—
OCTSEG	5.21±4.65	3.55±2.67	—	—	2.97±2.11	—	—
IRA	7.37±2.83	4.16±1.48	2.81±1.03	2.90±0.80	2.32±1.16	**1.21±0.63**	**1.18±0.56**
Proposed	**1.86± 1.17**	**1.59±0.54**	**1.36±0.36**	**1.57± 0.41**	**1.21±0.32**	1.32±0.53	1.38±0.44
Manual Expert 2	1.42±0.48	1.69±0.39	1.84± 0.44	1.76±0.44	1.51±0.43	2.21±0.91	2.31±0.88
Dice							
CASEREL	0.81±0.13	0.78±0.11	0.50±0.14	0.62±0.12	0.90±0.06	—	—
OCTSEG	0.77±0.18	0.76±0.22	—	—	0.91±0.06	—	—
IRA	0.60±0.15	0.61±0.12	0.45±0.11	0.64±0.10	0.92±0.03	0.88±0.05	**0.91±0.04**
Proposed	**0.88±0.05**	**0.92±0.06**	**0.87±0.05**	**0.86±0.03**	**0.96± 0.01**	**0.88±0.04**	0.89± 0.03
Manual Expert 2	0.88±0.04	0.92±0.03	0.83±0.04	0.84±0.03	0.95±0.01	0.81±0.07	0.84±0.05

Table 3.5: Layer Thickness Error in pixels and Dice coefficient (mean ± standard deviation) for 7 tissue regions on the NORMAL-2 dataset. The best result among the automated methods is indicated in bold.

3.4.4 Performance in the presence of Age-Related Macular Degeneration

The proposed method has been evaluated in the presence of drusen and Geographic Atrophy associated with AMD by employing a five-fold cross validation on the combined *NORMAL-1* and *AMD-1* dataset. In each fold, the test set consists of the 65 B-scans(21 Normal + 44 AMD) obtained from 2 OCT volumes from the *NORMAL-1* and 4 volumes from the *AMD-1* dataset. The CRF is trained on 263 B-scans (87 Normal+176 AMD) in each fold, obtained from the remaining 8 volumes from *NORMAL-1* and 16 OCT volumes from the *AMD-1* dataset. Since the GT for only the ILM, BM and RPE_{out} boundaries are available for the *AMD-1* dataset, our method has been evaluated on them. The quantitative results are reported in Tables 3.6, 3.7 and sample qualitative results on B-scans with AMD are presented in Fig. 3.9. Among the benchmark methods, only IRA segments the BM boundary, hence the regional metrics in Table 3.7 could only be compared against it. On the combined dataset, the U-BLE varies in the range of 1.18 to 2.38 pixels with a mean of 1.86 pixels (see Table 3.6) across all the three boundaries and the Dice is 0.98 for the ILM-BM region and 0.81 for the RPE layer.

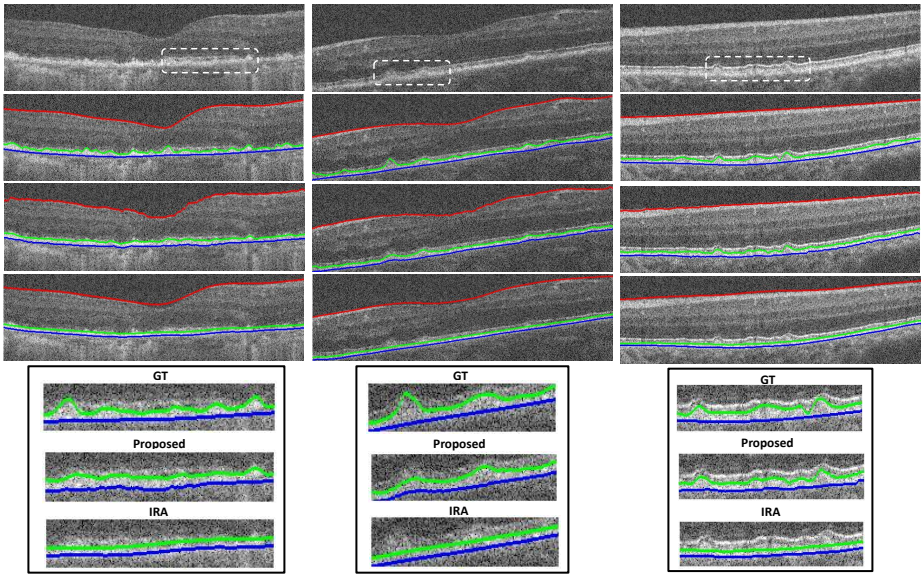

Figure 3.9: Qualitative results on 3 B-scans with AMD is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark.

	Dice		LTE (pixels)	
	ILM-BM	BM-RPE_{out}	ILM-BM	BM-RPE_{out}
AMD dataset alone				
IRA	0.96±0.07	0.74±0.13	4.01±2.18	3.73±1.53
Proposed	**0.98±0.01**	**0.79±0.09**	**2.57±1.12**	**3.45±1.61**
Manual Expert 2	0.98±0.01	0.83±0.05	2.60±0.75	2.72±0.87
Normal dataset alone				
IRA	**0.98±0.01**	0.84±0.07	2.48±1.16	**1.90±0.90**
Proposed	**0.98±0.01**	**0.85±0.05**	**1.84±0.77**	2.01±1.00
AMD + Normal dataset				
IRA	0.97±0.06	0.78±0.12	3.49±2.03	3.11±1.61
Proposed	**0.98±0.01**	**0.81±0.09**	**2.33±1.07**	**2.97±1.59**

Table 3.7: Layer Thickness Error in pixels and Dice coefficient (mean ± standard deviation) for 3 layer boundaries on the combined AMD and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold.

	Unsigned BLE (pixels)			Signed BLE (pixels)		
	ILM	*BM*	RPE_{out}	ILM	*BM*	RPE_{out}
AMD dataset alone						
CASEREL	**1.21±1.24**	—	2.93±3.55	0.22±1.34	—	0.01±4.00
OCTSEG	4.83±7.29	—	3.14±2.95	3.63±7.37	—	**-0.01±3.67**
IRA	2.62±5.13	4.27±5.42	2.96±5.45	-1.20±5.41	0.83±6.27	-0.05±5.92
Proposed	1.24±0.34	**2.20±1.11**	**2.82±2.29**	**0.02±0.71**	**0.43±1.66**	-0.17±3.15
Manual Expert 2	1.28±0.43	2.29±0.79	1.58±0.60	-0.25±0.75	-0.71±1.33	-0.03±1.13
Normal dataset alone						
Proposed	1.06±0.31	1.70±0.88	1.49±0.62	0.00±0.70	-0.09±1.56	0.00±1.26
AMD + Normal dataset						
CASEREL	**1.14±1.04**	—	2.39±2.99	0.15±1.16	—	0.05±3.36
OCTSEG	3.61±6.28	—	2.68±2.55	2.48±6.32	—	**0.03±3.08**
IRA	2.25±4.22	3.48±4.57	2.42±4.50	-0.79±4.47	0.51±5.24	0.04±4.85
Proposed	1.18±0.34	**2.03±1.07**	**2.38±2.01**	**0.02±0.71**	**0.26±1.64**	-0.11±2.68

Table 3.6: Boundary Localization Errors (mean ± std. deviation in pixels) on the combined *AMD* and *NORMAL-1* dataset. Best result among the automated methods in each column is indicated in bold.

3.4.5 Performance in the presence of Diabetic Macular Edema

To evaluate our method in the presence of fluid-filled regions associated with DME, a five-fold cross-validation has been performed on the combined *NORMAL-1* and *DME-1* dataset. The combined dataset was randomly divided into five parts. Each part consists of the B-scans from 4 OCT volumes; 2 volumes each from the *NORMAL-1* (21 OCT B-scans) and the *DME-1* dataset (22 B-scans). In each fold, the proposed method is tested on one part after learning a single CRF model to segment both the healthy and DME cases from the remaining 175 (87 Normal + 88 DME) B-scans. The performance of the proposed method on the combined dataset as well as the Normal and DME cases separately have been reported in Tables 3.8 to 3.11. Sample qualitative results on the B-scans with DME are presented in Fig. 3.10. The benchmark algorithms being unsupervised cannot be re-trained and their performance on the *NORMAL-1* dataset remains the same as presented in the Tables 3.2 and 3.3. On the combined dataset, the U-BLE varies in the range of 1.15 to 3.23 pixels with a mean of 2.04 pixels (see Table 3.8) and the Dice varies in the range of 0.75 to 0.92 with a mean of 0.84 (see Table 3.11) across all the layers.

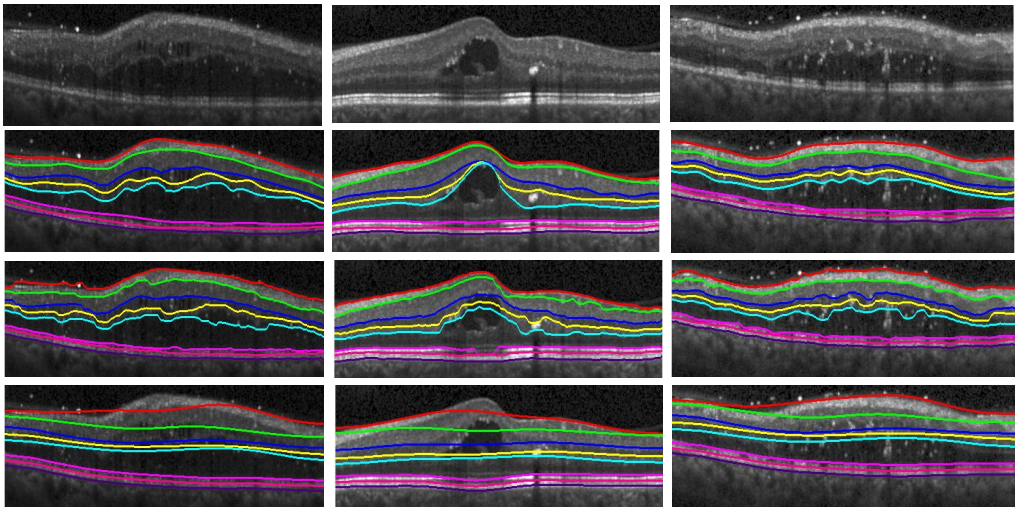

Figure 3.10: Qualitative results on 3 B-scans from the *DME-1* dataset is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark.

	ILM	NFL/GCL	IPL/INL	INL/OPL	OPL/ONL	IS/OS	*BM*	RPE_{out}
	DME dataset alone							
CASEREL	**1.14±0.28**	4.43±4.11	5.80±3.39	6.33±3.49	5.97±4.36	1.87±0.97	—	1.48±0.51
OCTSEG	2.44±3.36	10.61±9.19	7.66±6.97	—	6.06±5.24	**1.13±1.01**	—	**1.02±0.32**
IRA	4.29±5.69	12.92±8.67	9.23±7.17	7.05±4.98	6.09±4.20	2.56± 1.02	2.91±1.00	3.00±1.07
Proposed	1.22±0.44	**3.35±2.14**	**3.27±2.73**	**3.84±3.61**	**4.44±3.81**	1.44± 0.70	**1.34±0.43**	1.09±0.39
Manual Expert 2	1.27±0.41	1.77±0.69	2.12±1.66	2.21±1.46	2.49±1.63	1.25±0.53	1.27±0.50	1.25±0.46
	Normal dataset alone							
Proposed	1.08±0.32	1.82±0.89	1.64± 0.76	1.92± 0.71	1.99±0.94	1.14±0.49	1.39±0.66	1.54± 0.68
	Normal + DME combined							
CASEREL	**1.07±0.28**	3.68±3.56	5.22±2.94	5.75±2.98	5.54±4.28	1.57±0.99	—	1.58±0.57
OCTSEG	2.16±3.28	8.59±7.10	5.70±5.89	—	4.77±4.45	**1.12±1.27**	—	**1.31±1.25**
IRA	2.94±4.29	7.79± 8.08	5.48±6.36	4.44±4.42	4.19±3.66	1.78±1.11	2.44±1.12	2.20±1.20
Proposed	1.15±0.39	**2.59±1.81**	**2.46±2.17**	**2.89±2.78**	**3.23±3.04**	1.29±0.62	**1.36±0.56**	1.32±0.59

Table 3.8: Unsigned Boundary Localization Errors (mean ± standard deviation) on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold.

	ILM	NFL/GCL	IPL/INL	INL/OPL	OPL/ONL	IS/OS	*BM*	RPE_{out}
	DME dataset alone							
CASEREL	0.31±0.53	2.48±5.04	**0.00±5.42**	-0.83±5.29	-1.38±5.86	0.88±1.29	—	-0.05±0.68
OCTSEG	0.95±3.63	7.61±10.59	5.17±7.57	—	3.53±5.62	0.16±1.17	—	-0.09±0.71
IRA	1.64±5.70	4.72±12.72	4.34±8.61	2.41±5.37	1.78± 4.45	0.65±1.72	0.38±2.19	0.40±2.25
Proposed	**-0.05±0.75**	**0.29±3.20**	1.01±2.58	**0.73±3.41**	**1.35±4.17**	**-0.07±1.08**	**0.00±1.13**	**0.00± 0.84**
Manual Expert 2	-0.20±0.77	-0.30±1.06	-0.70±1.91	0.09±1.73	-0.62±1.97	-0.48±0.78	-0.44±0.94	-0.18±0.91
	Normal dataset alone							
Proposed	0.02±0.77	0.27±1.55	0.15±1.14	0.06±1.51	0.20±1.58	0.09±0.84	0.22±1.26	0.00±1.35
	Normal + DME combined							
CASEREL	0.16±0.58	1.96±4.28	**-0.58±4.82**	-1.05±4.83	-1.53± 5.76	0.47±1.26	—	0.03±0.86
OCTSEG	0.29±3.37	4.01±9.11	3.45±6.40	—	2.53±4.72	-0.03±1.31	—	-0.04±1.36
IRA	0.83±4.19	2.38± 9.46	2.19±6.55	1.22± 4.11	1.14±3.51	0.21±1.40	0.14±2.13	0.31±1.79
Proposed	**-0.02±0.76**	**0.28±2.52**	0.59±2.04	**0.40±2.66**	**0.78±3.21**	**0.01±0.97**	**0.11± 1.19**	**0.00± 1.12**

Table 3.9: Signed Boundary Localization Errors (mean ± standard deviation) on the combined DME and NORMAL-1 dataset. Best result among the automated methods in each column is indicated in bold.

	NFL	GCL-IPL	INL	OPL	ONL-IS	OS	RPE
	DME Dataset alone						
CASEREL	4.56±3.89	5.21±2.89	3.67±4.39	4.48±1.58	6.34±4.28	—	—
OCTSEG	9.67±6.94	5.48±3.17	—	—	6.09±5.03	—	—
IRA	10.23±4.76	5.86±2.30	4.01±4.22	**2.74±1.12**	6.14±3.92	**1.44±0.61**	**1.15± 0.33**
Proposed	**3.45±2.03**	**3.35±1.74**	**3.43±3.18**	2.95±1.42	**4.61±3.79**	1.64±0.79	1.41±0.51
Manual Expert 2	2.18±0.77	2.70±1.60	2.68±1.71	2.40±0.94	2.74±1.61	1.57± 0.51	1.50±0.43
	Normal dataset alone						
Proposed	2.06±0.85	2.21±0.73	2.12±0.66	2.43±1.10	2.24±1.05	1.65±0.61	1.76±0.64
	Normal + DME combined						
CASEREL	3.89±3.42	5.0±2.50	3.02± 3.30	4.71±1.57	5.83±4.16	—	—
OCTSEG	8.26±5.46	5.50±2.53	—	—	4.73±4.23	—	—
IRA	6.36±5.22	4.17±2.45	3.01±3.19	**2.61±1.09**	4.29±3.48	1.79±0.86	**1.52±0.77**
Proposed	**2.77±1.71**	**2.79±1.45**	**2.78±2.39**	2.69±1.30	**3.44±3.03**	**1.65±0.70**	1.58±0.60

Table 3.10: Mean Layer Thickness Error (mean ± standard deviation) in pixels for 7 tissue regions on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold.

	NFL	GCL-IPL	INL	OPL	ONL-IS	OS	RPE
	DME dataset alone						
CASEREL	0.78±0.14	0.74±0.14	0.54±0.16	0.57±0.15	0.87±0.07	—	—
OCTSEG	0.64±0.19	0.61±0.21	—	—	0.88± 0.07	—	—
IRA	0.49±0.17	0.55±0.17	0.44±0.20	0.51± 0.20	0.85±0.05	0.71±0.09	0.63±0.12
Proposed	**0.81±0.10**	**0.84±0.10**	**0.74±0.12**	**0.72±0.12**	**0.90±0.05**	**0.85±0.06**	**0.85±0.04**
Manual Expert 2	0.86±0.07	0.89± 0.05	0.80±0.06	0.72±0.09	0.88± 0.06	0.86± 0.05	0.84±0.05
	Normal Dataset alone						
Proposed	0.83±0.09	0.93±0.04	0.86±0.04	0.79±0.07	0.94± 0.02	0.87±0.05	0.86±0.06
	Normal + DME combined						
CASEREL	0.79± 0.14	0.78±0.13	0.57±0.16	0.58± 0.15	0.88±0.07	—	—
OCTSEG	0.63±0.16	0.70±0.19	—	—	0.89±0.07	—	—
IRA	0.64±0.20	0.73±0.22	0.65±0.25	0.65±0.21	0.89±0.06	0.77±0.11	0.74±0.14
Proposed	**0.82±0.09**	**0.88±0.09**	**0.80±0.11**	**0.75±0.10**	**0.92±0.04**	**0.86±0.05**	**0.85±0.05**

Table 3.11: Dice coefficient (mean ± standard deviation) for 7 tissue regions on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold.

3.4.6 Performance on DME cases alone

In this section, the proposed method has been evaluated on the *DME-1* dataset alone to compare its performance against some of the recent methods in [93], [99] that have been specifically designed to segment the OCT B-scans in the presence of DME through an explicit segmentation of the fluid-filled regions. A kernel regression (KR) based classification scheme was employed by the GTDP+KR method in [93] to explicitly segment the fluid-filled regions and the retinal layers which were further refined using a graph theory and dynamic programming (GTDP) framework.

A deep learning architecture called the *ReLayNet* was employed in [99] to segment both the fluid filled regions and layer boundaries. We followed the standard convention of splitting the *DME-1* dataset into the training and test sets as reported in [93], [99]. The B-scans from subjects 1-5 were used as the training set and the remaining B-scans from subjects 6-10 were used as the test set resulting in 55 B-scans in each set.

The mean Dice and the LTE metrics of the seven tissue regions are presented in Table 3.12. The results of the GTDP+KR has been reproduced from [93], while the performance of the *ReLayNet* and the second expert have been reproduced from [99]. The proposed method with a

mean Dice coefficient of 0.85 across the seven tissue layers, outperforms the GTDP+KR method which has a mean Dice of 0.82. The performance of the proposed method is below the DL based *ReLayNet* (mean Dice= 0.90). The reason for this could be that the fluid-filled regions lead to large deviations from the shape priors on the expected layer thickness and smoothness of the inner layers. An explicit modelling of the fluid filled regions as a separate auxiliary boundary between the OPL/ONL and the IS/OS boundaries similar to [96] can be explored in the future within our CRF framework to address this issue.

Nevertheless, the proposed method's performance still lies within the inter-observer variance in comparison to the manual markings of the second expert which has a mean Dice of 0.84. The performance of the second expert is marginally better than our method on the GCL-IPL and INL layers while our method outperforms it in the OPL, OS and the RPE layers.

	NFL	GCL-IPL	INL	OPL	ONL-IS	OS	RPE
Mean Dice coefficient							
GTDP+KR [93]	0.86	0.88	0.73	0.73	0.86	0.86	0.80
ReLayNet [99]	**0.90**	**0.94**	**0.87**	**0.84**	0.93	**0.92**	**0.90**
Expert 2	0.86	0.90	0.79	0.74	0.94	0.86	0.82
Proposed	0.86	0.88	0.77	0.76	**0.94**	0.88	0.87
Mean layer thickness error (pixels)							
GTDP+KR [93]	3.68	4.84	7.90	6.35	6.80	2.88	3.61
ReLayNet [99]	**1.50**	**1.20**	**1.00**	**1.31**	**1.35**	**0.62**	**0.92**
Expert 2	2.01	2.33	2.17	2.29	2.24	1.53	1.54
Proposed	2.56	2.54	2.39	2.13	2.25	1.42	1.42

Table 3.12: Mean Dice coefficient and the layer thickness error (in pixels) for 7 tissue regions evaluated on the *DME-1* dataset alone.

3.5 Discussions

A Matlab implementation of our method takes around 9 seconds to process each B-scan on a i7 processor with 8 GB RAM. In this work, each B-scan is segmented independently similar to the existing methods in [13], [14] and [93]. Although the proposed CRF framework can be directly extended to 3D by adding an inter-slice pairwise term (similar to the intra-layer pairwise terms defined in eq. 5) between the corresponding boundary points on the adjacent

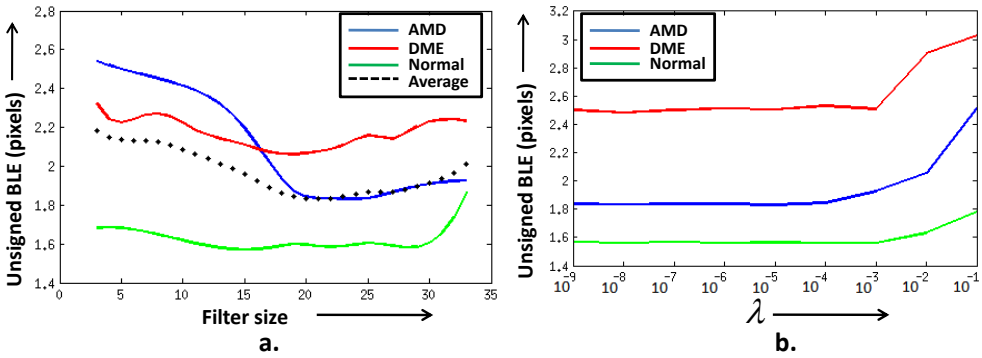

Figure 3.11: Variation in performance with respect to a) the filter size and b) the regularization weight λ. Lower values of U-BLE in pixels indicate better performance.

B-scans, it could not be evaluated due to the unavailability of the groundtruth markings for the consecutive B-scans. Due to the tedium involved in obtaining manual GT, the four public datasets (*NORMAL-1*, *NORMAL-2*, *DME-1* and *AMD-1*) only provide the GT for a few non-adjacent, linearly spaced B-scans from each OCT volume. Hence, a 3D CRF model could not be trained. Moreover, in retinal OCT imaging, the pixel resolution across the B-scans is approximately 10 times coarser than the intra-slice lateral resolution in most of the image acquisition settings (see for eg., Table 1). Hence, regularization across adjacent B-scans only has a marginal effect on the segmentation performance while adversely impacting the computational and memory requirements.

Next, we discuss the impact of the tunable hyperparameters on the performance followed by a discussion of the results presented in Section 3.4.

Effect of Hyperparameters on Performance : The proposed method has three tunable hyperparameters, the column spacing between the adjacent points on the boundary, the size of the convolutional filters $\mathbf{u_l}$ and $\mathbf{v_l}$ which capture the appearance of the layer boundaries and the intermediate tissue regions respectively, and the regularization weight λ in eq. 8 used during training. In all the experiments, the column spacing was empirically fixed to 4 pixels which resulted in 150 control points to represent each boundary. The intermediate boundary points were obtained through b-spline interpolation. This choice provided a good trade-off between the computational efficiency and the interpolation error.

The exact size of the convolutional filters and the regularization weight were fixed experimentally. The size of all the convolutional filters were kept equal to minimize the number

of tunable parameters. A small set of 15 B-scans were randomly selected from each of the *NORMAL-1*, *AMD-1* and the *DME-1* dataset. For each of the three sets, a separate CRF energy was trained using 8 B-scans and the remaining 7 B-scans were used as the validation set. At first λ was fixed to 10^{-4} and the filter size was varied in the range of 3 to 33 in steps of 2. Fig. 3.11 a. depicts the variation of the U-BLE in pixels against the varying filter size. The average performance across all the three datasets is depicted by the black dotted line. Overall, the performance of the proposed method was relatively stable with the U-BLE varying in the range of 1.6-2.6 pixels across the varying filter size for all the three sets. A filter size of 19×19 was found to be a robust choice across the normal and different pathological cases. Next, λ was varied in powers of 10 in the range of 10^{-9} to 10^{-1} keeping the filter size fixed at 19×19. Initially, the performance improved from 10^{-1} to 10^{-3} after which the U-BLE was relatively stable across all the three sets. Based on these experiments, the filter size and λ were fixed to 19×19 and 10^{-4} respectively, across all the experiments presented in Section 3.4.

	ILM	*NFL/GCL*	*IPL/INL*	*INL/OPL*	*OPL/ONL*	*IS/OS*	*BM*	RPE_{out}
NORMAL-1 (vs. IRA)	$1. \times 10^{-9}$	1×10^{-11}	0.0069	0.1056	7×10^{-4}	0.0090	8×10^{-9}	9×10^{-7}
NORMAL-2 (vs. OCTSEG)	0.3096	4×10^{-11}	8×10^{-9}	—	6×10^{-12}	3×10^{-5}	—	8×10^{-14}
NORMAL-1+DME (vs. CASEREL)	0.0011	5×10^{-9}	4×10^{-39}	3×10^{-40}	2×10^{-12}	2×10^{-7}	—	1×10^{-11}
NORMAL+AMD(vs. IRA)	8×10^{-6}	—	—	—	—	—	7×10^{-8}	0.8681

Table 3.13: p-values for the paired T-test between the U-BLE metric of the proposed method against the best performing method among CASERAL, OCTSEG, IRA. A p-value < 0.05 indicates statistically significant improvement.

Performance on *NORMAL-1* dataset: The results on the *NORMAL-1* dataset in Tables 3.2,3.3 illustrates the good performance of our method on OCT B-scans of healthy subjects. The U-BLE across the eight boundaries (see Table 3.2) varies in the range of 1.09 to 1.95 pixels with a mean value of 1.52 pixels which improves on the IRA (with a mean U-BLE of 1.77) by 16% which is the second best performing method. The improvement in performance over IRA is stastically significant for seven out of the eight boundaries with the exception of the INL/OPL boundary as indicated by the paired T-test values reported in Table 3.13 (first row). The S-BLE is close to 0 pixels across all boundaries indicating the absence of any significant bias towards over or under-estimating a boundary. The Dice coefficient for the seven retinal tissue layers (in Table 3.3) varies in the range of 0.8 to 0.94 with a mean value of 0.87 and is

consistently better than the other methods. In terms of LTE, the error of the proposed method is lower than the other methods on six out of the seven layers with a mean LTE of 2.00 pixels across all layers.

Our performance lies within the inter-expert variance with respect to the markings of the second expert on a subset of 28 images on the *NORMAL-1* dataset. The average U-BLE of our method over the 8 boundaries is 1.60 ± 0.30 pixels as compared to 1.91 ± 0.88 pixels for the second expert. Individually, our method performs better than the second expert on all except the NFL/GCL (l=2) boundary where our performance is comparable to that of the second expert (U-BLE of 1.75 as compared to 1.56 pixels).

Performance on *NORMAL-2* dataset: On the *NORMAL-2* dataset, our cross-testing performance is within the inter-expert variance, performing better than or equivalent to the second expert on all layers in terms of the Dice coefficient (see Table 3.5) and all except the NFL layer in terms of the LTE (1.86 pixels in comparison to 1.42 pixels for the second expert). The proposed method doesnot show any significant bias towards over or under-estimation of any boundary as indicated by a S-BLE value close to 0 pixels (in Table 3.4) for all the boundaries.

The good performance of our method on the *NORMAL-2* dataset even when trained on B-scans from the *NORMAL-1* dataset alone indicates the good generalizability of our method across different OCT scanners as the OCT volumes in the *NORMAL-1* and *NORMAL-2* dataset were acquired using Bioptingen and Spectralis SD-OCT scanners respectively.

In terms of the U-BLE metric reported in Table 3.4, OCTSEG is the best performing among the benchmark algorithms with a mean U-BLE of 2.23 pixels across the eight boundaries. The proposed method improves on the performance of OCTSEG by about 50% with a mean U-BLE of 1.11 pixels. The improvement in performance over OCTSEG is statistically significant for seven out of the eight boundaries excepth the ILM boundary as indicated by the paired T-test values in Table 3.13 (second row).

The better performance of our method with respect to the IRA (U-BLE of 3.04 pixels) illustrates the advantage of learning the energy over the handcrafted cost terms employed in IRA based on [91]. Moreover, while IRA performs the segmentation in 2 steps, the outer (1,7,8) layer boundaries followed by the inner (2-6) ones, our method extracts all the 8 boundaries in a single step.

Performance in the presence of AMD: The average U-BLE (see Table 3.6) of the proposed method on the ILM and RPE_{out} boundaries considering the AMD cases alone is 2.03 pixels which is only a slight improvement over the performance of CASEREL with a U-BLE of 2.07 pixels. However, the BM boundary which plays a crucial role in the detection of AMD is currently unavailable in the OCTSEG and CASEREL softwares. Considering all the three layers, our method outperforms IRA by 37% in terms of the average U-BLE across all the boundaries (average U-BLE of 2.08 pixels in comparison to 3.28 pixels for IRA) for the AMD cases. The improvement in performance over IRA is statistically significant for two out of the three boundaries except RPE_{out} as indicated by the paired T-test values reported in Table 3.13 (fourth row).

The S-BLE metric of our method is close to 0 for all the three boundaries and doesnot indicate any significant bias towards over or under-estimation. In terms of the Dice and LTE metric (see Table 3.7), the performance of the proposed method is similar to that of the second expert (LTE of 2.57 pixels compared to 2.60 for the second expert markings) on the tissue region between the ILM and BM boundaries. However, the performance drops by 4% for the RPE layer in terms of Dice indicating a scope for further improvement. This is consistent with the fact that AMD leads to drusen deposits in RPE layer.

Performance on combined Normal and DME cases: When the proposed method is jointly trained on the Normal and DME cases, it outperforms the three benchmark methods on each of the 8 boundaries both in terms of the Dice coefficient (see Table 3.11) and the U-BLE (see Table 3.8) metric considering the DME cases alone.

The CASEREL is the best performing among the benchmark algorithms and the proposed method improves on its performance by 35% (2.50 pixels in comparison to 3.86 pixels for CASEREL) in terms of the average U-BLE across all boundaries and 16% in terms of the Dice coefficient (0.81 of the proposed method in comparison to 0.70 for CASEREL). The improvement in performance over CASEREL is statistically significant as indicated by the paired T-test values reported in Table 3.13 (third row).

However, in comparison to the second expert, the performance of our method is slightly lower indicating a scope for further improvement. In terms of Dice, our performance is comparable to that of the second expert on the last four layers but drops by approximately 5% for each of the first three layers. This is consistent with the observation that the fluid-filled regions tend

to occur between the ILM and the INL/OPL boundaries. The absolute value of S-BLE is less than 1 pixel for all layers except IPL/INL and OPL/ONL for which there is a slight bias towards estimation of boundary to lie below the GT.

Effect of joint training on healthy Images: In Sections 3.4.4 and 3.4.5, the CRF models were learned on combined datasets consisting of both healthy and abnormal cases. This allows the method to be employed to segment new test cases for which any prior information on the presence of pathologies is unavailable. However, the healthy and abnormal B-scans differ widely in their appearance and there is also a large deviation in the distribution of the layer thickness and boundary smoothness statistics. Hence, the model trained on the combined datasets can adversely affect the segmentation performance on the healthy images. However, the results indicate that there is no significant decrease in performance on the *NORMAL-1* dataset. The average U-BLE of 1.52 pixels across the eight boundaries when trained on healthy images alone (in Table 3.2) drops to 1.56 pixels when trained jointly with DME cases (in Table 3.8). Similarly, considering the ILM, BM and RPE_{out} alone, the average U-BLE drops from 1.41 (in Table 3.2) to 1.42 pixels when the CRF is trained on the combined dataset with AMD cases (in Table 3.6).

3.6 Conclusions

The accurate segmentation of intra-retinal tissue layers plays an important role in the diagnosis of ocular diseases. In this work, we propose a supervised CRF framework for the joint multi-layer segmentation in macular OCT B-scans. The CRF energy consists of multiple cost terms to capture the appearance and the shape priors for each layer. It is linearly parameterized to allow a joint, end-to-end training of two convolutional filter banks and the relative weights of the shape priors by employing a SSVM formulation.

The proposed method has been extensively evaluated on 4 public datasets that cover a range of image resolution, scanners, image quality and contain B-scans of healthy as well as abnormal eyes suffering from AMD and DME. The quantitative and qualitative results demonstrate the better performance of the proposed method in comparison to three benchmark methods on both healthy and abnormal images. The improvement in performance over the IRA software

that employs a similar energy function demonstrates the effectiveness of learning the energy over handcrafting.

In case of the healthy images, our performance is within the inter-observer variability and generalizes well across B-scans acquired using different SD-OCT scanners as illustrated by the good cross-testing performance on the NORMAL-2 dataset after being trained on images from NORMAL-1 dataset alone. Our method can also be adapted to various pathologies asscociated with AMD and DME by re-training it on appropriate images. In each case, a single CRF model was learned on the combined healthy and abnormal dataset to allow the segmentation of a new test image without any prior information about the presence of pathologies in it. Though the proposed method outperforms the three benchmark methods in the presence of abnormalities, its performance is still lower than that of the second expert indicating a scope for further improvement.

Currently, the performance of our method when evaluated on the DME cases alone lies within the inter-observer variability. However, explicit modelling of the fluid filled regions within the CRF framework needs to be explored in the future to further improve the performance. Future work may also include evaluation of our method on OCT B-scans of the peri-papillary region. The proposed method can be utilized as an aid to the ophthalmologists in clinical practice and large-scale clincical studies for the quantitative analysis of structural changes in individual retinal layers.

Chapter 4

RACE-net: A Recurrent Neural Network for Biomedical Image Segmentation

The segmentation of anatomical structures in medical images plays an important role in the diagnosis and treatment of diseases. It is frequently used to visualize organs, extract quantitative clinical measurements, define the region of interest to localize pathologies and aid in the surgical or radiation treatment planning. Accurate segmentation algorithms can save the time and effort of medical experts and minimize the intra and inter-subject variability involved in manual segmentation. However, this task is often challenging due to the lack of sufficient contrast between the anatomy of interest and its background, large variability in its shape, variations in image quality across subjects or scanners and the presence of noise and non-uniform illumination.

In Chapters 2 and 3, we have explored a Conditional Random Field (CRF) based Energy Minimization framework for the extraction of the boundaries of the anatomical structures in the Color Fundus (CF) and Optical Coherence Tomography (OCT) images. However, the Level Set based Deformable Models (LDM) provide an alternative energy minimization based formulation for the segmentation task. Motivated by the recent success of deep learning in biomedical image segmentation, in this Chapter, we propose a deep Recurrent Neural Network (RNN) architecture, called the Recurrent Active Contour Evolution Network (RACE-net) which is inspired from the LDM. Given a rough localization of the anatomical structure to be segmented, RACE-net iteratively evolves it using a combination of a constant and a mean curvature velocity which are learned from a set of training images in an end-to-end manner. The key contributions of this chapter are: a) LDM is formulated as a novel RNN architecture; b) An appropriate loss

function is introduced to overcome the problems related to the re-initialization of the level set function in LDMs and the vanishing gradients during the training of the RNN; c) The constant and mean curvature velocities are modeled by a novel feed-forward architecture inspired from the multi-scale image pyramid; d) In addition to the segmentation of Optic Disc and Cup in CF images, we have also demonstrated the effectiveness of the proposed method on a variety of other structural segmentation tasks, such as the cell nuclei in histopathological images and the left atrium in 3D cardiac MRI volumes.

4.1 Background

The use of a fixed protocol and view during image acquisition and the overall similarity in the anatomical structures across subjects often allow for a rough localization of the structure of interest using simple image processing techniques and/or spatial priors. However, an accurate segmentation at a sub-pixel level is often essential for proper diagnosis. Traditionally, the LDMs (also known as the geometric or implicit active contour models) have been widely used for this task [114] in which an initial curve (or a surface in 3D) is iteratively evolved by a curve evolution velocity until it converges onto the desired boundary. Typically, the curve evolution velocity comprises image dependent and regularization terms. The image dependent terms drive the curve towards the desired boundary. On the other hand, the regularization terms such as the boundary length minimization [115] preserves its smoothness, while the balloon force [116] is used to drive the evolving curve into the concave regions in the object boundary.

The LDMs provide an Energy Minimization framework (see section 4.2.1) which allows an explicit modeling of the high level constraints on the boundary such as the trade-off between the intensity discontinuity [115], [117] at the boundary and its smoothness, or the regional homogeneity within and outside it [118]. However, LDMs cannot be used "off-the-shelf" as the curve evolution velocity has to be specifically handcrafted for each segmentation task, failing which, the curve can take a large number of iterations to converge, get entrapped near noise or spurious edges, or smooth out the sharp corners of the object of interest.

Recently, the Convolutional Neural Networks (CNNs) have been widely applied to various medical image segmentation tasks such as the extraction of neuronal structures in 2D electron microscopy [119], [102], prostate segmentation in MRI volumes [120] and, the Optic Disc and

vessel segmentation in fundus images [121]. Convolutional Neural Network (CNN)s learn a hierarchical feature representation of the images from large datasets in a supervised end-to-end manner, eliminating the need for handcrafted features. In the sliding window approach [119], [122], [123] a CNN is employed to classify the pixels into foreground or background at a patch level followed by a Conditional Random Field (CRF) [122] or LDM [123] based post-processing to incorporate the global context. Alternatively, the pixel labeling for the entire image is obtained in a single step by employing a deconvolutional architecture which consists of a "contraction" followed by an "expansion" stage [102], [120], [124]. The contracting stage has a series of convolutional layers with pooling layers in between which successively reduce the resolution of the feature channels. The expansion stage has one convolution layer corresponding to each contraction layer with upsampling operations in between to recover the original image resolution. Additional skip connections are also employed to directly connect the corresponding contraction and expansion layers.

CNNs lack an explicit way to capture high level, long range label dependencies between similar pixels and the spatial and appearance consistency of the segmentation labels [125]. This can result in poor object dilineation and small spurious regions in the segmentation output. Few RNN architectures have been explored to address these issues. RNNs generally employ a simple recurrent unit (RU) which comprises a single neuron with a feedback connection. Gated variants of the RUs such as LSTMs [126], [127] are commonly used to overcome the problem of vanishing gradients encountered during training over large time-steps. The multi-dimensional RNN (MDRNN) was explored in [128] to segment perimysium in skeletal muscle microscopy images. In MDRNN, the RUs are connected in a grid-like fashion with as many recurrent connections as there are spatial dimensions in the image. At each step, the RU receives the current pixel as input and the hidden states of the pixels explored in each direction in the previous step through feedback connections, thus recursively gathering information about all other pixels in the image. PyraMiD-LSTM [129] modified the toplology of the feedback connections in MDRNNs to improve its computational efficiency and applied it to segment neuronal electron microscopy images and MRI brain volumes.

Some Deep learning architectures have also attempted to combine RNN and CNNs in a *time-distributed* manner where the external input to the RNN at each time-step was provided by the output of the CNN. Such architectures have been employed in [126], [127] to segment the

intra-retinal tissue layers in OCT and 3D neuronal structures in electron microscopy images respectively. Both [126], [127] employed bi-directional RNNs to capture the context from both the past as well as the future time-steps. Moreover, multiple layers of RNNs were stacked together to obtain a deep RNN architecture. Alternatively, CNN and RNNs can also be integrated by replacing the RU by a CNN within the RNN architecture. Such an architecture was employed in [125] to formulate the mean-field based inference of a CRF as a RNN. The individual iterations of the mean-field algorithm was modeled as a CNN and the multiple mean-field iterations were implemented by using a simple feedback connection from the output of the entire CNN to its input, resulting in a RNN.

In comparison to CNNs, RNNs have received relatively less attention in the field of biomedical image segmentation. In this work, we explore a RNN architecture similar to [125] which combines the advantages of both LDM and CNN. In comparison to the LDMs, RACE-net learns the level set curve evolution velocity in an end-to-end manner. An attractive feature of the RACE-net is that the evolving curve was empirically found to converge onto the desired boundary over large distances in very few (5-7) time steps on a diverse set of applications. In contrast, LDMs either fail to converge over large distances or require a large number of iterations.

The RACE-net offers two main advantages over the CNNs. First, it provides a boundary-based alternative to the existing region-based pixel labeling approach of the CNNs. The RACE-net models a generalized level set curve evolution (see Section 4.2.1). Consequently, it can explicitly learn the high level dependencies between the points on the object boundary to preserve its overall shape and smoothness. At the same time, it can also maintain an optimal trade-off between the boundary discontinuity and the regional homogeneity of the structure in a convolutional feature space (which is learned in an end-to-end manner). Secondly, RACE-net can have a very complex network structure while utilizing very few learnable network parameters in comparison to the CNNs. Fewer network parameters serve to a) reduce the risk of over-fitting on small training datasets which is particularly useful in the medical domain where obtaining the ground truth markings is expensive; b) It reduces the computation time and memory requirements of the pre-trained architecture allowing it to be deployed on systems with limited resources [130].

4.2 Method

We begin by obtaining a generalized level set equation for a LDM in Section 4.2.1 to model it as a RNN. Next, we present a Feedforward Neural Network (FFNN) architecture which is designed to model each time-step of the curve evolution. The FFNN consists of a customized layer to compute the normal and curvature of the level set function (Section 4.2.2) and a novel CNN architecture to model the evolution velocities (Section 4.2.2.1). Finally, we show how the entire curve evolution is modeled as a RNN in Section 4.2.3 with an appropriate loss function (Section 4.2.3.1) for training.

4.2.1 A generalized PDE for curve evolution

Let I represent the image to be segmented. Consider a family of 2D planar curves (the discussion can be easily generalized to 3D surfaces) represented using an arc length parameterization as $C(s) = \{(x(s), y(s))|0 \leq s \leq l\}$ where $(x(s), y(s))$ are points on C and l denotes the length of the curve. The evolution of C, can be modeled by the Partial Differential Equation (PDE) $\frac{\partial}{\partial t}C(s,t) = V.\mathbf{n}(s,t)$, where t is the temporal parameter, $\mathbf{n}(s,t)$ represents the normal vector to C at $(x(s), y(s))$ and V is a velocity function defined over the spatial support of I. In order to define V in a meaningful manner, an Energy functional $E(C(s))$ is defined such that its minimum corresponds to the object boundary. $E(C(s))$ is minimized using functional gradient descent by iteratively evolving $C(s)$ in the *negative direction of the Euler-Lagrange* of E, i.e., $V = -\frac{d}{dC}E(C(s))$. Typically, $E(C(s))$ is composed of a linear combination of boundary and region-based cost terms such that

$$\underset{C}{\operatorname{argmin}}\ E(C(s)) = \gamma_1 . \oint_C h(x,y)ds + \gamma_2 . \iint_{R_c} q(x,y)dA, \tag{4.1}$$

where $h(x,y)$ and $q(x,y)$ are functions defined over the spatial support of I; γ_1 and γ_2 are the relative weights. The boundary cost term minimizes the line integral of $h(x,y)$ along the curve C. Both $h(x,y)$ and $q(x,y)$ can be composed of a weighted sum of multiple functions. For instance, $h(x,y)$ is often composed of an edge indicator function which is inversely proportional to the image gradient magnitude [115], [116] to drive C towards the edges. Setting $h(x,y) = 1$ as a constant scalar function serves to minimize the curve length and hence commonly used as a regularization term to smooth the boundary [116], [118].

The second term helps minimize the regional cost as it is the area integral of $q(x, y)$ in the region enclosed by C. Assigning $q(x, y) = \lambda_{in}.|I - c_{in}|^2 - \lambda_{out}.|I - c_{out}|^2$ leads to the Chan-Vese model [118], where λ_{in}, λ_{out} are the scalar weights. The terms c_{in} and c_{out} represent the mean intensities of the object and the background regions respectively, at each time step of the curve evolution. Setting $q(x, y) = 1(or - 1)$ serves to minimize (or maximize) the area enclosed by C and leads to the balloon force [116] which defines the default tendency of the curve to contract (or expand) in the absence of nearby edges in the image and speeds up the curve evolution.

The Euler-Lagrange for the boundary and the regional cost terms in eq. 4.1 are given by $\gamma_1.\{\langle \nabla h, \mathbf{n}\rangle + h.\kappa\}\mathbf{n}$ and $\gamma_2.q.\mathbf{n}$ respectively, where κ represents the curvature at $C(s)$ (See Appendix C for details). Thus, the curve evolution which minimizes eq. 4.1 is given by

$$\frac{\partial}{\partial t}C(s,t) = -\{\gamma_2.g.\mathbf{n} + \gamma_1.h.\kappa.\mathbf{n}\}, \tag{4.2}$$

where $g = \left(\frac{\gamma_1}{\gamma_2}.\langle \nabla h, \mathbf{n}\rangle + q\right)$. The curve evolution under the velocities $g.\mathbf{n}$ and $h.\kappa.\mathbf{n}$ are known as the constant flow and the mean curvature flow respectively.

During implementation, C is often represented using a level set function due to its ability to handle topological changes such as the merging or splitting of the boundary. The level set $\phi(x, y; t)$ is a scalar function which satisfies the following properties at every time-step t of the curve evolution: a) $\phi(x, y; t) > 0$ inside the region enclosed by the evolving curve $C(t)$; b) $\phi(x, y; t) < 0$ in the region outside $C(t)$; c) $C(t)$ is represented by the zero level set $C(t) = \{(x, y)|\phi(x, y; t) = 0\}$ of the level set function, and d) $\phi(x, y; t)$ should be differentiable with respect to the spatial co-ordinates x, y. Although, $\phi(x, y; t)$ can be any arbitary function that satisfies these properties, the Signed Distance Function (SDF) is commonly used to initialize the level set function in traditional LDMs for numerical stability. SDF has the desirable property that $|\nabla\phi(x, y)| = 1$ at all spatial co-ordinates x, y. However, $\phi(x, y; t)$ doesnot remain a SDF during the curve evolution and the evolving curve needs to be re-initialized to a SDF after every few iterations.

The curvature and the normal vectors in the level set representation can be shown to be $\kappa = div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)$ and $\mathbf{n} = -\frac{\nabla\phi}{|\nabla\phi|}$ [131]. Substituting these values in eq. 4.2 results in the equivalent level set equation for the curve evolution,

$$\frac{\partial\phi}{\partial t} = \alpha_1.g(I, \phi).|\nabla\phi| + \alpha_2.h(I).\kappa|\nabla\phi|, \tag{4.3}$$

where $\alpha_1 = -\gamma_2$ and $\alpha_2 = -\gamma_1$ are the scalar weights. The sign (positive or negative) of α_1 and α_2 determines the direction (inward or outward) of the curve evolution.

We propose a RNN architecture inspired from the PDE in eq. 4.3. Each time-step of the curve evolution is modeled as a FFNN by approximating the $g(I, \phi)$ and $h(I)$ functions by two separate CNNs. To simplify the network architecture, the $g(I, \phi)$ is learned directly and not restricted to the form $\left(\frac{\gamma_1}{\gamma_2}.\langle \nabla h, \mathbf{n} \rangle + q\right)$. This leads to a more generalized curve evolution model. Within the proposed RNN architecture, ϕ_t is evolved as

$$\phi_{t+1} = \alpha.\phi_t + \{\alpha_1.g(I, \phi) + \alpha_2.h(I).\kappa\}.|\nabla\phi|, \tag{4.4}$$

where α is a non-negative weight used to scale the level set function at each time-step. The parameters α, α_1 and α_2 are learned as a part of the network weights. The learned scale factor α was empirically found to be hepful in evolving the curve over large distances when the boundary initialization was far from the object boundary. We note that scaling ϕ_t by α preserves the zero level set ϕ_t [1] . Moreover, the non-negativity constraint on α ensures that the value of $\alpha.\phi_t$ remains positive in the region inside and negative outside the zero level set respectively. Thus, the conditions of the level set function are preserved during the curve evolution.

4.2.2 Single time-step of the Curve Evolution

Let ϕ_t be a single channel input to the CNN which denotes the evolving level set function at a time step t. The evolution of ϕ_t to ϕ_{t+1} is modeled as a FFNN depicted in Fig. 4.1 a. Both $g(I, \phi)$ and $h(I)$ are modeled using CNNs (details are described below in Section 4.2.2.1) which output a single channel feature map. We propose a customized network layer architecture $C(\phi)$ to compute the $|\nabla\phi|$ and κ from a given ϕ. The terms $g(I, \phi).\mid \nabla\phi \mid$ and $h(I).\kappa.\mid \nabla\phi \mid$ used in the right hand side of eq. 4.4 are obtained using a pixel-wise multiplication of $g(I, \phi)$ and $h(I)$ with the appropriate outputs from $C(\phi)$ as illustrated in Fig. 4.1 a.

The input ϕ_t and the computed feature maps $g(I, \phi).\mid \nabla\phi \mid$ and $h(I).\kappa.\mid \nabla\phi \mid$ are merged by concatenating them along the channel dimension, resulting in a three channel feature map. Next, a (1×1) convolution filter (across the three channels) with a linear activation function is applied to obtain the updated ϕ_{t+1}. Since, the merged feature map has three channels, the (1×1) convolution filter applied at each pixel position has three learnable weights which

[1] as multiplication of α to a 0 in ϕ_t will remain zero.

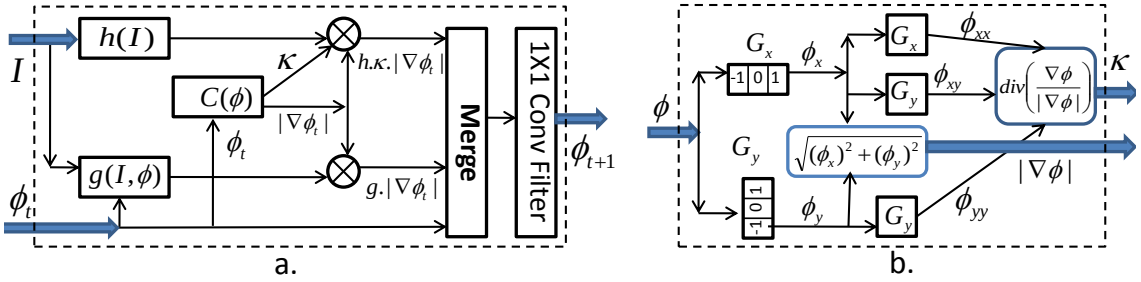

Figure 4.1: a. The Feedforward neural network (FFNN) architecture that models the evolution of the level set function ϕ in a single time step t. **b.** The details of the customized $C(\phi)$ module used within the FFNN in Fig. a. to compute the normal and curvature of ϕ.

correspond to the α, α_1 and α_2 in eq. 4.4. A non-negative constraint is applied only to the first weight of the convolution filter corresponding to α [2]

A detailed depiction of $C(\phi)$ is shown in Fig. 4.1 b. The two convolutional filters G_x, G_y are used to obtain the first and second order derivatives of ϕ. We *freeze* the weights of G_x, G_y to the one-dimensional gradient filter kernels depicted in Fig. 4.1 b to ensure that they are not modified by the backpropagation algorithm during training. Thereafter, the curvature $\kappa = \frac{\phi_{xx}.\phi_y^2 - 2.\phi_x.\phi_y.\phi_{x,y} + \phi_{yy}.\phi_x^2}{(\phi_x^2+\phi_y^2)^{3/2}}$ and $\mid \nabla\phi \mid$ are computed by two custom layers defined using the Theano library [132].

4.2.2.1 Network Architecture of $g(I,\phi)$ and $h(I)$

$g(I,\phi)$ and $h(I)$ are modeled using a similar CNN architecture but with separate network weights and inputs. While I is used as an input for $h(I)$, both I and ϕ_t are concatenated to obtain a multi-channel input for $g(I,\phi)$. Minimizing the number of network parameters can reduce the memory requirements and help prevent over-fitting. This is achieved by designing a novel CNN architecture depicted in Fig. 4.2 which is inspired from the multi-scale image pyramid used in image processing. In Fig. 4.2, $\{I_l \mid 0 \leq l \leq 3\}$ represents a four level image pyramid. I_0 is the input to the CNN, while the successive scales I_l are obtained by applying *average pooling* in a 2×2 neighborhood on I_{l-1}. The segmentation proceeds from a coarse to fine scale from $l = 3$ to 0. The output feature channels from the coarse scale I_l is upsampled

[2] the keras.constraints.NonNeg() function was modified during implementation.

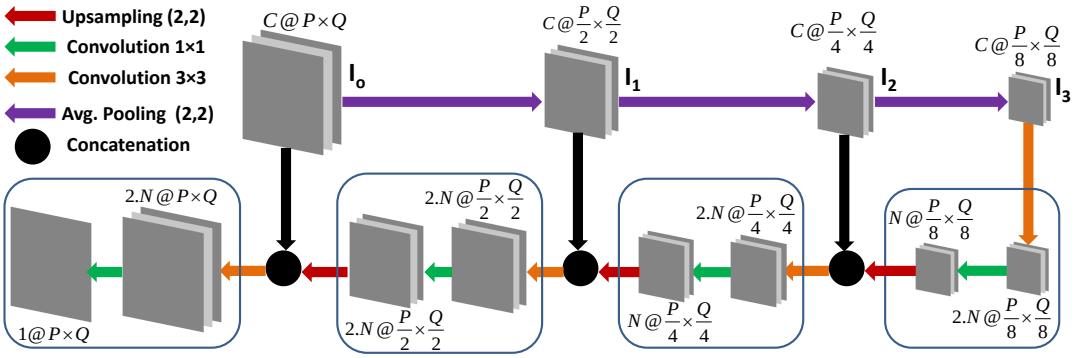

Figure 4.2: The proposed CNN architecture to model $h(I)$ and $g(I, \phi)$.

and concatenated with the image in the successive finer scale I_{l-1} to obtain the input for the convolutional layers in the $l-1^{th}$ scale. At the coarsest scale (l=3), only I_3 is used as the input.

The processing at each scale is done in two convolutional layers. The first convolutional layer (depicted by orange arrows) has $2.N$ filters to capture the patterns in the local neighborhood. The filter's receptive field size S_l is decreased from the coarse (l=3) to the finest scale (l=0), ie., $S_l \geq S_{l-1}$. The second convolutional layer in each scale (depicted by green arrows) has N, 1×1 ($\times 1$ in 3D) filters to reduce the dimensionality of its input feature channels from $2.N$ to N thereby *reducing the number of learnable filter weights in the next scale by half*. This is because the number of network parameters in a convolution layer is the product of filter kernel size and the number of input feature channels. Moreover, it adds an additional non-linearity to the network. An exception is the last layer which has only one (instead of N) filter to ensure a single feature channel as the final CNN output. Thus, the overall network size is controlled by a single tunable hyper-parameter N. All convolutional filters use the Rectified Linear Unit as the activation function. An exception is the final 1×1 convolutional filter at $l = 0$ which uses a linear activation to allow the RACE-net to learn arbitrary updates for ϕ in both positive (curve expands) or negative (curve contracts) direction. This was empirically found to be crucial in improving the segmentation performance.

The proposed architecture can be visualized as a modification to the U-net [102] where the encoder network is replaced by an image pyramid representation resulting in a drastic reduction of the network parameters.

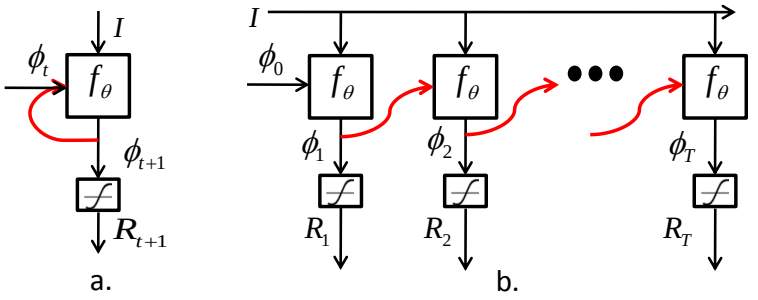

Figure 4.3: a. Curve Evolution as a RNN. **b.** The unrolling of the RNN in a. over time. The recurrent feedback connections are depicted in red.

4.2.3 Curve Evolution as a RNN

The FFNN in Fig. 4.1 a models one iteration of the curve evolution. At a time-step t, it takes the input image I and the current level set function ϕ_t as input to compute the evolved level set function ϕ_{t+1}. Let the function $f_\theta : \phi \times I \rightarrow \phi$ represent the complex, non-linear operation performed by the FFNN where θ denotes its learnable network parameters.

If ϕ_t evolves over a maximum of T time-steps during segmentation, then θ should be learned such that the final ϕ_T converges onto the desired boundary. We propose to do this by defining a RNN that can model the recursive equation: $\phi_{t+1} = f_\theta(\phi_t, I), 0 \leq t \leq T$. This is achieved by introducing a recurrent feedback connection which provides the output of f_θ as a feedback input to itself in the next iteration (see Fig. 4.3 a) for T time-steps. The feedback connection allows us to model all iterations of the curve-evolution within a RNN architecture which we refer to as the RACE-net. It is different from the most commonly used RNN architecture ([133] for example) where the output of an individual *neuron* is fed back onto itself to update its hidden state. In contrast, in the RACE-net architecture, the output of *an entire feed-forward network* is fed back to its input in the next time-step. An example of a similar RNN architecture can be found in [13] which modeled an iterative mean-field approximation based algorithm for CRF inference as a RNN. Each iteration of the update of the marginal probabilities of a fully connected CRF was modeled by a CNN and multiple iterations of the update were modeled by adding a feedback connection (similar to our method) from the output of the CNN to its input.

The RACE-net depicted in Fig. 4.3 a has the same network parameters θ as the FFNN in Fig. 4.1 a used to model a single time-step. However, the key difference between them is the

feedback connection in Fig. 4.3 from the output ϕ_{t+1} to its input ϕ_t in the next time-step. It allows the RACE-net to be trained in an end-to-end manner by defining a loss function over the entire curve evolution spanning multiple (T) time-steps. The initial level set function ϕ_0 is the input to the RACE-net, ϕ_T is its final output, while $\phi_t, 1 \leq t \leq T-1$ are the feature maps computed by f_θ in the intermediate time-steps of the RACE-net. A simple feedback connection without any sophisticated gated architecture (in contrast to the LSTMs) was employed in the RACE-net and auxiliary loss functions were incorporated at each individual time-step of the RNN during training to overcome the vanishing gradients problem.

A suitable loss function must maximize the regional overlap between the Ground Truth (GT) segmentation mask and the segmentation estimated by the RACE-net. Hence, the level set function ϕ_t is converted into a segmentation mask R_t within the the RACE-net architecture itself. Since ϕ_t is positive inside and negative outside the boundary curve, the binary segmentation mask of the foreground region can be obtained by thresholding $\phi_t > 0$. This hard thresholding operation is equivalent to applying the step function as an activation function to ϕ_t. Since step function is non-differentiable, it is approximated by a sigmoid function to derive a softmap that closely approximates a binary image, ie., $R_t = Sig(\phi_t)$. Both ϕ_t and R_t are the outputs of the RACE-net. While ϕ_t is fed back to provide the input to f_θ for the next time-step, R_t is an additional auxilliary output used to define the loss function for training. The final binary segmentation mask can be obtained by thresholding $R_T > 0.5$.

By *unrolling* the RNN through time, RACE-net can be viewed as a cascade of T feed-forward networks (f_θ) with identical parameters θ for each time-step (see Fig. 4.3 b). Thus, by choosing an appropriate T, an arbitrarily deep network can be obtained while keeping the number of shared network parameters constant. θ is learned in an end-to-end manner across the T time-steps using Back Propagation Through Time (BPTT) [134] which is logically equivalent to a simple backpropagation over the unrolled architecture of the RACE-net.

While g is a function of both I and ϕ, h is a function of I alone (see eq. 4.3). Thus, during implementation, h is computed only once before the feedback connection and used as an additional input at each time-step, to improve the computational efficiency.

4.2.3.1 Loss Function

In order to define an appropriate loss function to train the RACE-net, we need a metric to measure the regional overlap between the GT and the estimated segmentation masks. The Dice coefficient was used for this purpose. Let Y denote the binary ground truth (GT) segmentation mask for a training image I. Without any loss of generality, let R denote the segmentation mask obtained from the RACE-net for I at a particular time-step. The Dice coefficient between R and Y is defined as

$$Dice(R, Y) = \frac{2.|R \cap Y|}{|R| + |Y|} = \frac{2\sum_i r_i.y_i}{(\sum_i r_i + \sum_i y_i)}, \tag{4.5}$$

where r_i and y_i represents the value at the i^{th} pixel/voxel in R and Y respectively and the summations run over all the pixels/voxels in I. The Dice coefficient is differentiable with respect to r_i yielding the gradient

$$\frac{\partial}{\partial r_i}\frac{2\sum_i r_i.y_i}{(\sum_i r_i + \sum_i y_i)} = 2.\frac{y_i.(\sum_i r_i + \sum_i y_i) - (\sum_i r_i.y_i)}{(\sum_i r_i + \sum_i y_i)^2}. \tag{4.6}$$

The RACE-net should be trained to minimize the loss $S(R_T, Y) = 1 - Dice(R_T, Y)$ where $R_T = Sig(\phi_T)$ is the final segmentation mask obtained after the T time-steps. Though an iterative evolution of ϕ_t is commonly employed in the implementation of the LDMs, they lead to numerical instabilities and require ϕ_t to be re-initialized to a signed distance function (SDF) after every few iterations. Since $|\nabla\phi_t| = 1$ is a property of the SDF [135], an additional regularization term $|\nabla\phi_t - 1|_2^2$ is added to the loss function at each time-step to ensure that the ϕ_t remains close to a SDF during the curve evolution.

The RNN architectures are also susceptible to vanishing gradients. The gradient of the loss function at the final time-step T becomes extremely small when *backpropagated through time* to the earlier time-steps, thereby inhibiting the network's ability to learn curve evolutions across a long distance. Hence, an auxiliary loss function is added at each time-step to ensure that the network in time step t, receives multiple gradients backpropagated from the loss functions of all the following time-steps. Though the GT for the intermediate R_t is not available, we assume that the dice coefficient should successively increase with respect to Y as the curve evolves. Therefore, the auxiliary loss functions are defined as $\lambda_t S(R_t, Y)$ where λ_t represents the relative weightage of these terms such that $\lambda_t < \lambda_{t+1}$. Thus, the total loss L for the RNN

is defined as

$$L = \sum_{t=1}^{T} \left\{ \lambda_t S(R_t, Y) + \beta . ||\nabla \phi_t| - 1|_2^2 \right\}. \tag{4.7}$$

In our experiments, $\lambda_t = 10^{\{\lfloor -T/2 \rfloor + t\}}$ was decayed in powers of 10 (earlier time-steps had lower weights for $S(R_t, Y)$) while the relative weight $\beta = \lambda_1$ was fixed across all time-steps. The RACE-net is trained by minimizing L using BPTT [134] with the ADAM algorithm [136] to adapt the learning rate.

4.3 Results

The proposed method was evaluated on three different segmentation tasks encompassing a wide range of anatomical structures and imaging modalities to demonstrate its robustness and generalizability. The specific tasks considered were the segmentation of the Optic Disc and Cup in color fundus images, cell nuclei in histopathology images and the left atrium in cardiac MRI volumes. The left atrium segmentation was performed in 3D. The RACE-net was implemented for both 2D and 3D images in the Keras API [137] using Theano as the backend on a 12GB Nvidia Titan X GPU and Intel i7 processor. The details of each experiment is presented below.

4.3.1 Optic Disc and Cup Segmentation

Fundus photography is used to capture 2D color images of the retina which forms the interior surface of the eye. In fundus images, the Optic disc (OD) appears as a bright, roughly elliptical structure with a central depression called the Optic Cup (OC). Glaucoma is a chronic optic neuropathy which leads to the loss of retinal nerve fibers and results in the enlargement of the OC with respect to the OD. Hence, an accurate segmentation of the OD and OC is clinically important. OD segmentation suffers from vessel occlusion, indistinct gradient at boundaries and the presence of abnormalities such as Peripapillary atrophy. OC segmentation is more challenging as it is primarily characterized by the 3D depth information which is unavailable in the 2D color fundus images [138].

The proposed method was evaluated on the public DRISHTI-GS1 dataset [6] which consists of 50 training and 51 test images with ground truth (GT) OD and OC segmentation masks. The two structures were segmented sequentially using two different RACE-net architectures. At first, a square Region of Interest (ROI) (depicted in the first columns of Fig. 4.4 b, i-iv)

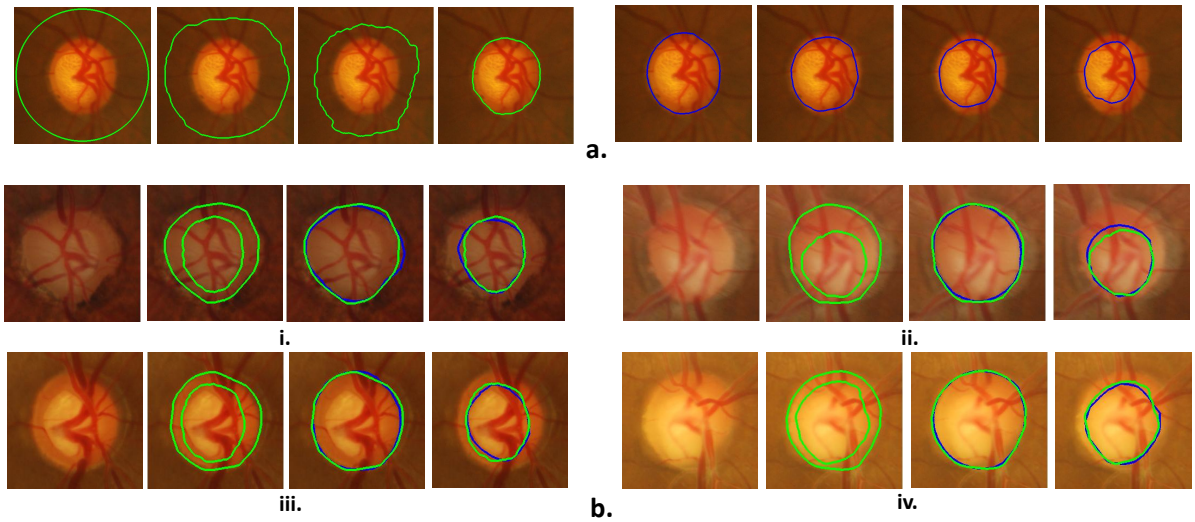

Figure 4.4: a. depicts the curve evolution of RACE-net over the intermediate time-steps, T=0, 2, 4, 6 for the OD boundary (in green) and over T=0, 1, 3, 5 for the OC boundary (in blue) from left to right. **Fig. b** depicts four sample results for OD and OC boundaries. In each subimage **i-iv**, the cropped region of interest is depicted in column 1, the result of the RACE-net for the OD and OC boundaries (in green) are depicted in column 2 followed by a comparison with the Ground truth markings (in blue) for OD (column 3) and OC (column 4).

was extracted from the entire fundus image using a method based on [138] which employed an intensity thresholding followed by a circular hough transform on the edge map extracted from the fundus image. The initial level set function was initialized as the largest circle within the ROI. Thereafter, the result of the OD segmentation was used to initialize the OC segmentation. The network parameters were fixed at N=6(6) and T=6(5) for OD (OC). The filter sizes S_l at each scale were fixed at 9×9, 7×7, 5×5 and 3×3 respectively from the coarse ($l = 3$) to the fine ($l = 0$) scale in both the architectures, resulting in a total of $26,322$ learnable parameters for each network. The number of network parameters is independent of T since they are shared across the time steps. The training dataset was augmented by applying horizontal, vertical translations and vertical flipping to each image. During testing, the RACE-net took 0.21 seconds to segment the OD and 0.19 seconds to segment the OC using a GPU.

The qualitative results depicted in Fig. 4.4 demonstrate the strength of the RACE-net. The curve evolution for OD and OC in Fig. 4.4 a indicates that the network architecture is able to learn to evolve inwards, stopping at the desired boundary while preserving the boundary

smoothness. The OC boundary does not correspond to a sharp intensity gradient demonstrating the strength of the method over traditional edge based LDMs.

The quantitative performance has been reported in Table 4.1. In addition to the Dice coefficient, the average Boundary Localization Error (BLE) [138], [6] has also been reported which measures the average distance (in pixels) between the GT and the estimated boundary points. It is defined as $BLE = \frac{1}{n}\sum_{\theta_1}^{\theta_n} |r_{gt}^{\theta} - r_{est}^{\theta}|$, where r_{est}^{θ} and r_{gt}^{θ} denote the Euclidean distances (in the radial direction) of the estimated and the GT boundary points from the centroid of the GT respectively, at orientation θ. The average BLE is computed based on the localisation errors at $n = 24$ equi-spaced boundary points. The desirable value for BLE should be close to 0 pixels. The RACE-net outperforms several existing methods. Both [36] and [74] employ a modified Chan-Vese based LDM to segment the OD while a supervised super-pixel feature classification based method is employed in [42] for both OD and OC segmentation. On the task of OD segmentation, the proposed architecture shows marginal improvement with respect to these methods and is at par with our MRF based joint OD and OC segmentation method [138] discussed in Chapter 2.

Since, the depth information is important in defining the OC boundary, [36] detects the bend in the blood vessels as they enter the OC, [74] employs stereo image pairs, while our method [138] discussed in Chapter 2 employs a coupled sparse dictionary based regression to obtain the depth estimates. In comparison, RACE-net outperforms these methods while directly using the raw RGB color channels as input without the need for explicit depth computation. The RACE-net also outperforms the existing CNN architectures proposed in [52], the U-net [102] and its modified version in [139] which was used to segment the OC .

The performance of the RACE-net architecture on the task of OD segmentation has also been evaluated on the MESSIDOR [5] dataset. It is a public dataset which comprises 1200 macula-centric images of subjects with Diabetic Retinopathy. The color fundus images were acquired from three different clinics in France at a 45^o Field of View (FOV) with different image resolutions of 1440×960, 2304×1536 and 2240×1488 respectively. The GT markings for the OD are available from the University of Huelva [70] which has been used to benchmark several OD segmentation algorithms [71],[29],[72]. The five-fold cross-validation performance of RACE-net has been reported in table 4.2. The Jaccard similarity coefficient J was used to benchmark the performance following the norms of the previously published work for a direct

Method	Optic Disc		Optic Cup	
	Dice	**BLE (pixels)**	**Dice**	**BLE**
Vessel Bend [36]	0.96/0.02	8.93/2.96	0.77/0.20	30.51/24.80
Graph cut prior [75]	0.94/0.06	14.74/15.66	0.77/0.16	26.70/16.67
Multiview [74]	0.96/0.02	8.93/2.96	0.79/0.18	25.28/18.00
Superpixel [42]	0.95/0.02	9.38/5.75	0.80/0.14	22.04/12.57
Joint OD-OC (CRF) [138]	**0.97/0.02**	6.61/3.55	0.83/0.15	18.61/13.02
CNN, Zilly et. al. [52]	94.7	9.4	0.83	16.5
CNN, U-net [102]	0.96/0.02	7.23/4.51	0.85/0.10	19.53/13.98
CNN, Sevastopolsky [139]	–	–	0.85	–
Proposed	**0.97/0.02**	**6.06/3.84**	**0.87/0.09**	**16.13/7.63**

Table 4.1: Segmentation performance on DRISHTI-GS1 test set. (mean/standard deviation)

comparison with their reported results. J is a monotonically increasing function of the Dice coefficient and also measures the extent of overlap between the set of pixels in the segmented region X and ground truth Y as

$$J(X,Y) = \frac{\mid X \cap Y \mid}{\mid X \cup Y \mid}. \tag{4.8}$$

The proposed method outperforms several existing methods with a J score of 0.92 including our CRF based method described in Chapter 2.

4.3.2 Cell Nuclei segmentation

The segmentation of the cell nuclei in hematoxylin and eosin ($H\&E$) stained histopathological images plays an important role in the detection of breast cancer. The cell nuclei appear as roughly elliptical purple blobs surrounded by a pink cytoplasm with the connective tissues appearing as wispy pink filaments (see Fig. 4.5). The challenges here include a large background clutter, intensity inhomogeneities, artifacts introduced during the slide preparation, staining or imaging and the variations in the morphology across a large number of cells in the image. The task unlike the previous case study involves the segmentation of an unknown number of regions in each image.

The proposed method was evaluated on 58 images consisting of 26 malignant and 32 benign cases provided in the UCSB Bio-segmentation Benchmark dataset [140]. Each image comprises

	Jaccard
Morales et al. [73]	0.82/0.14
Yu et al. [72]	0.84
Roychowdhury et. al. [31]	0.84
Aquino et al. [29]	0.86
Marin et. al. [32]	0.87
Giachetti et al.[71]	0.88
CRF (Chapter 2) [138]	0.87/0.22
RACE-net	0.92/0.02

Table 4.2: Performance of OD segmentation on the MESSIDOR dataset. (mean/standard deviation)

a 200×200 ROI along with a pixel-level binary mask provided as the GT. For a direct comparison with the recent work in [141], similar experimental setup and evaluation metrics were employed. The dataset was randomly divided into 24 images for training, 6 for validation and 28 for testing respectively. The network hypermeters were set to $N = 5$ and $T = 6$. The filter sizes were fixed to 9×9, 7×7, 5×5 and 3×3 respectively from the coarse to the fine scale. The training dataset was augmented by horizontal and vertical flipping, and rotating the images by $\pm 45°$. During testing, segmentation with the RACE-net took an average of 0.04 seconds using a GPU to process each image.

Method	**Precision(%)**	**Recall(%)**	**Accuracy(%)**	**F1-measure**
FCM [142]	71.63	84.86	88.84	0.7768
Watershed based [143]	70.26	87.01	88.56	0.7775
MI [144]	-	-	89.55	0.7733
DRLSE [135]	88.24	74.87		0.8042
CNN+SR [141]	82.41	86.04	92.45	0.8393
CNN [145]	-	-	86.88	-
CNN, U-net [102]	81.41	85.84	92.21	0.8333
Proposed	**85.29**	**88.38**	**93.82**	**0.8661**

Table 4.3: Cell nuclei Segmentation Performance on the UCSB Bio-segmentation Benchmark dataset

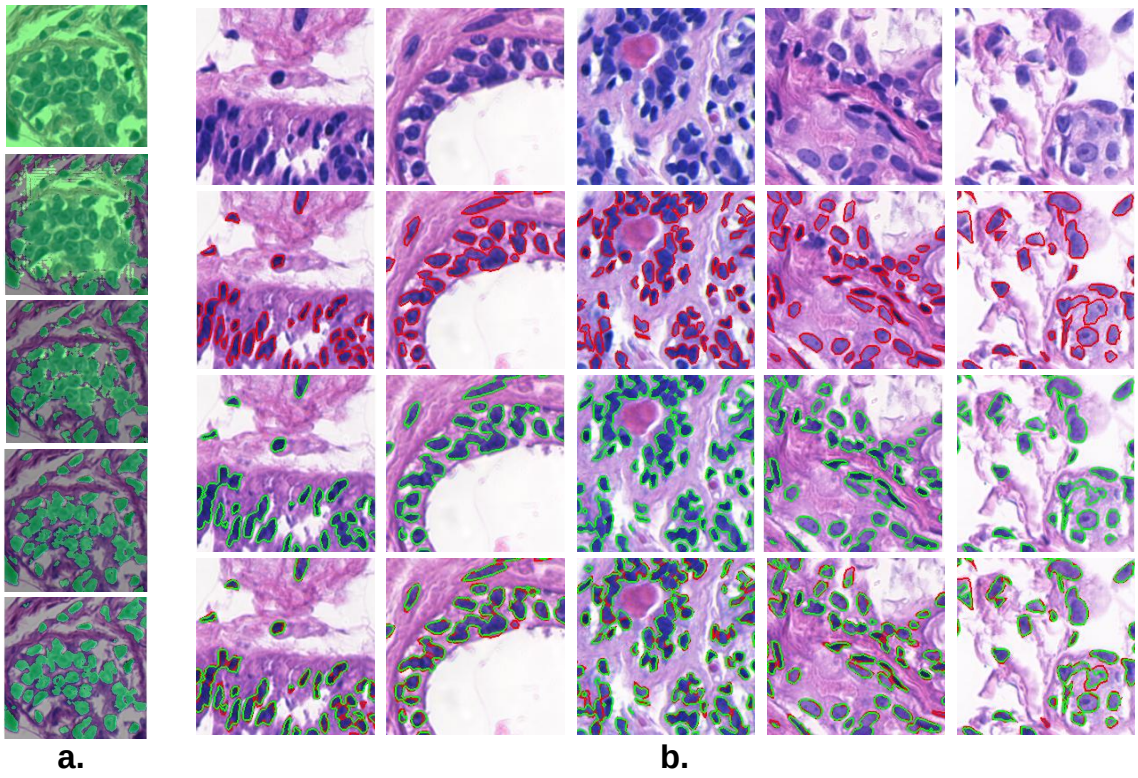

Figure 4.5: a. The curve evolution over time steps T=0,2,3,4,6 (from top to bottom). **Fig. b.** depicts the qualitative results of the RACE-net: 1^{st} **row:** input image, 2^{nd} **row:** Ground truth markings (in red), 3^{rd} **row:** result of RACE-net (in green), 4^{th} **row:** the result of RACE-net (in green) and the Ground truth markings (in red) are overlapped for comparison.

Raw RGB image channels were used as the input without any preprocessing and the border of the entire ROI was provided as the intitial boundary. Qualitative results depicted in Fig. 4.5 b demonstrate the strength of the proposed architecture. RACE-net effectively handles the topological changes associated with the splitting of the boundary curve to capture multiple cell nuclei. It is also capable of converging onto the desired boundary over large distances thereby eliminating the need for any preprocessing step for an accurate initialization. Fig. 4.5 a depicts the evolution of the level set curve across the time-steps. In the early time steps (T = 1-3), the level set curve progresses inwards, roughly segmenting the cell nuclei from the outer to the inner regions, while the final time-steps refine the object boundaries.

Results of quantitative assessment are reported in Table 4.2. The results show that the RACE-net outperforms the traditional image processing based methods employing Fuzzy C-means clustering [142], semi-supervised multi-image model (MI) [144] and the marker controlled

watershed [143]. It also achieves a 6% improvement in F1-measure against a deformable model [135] with distance regularized level set evolution (DRLSE) that employs handcrafted velocity terms. Unlike the proposed method, DRLSE failed to converge onto the cell boundaries over a large distance and Otsu thresholding was employed to initialize the level set. A comparison was also done against the existing CNN-based methods which have generally outperformed the traditional methods. The proposed method was found to perform better in this case as well. Moreover, while RACE-net employs the raw RGB images as input, the method in [141] employs a computationally expensive sparse coding based preprocessing step to roughly remove the background and accentuate the nuclei regions.

4.3.3 Left Atrium Segmentation

The accurate segmentation of the Left Atrium (LA) and the proximal pulmonary veins (PPVs) in cardiac MRI volumes plays an important role in the ablation therapy planning for the treatment of Atrial fibrillation, automated quantification of LA fibrosis and the construction of biophysical cardiac models. However, this task is non-trivial because of the large inter-subject morphological variations in the anatomy, overlapping image intensities between LA and the surrounding tissue, very thin myocardial walls and difficulty in demarcating LA from the left ventricle due to the different opening positions and low visibility of the mitral valve leaflets [146].

The proposed method was evaluated on the public STACOM 2013 left atrium segmentation challenge dataset [146], [147]. It consists of 10 training and 20 test cardiac MRI volumes. The binary segmentation masks for the LA and PPVs are provided for the training images, while an evaluation code written in ITK is provided for the test set. The MRI volumes were acquired using a 1.5T Philips Achieva scanner with an in-plane resolution of $1.25 \times 1.25\ mm^2$ and a slice thickness of $2.7\ mm$. The PPVs were defined as the segment of the pulmonary vein up to the first vessel branching or a maximum of 10 mm. We refer to [146] for further details of the protocols used in the acquisition of the images and GT.

Since all MRI volumes were acquired using the same view and resolution, the spatial priors on the anatomical structures were exploited to reduce the computational and memory requirements. An ROI was defined as a cuboid of dimensions $72 \times 104 \times 116$ centered at the image coordinates $(36, 52, 56)$ in each volume that contained the left atrium. Since the volumes were

not registered, a relatively large ROI had to be defined to ensure that the left atrium lies within it across all the images. On an average, the left atrium constituted only 8.05% of the ROI volume and appeared at varying spatial locations within the ROI. The level set surface ϕ_0 was initialized in a simple manner at the border of the ROI.

The proposed RACE-net architecture was extended to 3D by employing 3D convolutional filters within the CNNs that modeled the constant and mean curvature evolution velocities. The customized layer depicted in Fig. 4.1 b was also modified (written in Theano library [132]) to compute the gradient magnitude and the curvature by considering the x, y as well as the z directions. The network hyperparameters were fixed to $N = 8$ and $T = 7$. The convolutional filter sizes were fixed to $7 \times 7 \times 7$, $5 \times 5 \times 5$, $3 \times 3 \times 3$ and $3 \times 3 \times 3$ respectively from the coarse to fine scale.

The training dataset was augmented by applying random translations along the three directions and modifying the image intensity at a patch level by adapting the PCA based technique employed in [98]. In this method, the training volumes were divided into $3 \times 3 \times 3$ patches and projected onto a 27 dimensional PCA basis. Thereafter, a random Gaussian noise with zero mean and a standard deviation of 0.1 was added to each basis coefficient and the image patches were reconstructed.

The qualitative results of the proposed method is depicted in Fig. 4.6. A 3D rendering of the segmentation results in the intermediate time steps is presented in Fig. 4.6 a to visualize the surface evolution. In the first time step $T = 1$ itself, the initial surface can be seen to jump over a large distance from the border of the ROI to provide a moderately good localization of the left atrium. Thereafter, the shape of the structure is iteratively refined with large changes in the early iterations (T=2,3) followed by fine refinements in the later time steps. In Fig. 4.6 b, sample results are presented for 3 MRI test volumes where the 3D rendering of the proposed method is compared against the GT. These results demonstrate the ability of the method to handle sharp corners and protrusions in the object boundary due to the PPVs. The qualitative results on individual MRI slices depicted in Fig. 4.6 c indicate the ability of RACE-net to learn topological changes (splitting of the surface) near the PPV openings.

The quantitative assessment of results are presented in Table 4.3. The evaluation code provided by the STACOM 2013 left atrium segmentation challenge [147] was employed to compute the Dice coefficient and the surface-to-surface distance (S2S) in mm [146]. A good

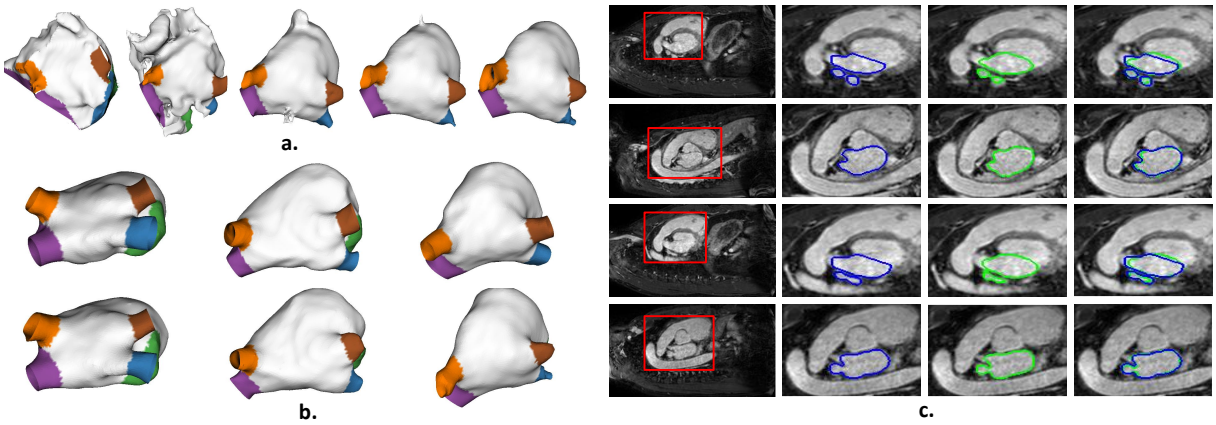

Figure 4.6: **a.** The curve evolution over time steps T=1,2,3,5 and 7 (from left to right). **b.** The 3D surface rendering of the segmented left atrium: 1^{st} row is the Ground truth and second row depicts the corresponding segmentation by the RACE-net. **c.** Qualitative results for the individual slices depicted in the 1^{st} column are provided. Only a small region marked by the red bounding box has been magnified for better visibility of the result. 2^{nd} column depicts the Ground Truth markings in blue, 3^{rd} column depicts the result of RACE-net depicted in green. The Ground truth and the result of the RACE-net are overlapped for comparison in the 4^{th} column.

segmentation is characterized by a high value for the Dice coefficient and low value for the S2S metric. The S2S metric is more sensitive to the local variations in the boundary as compared to dice which measures the global regional overlap. The proposed method was benchmarked against the performance of the other competing methods in the STACOM 2013 left atrium segmentation challenge [146] as well as the 3D U-net architecture [124].

The overall performance of the RACE-net is comparable to that of the 3D U-net. The RACE-net performed slighty better on the left atrium body both in terms of Dice and the S2S metrics. In case of the PPV, 3D U-net performed slightly better in terms of Dice while RACE-net performed slightly better in terms of the S2S. However, RACE-net provides significant gains in terms of the computational and memory requirements. We refer to Section 4.4 for the details.

In comparison to the participating methods in the STACOM 2013 left atrium segmentation challenge, our method achieved a dice of 0.91 for the left atrium body performing comparably to the second best performing method $LTSI\_VRG$, with UCL_1C achieving the best performance with a dice of 0.94. The PPVs offer a greater challenge for segmentation. Again the proposed method is the second best with a dice of 0.61 performing comparably to UCL_1C but below

Method	Left Atrium		PPVs	
	Dice	S2S (mm)	Dice	S2S (mm)
LUB_SSM	0.77/0.13	3.63/1.83	0.28/0.21	4.73/4.65
INRIA	0.78/0.26	3.66/4.59	0.42/0.28	4.66/7.74
LUB_SRG	0.84/0.12	2.67/1.75	0.51/0.31	2.80/2.63
TLEMCEN	0.85/0.07	2.40/0.96	0.37/0.30	3.68/3.20
LTSI_VSRG	0.87/0.03	2.22/0.32	0.48/0.19	2.54/0.54
LTSI_VRG	0.91/0.05	1.68/0.90	0.65/0.17	1.95/1.04
UCL_1C	0.94/0.02	1.09/0.31	0.61/0.23	1.62/0.61
3D U-net [124]	0.90/0.04	1.84/0.75	0.63/0.19	1.97/0.99
RACE-net	0.91/0.04	1.73/0.72	0.61/0.25	1.90/1.01
RACE-net ensemble	0.92/0.03	1.49/0.56	0.67/0.24	1.66/0.79
Human Expert 2	0.95/0.05	0.88/0.78	0.83/0.11	1.02/0.64

Table 4.4: 3D Left Atrium Segmentation Performance on the STACOM 2013 Challenge dataset (mean/standard deviation)

the performance of the $LTSI\_VRG$ method with a dice of 0.65. The biggest strength of the RACE-net over these methods is its computation speed; 1.72 seconds to process each MRI volume on an average using a GPU which is orders of magnitude lower than that reported by UCL_1C and $LTSI\_VRG$, namely 1200 and 3100 seconds respectively.

Incidentally, both UCL_1C and $LTSI\_VRG$ employed a multi-atlas approach. In these methods, multiple segmentation results were obtained by registering each atlas template to the test image followed by a majority voting scheme to obtain the final segmentation. Inspired from these methods, we evaluated an additional strategy which we refer to as the *RACE-net ensemble*. In this scheme, the data augmentation techniques used during the training were also applied to each test volume to obtain three additional images. The segmentation of all the four images were obtained using the RACE-net. Thereafter, the segmentation masks of the augmented images were brought back to the original co-ordinate space by reversing the translations employed during their construction. The average of the four segmentation masks was computed and thresholded at 0.5 to obtain the final binary segmentation map. Though the *RACE-net ensemble* increases the computation time, it significantly improves the robustness

of the method. While there is a slight improvement of 1% for the LA body in terms of dice, the performance of the PPV improves from 0.61 to 0.67 which outperforms all the existing methods.

4.4 Discussion

4.4.1 Performance

The results presented in the previous section show that the performance of the RACE-net architecture is comparable to the state of the art on a wide range of segmentation tasks involving different anatomical structures and imaging modality, outperforming the existing methods in most cases. RACE-net closely approximates the level set based active contour models within a deep learning framework. In the OD and OC segmentation task, it learned to extract the OC boundary though it is not characterized by a sharp intensity gradient and preserved the boundary smoothness constraints for both structures. On the task of cell nuclei segmentation, RACE-net demonstrated its ability to allow topological changes by allowing the boundary to split in order to capture multiple cell nuclei. RACE-net was also extended to 3D for left atrium segmentation, where the method showed its ability to converge to the desired boundary over large distances in only a few time-steps. A recent work [148] has explored an alternative way to incorporate CNNs within the active contour model framework to evolve a curve over small distances (30 pixels). It uses a parametric representation for the boundary and hence cannot handle topological changes. Furthermore, the velocity vectors at each boundary point are predicted independently at a local image patch level, requiring a handcrafted regularization step after each iteration to preserve the boundary smoothness.

The performance of RACE-net is summarized in Table 4.5. The BLE and the S2S were used as the Boundary error metrics for the OD, OC and the left atrium segmentation tasks respectively. A paired T-test was done to compare the mean performance of the proposed method against that of the U-net. On the task of the OD, cell nuclei and the left atrium segmentation, the p-value is < 0.05 for both the dice and the Boundary error metric indicating a statistically significant improvement in performance. However, in case of the OC and PPV, though our method's performance is marginally better than the U-net, the difference is not statistically significant.

	min	max	mean	median	standard deviation	p value (vs. U-net)
Dice						
Optic Disc	0.88	0.99	0.97	0.98	0.02	0.0042
Optic Cup	0.51	0.95	0.87	0.89	0.09	0.1016
Cell nuclei	0.84	0.91	0.87	0.87	0.02	<0.0001
Left Atrium	0.86	0.96	0.92	0.92	0.03	0.0239
PPV	0.20	0.88	0.67	0.70	0.16	0.4033
Boundary Error (pixels)						
Optic Disc	2.48	24.60	6.06	4.84	3.84	0.0095
Optic Cup	6.53	37.32	16.13	14.13	7.63	0.1537
Cell nuclei	–	–	–	–	–	–
Left Atrium	0.70	2.66	1.49	1.47	0.56	0.0236
PPV	0.87	3.58	1.66	1.41	0.79	0.1373

Table 4.5: Statistics for the Dice and Boundary Error of the proposed RACE-net on various tasks.

4.4.2 Failure cases

The qualitative examples of a few failure cases of the RACE-net are depicted in Fig. 4.7. Additionally, the results obtained using the U-net has been provided for comparison. Although RACE-net performs very well for OD segmentation in general, the failure case shown in Fig. 4.7 a is an exception which has the minimum dice (of 0.87) across all the test images in the DRISHTI-GS1 dataset. In this case, the RACE-net over-estimates the OD boundary and is attracted towards the thick blood vessels in the Inferior-nasal (bottom-right) quadrant. On the other hand, the U-net underestimates it and is attracted towards the pallor boundary in the inferior (bottom) sector. The pallor is a pale yellow region within the OD with maximum color contrast that lies close to the OC. The possible reason for failure in this case could be due to the poor contrast and the lack of a strong gradient at the OD boundary in the inferior sector.

OC segmentation from fundus images is a challenging task and is primarily guided by the intensity gradient at the pallor boundary and the bends in the thin blood vessels within the OD. However, the image depicted in Fig 4.7 b doesnot have a distinct pallor. Though the result of the RACE-net closely follows the GT in the inferior sector, it may have been attracted to the bend in the thin vessel in the superior (top-right) sector.

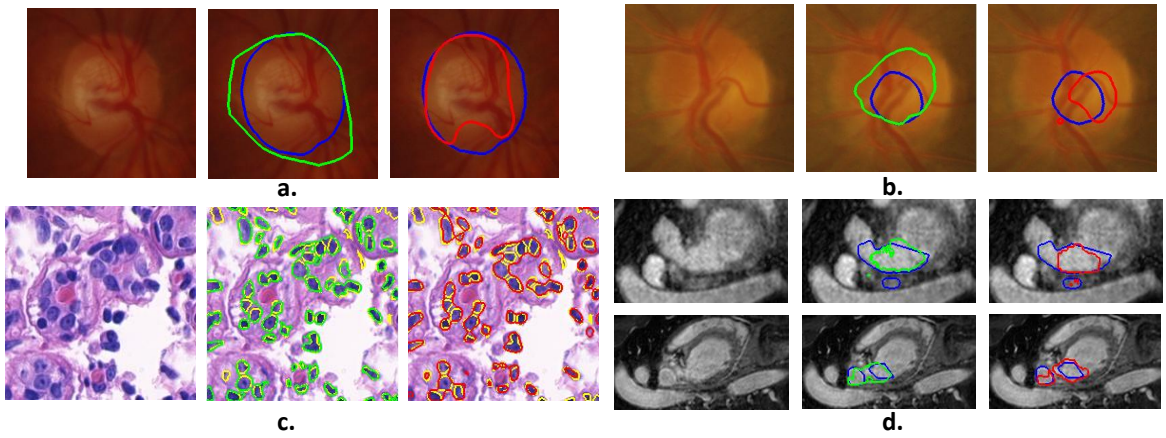

Figure 4.7: Examples of failure cases for: **a.** Optic disc, **b.** Optic cup, **c.** cell nuclei and **d.** left atrium. The Ground truth is marked in blue in **a.**, **b**, **d** and yellow in **c**. The results of RACE-net are marked in green in each case. The results obtained using U-net (marked in red in each case) are provided for comparison.

A failure case for the cell nuclei segmentation in histopathological images is depicted in Fig. 4.7 c. In this case, the performance of RACE-net is slightly better than that of the U-net (dice of 0.84 in comparison to 0.81). However, both the methods exhibit a tendency to miss the boundaries between the adjacent or overlapping cell nuclei.

In case of the left atrium segmentation, the errors in segmentation typically occur in slices near the PPV openings which are characterized by sharp corners and topological changes in the boundaries. Two such examples of failure cases are provided in Fig. 4.7 d. The example in the first row depicts a case where RACE-net under-estimates the boundary of the larger connected component and fails to detect the smaller one. In this case, there is a lack of contrast and distinct intensity gradient between the foreground and background regions. Another example (second row in Fig. 4.7 d) depicts a case of overestimation where instead of two connected components, RACE-net smoothes out the boundary resulting in a single larger region encompassing both.

4.4.3 Computational time and network size

RACE-net is ideal for deployment on systems with limited resources such as low memory and the unavailability of a GPU. Typically, the existing CNNs employ millions of parameters. For eg., the U-net architecture described in [102] employs $34,513,345$ parameters. However, the memory bandwith and storage requirements of a deep learning system is directly proportional

	Optic Disc	Optic Cup	Cell Nuclei	Left atrium
U-net	5.88	7.79	1.92	24.13
RACE-net	0.74	0.60	0.15	7.74

Table 4.6: Average time (in seconds) to segment each image using CPU.

to the number of network parameters. Therefore, the network parameters in RACE-net were kept to a minimum by applying two strategies. First, $g(I, \phi)$ and $h(I)$ functions in Fig. 4.1 were approximated by a novel CNN architecture inspired from the multi-scale image pyramid that employed very few parameters. Secondly, each time-step of the RACE-net shares the same network parameters thereby allowing for arbitarily complex architectures by adjusting T while keeping the number of shared network parameters constant. As a result, the RACE-net architecture for the cell nuclei segmentation (described in Section 4.3.2) employs only $8,394$ parameters. Similarly, for the OC segmentation task, the modified U-net architecure in [139] employed about 25 times the number of network parameters in the RACE-net ([139] employed $660,000$ compared to the $26,322$ network parameters in the RACE-net). On the 3D left atrium segmentation task, the RACE-net architecture detailed in Section 4.3.3 employed $71,862$ learnable netwrork parameters as compared to the 3D U-net [124] with $10,003,401$ parameters.

RACE-net runs moderately fast even in the absence of a GPU. To demonstrate this, the average segmentation time per image of the RACE-net is compared against the U-net [102] in Table 4.6. Both the networks were tested on all the three segmentation tasks without using a GPU. The 3D extension of the U-net in [124] was used for the left atrium segmentation. The results clearly illustrate the advantage of our method with a 3 to 12 times speedup across the different tasks. The low run time can be attributed to the fact that each time-step of the RACE-net is modeled using a very small FFNN that employs few computations. Moreover, the RACE-net only requires a few time-steps (5-7) to converge and uses a simple feedback connection instead of the computationally expensive gated architecture used in LSTMs [128] which further improves the computational efficiency. During training, the vanishing gradients problem is handled by using auxiliary loss functions at each time-step.

4.4.4 Boundary initialization and network hyperparameters

Similar to the traditional deformable models, the RACE-net requires an initial boundary localization ϕ_0 to be provided as an input in the first time step. In this work, very coarse boundary initializations were obtained using simple image processing techniques and leveraging prior spatial constraints on the expected location of the anatomical structure in the image. For example, in case of cell nuclei segmentation, the UCSB dataset contained 200×200 image patches and ϕ_0 was simply initialized at the edge of the input patch. In case of the 3D left atrium segmentation, spatial constraints were used to define a rough initial boundary. However since the MRI volumes were unregisteted, the left atrium constituted only around 8.05% of the region within the initial boundary on an average and appeared at varying spatial locations inside it. The proposed method demonstrated the ability to evolve the initial boundary over large distances to converge onto the desired boundary.

Moreover, RACE-net is also insensitive to the exact location of the structure with respect to the initial boundary. This is illustrated in Fig. 4.8 on the task of OD segmentation. Each row of Fig. 4.8 depicts a case where the OD is present in the center, bottom-left and bottom-right regions with respect to the initial bounary (at T=0) respectively. The RACE-net is able to converge onto the desired boundary in all the three cases.

The RACE-net architecture is defined using two hyperparameters, N and T that have to be empirically fine-tuned for each segmentation task. The size of the CNN used to model each iteration of the curve evolution is controlled by a single hyperparameter N which specifies the number of convolutional filters employed at each scale of the CNN. The depth of the CNN was kept fixed to 4 scales accross all the segmentation tasks. The RACE-net employs T recurrent feedback connections to model each time-step of the curve evolution. The value of T is kept the same for all images, both in the training and test set. Smaller values for T reduces training time and improves numerical stability of the back-propagation algorithm but may inhibit the networks ability to evolve the curve accross long distances. In our experiments, we found the optimal values for both T and N to vary in the range of 5-7 on a diverse set of segmentation tasks.

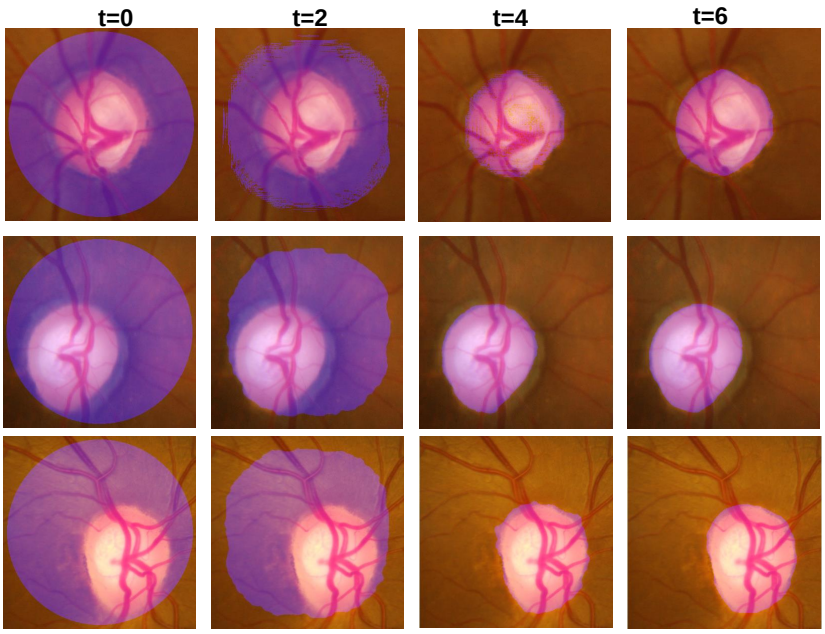

Figure 4.8: Each row depicts the evolution of the Curve for the segmentation of OD over time steps T=0,2,4,6 (from left to right). The proposed method is not sensitive to the exact location of the OD with respect to the initial boundary.

4.4.5 Effect of the regularization term in the Loss Function

SDF has the desirable property that $|\nabla\phi(x,y)| = 1$ at all spatial co-ordinates x, y. Inspired by the LDMs, we added an additional regularization term to minimize $\beta.||\nabla\phi_t| - 1|_2^2$ at each time-step of the loss function in eq. 4.7.

However, giving a large value to the weight β of the regularization term inhibits the ability of the RACE-net to evolve the boundary curve over large distances. As a result, the β was kept small in all our experiments and the evolving ϕ_t doesnot remain a SDF across the T time-steps. However, the regularization term was empirically found to stabilize the training loss and prevent overfitting. For example, on the task of OD segmentation, the dice corefficient improved from 0.95 to 0.97 on the test set when the regularization loss term was added. It's effect was even more drastic on the task of cell nuclei segmentation where the evolving boundary had to split to capture multiple cell nuclei in the histopathology images. The F1-score on the test set improved from 0.75 to 0.87 when the regularization term was added to the loss during training.

4.5 Conclusion

In this Chapter, a novel RNN architecture (RACE-net) was proposed for biomedical image segmentation which models the object boundaries as an evolving level set curve. Each time-step of the curve evolution is modeled using a FFNN that employs a customized layer to compute the normal and curvature of the level set function and a novel CNN architecture is used to model the curve evolution velocities. The size of the CNN is determined by a single hyper-parameter N which controls the number of convolutional filters at each scale. The recurrent feedback connections are used to model the T time steps of the curve evolution.

Since RACE-net can be viewed as a cascade of feedforward networks with shared network weights, an arbitarily deep network can be obtained to model complex long range, high-level dependencies by adjusting T while keeping the number of shared network parameters constant. Overfitting is avoided with fewer network parameters which is also critical for the deployment of RACE-net on devices with limited memory and computational resources. The entire RACE-net is trained in a supervised, end-to-end manner. An appropriate loss function was defined as a weighted sum of intermediate dice coefficients at each time-step to mitigate the vanishing gradients problem and a regularization term on the level set function was also incorporated to ensure its numerical stability.

Consistent performance of RACE-net on a diverse set of applications indicates its utility as a generic, off-the-shelf architecture for biomedical segmentation. Since RACE-net is based on the deformable models, it has the potential to incorporate explicit shape priors which presents a promising direction for work in the future.

Chapter 5

Applications in Retinal Disease Detection

In this Chapter, we explore different classification frameworks for the image-level detection of retinal diseases. Specifically, we focus on the tasks of the detection of glaucoma in Color Fundus (CF) images and Age-related Macular Degeneration (AMD) in 3D Optical Coherence Tomography (OCT) volumes.

In Section 5.1, we explore and compare two different strategies for the detection of glaucoma which are based on handcrafted features (Section 5.1.1) and deep learning (Section 5.1.2) respectively. Both the methods attempt to capture the structural changes through features derived from the Optic Disc (OD)-Optic Cup (OC) segmentation as well as the appearance based features directly derived from the CF image in the region around the OD.

In Section 5.2, we attempt to represent the normal retinal anatomy in an OCT using an atlas constructed by co-registering a set of OCT volumes of healthy subjects in Section 5.2.1.1. We attempt to characterize the irregularities and undulations caused by the deposition of drusen in the Retinal Pigment Epithelium Layer (RPE) layer in OCT as significant deviations from the Normative Atlas in Section 5.2.1.4.

5.1 Glaucoma detection in Fundus Images

Glaucoma is a chronic optic neuropathy which is the second leading cause of blindness in the world. It is asymptomatic in early stages and lead to a gradual but irreversible vision loss. Large scale population screening programs can help prevent its progression through early detection and timely treatment. An automated screening tool can aid in reducing the time and effort of ophthalmologists and allow them to serve more patients.

Glaucoma is characterized by the progressive degeneration of the retinal nerve fibers which leads to structural changes in the OD. The OD appears as a bright elliptical region (Fig. 5.1 b.) in CF images. It consists of a central depression called the OC which is surrounded by an annular region comprising of the retinal nerve fibers called the neuro-retinal rim. In glaucoma, the loss of retinal nerve fibers leads to the thinning of the neuro-retinal rim and consequently the enlargement of the OC. Clinical parameters such as the vertical Cup to Disk diameter ratio (CDR) and the ISNT rule [149] have been defined to quantitatively assess the OD region. According to the ISNT rule, the healthy OD is characterized by the maximum neuro-retinal rim thickness in the Inferior(I) sector followed by the Superior(S) sector followed by the Nasal(N) and the Temporal(T) sectors respectively (ie., the rim thickness distribution follows the order $I \geq S \geq N \geq T$ in healthy OD). However, this distribution is disturbed in the presence of glaucoma.

Existing work on the automatic assessment of glaucoma from fundus images either employ image-based features directly derived from the OD region in the CF images or features derived from the segmentation of OD and OC for glaucoma classification. The latter aids in capturing the structural deformation and CDR estimation is commonly employed. Most of these methods therefore report mean CDR estimation error, while a few methods [42] also report on the glaucoma classification performance. In [150], [151], an attempt was made to directly estimate the CDR without explicit segmentation of the OD and OC by correlating the vertical CDR values directly to the appearance of the CF image in the OD region using sparse representation. Though CDR is an important indicator for glaucoma, it is inadequate for accurate classification as it captures the retinal nerve fiber degeneration only along the vertical direction. Very few methods [152], [153] have considered additional features such as the disc size, rim thickness and ISNT rule compliance for glaucoma classification. However, small errors in segmentation accuracy can lead to significant changes in the clinical measurements. While additional cues such as genetic information [154] have significantly improved the classification performance, obtaining them is infeasible in a screening scenario.

Attempts to extract discriminative features at the pixel-level from the OD region without requiring OD-OC segmentation have also been reported. Some of the features that have been explored include: higher order spectra features [155]; energy of channels obtained from first level of wavelet decompositions [156]; and PCA coefficients of intensity, Fourier, and Gabor features

[157]. A major challenge in these methods is the need to handcraft appropriate features. Preliminary results of these methods show high performance, albeit on small datasets of 60 to 120 images [156][155],[157]. Being data-driven in nature, feature selection [156] and classification may tend to overfit due to the limited availability of annotated images.

Recently, the Deep Learning based methods have led to a significant improvement in the state of the art. In [158], a novel M-net architecture was explored to jointly segment the OD and OC in the log-polar domain and the CDR was extracted for glaucoma detection. The DENet architecture in [159] employed an ensemble of four independent neural networks whose predictions were fused to obtain the final decision. Two out of the four deep networks operate on the entire fundus image to extract image and segmentation based features respectively. The remaining two networks extract the convolutional features from a smaller region of interest around the OD which is provided as input to the two networks in the regular cartesian co-ordinates and the polar coordinates respectively.

In this work, we explore two different strategies for the image level detection of glaucoma which are based on handcrafted features and deep learning (DL) respectively. Both the methods combine complementary structural features derived from the OD-OC segmentation and the appearance features derived from the Region of Interest (ROI) in the fundus image.

The first method based on handcrafted features extracts a comprehensive set of features from the OD-OC segmentation to accurately capture the shape deformations characterizing glaucoma. To capture the appearance of the ROI in the CF images, we propose to use the Texture of Projection features [160] to capture the textural changes and Bag of Visual words to capture the color information in the different sectors of the OD. The details of this method have been discussed below in Section 5.1.1.

The second method explores a single Multi-task deep Convolutional Neural Network (CNN) architecture which jointly performs the three tasks of the segmentation of OD, OC and the image level prediction of glaucoma. Sharing the CNN features for multiple but related tasks improves the generalizability of the network by constraining it to learn meaningful features. Moreover, the proposed method achieves a performance comparable to the state of the art with a relatively small network architecture that employs far fewer network parameters in comparison to the existing methods such as DENet. The sharing of the CNN features for multiple tasks and the smaller network size ensures that the proposed method can be trained in

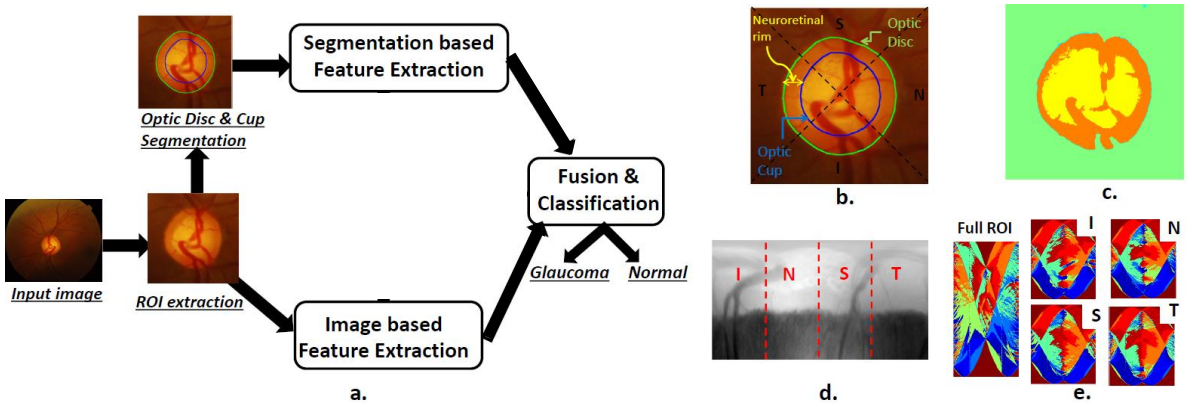

Figure 5.1: a. Block Diagram of the proposed method for glaucoma classification based on handcrafted features. **b.** OD and OC boundaries in a cropped fundus image; **c.** Color based clustering of b. used for BoW computation; **d.** polar representation of the red channel of b.; **e.** LBP of Radon transform of d. for ToP compuation.

the limited availability of data without over-fitting. This method also employs a combination of appearance as well as structural features obtained from the OD-OC segmentation within the CNN framework to improve the robustness of the glaucoma classification. The details of this method have been discussed below in Section 5.1.2.

5.1.1 Method based on Handcrafted Features

An overview of the proposed method is provided in Fig 5.1.a. First, the ROI around the OD is extracted using a simple image processing based method based on intesnity thresholding, vessel suppression using morphological operations and Hough Transform based circle detection. The details of this method have been previously disccused in Section 2.2.1 in Chapter 2. Finally, a square ROI of size $2.8R \times 2.8R$ (with a margin of 0.4 R on all sides) is extracted around the center of the circle detected using the Hough Transform, where R represents its radius. The ROI is resized to 400×400 pixels. Next, the OD and OC segmentations are obtained using the Conditional Random Field (CRF) based method proposed in Chapter 2. Thereafter, a 14-dimensional set of features is extracted from the OD and OC segmentations to capture the structural deformations in the OC. The details of these handcrafted segmentation based features is provided below in Section 5.1.1.1. Another complementary set of 90-D features is directly extracted from the ROI to capture the texture (65-D) and color information (25-D) in various

sectors of the OD. The details of the image based features is discussed below in Section 5.1.1.2. Finally, the two features were concatenated to obtain a 90+14=104 dimensional feature and a SVM classifier with RBF kernel (whose good performance on glaucoma classification has been established in [157],[156]) was employed for the binary classification of the fundus image into Normal and Glaucomatous categories. The hyperparameters of the RBF kernel SVM were set to $c = 10$ and $\gamma = 0.01$ by performing a grid-search over a validation set which was obtained by randomly partitioning the training set into 80% for training and 20% for validation.

5.1.1.1 Extraction of Segmentation based features

The following set of 14-dimensional(D) features were extracted from the OD-OC segmentations.

(a) Vertical OD diameter (1D).

(b) Vertical Cup-to-Disc Diameter Ratio (CDR) (1D).

(c) Rim to disc area ratio (1D).

(d) Ratio of horizontal to vertical CDR (1D).

(e) 6D features computed by $\frac{r_i - r_j}{\sigma_i + \sigma_j}$ with $i \neq j$ and $(i, j) \in \{(I, S, N, T)\}$.

(f) 4D features computed as $\sigma(r_k)$with $k \in \{I_N, I_T, S_N, S_T\}$.

I,S, N and T represents the Inferior, Superior, Nasal and Temporal sectors of the OD (see Fig. 5.1 b). The I and the S sectors are further sub-divided into Inferio-nasal (I_N), inferio-temporal (I_T) and superio-nasal (S_N), superio-temporal (S_T) quadrants respectively. The mean and standard deviation of the rim thickness between the OD and OC boundaries in the i^{th} sector are denoted by r_i and σ_i respectively.

Feature (a) was used as a measure of the OD size. Features (b), (c) were used to measure the Rim-to-Disc ratios to capture the extent of the enlargement of the OC with respect to the OD. The ISNT rule indicates that the presence of glaucoma leads to a change in the distribution of the rim thickness in the I, S, N and T sectors of the OD. This is captured using features (d),(e). In the features extracted in (d), larger rim thickness in the I and S sectors which characterize healthy images will lead to smaller value of the vertical CDR with respect to the

horizontal CDR. Features in (e) compute the pairwise difference in rim thickness distributions in the different sectors. Apart from an overall decrease in the rim thickness, glaucoma may also be characterized by a localized deformation in the rim, called notching which can clinically occur in the Inferior and Superior Sectors. The localized deformations are captured using the features extracted in (f). Each feature was normalized to zero mean and unit standard deviation.

5.1.1.2 Extraction of Image based features

The color and textural appearance of the ROI is captured using a combination of the Bag of Visual Words (BoW) [161] and the Texture of Projection (ToP)[160] features.

The ToP features are extracted from a give image patch by first representing it in the projection domain using the Radon Transform. Next, the Local Binary Pattern feature (LBP) [162] is computed to capture the textural information in the Radon sinograms which have been shown to correlate well with the structural changes in the image patch [160]. We found the ToP features extracted from the Red Channel of the ROI to be most discriminative for glaucoma classification.

The LBP [162] feature is computed as follows. For each pixel in the radon sinogram, its intensity is compared to each of its eight immediate neighbors in clockwise order and assigned a '0' if the center pixel's value is greater than the neighbor else '1' is assigned. This results in a 8 bit vector which is treated as a binary number and converted to its decimal equivalent in the range of 0-255. Next, a 255 bin histogram over the entire sinogram image is computed to obtain the global LBP feature.

Our BoW feature assigns each pixel in an image patch to one of the 5 clusters based on their color values and computes a 5-bin histogram (each bin corresponding to a cluster) to obtain a global feature. The details of the BoW feature extraction is discussed below.

Since both the ToP and the BoW features are histogram based, computing them over the entire ROI will lead to the loss of spatial information. Whereas, we wish to capture the changes in bilateral symmetry and rim thickness in the different sectors with these features. Hence, the ROI image is converted into the polar coordinates to obtain rectangular-regions corresponding to each sector (see Fig. 5.1d). Next, the ToP and the BoW features are computed for each of the I,S,N,T sectors as well as the entire ROI. Each of the 5 histogram features (1 for the entire

ROI + 4 for each of the four sectors) for both the BoW and the ToP features are individually normalized to sum to 1.

The final set of features extracted were :

(a) *Structural ToP features(65D)*: The red channel of the ROI was converted into polar co-ordinates (see Fig. 5.1d). ToP features followed by PCA based dimensionality reduction to 13D was computed on the entire ROI and each of the 4 sectors individually (Fig 5.1 e.) finally resulting in a $13 \times 5 = 65D$ feature.

(b) *Color BoW features (25D)*: Each pixel in the ROI was represented by a 2D feature vector of its color values in the R and L* channel of RGB and L*a*b* color models respectively. The 2D features were clustered into 5 visual words using k-means (Fig 5.1 c.) and their histogram was computed for the entire ROI and the 4 sectors independently.

5.1.2 Method based on Deep Learning

An outline of the proposed DL based method for glaucoma classification is depicted in Fig. 5.2. The ROI around the OD is extracted in an identical manner as discussed in Section 5.1.1. The ROI is provided as an input to the proposed Convolutional Neural Network (CNN) architecture which jointly segments the OD, OC as well as provides an image level prediction for glaucoma. Further details of the CNN architecture is discussed below in Section 5.1.3. The segmentation masks obtained from the CNN are further refined in a post-processing step (see Section 5.1.5) to improve the segmentation accuracy.

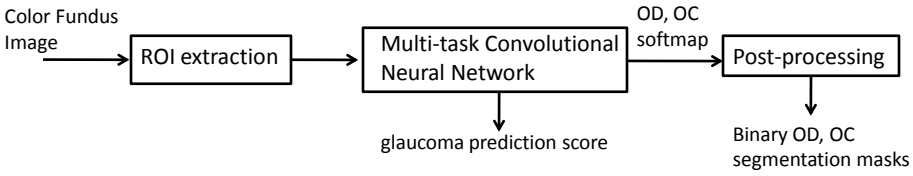

Figure 5.2: Outline of the proposed DL based method for glaucoma classification.

5.1.3 The CNN architecture

Fig. 5.3 depicts the architecture of the proposed Multi-task CNN network. The input ROI image is resized to 400×400 and fed into a Deconvolutional network similar to the U-net [102]. It consists of an encoder stage followed by a decoder. The encoder comprises a

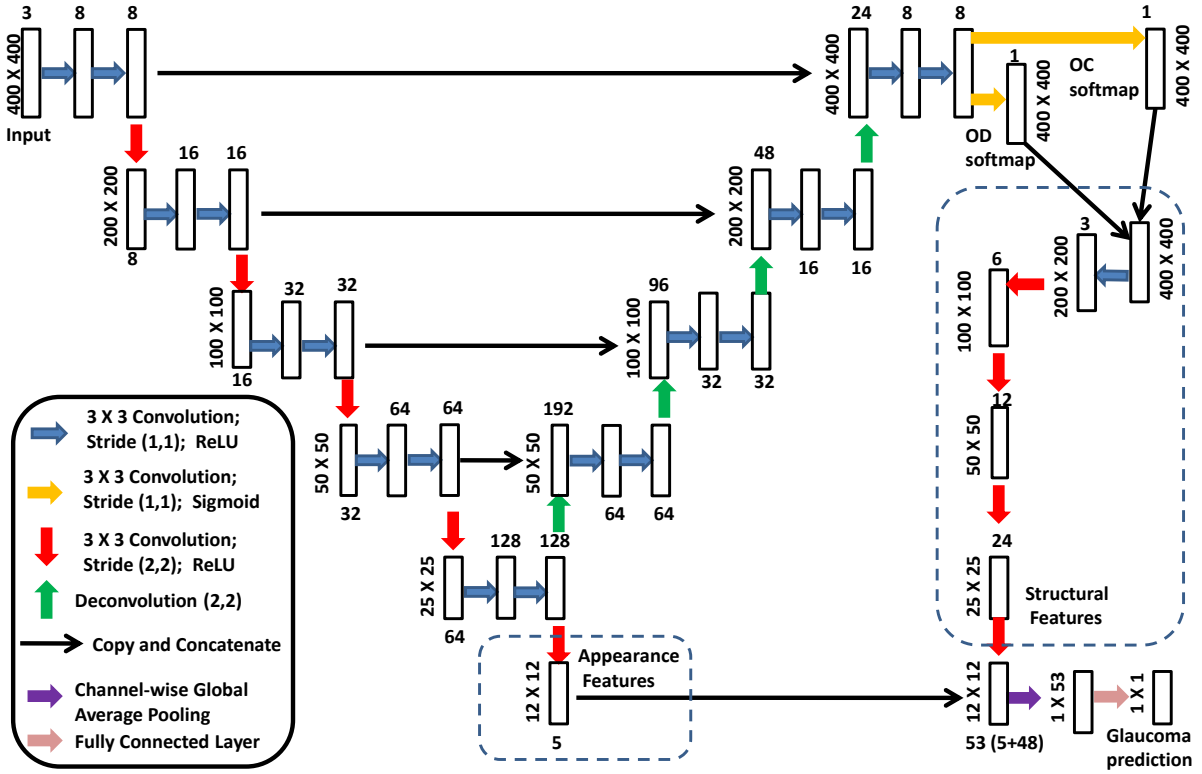

Figure 5.3: The proposed Multi-task Convolutional Neural Network architecture for joint segmentation of Optic disc, cup and image level classification of glaucoma using a combination of image appearance and structural features.

series of contraction blocks which extract feature channels at successive lower resolutions. Each contraction block consists of two 3×3 convolutional layers followed by a $(2, 2)$ Maxpooling with a stride of 2. While the spatial resolution of the extracted feature channels is halved at each scale, the number of feature channels is doubled to allow the architecture to learn complex structures effectively.

The decoder comprises a series of expansion blocks, one corresponding to each contraction block in the encoder. The input to each expansion block is obtained by concatenating the output feature channels of the corresponding contraction block in the encoder network with the output of the previous contraction block using skip connections. The expansion blocks consist of two 3×3 convolutional layers followed by a $(2, 2)$ upsampling layer. They successively upsample the feature channels to eventually obtain the features at the original image resolution. The

number of feature channels at each scale is halved to mantain the symmetry with the encoder architecture.

The final output of the decoder is fed into two separate 3×3 convolution layers with a sigmoid activation function to obtain the output segmentation masks for the OD and OC.

The classification part of the network uses a combination of image appearance and structural features (obtained from the OD-OC segmentation). The appearance features are obtained by reusing the $25 \times 25 \times 128$ output of the encoder part of the U-net architecture and applying a convolutional layer with five, 3×3 filters and a $(2, 2)$ stride to obtain a $12 \times 12 \times 5$ size feature channel.

In order to obtain the structural features, the output OD and OC segmentation masks are concatenated to obtain a 2 channel feature map and a series of 3×3 convolutional layers with a stride of (2,2) is successively applied to obtain the high-level structural features of size $12 \times 12 \times 48$. Finally, the appearance (5 channels) and the structural features (48 channels) are concatenated and a channel-wise Global average pooling is applied to obtain a 53-D feature vector. A fully connected layer with a single neuron is applied to the feature with a sigmoid function to obtain the probability of the presence of glaucoma.

5.1.4 Implementation details

A loss function has to be defined for each of the three outputs of the multi-task CNN network. The binary cross-entropy (BCE) was used as the loss function for the image level classification task. It is defined as

$$BCE(p, y) = -\left\{y.log(p) + (1 - y).log(1 - p)\right\}, \tag{5.1}$$

where y represents the ground truth (GT) target prediction and p represents the probability of glaucoma predicted by the proposed CNN.

The dice coefficient was used as a measure of the extent of overlap between the GT and the predicted segmentation masks. $L_{od} = 1 - Dice(R_{od}, Y_{od})$ and $L_{cp} = 1 - Dice(R_{cp}, Y_{cp})$ were used as the loss functions for the OD and OC segmentation tasks respectively, where R_{od} and R_{cp} denotes the segmentation masks computed by the network and Y_{od}, Y_{cp} represents the GT segmentation masks for the OD and OC respectively. The soft differentiable form of the Dice

metric (see Chapter 4, Section 4.2.3.1 for details) is defined as

$$Dice(R, Y) = \frac{2\sum_i r_i.y_i}{(\sum_i r_i + \sum_i y_i)}, \tag{5.2}$$

where r_i and y_i represents the value at the i^{th} pixel in R and Y respectively and the summations run over all the pixels in the image.

The proposed CNN architecture is trained using gradient backpropagation with the ADAM optimizer [136] to automatically adapt the learning rate. The batch size was fixed to 32 images and early stopping was used to terminate the training when the validation loss didnot improve for 60 consecutive epochs. The Data Augmentation plays an important role in the proper training of the CNN. Simple geometric transformations such as translation and rotation were incorporated to ensure that the proposed method is not sensitive to slight changes in the view and localization errors incurred during the ROI extraction step. Moreover, a PCA based technique similar to [98] was also employed to add some random noise to the input image. The class imbalance was handled by oversampling a larger number of glaucomatous images in comparison to the Normal images during training.

5.1.5 Post-processing

Since, the CNN poses the OD and OC segmentation as binary pixel-labeling problems, it doesnot explicitly constrain the segmented regions to be a single connected component or preserve the smoothness constraints on the OD and OC boundaries. In order, to address these issues, a post-processing step is employed to further refine the segmentation results obtained from the CNN. During post-processing, both the OD and OC segmentation softmaps are binarized by thresholding at 0.5. Thereafter, a morphological opening operation is employed to remove spurious small regions and smooth the boundaries. Finally, a connected component analysis is performed to remove all except the largest connected component in the binary segmentation mask. Though OC boundaries can have an arbitary shape, OD boundary can be closely approximated by an ellipse. Hence, an additional step is employed for the OD segmentation where an ellipse is fitted to it with the minimum least square error. Note, that the segmentation feature based pathway within the CNN which is used to predict the presence of glaucoma employs the original segmentation output of the CNN without the post-processing.

5.1.6 Experiments

Dataset: The performance of the proposed method has been evaluated using three public datasets, namely, DRISHTI-GS [6], RIM-ONE version 2 [68], and REFUGE [163], and two locally sourced private datasets: Private-Train and Private-Test. The REFUGE dataset consists of two separate sets of images for training and validation respectively which are acquired using different fundus cameras at different resolutions. The datasets were combined into two mutually exclusive sets and used for training and testing respectively. A summary of the training and test datasets is provided in Table 5.1. The datasets cover a range of ethnicity in population, imaging protocols, fundus cameras and image quality. The training dataset was randomly split and 80% of the data was used for training and the remaining 20% was used as the validation set. The proportion of the Normal to Glaucomatous cases was kept similar in the training and the validation set. The Private-Train and the Private-Test set dataset were collected from Aravind Eye Hospital, Madurai and Aravind Eye Hospital, Coimbatore respectively.

5.1.7 Results

Performance of Optic Disc and Cup Segmentation

The OD and OC segmentation performance of the proposed DL based architecture and the CRF based method detailed in Chapter 2 on the REFUGE-validation dataset is depicted in Table 5.2. Both the methods have a comparable performance with the CRF based method performing slightly better for OD and the DL based method performing marginally better on the task of OC segmentation respectively. Since OC is primarily defined by the depth information which is not explicitly available in the 2D CF images, the segmentation of OC is relatively more challenging in comparison to OD.

Glaucoma Classification Performance

The classification performance of the proposed methods based on handcrafted features and DL for each test dataset has been reported in Table 5.3. Additionally, the proposed method was re-trained and evaluated using segmentation and appearance based features alone to analyze their individual contribution towards the glaucoma classification performance. The corresponding ROC plots have been depicted in Fig. 5.4.

Dataset	# Normal	# Glaucoma	Total	Remarks	Ground Truth
Training					
DRISHTI-GS	31	70	101	Zeiss Visucam Fundus Camera, 2896×1944 pixels	Image-level decision; OD, OC Segmentation
REFUGE-train	360	40	400	Zeiss Visucam Fundus Camera, 2124×2056 pixels	Image-level decision; OD, OC Segmentation
Private-Train	500	500	1000	Zeiss Visucam Fundus Camera NM/FA, 2588×1958 pixels pixels	Image-level decision
Total	891	610	1501		
Testing					
Private-Test	145	169	314	Topcon TRC50EX Fundus Camera, 1900×1600	Image-level decision
RIM-ONE r2	255	200	455	NIDEK AFC-210 Fundus Camera, 2896×1944 pixels	Image-level decision
REFUGE-val	360	40	400	Canon CR-2 fundus camera, 1634×1634 pixels	Image-level decision; OD, OC Segmentation
Total	760	409	1169		

Table 5.1: Dataset Description

Considering the average AUC across all the datasets, the DL based method (AUC of 0.8675) outperforms the method based on handcrafted features (AUC of 0.8221). However, the method based on handcrafted features is able to achieve marginally better performance over the DL based method on the RIMONE dataset.

For the DL based method, the AUC values obtained after combining the appearance and the segmentation based features consistently performs better than using the individual features alone in all the three datasets. In case of the method based on handcrafted features, the AUC using the combined features is higher than the individual features in all except the REFUGE-val dataset where the segmentation features alone achieves a slightly higher AUC in comparison to the combined features.

	Optic Disc	Optic Cup
Proposed DL method	0.92	0.84
Proposed CRF method	0.94	0.83

Table 5.2: Performance of Optic Disc and Cup Segmentation. (mean Dice coefficient)

The overall performance using the combined features as well as the trends in the individual contributions of the segmentation and appearance features varies considerably across the three datasets as visualized in Fig. 5.4. For example, the segmentation based features (plotted in green) alone performs better than the appearance based features alone (plotted in blue) for both the methods on the Private Test Set. However, this trend is reversed on the RIM-ONE dataset for both the methods. The performance of the segmentation based features improves significantly for both the methods on the REFUGE-val dataset. The performance using the combined features (depicted in red) appears to be more robust to the changes in the dataset. These variations in trends may be due to the variations in image quality (as the datasets are acquired using different fundus cameras) and the inter-expert variations in the ground truth annotations across the three datasets.

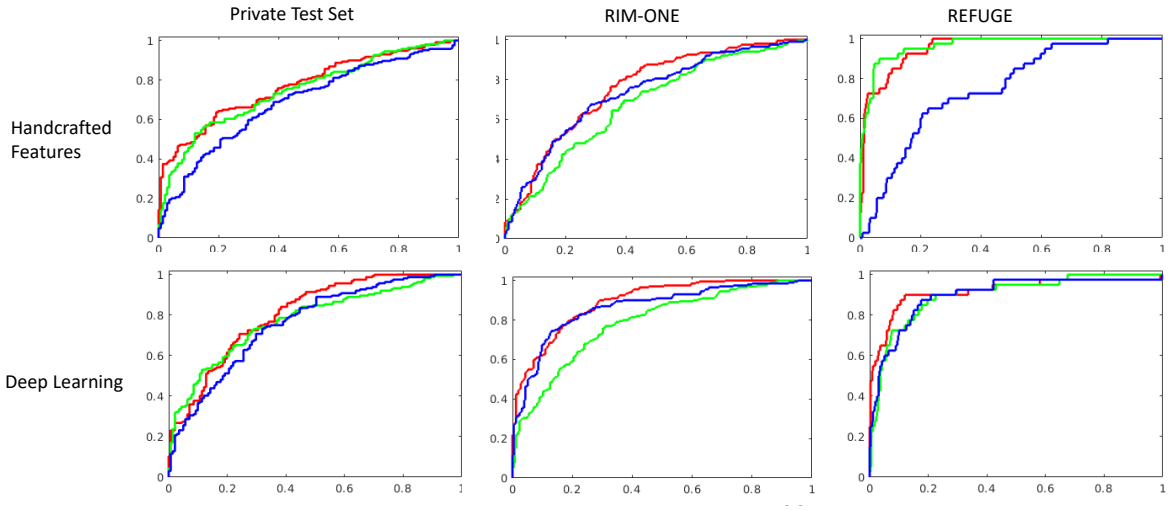

Figure 5.4: Area under the ROC curve for the proposed method based on handcrafted features (1^{st} row) and Deep Learning (2^{nd} row) on the private test set (1^{st} column), RIM-ONE (2^{nd} column) and the REFUGE (3^{rd} column) dataset respectively. In each plot, the ROC curve for the segmentation features (green), appearance features alone (blue) and the combined features (red) are depicted.

	Proposed-1 (Deep Learning)			Proposed-2 (Handcrafted Features)		
	Appearance Features	Segmentation Features	Combined	Appearance Features	Segmentation Features	Combined
Private-Test						
Sensitivity	0.7317	0.7256	0.6768	0.6890	0.5854	0.6220
Specificity	0.6809	0.7163	0.7589	0.6170	0.8298	0.8085
Balanced Acc.	0.7063	0.7210	0.7178	0.6530	0.7076	0.7152
AUC	0.7594	0.7754	0.7805	0.6887	0.7472	0.7556
RIM-ONE r2						
Sensitivity	0.7450	0.7700	0.7550	0.6750	0.6950	0.7950
Specificity	0.8706	0.6902	0.8531	0.7176	0.6078	0.6235
Balanced Acc.	0.8078	0.7301	0.8041	0.6963	0.6514	0.7093
AUC	0.8641	0.7860	0.8758	0.7334	0.6802	0.7570
REFUGE-val						
Sensitivity	0.8750	0.8750	0.9100	0.6500	0.9000	0.9250
Specificity	0.8250	0.8028	0.8906	0.7750	0.9361	0.8472
Balanced Acc.	0.8500	0.8389	0.9003	0.7125	0.9181	0.8861
AUC	0.8982	0.8983	0.9461	0.7472	0.9655	0.9536
Average across all Datasets						
Sensitivity	0.7129	0.7748	0.7806	0.6683	0.9035	0.7807
Specificity	0.7963	0.7460	0.8342	0.7566	0.6045	0.7597
Balanced Acc.	0.7546	0.7604	0.8074	0.7125	0.7540	0.7702
AUC	0.8114	0.8339	0.8675	0.7604	0.8253	0.8221

Table 5.3: Individual Contribution of the Segmentation and Appearance based features.

The performance of the proposed methods have also been benchmarked against some state of the art methods. The Resnet-50 [164] and the VGG-16 [165] DL architectures were fine-tuned for this task using the same dataset that was used to train our proposed methods. In order to fine-tune the two networks, the final classification layer was replaced by a Fully connected layer with a single output using a sigmoid activation similar to our proposed DL based architecture. The fine-tuning was carried out in two steps. First, the weights of the convolutional layers were fixed to the pre-trained weights trained on natural images (from the imagenet dataset) and only the weights of the final classification layer were learned using a learning rate of 10^{-2}. Next, the weights of the convolutional layers were unfrozen and the entire network was trained in an end-to-end manner with a smaller learning rate of 10^{-4}. The Binary cross-entropy loss was employed for fine-tuning these models. We also benchmarked our performance against an existing DL based method called the DENet [159] and a handcrafted wavelet feature based method [156] which have been specifically designed for the task of glaucoma classification. The DENet employed an ensemble of four independent neural networks whose predictions were fused to obtain the final decision.

The performance of the proposed methods after combining both the appearance and segmentation based features have been compared against these methods in Table 5.4. Considering the average over all datasets, the performance of the proposed method based on handcrafted features is better than the wavelet based features and DENet architecture with an AUC of 0.8221 but is lower than the fine-tuned Resnet-50 and VGG-16 architectures.

The proposed DL based method has the best performance among all the methods with an AUC of 0.8675 with only a marginal improvement over the fine-tuned Resnet-50. However, in contrast to the larger Resnet architecture with around 25 million network parameters, the proposed DL based method is able to achieve comparable performance using a much smaller network with 609, 170 network parameters. As a result, the proposed method is more memory efficient and less prone to overfitting. Thus, the proposed method was trained from scratch without the need to pre-train it on larger datasets.

5.1.8 Conclusion

In this work, we have explored and compared two classification frameworks for the image level detection of glaucoma in CF images based on handcrafted features and DL respectively.

	Proposed-1 (Deep Learning)	Proposed-2 (Handcrafted)	VGG-16	RESnet-50	DE-net	Wavelet
Private-Test						
Sensitivity	0.6768	0.6220	0.6668	0.7186	—	0.7622
Specificity	0.7589	0.8085	0.8056	0.7022	—	0.4894
Balanced Acc.	0.7178	0.7152	0.7362	0.7104	—	0.6258
AUC	0.7805	0.7556	0.7930	0.7705	—	0.6696
RIM-ONE						
Sensitivity	0.7550	0.7950	0.7850	0.7850	0.7576	0.4300
Specificity	0.8531	0.6235	0.8271	0.9216	0.6923	0.8275
Balanced Acc.	0.8041	0.7093	0.8060	0.8533	0.7250	0.6287
AUC	0.8758	0.7570	0.8640	0.9201	0.7821	0.6438
REFUGE-val						
Sensitivity	0.9100	0.9250	0.7250	0.7250	0.7750	0.8250
Specificity	0.8906	0.8472	0.9250	0.9889	0.7395	0.6222
Balanced Acc.	0.9003	0.8861	0.8250	0.8569	0.7573	0.7639
AUC	0.9461	0.9536	0.8660	0.8956	0.8343	0.7236
Average across all Datasets						
Sensitivity	0.7806	0.7807	0.7256	0.7429	0.7663	0.6724
Specificity	0.8342	0.7597	0.8526	0.8709	0.7160	0.6464
Balanced Acc.	0.8074	0.7702	0.7891	0.8069	0.7412	0.6728
AUC	0.8675	0.8221	0.8410	0.8621	0.8082	0.6790

Table 5.4: Benchmarking Glaucoma Classification Performance against the state of the art.

Both the methods employed a combination of appearance features and structural features to obtain robust predictions accross a wide range of datasets that covered a range of ethnicity in population, imaging protocols, fundus cameras and image quality. While the appearance

features were directly extracted from the ROI image, the structural features were extracted from the Optic Disc and Cup segmentation.

The method based on handcrafted features employed the CRF based OD and OC segmentation method detailed in Chapter 2 and extracted a 104 dimensional feature which achieved an AUC of 0.82 on the glaucoma classification task using a RBF kernel SVM classifier. The proposed DL based method employed a Multi-task CNN architecture to jointly segment the OD, OC and predict the presence of glaucoma. The features of the CNN are shared across all the three related tasks thereby reducing the computational requirements and improving the generalizability of the learned features. The proposed architecture employs $609,170$ network parameters and is significantly smaller in comparison to the existing state of the art architectures such as DENet [159]. Fewer network parameters ensure that the network doesnot overfit when trained from scratch on small datasets and also reduces the memory requirements. Though the images in training and test set are acquired using different fundus cameras and at different resolutions, the proposed method achieved an AUC of 0.87 on the test set illustrating its potential as a mass screening tool for the early detection of glaucoma.

5.2 AMD detection in 3D OCT Volumes

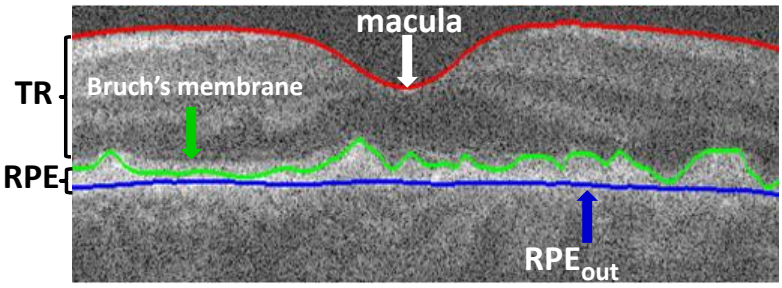

Figure 5.5: OCT B-scan of an AMD case with the three relevant layer boundaries.

Optical Coherence Tomography (OCT) is an important 3D retinal imaging modality that provides the cross-sectional view of the various intra-retinal tissue layers [1]. The thickness and morphology of the various layers are directly correlated to the health of the eye. In Age Related Macular Degenration (AMD), the drusen deposits occur in the Retinal Pigment Epithelium

[1]We refer to the discussion in Chapter 1, Section 1.1.2, pages 7-8 for details of the OCT imaging and the seven intra-retinal tissue layers.

(RPE) layer leading to irregularities and undulations in the Bruch's membrane (see Fig. 5.5) around the macular region. Using the terminology from [166], we refer to the region between the ILM and Bruch's membrane as the Total Retina (TR).

Atlas plays an important role in medical image analysis by providing a standard coordinate frame to represent the anatomy. It is often in the form of a mean intensity template (MT) obtained by the average of a set of co-registered images along with the probability maps (P-maps) which give the probability of observing a particular tissue or structure at a given location (see Fig. 5.6). The P-maps are computed from a set of annotated images by registering them to the MT and transferring their manual expert labelings into the Atlas space. Normalization of images to a single coordinate frame via registration to the MT is useful in applications such as the localization of anatomical structures, disease detection and therapy planning.

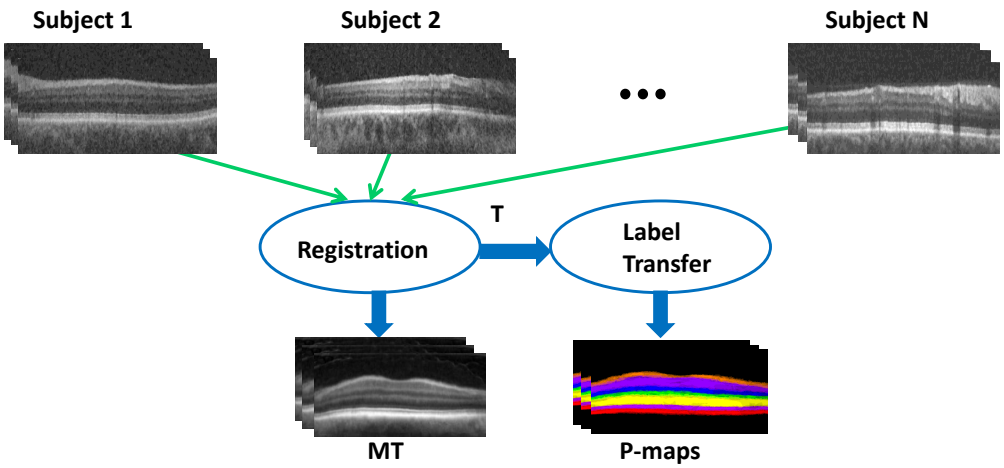

Figure 5.6: The normative OCT atlas consists of a mean intensity template (MT) obtained by the average of a set of co-registered healthy OCT Volumes and probability maps (P-maps) which give the probability of observing a particular tissue layer at a given location.

The construction of a Normative OCT atlas has received little attention. The problem of the inter-subject registration of the OCT images was explored in [167] where after an initial rigid alignment, each A-scan in the moving image was individually registered to the corresponding A-scan in the reference image in the axial direction. In comparison, in this chapter we explore a free-form deformation in 3D. In contrast to brain imaging where multiple atlases are publicly available [168], no retinal OCT atlas is available so far. The main challenges in the construction of an OCT atlas include the presence of speckle noise, vessel shadows and the inter or intra-

scanner intensity variations across the A-scans. The curvature of the retinal surface also leads to large shifts in the retinal tissue across the B-scans.

Our contribution in this work is two-fold. First, we construct a Normative OCT atlas for 3D macular OCT Volumes. Next, the utility of the Atlas is demonstrated on the task of detection of AMD in OCT volumes by leveraging the deviations in the local similarity between the MT and the registered test OCT Volumes.

5.2.1 Method

An overview of the atlas construction method is presented in Section 5.2.1.1. The application of the atlas to a coarse localization of the intra-retinal layers and AMD detection (see Fig. 5.7) is detailed in Sections 5.2.1.5 and 5.2.1.4 respectively.

The preprocessing step comprises resizing, denoising, intensity standardization and retinal curvature flattening of the OCT volumes. Each B-scan in the OCT volume is resized to normalize the pixel dimensions to 3.6 μm by 8.6 μm. Denoising is done via speckle reducing anisotropic diffusion [104](30 iterations, timestep=0.1) and intensity standardization is achieved with the method proposed in [105].

Retinal flattening is performed in two steps. First, each B-scan is individually flatenned using the method described in Chapter 3 (Section 3.2.1). In the second step, the B-scans are aligned across the volume. Each B-scan is sequentially aligned to its previous slice by an exhaustive search of the axial pixel translations that maximizes the normalized cross correlation (NCC) between them.

5.2.1.1 Atlas Construction

Let us consider a set of N OCT volumes, each of size $X \times Y \times Z$ with X voxels along the axial direction in each A-scan, Y A-scans in each B-scan (slice) and Z B-scans in the entire volume. Let $\Omega = \{(x, y, z) | 1 \leq x \leq X, 1 \leq y \leq Y, 1 \leq z \leq Z \text{ and } x, y, z \in \mathbb{Z}\}$ represent the set of all voxel locations in the entire image domain where $\mathbb{Z}$ denotes the set of integers.

Our objective is to compute a MT which approximates the group mean of the N images. This involves a groupwise registration of the N OCT volumes followed by computing their average to obtain the MT. Although some methods [169], [170] have been explored to simultaneously co-register multiple images to a common reference coordinate system, their computational and

Algorithm 1: Atlas Construction

Input : Set of Normal OCT volumes

$\{I_1, I_2 ... I_N\}$

Output: Atlas Mean Template $\mathcal{M}_T$

$\mathcal{M}_0 \leftarrow$ Select_Initial_Template$(I_1, I_2 ... I_N)$

for $t \leftarrow 1$ **to** T **do**

  $\hat{I}_i \leftarrow$Pairwise_Register $(I_i, \mathcal{M}_{t-1})$,

  $\forall 1 \leq i \leq N$

  $\mathcal{M}_t \leftarrow$Weighted_Average$(\hat{I}_1, ... \hat{I}_N)$

end

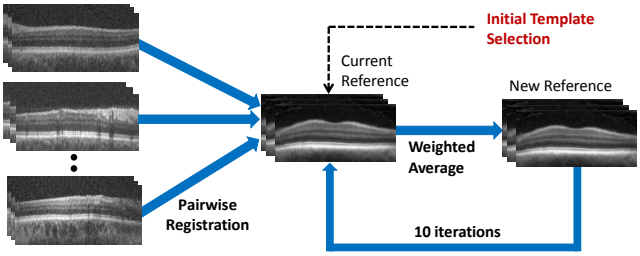

Figure 5.7: Atlas Construction pipeline.

memory requirements do not scale well for a large number of image volumes. Therefore, we employed an iterative method similar to [171], [172] that employs N pairwise registrations in each iteration (see Fig. 5.7). At first, one of the OCT volumes is selected as an initial estimate of the MT denoted by $\mathcal{M}_0$. Thereafter, in each iteration denoted by t, two steps are alternatively performed as depicted in the *Algorithm 1*. First, each of the N OCT Volumes I_i is transformed to the Atlas coordinate space by a pairwise registration to $\mathcal{M}_{t-1}$ to obtain $\hat{I}_i$. Next, the estimate of the MT is recomputed from the transformed images as a weighted average of the intensities at each voxel location $\mathbf{p} \in \Omega$. Thus, $\mathcal{M}_t$ is computed as $\mathcal{M}_t(\mathbf{p}) = \sum_{i=1}^{N} w_i(\mathbf{p}) \times \hat{I}_i$, where $w_i(\mathbf{p})$ denotes the scalar weights. The weighted average reduces to the mean of the registered images (which was used in [171]) by fixing $w_i(\mathbf{p})$ to $1/N$ for each image $\hat{I}_i$ and voxel location $\mathbf{p}$.

However, this iterative method is sensitive to the initialization of $\mathcal{M}_0$ which should selected as an image close to the group mean. Moreover, the large inter-subject variations can result in a blurred $\mathcal{M}_t$, particularly in the initial iterations. In order to prevent blurring, we have employed the method based on [172] to adapt the $w_i(\mathbf{p})$ for each spatial location across each volume. A local patch centered around the location $\mathbf{p}$ is used to compute the local similarity between $\hat{I}_i(\mathbf{p})$ and $\mathcal{M}_{t-1}(\mathbf{p})$ using the Euclidean distance denoted by $d\left(\hat{I}_i(\mathbf{p}), \mathcal{M}_{t-1}(\mathbf{p})\right)$. The weights are defined to be inversely proportional to the distance so that the image patches similar to the $\mathcal{M}_{t-1}(\mathbf{p})$ are given a higher weight. Moreover, the weights are exponentially decayed to ensure that the distant $\hat{I}_i(\mathbf{p})$ have negligible weights. Thus,

$$w_i(\mathbf{p}) = exp\left(-\frac{d\left(\hat{I}_i(\mathbf{p}), \mathcal{M}_{t-1}(\mathbf{p})\right)}{r}\right), \tag{5.3}$$

where $r = 1 + \Delta r.\frac{t}{T}$ controls spread of the distribution of the weights. Initially, the value of r is kept small to ensure that only few images which are similar to the MT are used to update it and preserve the sharpness of the layer boundaries. As the iterations progress, r is slowly increased to gradually encourage an equal weighting for all $\hat{I}_i(\mathbf{p})$. $\Delta r = \underset{\mathbf{p},i}{\text{argmax}} \quad d\left(\hat{I}_i(\mathbf{p}), \mathcal{M}_0(\mathbf{p})\right)$ is defined as the maximum euclidean distance encountered between $\mathcal{M}_0$ and the images $\hat{I}_i$ across all the spatial locations at $t = 1$. The weights at each location $\mathbf{p}$ are normalized as $w_i(\mathbf{p}) = \frac{w_i(\mathbf{p})}{\sum_{j=1}^{N} w_j(\mathbf{p})}$ to sum to 1 before computing the weighted average. The final MT was obtained after $T = 10$ iterations and the patch size used to compute $w_i(\mathbf{p})$ was decreased from $19 \times 19 \times 19$ to $1 \times 1 \times 1$ in steps of 2 at each iteration.

Our implementation differs from [172] in two ways: a) The initial template $\mathcal{M}_0$ is selected as the volume that lies closest to the group mean of the N images as discussed below in Section 5.2.1.3; b) Since size of the OCT volumes ($216 \times 770 \times 100$) is much larger than the typical size encountered in neuroimaging, we replace the Demon's registration approach in [172] with a discrete registration framework (discussed below in Section 5.2.1.2) for the pairwise registrations to improve the computational efficiency.

5.2.1.2 The pairwise registration algorithm

The pairwise registration between a fixed (I_f) and moving (I_m) OCT volume is performed in two steps. Since, registration along the axial direction is critical for the alignment of the anatomical layers, an initial rigid translation is performed by maximizing the global NCC of the entire volume using an exhaustive search of pixel translations along the axial direction.

This is followed by a discrete, non-parametric, free-form registration where each voxel at a location $\mathbf{p}$ in I_m is displaced by a vector $\mathbf{u_p} = [u_x(\mathbf{p}), u_y(\mathbf{p})), u_z(\mathbf{p})]$ in order to register it to I_f. Let l_{max} denote the maximum amount of displacement allowed in each of the three directions and the displacements be discretized to take only integer values. Thus, $\mathbf{u_p}$ can take values from a discrete, finite Label set $\mathcal{L} = \{-l_{max} \leq l \leq +l_{max}, l \in \mathbb{Z}\}^3$. The set of all the random variables $\mathbf{U} = \{\mathbf{u_p} | p \in \Omega\}$ forms a random (deformation) field. A feasible labeling for

$\mathbf{U}$ denoted by $\mathbf{u}$ can be obtained by assigning a label from $\mathcal{L}$ to each random variable $\mathbf{u_p}$. The deformation field for the registration can be obtained by finding the feasible labeling $\mathbf{u}^*$ that minimizes the Markov Random Field Energy,

$$\mathbf{u}^* = \underset{\boldsymbol{u}}{\operatorname{argmin}} \sum_{p \in \Omega} S(I_f, I_m, \boldsymbol{u_p}) + \alpha . |\nabla \boldsymbol{u}|^2. \tag{5.4}$$

The first term in eq. 5.4 imposes a unary cost $S(I_f, I_m, \boldsymbol{u}_p)$ defined over each $\mathbf{u_p}$ and measures the dissimilarity between the local $3 \times 3 \times 3$ image patches centered at $I_f(\mathbf{p})$ and $I_m(\mathbf{p} + \mathbf{u_p})$ using the negative of the NCC similarity metric. The second term in eq.5.4 is a regularization term that ensures the smoothness of the deformation field. The scalar α is the relative weight. $|\nabla \boldsymbol{u}|^2$ can be approximated using pairwise cost terms defined between each random variable u_p at location $\mathbf{p}$ and its immediate 6-connected neighbors in 3D denoted by $\mathcal{N}_p$ as $\sum_{\mathbf{p} \in \Omega} \sum_{\mathbf{q} \in \mathcal{N}_p} ||\mathbf{u_p} - \mathbf{u_q}||^2$ which measures the magnitude of the vector differences [173].

However, solving the above MRF formulation is computationally infeasible due to the large number of nodes (one corresponding to each voxel location) and a big label set. For example, if we consider a maximum displacement of ± 7 voxels in each direction (ie., $l_{max} = 7$) then the size of the label set $|\mathcal{L}| = 15^3 = 3375$. Moreover, in our case the number of nodes in the MRF is $216 \times 770 \times 100 = 16632000$. Therefore, instead of the TRW-S algorithm which was employed in Chapters 2 and 3, we use the method proposed in [174] which provides a very fast but approximate solution to the MRF inference problem. The eq. 5.4 can be re-written using an auxiliary deformation field $\mathbf{v}$ as

$$\underset{\mathbf{u},\mathbf{v}}{\operatorname{argmin}} \sum_{p \in \Omega} S(I_f, I_m, \mathbf{v}_p) + \frac{1}{\lambda} ||\mathbf{v} - \mathbf{u}||_2^2 + \alpha . |\nabla \boldsymbol{u}|^2. \tag{5.5}$$

Now, eq. 5.5 is solved using an iterative approach by alternatively optimizing with respect to $\mathbf{u}$ and $\mathbf{v}$ until convergence. The second term in the equation ensures that $\mathbf{v} \approx \mathbf{u}$ and it is separable at each voxel position $\boldsymbol{p}$ as $\frac{1}{\lambda}||\mathbf{v} - \mathbf{u}||_2^2 = \sum_{p \in \Omega} \frac{1}{\lambda}||\mathbf{v_p} - \mathbf{u_p}||_2^2$. When $\mathbf{u}$ is fixed, the above optimization problem reduces to $\underset{\boldsymbol{v}}{\operatorname{argmin}} \sum_{p \in \Omega} \left\{ S(I_f, I_m, \boldsymbol{v}_p) + \frac{1}{\lambda}||\mathbf{v_p} - \mathbf{u_p}||_2^2 \right\}$. The optimal $\boldsymbol{v}$ can be simply solved by adding $\frac{1}{\lambda}||\mathbf{v_p} - \mathbf{u_p}||_2^2$ to the Unary Cost term $S(I_f, I_m, \mathbf{v}_p)$ and taking the argmin for each $\boldsymbol{p}$ across all possible displacements independent of it's neighbors. The value of the scalar weight λ is gradually decreased in each iteration to ensure convergence. Similarly, when $\mathbf{v}$ is fixed, the above optimization reduces to $\underset{\mathbf{u}}{\operatorname{argmin}} \frac{1}{\lambda}||\mathbf{v} - \mathbf{u}||_2^2 + \alpha . |\nabla \boldsymbol{u}|^2$. The

optimal $\mathbf{u}$ is approximated by simply smoothing $\mathbf{v}$ with a Gaussian filter. In our experiments, the σ of the gaussian filter was set to 1 which indirectly controls the weight α of the pairwise term in eq. 5.4. We performed the deformable registration at 3 scales. l_{max} was restricted to ± 7, ± 4 and ± 2 voxels from the coarse to fine scale respectively.

The deformation field should be invertible in order to avoid being dependent on the choice of the fixed (I_f)and the moving (I_m) images for the pairwise registration. The symmetric deformation field is obtained by registering the two images in both forward and backward directions to obtain the two deformation fields $\boldsymbol{u_f}$ and $\boldsymbol{u_b}$ respectively. Then, following [174] the symmetric deformation fields are obtained by iteratively updating them for 10 iterations as

$$\begin{aligned} \boldsymbol{u_f}^{n+1} &= 0.5(\boldsymbol{u_f}^{n} - \boldsymbol{u_b}^{n}(\boldsymbol{p} + \boldsymbol{u_f}^{n})) \\ \boldsymbol{u_b}^{n+1} &= 0.5(\boldsymbol{u_b}^{n} - \boldsymbol{u_f}^{n}(\boldsymbol{p} + \boldsymbol{u_b}^{n})) \end{aligned} \tag{5.6}$$

5.2.1.3 Initial Template Selection

The selection of the initial template $\mathcal{M}_0$ is crucial for obtaining an unbiased MT and it should be chosen as an image close to the group mean of the N OCT volumes. We employed a method based on [175] for this purpose.

At first, the set of N images is represented as an undirected complete graph by constructing an adjacency matrix denoted by D using the pairwise dissimilarities $d_{i,j}$ between each pair of OCT Volumes I_i and I_j. The dissimilarity measure $d_{i,j}$ is defined as the registration error $E(\boldsymbol{u^*}; I_i, I_j)$ where $\mathbf{u}^*$ represents the displacement field obtained after registering I_j to I_i. In contrast, the bending energy of $\boldsymbol{u^*}$ alone was used in [175] to define the distance neglecting the similarity between the two image. Since, the registration of each pair of images is computationally expensive, the images were resized by a factor 0.5 before the registration to obtain an approximate estimate of $d_{i,j}$. The diagonal elements of D represent the dissimilarity of the image to itself and hence $d_{i,i} = 0$. Moreover, since a symmetric registration is employed, $d_{i,j} \approx d_{j,i}$, requiring us to compute the registration errors for the elements in the upper triangular matrix of D which are replicated to the corresponding elements in the lower triangular part of D.

However, the dissimilarity values $d_{i,j}$ is not a metric since the triangular inequality constraint $d_{i,j} \leq d_{i,k} + d_{k,j}$ for any three images I_i,I_j and I_k may not be satisfied. Therefore, based on the

work in [175], the Multidimensional Scaling (MDS) was applied to map each of the N images I_i to a point $\mathbf{x_i}$ in a N-dimensional metric space such that the euclidean distances between every pair of points best approximates the dissimilarity between them, ie., $d_{i,j} \approx ||\mathbf{x_i} - \mathbf{x_j}||_2$. The MDS is computed in two steps. First, the matrix of the inner product matrix $B = X^T.X$ is obtained from the adjacency matrix D where $X = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}]$. The $(i,j)^{th}$) value in B is computed as:

$$b_{i,j} = -\frac{1}{2}\left[d_{i,j}^2 - \frac{1}{N}.\sum_j d_{i,j}^2 - \frac{1}{N}.\sum_i d_{i,j}^2 + \frac{1}{N^2}.\sum_j\sum_i d_{i,j}^2\right] \tag{5.7}$$

Next, an eigen decomposition of $B = V.\Lambda.V^T$ is performed to obtain $X = \Lambda^{1/2}.V$. Finally, the mean of the MDS feature vectors of all images are computed and the image with the minimum euclidean distance from the mean feature vector is selected as $\mathcal{M}_0$.

5.2.1.4 AMD Classification

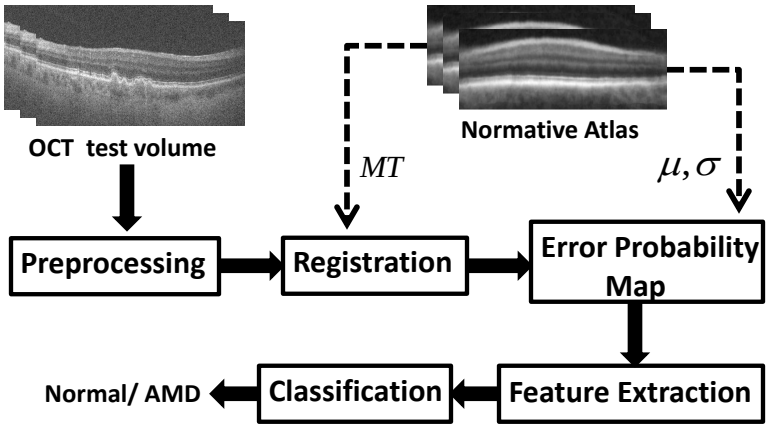

Figure 5.8: AMD classification pipeline.

The OCT atlas can aid in the detection of ocular diseases by enabling the extraction of a clinically relevant Region of Interest (ROI). Moreover, the abnormal images when registered to a Normative atlas would tend to have higher registration errors which can be leveraged to extract relevant features for disease classification.

We demonstrate this on the task of AMD detection. An atlas based approach has not been explored for AMD detection so far. Existing methods extract the RPE layer to define the (ROI) and employ handcrafted features such as Local Binary Patterns (LBP) [176], Histogram

of Oriented Gradients (HoG) [177] or Bag of visual words (BoW) [178] to detect AMD. The local features are aggregated using histogram [178] and employ multi-scale [177] or spatial pyramid [176] approaches to encode the spatial context. The details of the proposed atlas based method are outlined in Fig. 5.7 and described below.

ROI Extraction : A test OCT volume is first preprocessed and registered to the MT to obtain I_{tst}. Since the morphological changes in AMD are found in the area surrounding the macula in the RPE layer, a smaller ROI is extracted in the atlas space for computational efficiency. The ROI is defined to span 31 B-scans and 180 A-scans centered at the macula. In the axial direction, the ROI is restricted to a small region around the RPE layer (rows 124-210 in the atlas space).

Error Probability Map Extraction : A local similarity measure for I_{tst} is evaluated at each voxel $\boldsymbol{p}$ within the ROI by extracting a $3 \times 3 \times 3$ patch $I_{tst}(p)$ around p and computing the NCC similarity $S(I_{tst}(\boldsymbol{p}))$ with respect to the corresponding patch in the MT. The AMD cases (volumes) are expected to have smaller values for $S(I_{tst}(\boldsymbol{p}))$ within the ROI.

However, the absolute value of the $S(I_{tst}(\boldsymbol{p}))$ may not be meaningful. Depending on the complexity of the anatomical structure at a voxel location, the distribution of the local similarity of the Normal OCT volumes also varies across $\boldsymbol{p}$. Hence the $S(I_{tst}(\boldsymbol{p}))$ values are used to derive an error probability map $P_{err}(I_{tst})$. The $P_{err}(I_{tst})$ at location $\boldsymbol{p}$ measures the probability that $S(I_{tst}(\boldsymbol{p}))$ is sampled from the distribution of similarity values at $\boldsymbol{p}$ that is encountered in the Normal OCT volumes. Assuming a Gaussian distribution for the similarity at each $\boldsymbol{p}$ independently, a mean ($\mu(\boldsymbol{p})$) and standard deviation $\sigma(\boldsymbol{p})$ map is generated for the ROI from the Normal OCT volumes used in the atlas construction. During testing, $P_{err}(I_{tst})$ is computed as $\frac{1}{\sqrt{2\pi\sigma(\boldsymbol{p})^2}} exp\{\frac{(S(I_{tst}(\boldsymbol{p}))-\mu(\boldsymbol{p})).^2}{2\sigma(\boldsymbol{p})^2}\}$.

Feature Extraction and Classification : To obtain a global feature representation from the P_{err}, it is scaled to $[0, 1]$ across the samples. Thereafter, a 10 bin (uniform bins of width 0.1,centered at 0.05 to 0.95) histogram is computed. Though the abnormality can occur at any column position in the ROI, it should lie close to the bruch's membrane. Hence, the spatial location of the error probabilities along the rows may be important for classification. To incorporate this information, the ROI was further sub-divided into 6 equal row-wise sections and a separate histogram of the P_{err} was also computed for each sub-region. Finally, all histogram features were concatenated resulting in a $(10 + 10 \times 6 =)$70-D feature. A binary

linear SVM classifier was employed for AMD classifcation.

5.2.1.5 Coarse Retinal Tissue Localization

Apart from abnormality detection, the Normative Atlas can also be used to obtain a coarse localization of the different intra-retinal tissue layers by computing their P-maps during training. The P-maps for each layer is obtained using a set of OCT images with manually segmented layers. Each OCT image is independently registered to the Atlas reference MT and its displacement field is applied to deform the corresponding groundtruth segmentation label maps. Finally, for each voxel in the atlas space, a probability value is computed for each label as the fraction of the registered OCT images that have the particular label at that location.

During testing, a probabilistic segmentation is obtained by registering the atlas MT to a test OCT volume and applying the corresponding deformation field to the atlas P-maps for each layer.

5.2.2 Experimental Setup

Dataset: The proposed method for the detection of AMD in OCT volumes was evaluated on the public AMD-3D dataset [166]. It consists of 290 (102 Normal, 188 AMD cases) 3D macular OCT volumes acquired from 4 clinics at a size of $512 \times 1000 \times 100$ with a voxel resolution of $(3.25, 6.7, 67)$ $\mu m/voxel$ along the axial, lateral and azimuthal directions.

In order to evaluate the performance of the proposed method for retinal tissue localization, two additional public datasets, NORMAL-1 [13] and NORMAL-2 [89] were employed. Both NORMAL-1 and NORMAL-2 datasets contain 10 OCT volumes each, acquired using different OCT scanners at varying resolutions (See Chapter 3, Table 3.1 for further details). The two datasets provide the GT markings of seven retinal layers for only a few non-consecutive linearly spaced B-scans per volume (10 B-scans per volume in NORMAL-1 and 11 B-scans per volume in NORMAL-2).

5.2.3 Results

AMD classification The μ and σ used to compute the P_{err} was computed from the 40 volumes from AMD-3D dataset used in the atlas construction. These volumes were excluded

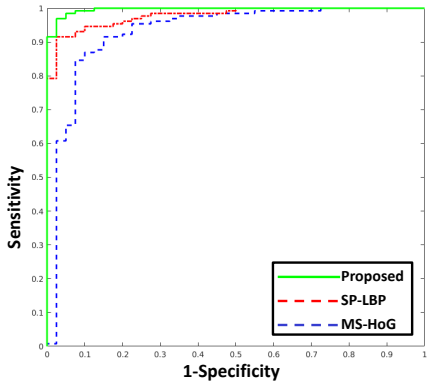

Figure 5.9: ROC for AMD detection.

	Sens.	Spes.	Acc.	AUC
MS-HoG [177]	0.90	0.85	0.90	0.930
SP-LBP [176]	0.93	0.92	0.92	0.978
BoW [178]	0.96	0.92	0.94	0.984
Proposed	0.97	0.98	0.98	0.996

Table 5.5: AMD classification performance.

and the remaining OCT volumes in the AMD-3D dataset were randomly divided into a separate training set of 80 (22 Normal and 58 AMD cases) volumes and a test set of 170(40 Normal and 130 AMD cases) volumes respectively. The cost paramter (C) of the linear SVM classifier which controls the relative weight of the L2-regularization was set to 100 and the mis-classification penalty for each class was weighted during training to handle the class imbalance. The classification performance is depicted in Table 5.5 and the corresponding ROC plot in Fig. 5.9. The good performance of our method with a linear classifier demonstrates the discriminative capacity of the proposed feature. Our atlas based feature outperforms the existing Bag of Words, multiscale HoG (MS-HoG) and the pyramid based SP-LBP features explored in [178], [177] and [176] respectively.

Intra-retinal layer segmentation To evaluate the localization of the retinal layers, the P-maps were computed using the NORMAL-1 dataset. Since, GT for only a few B-scans are available, the P-maps for each B-scan in the atlas MT was computed individually by registering to it, the nearest B-scan (with respect to the distance from macula) from each volume in NORMAL-1. The P-maps for the seven tissue layers are depicted in Fig. 5.10. The localization of all the seven layers were evaluated on the NORMAL-2 dataset and benchmarked against three open source OCT segmentation softwares, the Iowa Reference Algorithm (IRA) version 3.8.0 based on [91], CASEREL based on [13] and OCTSEG based on [112]. Additionally, the performance of our CRF based method proposed in Chapter 3 has been reproduced for comparison. CASEREL provides segmentation of seven boundaries (except the bruch's membrane) and OCTSEG segments six boundaries (except the bruch's membrane and INL/OPL). The quan-

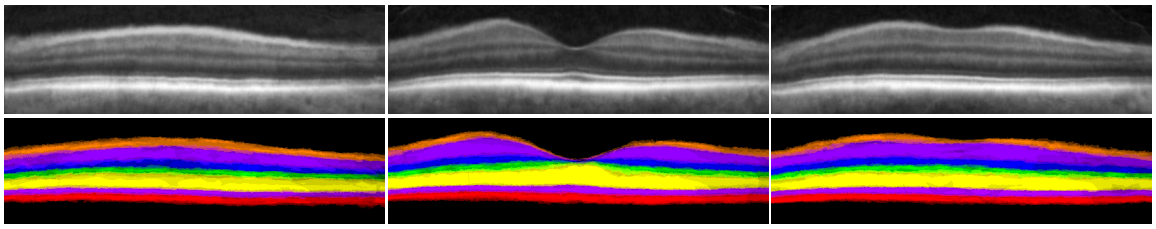

Figure 5.10: (left to right) the mean intensity atlas (1^st row) and the corresponding probability maps for the seven tissue layers for the 30^{th}, 50^{th} and 60^{th} B-scans.

titative results in Table 5.6 show that the performance of the proposed Atlas based method is comparable to the existing methods but consistently outperformed by the proposed CRF based method. The results illustrate that the proposed Atlas based method can provide a reasonably good localization of each retinal layer and can be used to provide a good initialization to deformable models such as [85] to further fine-tune the layer segmentation.

	NFL	GCL-IPL	INL	OPL	ONL-IS	OS	RPE
CASEREL	0.81±0.13	0.78±0.11	0.50±0.14	0.62±0.12	0.90±0.06	—	—
OCTSEG	0.77±0.18	0.76±0.22	—	—	0.91±0.06	—	—
IRA	0.60±0.15	0.61±0.12	0.45±0.11	0.64±0.10	0.92±0.03	0.88±0.05	0.91±0.04
CRF (Chapter 3)	**0.88±0.05**	**0.92±0.06**	**0.87±0.05**	**0.86±0.03**	**0.96± 0.01**	**0.88±0.04**	0.89± 0.03
Proposed	0.79±0.08	0.89±0.07	0.81±0.07	0.75±0.08	0.93±0.03	0.74±0.11	0.83±0.06

Table 5.6: Dice coefficients (mean ± standard deviation) for the segmentation of 7 retinal tissue layers on the NORMAL-2 dataset.

5.2.4 Conclusion

In this Section, we have constructed an atlas for normal OCT volumes which can have many potential applications such as the extraction of a clinically relevant ROI for analysis, characterizing abnormalities as deviations from the Normative atlas or providing rough localization of the intra-retinal tissue layers. The probabilistic segmentation maps (P-maps) of seven retinal tissue regions have been obtained which can be transferred onto the test images by registering the Atlas MT to them. It achieved an average dice of 0.82 across the seven layers. Better registration algorithms specifically adapted for OCT volumes can be investigated in the future to further improve on the segmentation performance. We have also proposed a novel classification scheme for AMD which demonstrates good performance with an accuracy of 98% and

area under the ROC curve of 0.996 on a test set of 170 OCT volumes. Investigating methods to employ the Normative Atlas to more challenging pathologies such as cysts caused by Diabetic Macular Edema(DME) is a promising direction for future work.

Chapter 6

Conclusions

In this Chapter, we conclude this thesis by summarizing our contributions and discussing some possible directions for future research.

6.1 Summary

This thesis introduces various strategies for the extraction of boundaries of anatomical structures in retinal images with potential applications in the automated detection and quantification of ocular diseases. In this context, we have explored novel Conditional Random Field (CRF) frameworks for the simultaneous extraction of Optic Disc (OD) and Optic Cup (OC) in Color Fundus (CF) and the eight intra-retinal layer boundaries in the cross-sectional Optical Coherence Tomography (OCT) images. These structures play a critical role in the detection and tracking the progression of retinal diseases such as glaucoma and Age-related Macular Degeneration (AMD).

Annotated medical datasets are a valuable resource in the medical image analysis community and I have partially contributed towards the creation of the DRISHTI-GS dataset [6] which consists of 101 CF images with groundtruth segmentation masks of the OD, OC and the image level diagnosis of glaucoma which were obtained by taking the majority voting of the annotations collected from four medical experts.

A novel Recurrent Neural Network (RNN) called the Recurrent Active Contour Evolution Network (RACE-net) has also been explored for the segmentation task which models the level set based deformable models within a deep learning framework. Apart from OD and OC segmentation, the effectiveness of the RACE-net architecture has been demonstrated on a

diverse set of tasks including the segmentation of cell nucleus in histopathology and left atrium in cardiac MRI volumes. Finally, we have explored classification frameworks for the image-level detection of Glaucoma from CF images and AMD from OCT volumes respectively. Two strategies based on handcrafted features and deep learning respectively, have been explored for the image-level detection of glaucoma in CF images. Both the methods attempt to capture the structural changes through features derived from the OD-OC segmentation as well as the appearance using features directly derived from the region of interest around the OD. A solution for the detection of AMD from 3D OCT volumes has also been proposed which attempts to model the structural changes in the Bruch's membrane as deviations from a Normative OCT atlas. Thus, the outcomes of this thesis can be summarized as follows.

1. A CRF framework for the joint extraction of OD and OC boundaries in retinal CF images. The method estimates depth from a single 2D CF image alone during testing to model the OC boundary.

2. An end-to-end trained CRF framework for the simultaneous extraction of multiple intra-retinal layer boundaries in OCT images. The framwework incorporates both shape priors and appearance based cost terms which are jointly trained within a Structural Support Vector Machine formulation.

3. A generic RNN Architecture for boundary based segmentation that models the level set based curve evolutions within a deep learning framework. The method has the potential of being used for the segmentation of a wide range of anatomical structures in different medical imaging modalities.

4. CAD solutions for the image-level detection of glaucoma in CF and AMD in 3D OCT Volumes. Two strategies based on handcrafted features and deep learning have been explored for glaucoma classification. The irregularities in the Bruchs membrane due to AMD has been modeled as significant deviations from a Normative Atlas.

6.2 Future Directions

We briefly discuss few promising directions for future research based on this thesis below.

1. In Chapter 3, we explored a CRF framework for the joint extraction of the eight intra-retinal boundaries in OCT B-scans. Instead of handcrafting the Cost terms for the CRF individually, a Structural Support Vector Machine (SSVM) formulation was proposed to learn the appearance of the retinal tissue and relative weights of the shape priors in an end-to-end manner. While the proposed method performed well on healthy and AMD cases, it's performance degraded in the presence of large fluid-filled regions in the advanced Diabetic Macular Edema (DME) cases due to significant deviations from the expected shape prior. An explicit modelling of the fluid filled regions as a separate auxiliary boundary between the OPL/ONL and the IS/OS boundaries similar to [96] can be explored in the future within our CRF framework to address this issue. Another promising direction would be to incorporate Deep Learning to learn the Cost terms for the CRF in an end to end manner [179].

2. In Chapter 4, we proposed the RACE-net architecture which modeled the level set based deformable models within a deep learning framework. Shape priors play an important role in the segmentation of anatomical structures with indistinct boundaries, particularly in noisy medical imaging modalities such as ultrasound. Various methods have been explored to explicitly incorporate the shape information in the classical level set based deformable models [180]. Thus, extending the RACE-net architecture to incorporate shape priors presents a promising direction for future work.

3. In Chapter 5, we explored the strategy to model the structural changes as deviations from a Normative Atlas. The Normative Atlas was constructed by co-registering a set of healthy OCT volumes. Thereafter, the abnormalities in a given OCT volume were modeled as deviations from the Atlas characterized by its registration error to the Mean Intensity Template of the Atlas. We illustrated the viability of this strategy on the task of the image level detection of AMD. A promising direction for future work would be to extend the method to the more challenging task of the detection and localization of cysts (fluid-filled regions) in the OCT Volumes that characterize DME. This would require more accurate registration algorithms that can handle large deformations in the structure. Deep learning based registration methods such as the Spatial Transformer Networks [181], [182] can be explored for this purpose.

Appendix A

Conditional Random Field Inference

In Chapters 2 and 3, the simultaneous extraction of multiple boundaries of relevant anatomical structures in retinal images was posed as a novel Conditional Random Field (CRF) formulation. In both cases, the images were segmented by finding the set of boundaries corresponding to the minimum CRF energy using the Sequential Sequential Tree Re-weighted Message Passing (TRW-S) Algorithm [67]. Here, we discuss the details of the CRF inference problem and how it can be solved using the TRW-S algorithm.

First, the Conditional Random Field inference problem is described for an arbitary undirected graphical model (of clique size 2) in Section A.1. Next, the Belief Propagation Algorithm [183] is discussed in Section A.2 to solve the CRF inference for tree structured undirected graphs in an efficient manner. Finally, in Section A.3, the TRW-S algorithm is discussed as an extension of the belief propagation to handle undirected graphs with cycles .

A.1 Conditional Random Field

Let $G(V, E)$ represent an arbitary undirected graph where $V = \left\{v_1, v_2 ... v_{|V|}\right\}$ represents the set of $|V|$ nodes and $E = \left\{e_1, e_2, ... e_{|E|}\right\}$ represents the set of $|E|$ edges in G. Each edge $e_p \in E$ is defined by the pair of nodes (v_i, v_j) it connects $(v_i, v_j \in V)$. A discrete random variable x_n is defined for each node $v_n \in V$. Each x_n can be assigned a value from a discrete label set $L = \left\{l_1, l_2, ... l_{|L|}\right\}$ consisting of $|L|$ values with some probability. The set of all random variables $X = \{x_n | 1 \leq n \leq |V|\}$ forms a random field. A feasible labeling $\mathbf{x} \in L^{|V|}$ for the Random Field X can be obtained by assigning an arbitary label from L to each random variable x_n. In a CRF,

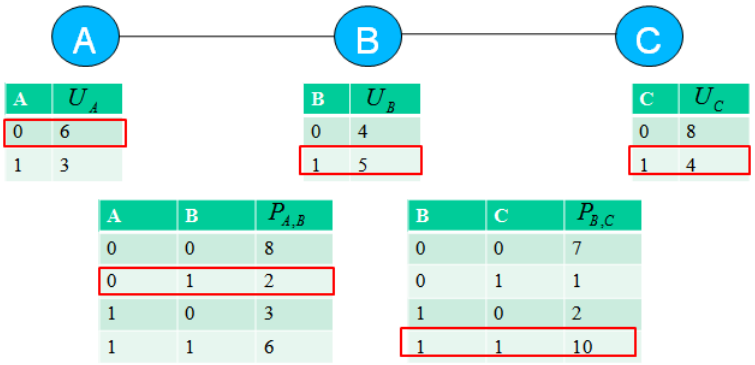

Figure A.1: An example of a Conditional Random Field with corresponding Unary and Pairwise cost tables.

an Energy $E(\mathbf{x})$ is assigned to each feasible labeling $\mathbf{x}$. The objective of the CRF inference optimization problem is to find the labeling $\mathbf{x}^*$ that minimizes E, ie., $\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}}\, E(\mathbf{x})$.

The CRF energy has a specific structure and satisfies the Markovian property which states that the labeling of a random variable x_n becomes independent of all other nodes in the graph, once the labelings of it's immdediate neighboring nodes are fixed, ie., $P(x_n|X) = P(x_n|\mathcal{N}(x_n))$, where $\mathcal{N}(x_n)$ represents the neighboring random variables of x_n in the graph. A Unary cost $\mathcal{U}_n(x_n)$ is associated with each node in the graph. Similarly, pairwise costs denoted by $\mathcal{P}_{i,j}(x_i, x_j)$ is associated with each edge $(v_i, v_j) \in E$. The total CRF energy is defined as the sum of all the Unary and Pairwise cost terms for a particular labeling,

$$E(\mathbf{x}) = \sum_{i=1}^{N} \mathcal{U}_n(x_n) + \sum_{(x_i,x_j)\in E} \mathcal{P}_{i,j}(x_i, x_j). \tag{A.1}$$

The above discussion is illustrated with an example in Fig. A.1. The undirected graphical model has $|V| = 3$ nodes where $V = \{A, B, C\}$. The random variables at each node can be assigned a binary label from the label set $L = \{0, 1\}$. A Unary cost table $\{\mathcal{U}_i | i = A, B, C\}$ is defined for each node in the graph. Similarly, pairwise cost tables are defined for each edge (A, B) and (B, C) (here $|E| = 2$) in the undirected graph. Each row in the Unary and Pairwise cost tables correspond to one of the possible label assignments to the nodes related to that table. Thus, in general, there are $|V|$ Unary cost tables, each with $|L|$ (2 in our example) rows. Similarly, there are $|E|$ (2 in our example) Pairwise cost tables, each with $|L| \times |L|$ rows ($2 \times 2 = 4$ in our example). The total CRF energy for a given label assignment is given by the sum of the corresponding entries in each of the Unary and Pairwise cost tables. For example,

the labeling $(A = 0, B = 1, C = 1)$ will have the energy $E(A = 0, B = 1, C = 1) = U_A(A = 0) + U_B(B = 1) + U_C(C = 1) + \mathcal{P}_{A,B}(A = 0, B = 1) + \mathcal{P}_{B,C}(B = 1, C = 1) = 6 + 5 + 4 + 2 + 10 = 27$. The CRF inference problem refers to finding the optimal label assignment for A, B, C which has the minimum value for $E(A, B, C)$. In general, a bruteforce algorithm would require the computation of the CRF energy for all possible $|L|^{|V|}$ label assignments and finding the minimum. However, this operation is computationally infeasible for large graphs and efficient algorithms such as TRW-S have been proposed to solve the CRF inference problem.

In practice, the Unary and Pairwise cost tables are task-specific and needs to be handcrafted or learned for a given task. For example, in Chapter 2, the equations 2.6 - 2.8 are employed to compute the Unary and Pairwise cost tables for the task of joint OD and OC segmentation in fundus images. Similarly, in Chapter 3, the equations 3.4 - 3.7 define the cost tables for the task of joint extraction of intra-retinal tissue layer boundaries in OCT B-scans. (The parameters in these equations such as the weights of the convolutional filters and the relative weights of the shape priors were learned in an end-to-end manner using a Structural Support Vector Formulation.) *The CRF inference algorithms only require the Unary, Pairwise cost tables and the graph connectivity information to be provided as input and donot require the initialization of the layer boundaries* (unlike deformable models).

A.2 Belief Propagation

The belief propagation algorithm mantains a $|L|$ length vector called a *message* for each edge in the graph. Initially, all *message* vectors are initialized to zero. The *message* from a node s to it's neighboring node t is denoted by $m_{s \to t}$ and is iteratively updated using the equation:

$$m_{s \to t}(x_t = l) = \min_{\mathbf{x_s}} \left\{ \mathcal{U}_s(x_s) + \mathcal{P}_{s,t}(x_s, x_t) + \sum_{u \in \mathcal{N}(s) - t} m_{u \to s}(x_s) \right\}, \tag{A.2}$$

where u denotes all nodes adjacent to s except t. The messages are computed iteratively until convergence. After convergence of the messages, a *belief* vector denoted by $\mathbf{b}_n$ is computed for each node x_n as

$$\mathbf{b}_n(x_n = l) = U_n(x_n = l) + \sum_{p \in \mathcal{N}(x_n)} m_{p \to n}(x_n = l), \tag{A.3}$$

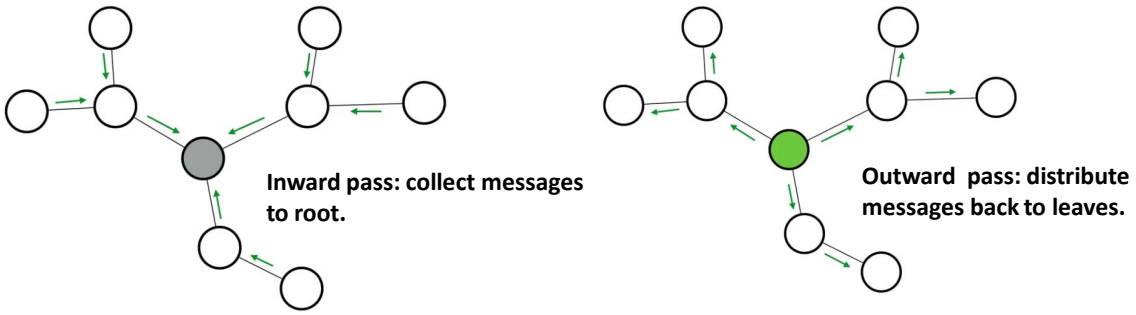

Figure A.2: Ordering of message updates for the Belief Propagation Algorithm for a tree structured graph.

where p denotes all nodes in the graph adjacent to the node n. The *belief* vector $\mathbf{b}_n$ is a $|L|$ length vector which indicates the cost of assigning a particular label to the node. Finally, each node is assigned the label corresponding to the minimum cost in it's *belief* vector.

The order in which the nodes are considered during the belief propagation algorithm is crucial for its fast convergence. In case of a tree structured graph, the Belief propagation algorithm is guaranteed to converge in just two passes of message updates over all the nodes (an inward followed by an outward pass). An arbitary node can be selected as the root node (since the graph is undirected). During the inward pass, the messages are updated from the leaf nodes towards the selected root node. This is followed by an outward pass where the messages are sequentially updated from the root node towards the leaf nodes. Thus, there are two message updates per edge, one during the inward and the other during the outward pass. An example of an inward and outward pass is illustrated in Fig. A.2. The belief propagation algorithm can also be applied to undirected graphs with cycles. However, in this case the algorithm may take multiple passes of message updates and the *message* vectors are not guaranteed to converge. Alternatively, more efficient algorithms such as TRW-S can be applied in such cases to guarantee convergence over loopy graphs.

A.3 Tree-Reweighted Message Passing -Sequential Algorithm

The TRW-S algorithm decomposes a given undirected loopy graph $G(V, E)$ into a set of sub-trees $\mathcal{T} = \{T_1(V_1, E_1), T_2(V_2, E_2)...T_M(V_M, E_M)\}$ such that $\left\{\bigcup_{i=1}^{M} V_i\right\} = V$ and $\left\{\bigcup_{i=1}^{M} E_i\right\} = E$. However, the same node or edge can occur in multiple T_i, ie., the sets V_i and E_i may not

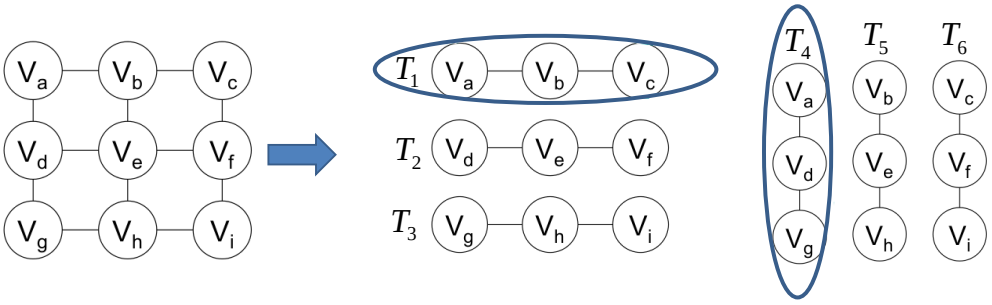

Figure A.3: A grid structured loopy graph with nine nodes is decomposed into 6 sub-trees: $\{T_1, T_2 ... T_6\}$. Each node is a member of two sub-trees and each edge occurs in a single sub-tree.

be mututally exclusive. The optimal label assignment for each sub-tree T_i can be efficiently computed using the belief propagation algorithm discussed above. Let $\mathcal{T}_n \subset \mathcal{T}$ denote the subset of all trees that contain the node v_n. Computing the optimal labeling for each sub-tree $T_i \in \mathcal{T}_n$ independently leads to multiple possible belief vectors b_n for each node v_n (one possible vector for each sub-tree T_i). The TRW-S algorithm employs an iterative strategy to ensure the label consistency across the different T_i.

We illustrate the TRW-S algorithm with an example. A grid structured loopy graph $G(V, E)$ (similar to the structure employed in our work) is depicted in Fig. A.3. It can be decomposed into a set of six sub-trees $\mathcal{T} = \{T_1, T_2 ... T_6\}$ along each row and column. The node v_a is shared by the set of trees $\mathcal{T}_a = \{T_1, T_4\}$.

The entries in the unary cost table $\mathcal{U}_n^{(i)}$ for each node v_n in the tree $T_i \in \mathcal{T}_n$ is initialized as $\mathcal{U}_n^{(i)} = \frac{1}{|\mathcal{T}_n|}.\mathcal{U}_n$, where $\mathcal{U}_n$ denotes the unary cost for v_n in the original graph $G(V, E)$ and $|\mathcal{T}_n|$ represents the number of trees in the set $\mathcal{T}_n$. This normalization ensures that the sum of the unary cost terms from all the sub-trees containing that node is equal to the unary cost terms of the original graph ie.,$\mathcal{U}_n = \sum_{T_i \in \mathcal{T}_n} \mathcal{U}_n^{(i)}$. A similar normalization is also performed for the pairwise terms corresponding to each edge in $G(V, E)$. Since, in our example each edge in $G(V, E)$ occurs in only one of the sub-trees, the pairwise cost tables for each T_i is identical to the corresponding pairwise tables in $G(V, E)$ while the unary cost values are halved as each node belongs to two sub-trees.

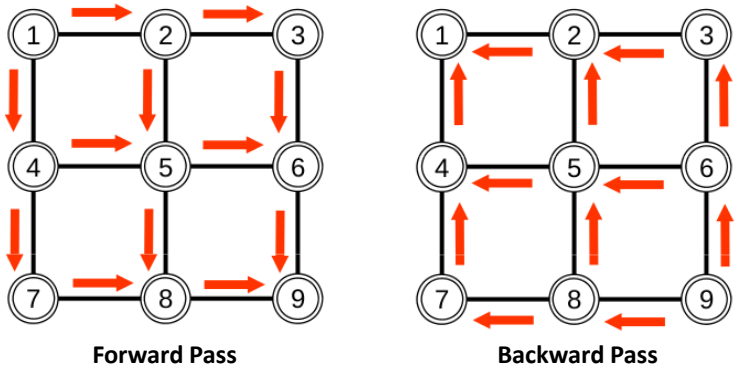

Figure A.4: The monotonic chain ordering of nodes during a forward and backward pass of the TRW-S algorithm. The numbering of each node indicates the order in which it is considered in the TRW-S algorithm.

The key steps of the TRW-S algorithm are as follows:

1. Consider a node v_n in $G(V, E)$ and find the set of trees $\mathcal{T}_n$ that contains v_n.
2. Apply belief propagation algorithm to each tree $T_i \in \mathcal{T}_n$ and compute the $|\mathcal{T}_n|$ belief vectors $\mathbf{b}_i$, one corresponding to each tree T_i.
3. Perform "node averaging" of the belief vectors as $\hat{\mathbf{b}}_n = \frac{1}{|\mathcal{T}_n|} \cdot \sum_{T_i \in \mathcal{T}_n} \mathbf{b}_i$.
4. Update the Unary cost table for v_n in each tree T_i by the average belief vector $\hat{\mathbf{b}}_n$ and repeat steps 1-3 until the convergence of $\hat{\mathbf{b}}_n$ for all nodes.

It has been shown in [67], that the TRW-S algorithm is guaranteed to converge if the selection of the nodes v_n in step 1 of the above algorithm follows a specific sequential order. The nodes are considered in a order called the monotonic chain. Once all the nodes have been visited, the ordering is reversed alternatively (called a forward and backward pass respectively). The ordering for the grid structured graphs during the forward and backward pass is depicted in Fig. A.4.

Appendix B

Structural Support Vector Machine

In Chapter 3, the extraction of the eight intra-retinal layer boundaries in Optical Coherence Tomography (OCT) images was posed as a Conditional Random Field (CRF). Instead of handcrafting the data and pairwise cost terms for the CRF individually, they were learned in an end-to-end manner by posing it as a Structural Support Vector Machine (SSVM) formulation (see Section 3.2.4). In this Appendix, we provide the details of the Block Co-ordinate Frank-Wolfe (BCFW) Algorithm which was employed to train the SSVM as proposed in [108].

B.1 The Structural Support Vector Machine formulation

Let $\left\{(\mathbf{I}^k, \mathbf{x}^k)\right\}_{k=1}^{M}$ denote a set of M training samples, where $\mathbf{I}^k$ is an input sample and $\mathbf{x}^k$ denotes the corresponding *optimal* structured output. Each $\mathbf{I}^k$ can be assigned an output vector $\mathbf{x}$ from a feasible set of output vectors denoted by $\mathcal{Y}_k$ with some loss defined by the function $\Delta(\mathbf{x}^{(k)}, \mathbf{x})$. The joint feature function $F(\mathbf{I}, \mathbf{x})$ provides a d dimension feature representation that captures the relevant information between any input-output pair $(\mathbf{I}, \mathbf{x})$. The goal of structural support vector machine training is to find a linear hyperplane parameterized by θ such that:

$$\begin{aligned} \underset{\theta, \xi \geq 0}{\operatorname{argmin}} \quad & \frac{\lambda}{2} \mid\mid \theta \mid\mid^2 + \frac{1}{M} \sum_{k=1}^{M} \xi_k \\ \text{s.t.} \quad & \theta^\top . \Psi_k(\mathbf{x}) \geq \Delta(\mathbf{x}^{(k)}, \mathbf{x}) - \xi_k \quad \forall k, \forall \mathbf{x} \in \mathcal{Y}_k, \end{aligned} \tag{B.1}$$

where $\Psi_k(\mathbf{x}) = F(\mathbf{x}, \mathbf{I}^k) - F(\mathbf{x}^k, \mathbf{I}^k)$and ξ_k are the slack variables. For each training sample, there are an exponentially large number of feasible output vectors in $\mathcal{Y}_k$ resulting in a total of $c =$

$\sum_k |\mathcal{Y}_k|$) constraints. Although eq. B.1 is a convex Quadratic Programming (QP) optimization problem, it cannot be solved directly due to the extremely large number of constraints.

The constraints in eq. B.1 for each I_k can be rearranged as $\xi_k \geq \Delta(\mathbf{x}^{(k)}, \mathbf{x}) - \theta^\top.\Psi_k(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{Y}_k$. Each of these $|\mathcal{Y}_k|$ constraints can be replaced by a single *most violating constraint* of the form $\xi_k = H_k(\mathbf{x}, \theta)$, where H_k is obtained by solving the *max oracle* optimization problem $H_k(\mathbf{x}, \theta) = \underset{\mathbf{x}}{\text{argmax}} \quad \Delta(\mathbf{x}^{(k)}, \mathbf{x}) - \theta^\top.\Psi_k(\mathbf{x})$. By substituting $\xi_k = H_k(\mathbf{x}, \theta)$ in the objective function of eq. B.1, we get an equivalent unconstrained optimization problem,

$$\underset{\theta}{\text{argmin}} \qquad \frac{\lambda}{2} \mid\mid \theta \mid\mid^2 + \frac{1}{M} \sum_{k=1}^{M} H_k(\mathbf{x}). \tag{B.2}$$

B.1.1 The dual optimization problem

The Lagrange dual of the formulation in eq. B.2 is given by

$$\begin{aligned} &\underset{\alpha \in \mathbb{R}^c, \alpha \geq 0}{\text{argmin}} && \frac{\lambda}{2} \mid\mid A\alpha \mid\mid^2 - b^\top.\alpha \\ &\text{s.t.} && \sum_{y \in \mathcal{Y}_k} \alpha_k(\mathbf{x}) = 1, \forall k, \forall x \in \mathcal{Y}_k. \end{aligned} \tag{B.3}$$

Here, $\alpha \in \mathbb{R}^c$ is the dual variable with $c = \sum_{k=1}^{M} |\mathcal{Y}_k|$ dimensions, one corresponding to each constraint in B.1. The matrix $A \in \mathbb{R}^{d \times C}$ and the vector $b \in \mathbb{R}^C$ respectively. Using the KKT optimality conditions, the relationship between the primal and dual variables can be derived to be $\theta = A.\alpha$. (A detailed derivation is provided in [108].) The eq. B.3 can be re-written as

$$\begin{aligned} &\underset{\alpha_k \in \mathbb{R}^{|\mathcal{Y}_k|}, \alpha_k \geq 0, \forall k}{\text{argmin}} && \frac{\lambda}{2} \mid\mid \sum_{k=1}^{M} A_k(\mathbf{x})\alpha_k(\mathbf{x}) \mid\mid^2 - \sum_{k=1}^{M} b_k^\top(\mathbf{x}).\alpha_k(\mathbf{x}) \\ &\text{s.t.} && \sum_{\mathbf{x} \in \mathcal{Y}_k} \alpha_k(\mathbf{x}) = 1, \forall k, \forall x \in \mathcal{Y}_k, \end{aligned} \tag{B.4}$$

where $\alpha = [\alpha_1(\mathbf{x}) \quad \alpha_2(\mathbf{x})...\alpha_k(\mathbf{x})...\alpha_M(\mathbf{x})]$ and each $\alpha_k(\mathbf{x}) \in \mathbb{R}^{|\mathcal{Y}_k|}$ corresponds to the image $\mathbf{I}_k$. Similarly, $b = [b_1(\mathbf{x})b_2(\mathbf{x})...b_k(\mathbf{x})...b_M(\mathbf{x})]$, where each $b_k(\mathbf{x}) \in \mathbb{R}^{|\mathcal{Y}_k|}$ is of the form $\left[\frac{1}{M}\Delta(\mathbf{x}^{(k)}, \mathbf{x}) | \forall \mathbf{x} \in \mathcal{Y}_k\right]$. The matrix A can be grouped into $A = [A_1(\mathbf{x})A_2(\mathbf{x})...A_k(\mathbf{x})...A_M(\mathbf{x})]$ where each submatrix $A_k(\mathbf{x}) \in \mathbb{R}^{d \times |\mathcal{Y}_k|}$ contains the columns of form $\left\{\frac{1}{\lambda.M}\Psi_k(\mathbf{x}) | \forall x \in \mathcal{Y}_k\right\}$ corresponding to $\mathbf{I}_k$.

B.2 The Frank-Wolfe Optimization Algorithm

An efficient Block Coordinate version of the Franke-Wolfe (BCFW) algorithm was proposed in [108] to optimize eq. B.4. The BCFW is an iterative method where at each iteration, the function is minimized (using a single Frank-Wolfe update step) along a subset of dimensions $\alpha_k(\mathbf{x})$ corresponding to a randomly selected image $\mathbf{I}_k$ while keeping the other dimensions fixed. Since all α except $\alpha_k(\mathbf{x})$ is fixed at a given iteration, the optimization problem in eq. B.4 reduces to

$$\begin{aligned} \underset{\alpha_k \geq 0}{\text{argmin}} \quad & \frac{\lambda}{2} \mid\mid A_k(\mathbf{x})\alpha_k(\mathbf{x}) \mid\mid^2 - b_k^\top(\mathbf{x}).\alpha_k(\mathbf{x}) \\ \text{s.t.} \quad & \sum_{\mathbf{x} \in \mathcal{Y}_k} \alpha_k(\mathbf{x}) = 1, \forall x \in \mathcal{Y}_k. \end{aligned} \tag{B.5}$$

The Frank-Wolfe based update step can be applied to the above eq. since the objective function is convex (of quadratic form) and the constraints define a bounded feasible set. Let $f(\alpha_k) = \frac{\lambda}{2} \mid\mid A_k(\mathbf{x})\alpha_k(\mathbf{x}) \mid\mid^2 - b_k^\top(\mathbf{x}).\alpha_k(\mathbf{x})$ denote the objective function in eq. B.5. At each iteration t, $\alpha_k^{(t)}$ is updated to $\alpha_k^{(t+1)}$ as follows:

1. **Find update direction:** A local linear approximation of $f(\alpha_k)$ around the point $\alpha_k^{(t)}$ is obtained by the hyperplane $\mathbf{s}_k^\top.\nabla f(\alpha_k)$. Let $\mathbf{s}_k^{(t)}$ denote the point within the feasible region of α_k that minimizes the value on this hyperplane. Mathematically, $\mathbf{s}_k^{(t)} = \underset{\mathbf{s_k}}{\text{argmin}} \quad \mathbf{s}_k^\top.\nabla f(\alpha_k) \quad s.t. \quad \sum_{y \in \mathcal{Y}_k} \alpha_k(\mathbf{x}) = 1$. Since, the objective function as well as the constraints for this optimization problem is linear, it can be solved using an LP solver. The direction of the update for $\alpha_k^{(t+1)}$ is along the line joining the points $\alpha_k^{(t)}$ and $\mathbf{s}_t$.

2. **Compute optimal step size:** The optial step-size γ for the update can be found by searching for the point along the line joining the points $\alpha_k^{(t)}$ and $\mathbf{s}_k^{(t)}$ that has the minimum value for $f(\alpha_k)$, ie., $\gamma = \underset{\gamma \in [0,1]}{\text{argmin}} \quad \gamma f\left((1-\gamma).\alpha_k^{(t)} + \gamma.s_k^{(t)}\right)$. In the case of eq. B.5, the optimal step size has a closed form analytical solution given by $\gamma_{opt} = \frac{\left\langle \alpha_{\mathbf{k}}^{(\mathbf{t})} - \mathbf{s}_k^{(t)}, \nabla f(\alpha_k^{(t)}) \right\rangle}{\lambda \mid\mid A(\alpha_k^{(t)} - \mathbf{s}_k^{(t)}) \mid\mid^2}$

3. **Update $\alpha_k^{(t)}$:** $\alpha_k^{(t+1)} = (1-\gamma).\alpha_k^{(t)} + \gamma.s_k^{(t)}$.

4. Repeat steps 1-3 until convergence or the maximum number of iterations is reached.

However, the above algorithm cannot be directly implemented in the dual space due to the exponentially large number of dimensions $|\mathcal{Y}_k|$ (see Section B.1.1 above) of the vectors

α_k, $\mathbf{b}_k$ and the number of columns in the matrix A_k. To address this issue, the equivalent update equations for the primal variable were derived in [108] using the relation $\theta = A.\alpha$ and an efficient algorithm called the Primal-Dual Block Co-ordinate Frank-Wolfe method (see Algorithm 2) was proposed. Although this algorithm optimizes the dual of SSVM formulation, only the primal variables are mantained explicitly during implementation. In Algorithm 2, the required parameters θ is updated iteratively while an additional set of vectors l_k and θ_k are mantained for each training sample separately for efficient computation.

Algorithm 2: Primal-Dual Frank-Wolfe Algorithm for Structural SVM

Initialize $\theta^{(\mathbf{0})}$ to a random vector and $l^{(0)} = 0$.

Initialize vectors $\theta_k^{(0)} = 0$ and $l_k^{(0)} = 0$ for each training sample $(\mathbf{I}_k, \mathbf{x}_k)$

for $t \leftarrow 1$ **to** T **do**

  Pick an image $\mathbf{I}_k$ at random from $\{\mathbf{I}_1, \mathbf{I}_2, ...\mathbf{I}_M\}$.

  $\mathbf{x_k^*} = H_k(\mathbf{x}, \theta_t)$. `// most violating constraint`

  $\theta_s = \frac{1}{\lambda.M}\psi_k(\mathbf{x_k^*})$ and $l_s = \frac{1}{M}\Delta(\mathbf{x_k^*}, \mathbf{x})$.

  $\gamma = \frac{\lambda(\theta_k - \theta_s)^\top \theta^{(t)} - l_k^{(t)} + l_s}{\lambda ||\theta_k^{(t)} - \theta_s||^2}$ and clip to [0,1].

  Update $\theta_k^{(t+1)} = (1-\gamma)\theta_k^{(t)} + \gamma\theta_s$ and $l_k^{(t+1)} = (1-\gamma)l_k^{(t)} + \gamma l_s$

  Update $\theta^{(t+1)} = \theta^{(t)} + \theta_k^{t+1} - \theta_k^{(t)}$ and $l^{(t+1)} = l^{(t)} + l_k^{t+1} - l_k^{(t)}$

end

Appendix C

Level Set based Deformable Models

Deformable Model based Segmentation

The Recurrent Active Contour Evolution Network (RACE-net) described in Chapter 4 aims to model the level set equation $\frac{\partial \phi}{\partial t} = \alpha_1.g(I,\phi).|\nabla\phi| + \alpha_2.h(I).\kappa|\nabla\phi|$ (see eq. 4.3 in Chapter 4 for details) which provides a generalized equation for the geometric active contours. The equation was dervied in two steps. First, the Partial Differential Equation (PDE) required to minimize the Energy functional in eq. 4.1 was derived in eq. 4.2 by using the Euler-Lagrange of the constituent boundary and regional cost terms. In the second step, the PDE in eq. 4.2 was converted into the level set representation.

In this Appendix, a detailed derivation for the Euler-Lagrange equations for the boundary and the regional cost terms is presented in Sections C.1 and C.2 respectively. These equations were employed to obtain the curve evolution PDE in eq. 4.2. The derivations are based on [184].

In Section C.3, we show that a PDE of form $\frac{\partial C}{\partial t} = V.\mathbf{n}$ in the parametric curve representation is equivalent to the level set equation $\frac{\partial \phi}{\partial t} = V.|\nabla\phi|$ in the level set representation. This was employed in Chapter 4 to convert the PDE in eq. 4.2 to its corresponding level set equation presented in eq. 4.3. We refer the readers to [131] for further details.

C.1 E-L for the Boundary Cost

Let $C(s)$ denote a parametric representation of a curve. In this section, we derive the Euler-Lagrange for the energy functional $E_l(C(s)) = \int_C h\, ds$ where $h(x,y)$ is a scalar function defined

over the image domain. By the definition of line integral,

$$\int_C h ds = \int_0^l h(x,y).|C_s| ds \quad \text{where} \quad |C_s| = \sqrt{x_s^2 + y_s^2}. \tag{i}$$

x_s and y_s represent the partial derivatives $\frac{\partial x}{\partial s}$and $\frac{\partial y}{\partial s}$ respectively. *This shorthand notation for the partial derivatives has been used throught this document.* The Euler-Lagrange equation is given by

$$\begin{aligned} \frac{\partial g}{\partial x} - \frac{d}{ds}\left(\frac{\partial g}{\partial x_s}\right) = 0 \\ \frac{\partial g}{\partial y} - \frac{d}{ds}\left(\frac{\partial g}{\partial y_s}\right) = 0. \end{aligned} \tag{ii}$$

In this case, $g = \left[h(x,y).\left(x_s^2 + y_s^2\right)^{\frac{1}{2}}\right]$. Next, we show the details of the derivative with respect to x (first row of eq. (ii)) and the derivatives with respect to y can be computed in a similar manner. The first term [1],

$$\begin{aligned} \frac{\partial g}{\partial x} &= \frac{\partial}{\partial x}\left[h(x,y).\left(x_s^2 + y_s^2\right)^{\frac{1}{2}}\right] = h(x,y).\frac{\partial}{\partial x}\left(x_s^2 + y_s^2\right)^{\frac{1}{2}} + \left(x_s^2 + y_s^2\right)^{\frac{1}{2}}.\frac{\partial}{\partial x}h(x,y) \\ &= 0 + \left(x_s^2 + y_s^2\right)^{\frac{1}{2}}.\frac{\partial}{\partial x}h(x,y) = |C_s|.\frac{\partial}{\partial x}h(x,y) \end{aligned} \tag{iii}$$

Now, $\frac{\partial g}{\partial x_s}$ in the second term of eq. (ii) can be simplified as follows:

$$\begin{aligned} \frac{\partial g}{\partial x_s} &= \frac{\partial}{\partial x_s}\left[h(x,y).\left(x_s^2 + y_s^2\right)^{\frac{1}{2}}\right] = h(x,y).\frac{\partial}{\partial x_s}\left(x_s^2 + y_s^2\right)^{\frac{1}{2}} + \left(x_s^2 + y_s^2\right)^{\frac{1}{2}}.\frac{\partial}{\partial x_s}h(x,y) \\ &= h(x,y).\frac{\partial}{\partial x_s}\left(x_s^2 + y_s^2\right)^{\frac{1}{2}} + 0 = h(x,y).\frac{x_s}{\left(x_s^2 + y_s^2\right)^{\frac{1}{2}}} = h(x,y).\frac{x_s}{|C_s|} \end{aligned} \tag{iv}$$

By substituting eq. (iv) we get,

$$\begin{aligned} \frac{d}{ds}\left(\frac{\partial g}{\partial x_s}\right) &= \frac{d}{ds}\left[h(x,y).\frac{x_s}{|C_s|}\right] \\ &= h(x,y).\frac{d}{ds}\left(\frac{x_s}{|C_s|}\right) + \left\{\frac{d}{ds}h(x,y)\right\}.\frac{x_s}{|C_s|} \end{aligned} \tag{v}$$

Next, we simplify each term in the R.H.S of eq. (v) individually. The first term can be simplified by substituting the definition of curvature $\kappa = \frac{(x_s.y_{ss} - x_{ss}.y_s)}{(x_s^2 + y_s^2)^{\frac{3}{2}}}$ as follows

$$\begin{aligned} h(x,y).\frac{d}{ds}\left(\frac{x_s}{|C_s|}\right) &= h(x,y).\frac{d}{ds}\left(\frac{x_s}{\left(x_s^2 + y_s^2\right)^{\frac{1}{2}}}\right) \\ &= h(x,y).\left(\frac{-y_s(x_s.y_{ss} - x_{ss}.y_s)}{\left(x_s^2 + y_s^2\right)^{\frac{3}{2}}}\right) = -h.y_s.\kappa \end{aligned} \tag{vi}$$

[1] During the computation of the partial derivatives, x and x_s have to be treated as independent variables.

The second term in eq. (v) can be simplified by applying the chain rule as

$$\left\{\frac{d}{ds}h(x,y)\right\}.\frac{x_s}{|C_s|} = \frac{x_s}{|C_s|}.\left[\left(\frac{\partial h}{\partial x}.\frac{\partial x}{\partial s}\right)+\left(\frac{\partial h}{\partial y}.\frac{\partial y}{\partial s}\right)\right] = \frac{x_s}{|C_s|}.\langle\nabla h,\mathbf{T}\rangle\,, \tag{vii}$$

where $\mathbf{T} = [\frac{\partial x}{\partial s}, \frac{\partial y}{\partial s}]$ is the tanget vector and the gradient $\nabla h = \left[\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y}\right]$. Finally, the Euler-Lagrange with respect to x is obtained by substituting the values from eq. (iii), (v), (vi) and (viii) into eq. (ii),

$$|C_s|.\frac{\partial}{\partial x}h(x,y) - \frac{x_s}{|C_s|}.\langle\nabla h,\mathbf{T}\rangle + h\kappa y_s = 0. \tag{viii}$$

In a similar way, the Euler-Lagrange with respect to y can be shown to be

$$|C_s|.\frac{\partial}{\partial y}h(x,y) - \frac{y_s}{|C_s|}.\langle\nabla h,\mathbf{T}\rangle + h\kappa.(-x_s) = 0 \tag{ix}$$

The functional derivative obtained in eq. (viii) and (ix) can be represented in the vector notation as

$$|C_s|.\nabla h(x,y) - \frac{\langle\nabla h,\mathbf{T}\rangle}{|C_s|}.\mathbf{T} + h(x,y).\kappa.\mathbf{n}. \tag{x}$$

Let a PDE $\frac{\partial C(s)}{\partial t} = -V(s,t)$ represent a curve evolution where $V(s,t)$ is obtained by the Euler-Lagrange of an Energy functional. The velocity $V(s,t)$ can be projected onto the normal $\mathbf{n}$ and tangent $\mathbf{T}$ vectors of $C(s)$ at s. The normal component alone is responsible for the deformation of the curve during evolution while the tangential component only moves the points along the curve. The second term in eq. (x) only has a tangential component and hence can be neglected. Similarly, the first term should be projected onto its normal component. Moreover, by an intelligent parameterization of $C(s)$, by the arc length, ie., $s \in [0,l]$ where l denotes the length of the curve, $|C_s| = 1$. Under these assumptions, eq. (x) reduces to

$$\nabla h(x,y).\mathbf{n} + h(x,y).\kappa.\mathbf{n} \tag{xi}$$

C.2 E-L for the Regional Cost

In this section, we derive the Euler Lagrange for the functional $E_{area} = \int_{R_C} q(x,y)dxdy$ to minimize the sum of the function q over all points within the area enclosed by the curve C. The region integration is converted into line integrations along its enclosing boundary by applying

the Green's theorem which states that $\int_{R_C}\left(\frac{\partial P}{\partial x}\right) - \left(\frac{\partial Q}{\partial y}\right) dxdy = \int_C Pdx + Qdy$. In order to arrange E_{area} into the required form, it is decomposed as

$$\int_{R_C} q(x,y)dxdy = \int_{R_C} \left\{ \left(\frac{1}{2}q(x,y)\right) - \left(-\frac{1}{2}q(x,y)\right) \right\} dxdy \tag{xii}$$

Let $Q = \frac{1}{2}\int_0^x q(z,y)dz$. Then, $\frac{\partial Q}{\partial x} = \frac{1}{2}\frac{\partial}{\partial x}\left[\int_0^x q(z,y)dz\right] = \frac{1}{2}q(x,y)$. This is based on the second fundamental theorem of calculus. Similarly, if $P = -\frac{1}{2}\int_0^y q(x,z)dz$, it can be shown that $\frac{\partial P}{\partial y} = -\frac{1}{2}q(x,y)$. Substituting these values in eq. xii and applying the Green's theorem,

$$\begin{aligned} \int_{R_C} q(x,y)dxdy &= \int_{R_C}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dxdy = \int_C Pdx + Qdy \\ &= \int_0^l \left(P\frac{dx}{ds} + Q\frac{dy}{ds}\right)ds = \int_0^l \left(P.x_s + Q.y_s\right)ds \end{aligned} \tag{xiii}$$

The Euler-Lagrange is obtained by substituting $g = P(x,y).x_s(s) + Q(x,y).y_s$ in eq. (ii). Now, $\frac{\partial g}{\partial x} = x_s.\frac{\partial P}{\partial x} + y_s.\frac{\partial Q}{\partial x}$ and $\frac{\partial g}{\partial x_s} = P$.

$$\begin{aligned} \frac{\partial g}{\partial x} - \frac{d}{ds}\left(\frac{\partial g}{\partial x_s}\right) &= 0 \\ \Rightarrow x_s.\frac{\partial P(x,y)}{\partial x} + y_s.\frac{\partial Q(x,y)}{\partial x} - \frac{d}{ds}(P(x,y)) &= 0 \\ \Rightarrow \left(x_s.\frac{\partial P(x,y)}{\partial x} - \frac{d}{ds}(P(x,y))\right) + y_s.\frac{\partial Q(x,y)}{\partial x} &= 0 \end{aligned} \tag{xiv}$$

Applying Chain rule, $\frac{d}{ds}(P(x,y)) = \frac{\partial P}{\partial x}.\frac{\partial x}{\partial s} + \frac{\partial P}{\partial y}.\frac{\partial y}{\partial s} = \frac{\partial P}{\partial x}.x_s + \frac{\partial P}{\partial y}.y_s$. Substituting the value in eq. (xiv), we get

$$\begin{aligned} \left(x_s.\frac{\partial P}{\partial x} - \frac{\partial P}{\partial x}.x_s - \frac{\partial P}{\partial y}.y_s\right) + y_s.\frac{\partial Q}{\partial x} &= 0 \\ \Rightarrow \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right).y_s &= 0 \\ \Rightarrow q(x,y).y_s &= 0 \end{aligned} \tag{xv}$$

Similarly, we can show that

$$\frac{\partial g}{\partial y} - \frac{d}{ds}\left(\frac{\partial g}{\partial y_s}\right) = \left(-\frac{\partial Q}{\partial x} + \frac{\partial P}{\partial y}\right).x_s = q(x,y). - x_s = 0 \tag{xvi}$$

The vector $\mathbf{n} = [y_s, -x_s]$ is the unit normal vector at $(x(s), y(s))$ since its dot product with the unit tangent vector given by $\mathbf{T} = [x_s, y_s]$ is 0. Thus, using the vector representation, the functional derivatives in eq. (xv) and (xvi) can be re-written as $q(x,y).\mathbf{n}$.

C.3 Curve Evolution in Level Set Representation

In this section we derive the level set representation of the curve evolution from its parametric representation. Let $\frac{\partial C}{\partial t} = V(x,y).\mathbf{n}$ represent a curve evolution PDE, where $V(x,y)$ represents a scalar velocity field defined over the image domain. Let ϕ denote the corresponding level set function for the curve C.

By definition, along the boundary curve, $\phi(x,y;t) = 0$. Taking derivative with respect to t on both sides we get,

$$
\begin{aligned}
& \frac{d}{dt}\phi(x,y,t) = 0 \\
\Rightarrow \quad & \phi_x.x_t + \phi_y.y_t + \phi_t = 0 \qquad \text{... by applying chain rule} \\
\Rightarrow \quad & \left\langle (\phi_x,\phi_y)^T, (x_t,y_t)^T \right\rangle = -\phi_t \qquad \text{... using the vector notation of dot product} \\
\Rightarrow \quad & -\phi_t = \langle \nabla\phi, C_t \rangle \qquad \text{... substituting } \nabla\phi = (\phi_x,\phi_y)^T \text{ and } C_t = (x_t,y_t)^T \\
\Rightarrow \quad & -\phi_t = \langle \nabla\phi, V.\mathbf{n} \rangle \qquad \text{... substituting } C_t = V.\mathbf{n} \\
\Rightarrow \quad & -\phi_t = \left\langle \nabla\phi, V.-\frac{\nabla\phi}{\mid \nabla\phi \mid} \right\rangle \qquad \text{... in level set representation, } \mathbf{n} = -\frac{\nabla\phi}{\mid \nabla\phi \mid} \\
\Rightarrow \quad & -\phi_t = -V.\left\langle \nabla\phi, \frac{\nabla\phi}{\mid \nabla\phi \mid} \right\rangle \\
\Rightarrow \quad & \phi_t = V. \mid \nabla\phi \mid \qquad \text{(xvii)}
\end{aligned}
$$

Thus, given a curve evolution PDE $\frac{\partial C}{\partial t} = V(x,y).\mathbf{n}$, the corresponding level set equation is given by $\frac{\partial \phi}{\partial t} = V(x,y) \mid \nabla\phi \mid$.

Appendix D

Public Datasets

We would like to acknowledge the creators of the following public datasets which have been used throughout this thesis.

- INSPIRE [60] (https://medicine.uiowa.edu/eye/inspire-datasets).
- RIM-ONE [68] (http://medimrg.webs.ull.es/).
- DRIONS-DB [69] (http://www.ia.uned.es/~ejcarmona/DRIONS-DB.html).
- MESSIDOR [5] (http://www.adcis.net/en/third-party/messidor/)with Ground Truth segmentation masks provided by University of Huelva [70] (http://uhu.es/retinopathy/muestras2.php).
- NORMAL-1 [13] provided by Prof. Sina Farsiu (sina.farsiu@duke.edu) upon request.
- NORMAL-2 (OCTRIMA) [89] provided by Prof. Delia Cabrera DeBuc (DCabrera2@med.miami.edu) upon request.
- AMD-1 [14] (http://people.duke.edu/~sf59/Chiu_IOVS_2011_dataset.htm).
- DME-1 [93] (http://people.duke.edu/~sf59/Chiu_BOE_2014_dataset.htm).
- AMD-3D [166] (http://people.duke.edu/~sf59/RPEDC_Ophth_2013_dataset.htm).
- UCSB Bio-segmentation Benchmark [140] (https://bioimage.ucsb.edu/research/bio-segmentation).
- STACOM 2013 MRI left atrium segmentation dataset [146] (https://www.cardiacatlas.org/challenges/left-atrium-segmentation-challenge/).
- REFUGE [163] (https://refuge.grand-challenge.org/).

Publications

Part of the work described in this thesis has previously been presented in the following publications.

Journal

1. **Arunava Chakravarty**, Jayanthi Sivaswamy, "RACE-net: A Recurrent Neural Network for Biomedical Image Segmentation", *IEEE journal of biomedical and health informatics*, IEEE, vol. 23, no. 3, pp. 1151-1162, May 2019.

2. **Arunava Chakravarty**, Jayanthi Sivaswamy, "A Supervised Joint Multi-layer Segmentation Framework for Retinal Optical Coherence Tomography Images using Conditional Random Field", *Computer Methods and Programs in Biomedicine*, Elsevier, vol 165, pp. 235-250, 2018.

3. **Arunava Chakravarty**, Jayanthi Sivaswamy, "Joint optic disc and cup boundary extraction from monocular fundus images", *Computer Methods and Programs in Biomedicine*, Elsevier, vol 147, pp, 51-61, 2017.

Conference

4. **Arunava Chakravarty**, Divya Jyothi Gaddipati, Jayanthi Sivaswamy, "Construction of a Retinal Atlas for Macular OCT Volumes", *International Conference Image Analysis and Recognition (ICIAR)*, Povoa de Varzim, Portugal, Springer Cham LNCS, vol. 10882, pp. 650-658, 2018.

5. **Arunava Chakravarty**, Jayanthi Sivaswamy, "End-to-End Learning of a Conditional Random Field for Intra-retinal Layer Segmentation in Optical Coherence Tomography", *Annual Conference on Medical Image Understanding and Analysis (MIUA)*, Edinburgh, United Kingdom, Springer Cham CCIS,vol. 723, pp. 3-14, 2017.

6. **Arunava Chakravarty**, Jayanthi Sivaswamy, "Glaucoma Classification with a Fusion of Segmentation and Image-based Features", *IEEE International Symposium on Biomedical Imaging (ISBI)*, Prague, Czech Republic, IEEE, pp. 689-692, 2016.

7. **Arunava Chakravarty**, Jayanthi Sivaswamy, "Coupled sparse dictionary for depth-based cup segmentation from single color fundus image", *International Conference on Medical Image Com-*

puting and Computer-Assisted Intervention (MICCAI), Massachusetts, USA, Springer Cham LNCS, vol. 8673, pp. 747-754, 2014.

8. **Arunava Chakravarty**, Jayanthi Sivaswamy, "A Deep Learning based Joint Segmentation and Classification Framework for Glaucoma Assesment in Retinal Color Fundus Images, arxiv pre-print http://arxiv.org/abs/1808.01355, 2018.

Other publications broadly related to the thesis area are as follows:

Journal

9. Lipi Chakrabarty, Gopal Datt Joshi, **Arunava Chakravarty**, Ganesh V Raman, S.R. Krishnadas, Jayanthi Sivaswamy, "Automated detection of glaucoma from topographic features of the optic nerve head in color fundus photographs", *Journal of glaucoma*, Wolters Kluwer, vol. 25(7), pp. 590-597, 2016.

10. Jayanthi Sivaswamy, S Krishnadas, **Arunava Chakravarty**, Gopal Datt Joshi, A. Syed Tabish , "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis", *JSM Biomedical Imaging Data Papers*, JSciMed Central, 2015.

Conference

11. Ujjwal, **Arunava Chakravarty**, Jayanthi Sivaswamy, "An assistive annotation system for retinal images", *IEEE International Symposium on Biomedical Imaging (ISBI)*, New York, USA IEEE, pp. 1506-1509, 2015.

12. Ujjwal, K. Sai Deepak, **Arunava Chakravarty**, Jayanthi Sivaswamy, "Visual saliency based bright lesion detection and discrimination in retinal images", *IEEE International Symposium on Biomedical Imaging (ISBI)*, San Francisco, USA, IEEE, pp. 1436-1439, 2013.

13. Mark JJP van Grinsven, **Arunava Chakravarty**, Jayanthi Sivaswamy, Thomas Theelen, Bram van Ginneken, Clara I Sanchez, "A Bag of Words approach for discriminating between retinal images containing exudates or drusen.", *IEEE International Symposium on Biomedical Imaging (ISBI)*, San Francisco, USA, IEEE, pp. 1444-1447, 2013.

14. **Arunava Chakravarty**, Jayanthi Sivaswamy, "A novel approach for quantification of retinal vessel tortuosity using quadratic polynomial decomposition", *Indian Conference on Medical Informatics and Telemedicine (ICMIT)*, IIT Kharagpur, IEEE, pp. 7-12, 2013.

Bibliography

[1] Rupert RA Bourne, Seth R Flaxman, Tasanee Braithwaite, Maria V Cicinelli, Aditi Das, Jost B Jonas, Jill Keeffe, John H Kempen, Janet Leasher, Hans Limburg, et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health*, 5(9):e888–e897, 2017.

[2] Donatella Pascolini and Silvio Paolo Mariotti. Global estimates of visual impairment: 2010. *British Journal of Ophthalmology*, 96(5):614–618, 2012.

[3] V Thulasi Bai, V Murali, R Kim, and SK Srivatsa. Teleophthalmology-based rural eye care in india. *Telemedicine and e-Health*, 13(3):313–321, 2007.

[4] Yu Cui Neilsen De Souza, Stephanie Looi, Prakash Paudel, Lakshmi Shinde, Krishna Kumar, Rajbir Berwal, Rajesh Wadhwa, Vinod Daniel, Judith Flanagan, and Brien Holden. The role of optometrists in india: An integral part of an eye health team. *Indian journal of ophthalmology*, 60 (5):401, 2012.

[5] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, Béatrice Charton, and Jean-Claude Klein. Feedback on a publicly distributed image database: The messidor database. *Image Analysis and Stereology*, 33(3):231–234, 2014. URL http://www.adcis.net/en/Download-Third-Party/Messidor.html.

[6] Jayanthi Sivaswamy, SR Krishnadas, Arunava Chakravarty, GD Joshi, A Syed Tabish, et al. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomedical Imaging Data Papers*, 2(1):1004, 2015.

[7] Maximilian WM Wintergerst, Christian K Brinkmann, Frank G Holz, and Robert P Finger. Undilated versus dilated monoscopic smartphone-based fundus photography for optic nerve head evaluation. *Scientific reports*, 8(1):10228, 2018.

[8] Kai Jin, Haitong Lu, Zhaoan Su, Chuming Cheng, Juan Ye, and Dahong Qian. Telemedicine screening of retinal diseases with a handheld portable non-mydriatic fundus camera. *BMC ophthalmology*, 17(1):89, 2017.

[9] Michael David Abràmoff, Yiyue Lou, Ali Erginay, Warren Clarida, Ryan Amelon, James C Folk, and Meindert Niemeijer. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative ophthalmology & visual science*, 57(13):5200–5206, 2016.

[10] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[11] Yehui Yang, Tao Li, Wensi Li, Haishan Wu, Wei Fan, and Wensheng Zhang. Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 533–540. Springer, 2017.

[12] Raheleh Kafieh, Hossein Rabbani, and Saeed Kermani. A review of algorithms for segmentation of optical coherence tomography from retina. *Journal of Medical Signals and Sensors*, 3(1):45, 2013.

[13] Stephanie J Chiu, Xiao T Li, Peter Nicholas, Cynthia A Toth, Joseph A Izatt, and Sina Farsiu. Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation. *Optics Express*, 18(18):19413–19428, 2010.

[14] Stephanie J Chiu, Joseph A Izatt, Rachelle V O'Connell, Katrina P Winter, Cynthia A Toth, and Sina Farsiu. Validated automatic segmentation of amd pathology including drusen and geographic atrophy in sd-oct images. *Investigative Ophthalmology & Visual Science*, 53(1):53–61, 2012.

[15] GN Girish, VA Anima, Abhishek R Kothari, PV Sudeep, Sohini Roychowdhury, and Jeny Rajan. A benchmark study of automated intra-retinal cyst segmentation algorithms using optical coherence tomography b-scans. *Computer methods and programs in biomedicine*, 153:105–114, 2018.

[16] Ehsan Shahrian Varnousfaderani, Jing Wu, Wolf-Dieter Vogl, Ana-Maria Philip, Alessio Montuoro, Roland Leitner, Christian Simader, Sebastian M Waldstein, Bianca S Gerendas, and Ursula Schmidt-Erfurth. A novel benchmark model for intelligent annotation of spectral-domain optical coherence tomography scans using the example of cyst annotation. *Computer Methods and Programs in Biomedicine*, 130:93–105, 2016.

[17] Xiayu Xu, Kyungmoo Lee, Li Zhang, Milan Sonka, and Michael D Abràmoff. Stratified sampling voxel classification for segmentation of intraretinal and subretinal fluid in longitudinal clinical oct data. *IEEE transactions on medical imaging*, 34(7):1616–1623, 2015.

[18] Serge Resnikoff, Donatella Pascolini, Daniel Etya'Ale, Ivo Kocur, Ramachandra Pararajasegaram, Gopal P Pokharel, and Silvio P Mariotti. Global data on visual impairment in the year 2002. *Bulletin of the world health organization*, 82:844–851, 2004.

[19] Harry A Quigley and Aimee T Broman. The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology*, 90(3):262–267, 2006.

[20] Lingam Vijaya, Ronnie George, Pradeep G Paul, Mani Baskaran, Hemamalini Arvind, Prema Raju, S Ve Ramesh, Govindasamy Kumaramanickavel, and Catherine McCarty. Prevalence of open-angle glaucoma in a rural south indian population. *Investigative ophthalmology & visual science*, 46(12):4461–4467, 2005.

[21] Sunny Y Shen, Tien Y Wong, Paul J Foster, Jing-Liang Loo, Mohamad Rosman, Seng-Chee Loon, Wan Ling Wong, Seang-Mei Saw, and Tin Aung. The prevalence and types of glaucoma in malay people: the singapore malay eye study. *Investigative ophthalmology & visual science*, 49(9): 3846–3851, 2008.

[22] Jost Bruno Jonas, Gabriele Charlotte Gusek, and GO Naumann. Optic disc, cup and neuroretinal rim size, configuration and correlations in normal eyes. *Investigative ophthalmology & visual science*, 29(7):1151–1158, 1988.

[23] Marc Dinkin, Michelle Banks, and Joseph F Rizzo. Imaging the nerve fiber layer and optic disc. In *Pediatric Ophthalmology, Neuro-Ophthalmology, Genetics*, pages 99–118. Springer, 2008.

[24] Patricia Hrynchak, Natalie Hutchings, Deborah Jones, and Trefford Simpson. A comparison of cup-to-disc ratio measurement in normal subjects using optical coherence tomography image analysis of the optic nerve head and stereo fundus biomicroscopy. *Ophthalmic and Physiological Optics*, 24 (6):543–550, 2004.

[25] Juan Xu, Hiroshi Ishikawa, Gadi Wollstein, Richard A Bilonick, Kyung R Sung, Larry Kagemann, Kelly A Townsend, and Joel S Schuman. Automated assessment of the optic nerve head on stereo disc photographs. *Investigative ophthalmology & visual science*, 49(6):2512–2517, 2008.

[26] J Xu, O Chutatape, C Zheng, and PCT Kuan. Three dimensional optic disc visualisation from stereo images via dual registration and ocular media optical correction. *British journal of ophthalmology*, 90(2):181–185, 2006.

[27] Sribalamurugan Sekhar, Waleed Al-Nuaimy, and Asoke K Nandi. Automated localisation of retinal optic disk using hough transform. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1577–1580. IEEE, 2008.

[28] Alauddin Bhuiyan, Ryo Kawasaki, Tien Yin Wong, and Rao Kotagiri. A new and efficient method for automatic optic disc detection using geometrical features. In *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pages 1131–1134. Springer, 2009.

[29] Arturo Aquino, Manuel Emilio Gegúndez-Arias, and Diego 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.

[30] Behdad Dashtbozorg, Ana Maria Mendonça, and Aurélio Campilho. Optic disc segmentation using the sliding band filter. *Computers in Biology and Medicine*, 56:1–12, 2015.

[31] Sohini Roychowdhury, Dara D Koozekanani, Sam N Kuchinka, and Keshab K Parhi. Optic disc boundary and vessel origin segmentation of fundus images. *IEEE Journal of Biomedical and Health Informatics*, 20(6):1562–1574, 2016.

[32] Diego Marin, Manuel E Gegundez-Arias, Angel Suero, and Jose M Bravo. Obtaining optic disc center and pixel region by automatic thresholding methods on morphologically processed fundus images. *Computer methods and programs in biomedicine*, 118(2):173–185, 2015.

[33] M Partha Sarathi, Malay Kishore Dutta, Anushikha Singh, and Carlos M Travieso. Blood vessel inpainting based technique for efficient localization and segmentation of optic disc in digital fundus images. *Biomedical Signal Processing and Control*, 25:108–117, 2016.

[34] James Lowell, Andrew Hunter, David Steel, Ansu Basu, Robert Ryder, Eric Fletcher, and Lee Kennedy. Optic nerve head segmentation. *IEEE Transactions on medical Imaging*, 23(2):256–264, 2004.

[35] DWK Wong, J Liu, JH Lim, X Jia, F Yin, H Li, and TY Wong. Level-set based automatic cup-to-disc ratio determination using retinal fundus images in argali. In *IEEE International Conference Engineering in Medicine and Biology Society*, pages 2266–2269, 2008.

[36] Gopal Datt Joshi, Jayanthi Sivaswamy, and SR Krishnadas. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE transactions on medical imaging*, 30(6):1192–1205, 2011.

[37] J. R. H. Kumar, A. K. Pediredla, and C. S. Seelamantula. Active discs for automated optic disc segmentation. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 225–229. IEEE, Dec 2015.

[38] Balazs Harangi, Rashid Jalal Qureshi, Adrienne Csutak, Tunde Peto, and Andras Hajdu. Automatic detection of the optic disc using majority voting in a collection of optic disc detectors. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1329–1332. IEEE, 2010.

[39] Huiqi Li and Opas Chutatape. Boundary detection of optic disk by a modified asm method. *Pattern Recognition*, 36(9):2093–2104, 2003.

[40] Fengshou Yin, Jiang Liu, Sim Heng Ong, Ying Sun, Damon WK Wong, Ngan Meng Tan, Carol Cheung, Mani Baskaran, Tin Aung, and Tien Yin Wong. Model-based optic nerve head segmentation on retinal fundus images. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2626–2629. IEEE, 2011.

[41] Michael D Abramoff, Wallace LM Alward, Emily C Greenlee, Lesya Shuba, Chan Y Kim, John H Fingert, and Young 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.

[42] Jun Cheng, Jiang Liu, Yanwu Xu, Fengshou Yin, Damon Wing Kee Wong, Ngan-Meng Tan, Dacheng Tao, Ching-Yu Cheng, Tin Aung, and Tien Yin Wong. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE transactions on Medical Imaging*, 32(6):1019–1032, 2013.

[43] Paul L Rosin, David Marshall, and James E Morgan. Multimodal retinal imaging: new strategies for the detection of glaucoma. In *Proceedings. International Conference on Image Processing*, volume 3, pages III–III. IEEE, 2002.

[44] Mei-Ling Huang, Hsin-Yi Chen, and Jian-Jun Huang. Glaucoma detection using adaptive neuro-fuzzy inference system. *Expert Systems with Applications*, 32(2):458–468, 2007.

[45] Juan Xu, Opas Chutatape, Eric Sung, Ce Zheng, and Paul Chew Tec Kuan. Optic disk feature extraction via modified deformable model technique for glaucoma analysis. *Pattern recognition*, 40(7):2063–2076, 2007.

[46] Toshiaki Nakagawa, Yoshinori Hayashi, Yuji Hatanaka, Akira Aoyama, Takeshi Hara, Akihiro Fujita, Masakatsu Kakogawa, Hiroshi Fujita, and Tetsuya Yamamoto. Three-dimensional recon-

struction of optic nerve head from stereo fundus images and its quantitative estimation. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6747–6750. IEEE, 2007.

[47] Mohammad Saleh Miri, Michael D Abràmoff, Kyungmoo Lee, Meindert Niemeijer, Jui-Kai Wang, Young H Kwon, and Mona K Garvin. Multimodal segmentation of optic disc and cup from sd-oct and color fundus photographs using a machine-learning graph-based approach. *IEEE transactions on medical imaging*, 34(9):1854–1866, 2015.

[48] Gopal Datt Joshi, Jayanthi Sivaswamy, Kundan Karan, and SR Krishnadas. Optic disk and cup boundary detection using regional information. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 948–951. IEEE, 2010.

[49] Shibal Bhartiya, Ritu Gadia, Harinder S Sethi, and Anita Panda. Clinical evaluation of optic nerve head in glaucoma. *Journal of Current Glaucoma Practice*, 4(3):115–132, 2010.

[50] Wong Wing Kee Damon, Jimmy Liu, Tan Ngan Meng, Yin Fengshou, and Wong Tien Yin. Automatic detection of the optic cup using vessel kinking in digital retinal fundus images. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1647–1650. IEEE, 2012.

[51] Yuji Hatanaka, Yuuki Nagahata, Chisako Muramatsu, Susumu Okumura, Kazunori Ogohara, Akira Sawada, Kyoko Ishida, Tetsuya Yamamoto, and Hiroshi Fujita. Improved automated optic cup segmentation based on detection of blood vessel bends in retinal fundus images. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 126–129. IEEE, 2014.

[52] Julian G Zilly, Joachim M Buhmann, and Dwarikanath Mahapatra. Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images. In *International Workshop on Machine Learning in Medical Imaging*, pages 136–143. Springer, 2015.

[53] 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 *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3260–3264, Aug 2016.

[54] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424. ACM Press/Addison-Wesley Publishing Co., 2000.

[55] Brian V Funt, Mark S Drew, and Michael Brockington. Recovering shading from color images. In *European Conference on Computer Vision*, pages 124–132. Springer, 1992.

[56] Tsai Ping-Sing and Mubarak Shah. Shape from shading using linear approximation. *Image and Vision computing*, 12(8):487–498, 1994.

[57] Shireen Y. Elhabian. Hands on shape from shading. Technical report, University of Louisville-Electrical and Computer Engineering Department, May 2008.

[58] H Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[59] Jitendra Malik, Serge Belongie, Jianbo Shi, and Thomas Leung. Textons, contours and regions: Cue integration in image segmentation. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 918–925. IEEE, 1999.

[60] Li Tang, Mona K Garvin, Kyungmoo Lee, Wallace LW Alward, Young H Kwon, and Michael D Abramoff. Robust multiscale stereo matching from fundus images with radiometric differences. *IEEE transactions on pattern analysis and machine intelligence*, 33(11):2245–2258, 2011.

[61] David R Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664, 2004.

[62] Jianchao Yang, Zhaowen Wang, Zhe Lin, Scott Cohen, and Thomas Huang. Coupled dictionary training for image super-resolution. *IEEE transactions on image processing*, 21(8):3467–3478, 2012.

[63] Shujian Yu, Weihua Ou, Xinge You, Yi Mou, Xiubao Jiang, and Yuanyan Tang. Single image rain streaks removal based on self-learning and structured sparse representation. In *2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pages 215–219. IEEE, 2015.

[64] Weihua Ou, Xinge You, Dacheng Tao, Pengyue Zhang, Yuanyan Tang, and Ziqi Zhu. Robust face recognition via occlusion dictionary learning. *Pattern Recognition*, 47(4):1559–1572, 2014.

[65] Zhenyu He, Shuangyan Yi, Yiu-Ming Cheung, Xinge You, and Yuan Yan Tang. Robust object tracking via key patch sparse representation. *IEEE transactions on cybernetics*, 47(2):354–364, 2016.

[66] Michal Aharon, Michael Elad, Alfred Bruckstein, et al. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11): 4311, 2006.

[67] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.

[68] Francisco Fumero, Silvia Alayón, José L Sanchez, Jose Sigut, and M Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. In *2011 24th international symposium on computer-based medical systems (CBMS)*, pages 1–6. IEEE, 2011.

[69] Enrique J Carmona, Mariano Rincón, Julián García-Feijoó, and José M Martínez-de-la Casa. Identification of the optic nerve head with genetic algorithms. *Artificial Intelligence in Medicine*, 43(3):243–259, 2008.

[70] *Expert system for early automatic detection of diabetic retinopathy by analysis of digital retinal images: http://uhu.es/retinopathy/eng/bd.php*, 2012.

[71] Andrea Giachetti, Lucia Ballerini, and Emanuele Trucco. Accurate and reliable segmentation of the optic disc in digital fundus images. *Journal of Medical Imaging*, 1(2):024001, 2014.

[72] Honggang Yu, E Simon Barriga, Carla Agurto, Sebastian Echegaray, Marios S Pattichis, Wendall Bauman, and Peter 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.

[73] Sandra Morales, Valery Naranjo, Jesús Angulo, and Mariano Alcañiz. Automatic detection of optic disc based on pca and mathematical morphology. *IEEE transactions on medical imaging*, 32 (4):786–796, 2013.

[74] Gopal Datt Joshi, Jayanthi Sivaswamy, and SR Krishnadas. Depth discontinuity-based cup segmentation from multiview color retinal images. *IEEE Transactions on Biomedical Engineering*, 59 (6):1523–1531, 2012.

[75] Yuanjie Zheng, Dwight Stambolian, Joan O'Brien, and James 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.

[76] Sandra Morales, Valery Naranjo, David Perez, Amparo Navea, and Mariano Alcaniz. Automatic detection of optic disc based on pca and stochastic watershed. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 2605–2609. IEEE, 2012.

[77] Thomas Walter, J-C Klein, Pascale Massin, and Ali Erginay. A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *IEEE transactions on medical imaging*, 21(10):1236–1243, 2002.

[78] Swamidoss Issac Niwas, Weisi Lin, Xiaolong Bai, Chee Keong Kwoh, C-C Jay Kuo, Chelvin C Sng, Maria Cecilia Aquino, and Paul TK Chew. Automated anterior segment oct image analysis for angle closure glaucoma mechanisms classification. *Computer Methods and Programs in Biomedicine*, 130:65–75, 2016.

[79] Désiré Sidibé, Shrinivasan Sankar, Guillaume Lemaître, Mojdeh Rastgoo, Joan Massich, Carol Y Cheung, Gavin SW Tan, Dan Milea, Ecosse Lamoureux, Tien Y Wong, et al. An anomaly detection approach for the identification of dme patients using spectral domain optical coherence tomography images. *Computer Methods and Programs in Biomedicine*, 139:109–117, 2017.

[80] Raed Behbehani, Abdullah Abu Al-Hassan, Ali Al-Salahat, Devarajan Sriraman, JD Oakley, and Raed Alroughani. Optical coherence tomography segmentation analysis in relapsing remitting versus progressive multiple sclerosis. *PloS One*, 12(2):e0172120, 2017.

[81] Ferdinand G Schlanitz, Christian Ahlers, Stefan Sacu, Christopher Schütze, Marcos Rodriguez, Sabine Schriefl, Isabelle Golbaz, Tobias Spalek, Geraldine Stock, and Ursula Schmidt-Erfurth. Performance of drusen detection by spectral-domain optical coherence tomography. *Investigative Ophthalmology & Visual Science*, 51(12):6715–6721, 2010.

[82] S Mojtaba Golzan, Alberto Avolio, and Stuart L Graham. Minimising retinal vessel artefacts in optical coherence tomography images. *Computer Methods and Programs in Biomedicine*, 104(2): 206–211, 2011.

[83] Hiroshi Ishikawa, Daniel M Stein, Gadi Wollstein, Siobahn Beaton, James G Fujimoto, and Joel S Schuman. Macular segmentation with optical coherence tomography. *Investigative Ophthalmology & Visual Science*, 46(6):2012–2017, 2005.

[84] Tapio Fabritius, Shuichi Makita, Masahiro Miura, Risto Myllylä, and Yoshiaki Yasuno. Automated segmentation of the macula by optical coherence tomography. *Optics Express*, 17(18):15659–15669, 2009.

[85] Jelena Novosel, Gijs Thepass, Hans G Lemij, Johannes F de Boer, Koenraad A Vermeer, and Lucas J van Vliet. Loosely coupled level sets for simultaneous 3d retinal layer segmentation in optical coherence tomography. *Medical Image Analysis*, 26(1):146–158, 2015.

[86] Florence Rossant, Isabelle Bloch, Itebeddine Ghorbel, and Michel Paques. Parallel double snakes. application to the segmentation of retinal layers in 2d-oct for pathological subjects. *Pattern Recognition*, 48(12):3857–3870, 2015.

[87] Azadeh Yazdanpanah, Ghassan Hamarneh, Benjamin R Smith, and Marinko V Sarunic. Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach. *IEEE Transactions on Medical Imaging*, 30(2):484–496, 2011.

[88] Vedran Kajić, Boris Považay, Boris Hermann, Bernd Hofer, David Marshall, Paul L Rosin, and Wolfgang Drexler. Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis. *Optics Express*, 18(14):14730–14744, 2010.

[89] Jing Tian, Boglárka Varga, Gábor Márk Somfai, Wen-Hsiang Lee, William E Smiddy, and Delia Cabrera DeBuc. Real-time automatic segmentation of optical coherence tomography volume data of the macular region. *PloS One*, 10(8):e0133908, 2015.

[90] Mona Haeker, Michael Abramoff, Randy Kardon, and Milan Sonka. Segmentation of the surfaces of the retinal layer from oct images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 800–807. Springer, 2006.

[91] Mona Kathryn Garvin, Michael David Abràmoff, Xiaodong Wu, Stephen R Russell, Trudy L Burns, and Milan Sonka. Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Transactions on Medical Imaging*, 28(9): 1436–1447, 2009.

[92] Pascal A Dufour, Lala Ceklic, Hannan Abdillahi, Simon Schroder, Sandro De Dzanet, Ute Wolf-Schnurrbusch, and Jens Kowal. Graph-based multi-surface segmentation of oct data using trained hard and soft constraints. *IEEE Transactions on Medical Imaging*, 32(3):531–543, 2013.

[93] Stephanie J Chiu, Michael J Allingham, Priyatham S Mettu, Scott W Cousins, Joseph A Izatt, and Sina Farsiu. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomedical Optics Express*, 6(4):1172–1194, 2015.

[94] Md Akter Hussain, Alauddin Bhuiyan, Andrew Turpin, Chi D Luu, R Theodore Smith, Robyn H Guymer, and Ramamohanrao Kotagiri. Automatic identification of pathology-distorted retinal layer boundaries using sd-oct imaging. *IEEE Transactions on Biomedical Engineering*, 64(7): 1638–1649, 2017.

[95] Fei Shi, Xinjian Chen, Heming Zhao, Weifang Zhu, Dehui Xiang, Enting Gao, Milan Sonka, and Haoyu Chen. Automated 3-d retinal layer segmentation of macular optical coherence tomography images with serous pigment epithelial detachments. *IEEE Transactions on Medical Imaging*, 34 (2):441–452, 2015.

[96] Jelena Novosel, Koenraad A Vermeer, Jan H de Jong, Ziyuan Wang, and Lucas J van Vliet. Joint segmentation of retinal layers and focal lesions in 3-d oct data of topologically disrupted retinas. *IEEE Transactions on Medical Imaging*, 36(6):1276–1286, 2017.

[97] Leyuan Fang, David Cunefare, Chong Wang, Robyn H Guymer, Shutao Li, and Sina Farsiu. Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search. *Biomedical optics express*, 8(5):2732–2744, 2017.

[98] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[99] Abhijit Guha Roy, Sailesh Conjeti, Sri Phani Krishna Karri, Debdoot Sheet, Amin Katouzian, Christian Wachinger, and Nassir Navab. Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomedical optics express*, 8(8): 3627–3642, 2017.

[100] Avi Ben-Cohen, Dean Mark, Ilya Kovler, Dinah Zur, Adiel Barak, Matias Iglicki, and Ron Soferman. Retinal layers segmentation using fully convolutional network in oct images. 2017. URL https://www.rsipvision.com/wp-content/uploads/2017/06/Retinal-Layers-Segmentation.pdf.

[101] Mike Pekala, Neil Joshi, David E Freund, Neil M Bressler, Delia Cabrera DeBuc, and Philippe M Burlina. Deep learning based retinal oct segmentation. *arXiv preprint arXiv:1801.09749*, 2018.

[102] Olaf Ronneberger, Philipp Fischer, and Thomas 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.

[103] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[104] Yongjian Yu and Scott T Acton. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 11(11):1260–1270, 2002.

[105] László G Nyúl, Jayaram K Udupa, and Xuan Zhang. New variants of a method of mri scale standardization. *IEEE Transactions on Medical Imaging*, 19(2):143–150, 2000.

[106] SM Lee, JH Xin, and S Westland. Evaluation of image similarity by histogram intersection. *Color Research & Application*, 30(4):265–274, 2005.

[107] Thomas Finley and Thorsten Joachims. Training structural svms when exact inference is intractable. In *Proceedings of the 25th International Conference on Machine Learning*, pages 304–311. ACM, 2008.

[108] Simon Lacoste-Julien, Martin Jaggi, Mark Schmidt, and Patrick Pletscher. Block-coordinate frank-wolfe optimization for structural svms. In *30th International Conference on Machine Learning*, volume 28, pages I–53–I–61. JMLR.org, 2013.

[109] Pangyu Teng. Caserel - an open source software for computer-aided segmentation of retinal layers in optical coherence tomography images, 2013. URL http://pangyuteng.github.io/caserel/.

[110] Kyungmoo Lee, Michael D. Abramoff, Mona Garvin, Milan Sonka, et al. The iowa reference algorithms, version 3.8.0 (retinal image analysis lab, iowa institute for biomedical imaging, iowa city, ia), 2014. URL http://www.iibi.uiowa.edu/content/iowa-reference-algorithms-human-and-murine-oct-retinal-layer-analysis-and-display.

[111] Markus Mayer et al. Octseg, version 4.0 (pattern recognition lab, friedrich-alexander-universität erlangen-nürnberg), 2016. URL https://www5.cs.fau.de/research/software/octseg/.

[112] Markus A Mayer, Joachim Hornegger, Christian Y Mardin, and Ralf P Tornow. Retinal nerve fiber layer segmentation on fd-oct scans of normal subjects and glaucoma patients. *Biomedical Optics Express*, 1(5):1358–1383, 2010.

[113] Adeel M Syed, Taimur Hassan, M Usman Akram, Samra Naz, and Shehzad Khalid. Automated diagnosis of macular edema and central serous retinopathy through robust reconstruction of 3d retinal surfaces. *Computer Methods and Programs in Biomedicine*, 137:1–10, 2016.

[114] Tim McInerney and Demetri Terzopoulos. Deformable models in medical image analysis: a survey. *Medical image analysis*, 1(2):91–108, 1996.

[115] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International journal of computer vision*, 22(1):61–79, 1997.

[116] Laurent D Cohen and Isaac Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE Transactions on pattern analysis and machine intelligence*, 15(11): 1131–1147, 1993.

[117] Chenyang Xu and Jerry L Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on image processing*, 7(3):359–369, 1998.

[118] Tony F Chan and Luminita A Vese. Active contours without edges. *IEEE Transactions on image processing*, 10(2):266–277, 2001.

[119] Dan Ciresan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, pages 2843–2851, 2012.

[120] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.

[121] Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbeláez, and Luc Van Gool. Deep retinal image understanding. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 140–148. Springer, 2016.

[122] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.

[123] Kenny H Cha, Lubomir Hadjiiski, Ravi K Samala, Heang-Ping Chan, Elaine M Caoili, and Richard H Cohan. Urinary bladder segmentation in ct urography using deep-learning convolutional neural network and level sets. *Medical physics*, 43(4):1882–1896, 2016.

[124] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.

[125] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1529–1537, 2015.

[126] Karthik Gopinath, Samrudhdhi B Rangrej, and Jayanthi Sivaswamy. A deep learning framework for segmentation of retinal layers from oct images. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 888–893. IEEE, 2017.

[127] Jianxu Chen, Lin Yang, Yizhe Zhang, Mark Alber, and Danny Z Chen. Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In *Advances in neural information processing systems*, pages 3036–3044, 2016.

[128] Yuanpu Xie, Zizhao Zhang, Manish Sapkota, and Lin Yang. Spatial clockwork recurrent neural network for muscle perimysium segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 185–193. Springer, 2016.

[129] Marijn F Stollenga, Wonmin Byeon, Marcus Liwicki, and Juergen Schmidhuber. Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation. In *Advances in neural information processing systems*, pages 2998–3006, 2015.

[130] Yu Cheng, Duo Wang, Pan Zhou, and Tao Zhang. Model compression and acceleration for deep neural networks: The principles, progress, and challenges. *IEEE Signal Processing Magazine*, 35 (1):126–136, 2018.

[131] Guillermo Sapiro. *Geometric partial differential equations and image analysis*. Cambridge university press, 2006. pg. 74–76.

[132] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, May 2016. URL http://arxiv.org/abs/1605.02688.

[133] Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Eleventh Annual Conference of the International Speech Communication Association*, pages 1045–1048, 2010.

[134] Paul J Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.

[135] Chunming Li, Chenyang Xu, Changfeng Gui, and Martin D Fox. Distance regularized level set evolution and its application to image segmentation. *IEEE transactions on image processing*, 19 (12):3243–3254, 2010.

[136] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint:1412.6980*, 2014.

[137] François Chollet et al. Keras. https://github.com/keras-team/keras, 2015.

[138] Arunava Chakravarty and Jayanthi Sivaswamy. Joint optic disc and cup boundary extraction from monocular fundus images. *Computer methods and programs in biomedicine*, 147:51–61, 2017.

[139] Artem Sevastopolsky. Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. *Pattern Recognition and Image Analysis*, 27 (3):618–624, 2017.

[140] Elisa Drelie Gelasca, Boguslaw Obara, Dmitry Fedorov, Kristian Kvilekval, and BS Manjunath. A biosegmentation benchmark for evaluation of bioimage analysis methods. *BMC bioinformatics*, 10(1):368, 2009.

[141] Xipeng Pan, Lingqiao Li, Huihua Yang, Zhenbing Liu, Jinxin Yang, Lingling Zhao, and Yongxian Fan. Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks. *Neurocomputing*, 229:88–99, 2017.

[142] Jing Rui Tang, Nor Ashidi Mat Isa, and Ewe Seng Ch'ng. A fuzzy-c-means-clustering approach: Quantifying chromatin pattern of non-neoplastic cervical squamous cells. *PloS one*, 10 (11):e0142830, 2015.

[143] Felix Buggenthin, Carsten Marr, Michael Schwarzfischer, Philipp S Hoppe, Oliver Hilsenbeck, Timm Schroeder, and Fabian J Theis. An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy. *BMC bioinformatics*, 14(1):297, 2013.

[144] Yan Nei Law, Hwee Kuan Lee, Michael K Ng, and Andy M Yip. A semisupervised segmentation model for collections of images. *IEEE transactions on Image processing*, 21(6):2955–2968, 2012.

[145] Nuh Hatipoglu and Gokhan Bilgin. Classification of histopathological images using convolutional neural network. In *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2014.

[146] Catalina Tobon-Gomez, Arjan J Geers, Jochen Peters, Jürgen Weese, Karen Pinto, Rashed Karim, Mohammed Ammar, Abdelaziz Daoudi, Jan Margeta, Zulma Sandoval, et al. Benchmark for algorithms segmenting the left atrium from 3d ct and mri datasets. *IEEE transactions on medical imaging*, 34(7):1460–1473, 2015.

[147] "Catalina Toboz-Gomez". *Left Atrium Segmentation Challenge*, 2012. URL http://www.cardiacatlas.org/challenges/left-atrium-segmentation-challenge/.

[148] Christian Rupprecht, Elizabeth Huaroc, Maximilian Baust, and Nassir Navab. Deep active contours. *arXiv preprint arXiv:1607.05074*, 2016.

[149] Noga Harizman, Cristiano Oliveira, Allen Chiang, Celso Tello, Michael Marmor, Robert Ritch, and Jeffrey M Liebmann. The isnt rule and differentiation of normal from glaucomatous eyes. *Archives of ophthalmology*, 124(11):1579–1583, 2006.

[150] Jun Cheng, Fengshou Yin, Damon Wing Kee Wong, Dacheng Tao, and Jiang Liu. Sparse dissimilarity-constrained coding for glaucoma screening. *IEEE Transactions on Biomedical Engineering*, 62(5):1395–1403, 2015.

[151] Jun Cheng, Zhuo Zhang, Dacheng Tao, Damon Wing Kee Wong, Jiang Liu, Mani Baskaran, Tin Aung, and Tien Yin Wong. Similarity regularized sparse group lasso for cup to disc ratio computation. *Biomedical optics express*, 8(8):3763–3777, 2017.

[152] Jagadish Nayak, Rajendra Acharya, P Subbanna Bhat, Nakul Shetty, and Teik-Cheng Lim. Automated diagnosis of glaucoma using digital fundus images. *Journal of medical systems*, 33(5):337, 2009.

[153] Zhuo Zhang, Chee Keong Kwoh, Jiang Liu, Fengshou Yin, Adrianto Wirawan, Carol Cheung, Mani Baskaran, Tin Aung, and Tien Yin Wong. Mrmr optimized classification for automatic glaucoma diagnosis. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 6228–6231. IEEE, 2011.

[154] Jiang Liu, Yanwu Xu, Jun Cheng, Zhuo Zhang, Damon Wing Kee Wong, Fengshou Yin, and Tien Yin Wong. Multiple modality fusion for glaucoma diagnosis. In *The international conference on health informatics*, pages 5–8. Springer, 2014.

[155] U Rajendra Acharya, Sumeet Dua, Xian Du, Chua Kuang 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.

[156] Sumeet Dua, U Rajendra Acharya, Pradeep Chowriappa, and S Vinitha Sree. Wavelet-based energy features for glaucomatous image classification. *IEEE transactions on information technology in biomedicine*, 16(1):80–87, 2012.

[157] Rüdiger Bock, Jörg Meier, Georg Michelson, László G Nyúl, and Joachim Hornegger. Classifying glaucoma with image-based features from fundus photographs. In *Joint Pattern Recognition Symposium*, pages 355–364. Springer, 2007.

[158] Huazhu Fu, Jun Cheng, Yanwu Xu, Damon Wing Kee Wong, Jiang Liu, and Xiaochun Cao. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE transactions on medical imaging*, 37(7):1597–1605, 2018.

[159] Huazhu Fu, Jun Cheng, Yanwu Xu, Changqing Zhang, Damon Wing Kee Wong, Jiang Liu, and Xiaochun Cao. Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE transactions on medical imaging*, 37(11):2493–2501, 2018.

[160] NV Medathati and Jayanthi Sivaswamy. Local descriptor based on texture of projections. In *Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing*, pages 398–404. ACM, 2010.

[161] Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, pages 1–2. Prague, 2004.

[162] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.

[163] Refuge dataset, 2018. URL https://refuge.grand-challenge.org/.

[164] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[165] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[166] Sina Farsiu, Stephanie J Chiu, Rachelle V O'Connell, Francisco A Folgar, Eric Yuan, Joseph A Izatt, Cynthia A Toth, Age-Related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography Study Group, et al. Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography. *Ophthalmology*, 121(1):162–172, 2014.

[167] Min Chen, Andrew Lang, Elias Sotirchos, Howard S Ying, Peter A Calabresi, Jerry L Prince, and Aaron Carass. Deformable registration of macular oct using a-mode scan similarity. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 476–479. IEEE, 2013.

[168] FSL atlases. https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases/. [Online; accessed 16-Jan-2018].

[169] Serdar K Balci, Polina Golland, Martha Elizabeth Shenton, and William Mercer Wells. Free-form b-spline deformation model for groupwise registration. In *Statistical Registration Workshop, MICCAI*. Springer, 2007.

[170] Kanwal K Bhatia, Joseph V Hajnal, Basant K Puri, A David Edwards, and Daniel Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, pages 908–911. IEEE, 2004.

[171] Sarang Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.

[172] Guorong Wu, Hongjun Jia, Qian Wang, and Dinggang Shen. Sharpmean: groupwise registration guided by sharp mean image and tree-based registration. *NeuroImage*, 56(4):1968–1981, 2011.

[173] Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab, and Nikos Paragios. Dense image registration through mrfs and efficient linear programming. *Medical image analysis*, 12(6):731–741, 2008.

[174] Mattias P Heinrich, Bartlomiej W Papież, Julia A Schnabel, and Heinz Handels. Non-parametric discrete registration with convex optimisation. In *International Workshop on Biomedical Image Registration*, pages 51–61. Springer, 2014.

[175] Hyunjin Park, Peyton H Bland, Alfred O Hero, and Charles R Meyer. Least biased target selection in probabilistic atlas construction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 419–426. Springer, 2005.

[176] Yu-Ying Liu, Mei Chen, Hiroshi Ishikawa, Gadi Wollstein, Joel S Schuman, and James M Rehg. Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Medical image analysis*, 15(5):748–759, 2011.

[177] Pratul P Srinivasan, Leo A Kim, Priyatham S Mettu, Scott W Cousins, Grant M Comer, Joseph A Izatt, and Sina Farsiu. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomedical optics express*, 5 (10):3568–3577, 2014.

[178] Freerk G Venhuizen, Bram van Ginneken, Bart Bloemen, Mark JJP van Grinsven, Rick Philipsen, Carel Hoyng, Thomas Theelen, and Clara I Sánchez. Automated age-related macular degeneration classification in oct using unsupervised feature learning. In *Medical Imaging 2015: Computer-Aided Diagnosis*, volume 9414, page 94141I. International Society for Optics and Photonics, 2015.

[179] Anurag Arnab, Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Måns Larsson, Alexander Kirillov, Bogdan Savchynskyy, Carsten Rother, Fredrik Kahl, and Philip HS Torr. Conditional random fields meet deep neural networks for semantic segmentation: Combining probabilistic graphical models with deep learning for structured prediction. *IEEE Signal Processing Magazine*, 35(1):37–52, 2018.

[180] Mikael Rousson and Nikos Paragios. Shape priors for level set representations. In *European Conference on Computer Vision*, pages 78–92. Springer, 2002.

[181] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.

[182] Bob D de Vos, Floris F Berendsen, Max A Viergever, Marius Staring, and Ivana Išgum. End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 204–212. Springer, 2017.

[183] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:236–239, 2003.

[184] Amar Mitiche and Ismail Ben Ayed. *Variational and level set methods in image segmentation*, volume 5. Springer Science & Business Media, 2010. pg. 17–19.