

LEARNING REPRESENTATIONS FOR RETINAL
OPTICAL COHERENCE TOMOGRAPHY IMAGE
ANALYSIS

Sri Phani Krishna Karri

LEARNING REPRESENTATIONS FOR RETINAL OPTICAL COHERENCE TOMOGRAPHY IMAGE ANALYSIS

*Thesis submitted to the
Indian Institute of Technology Kharagpur
for award of the degree*

of

Doctor of Philosophy

by

Sri Phani Krishna Karri

Under the guidance of

**Dr. Jyotirmoy Chatterjee
Dr. Debjani Chakraborty**



**School of Medical Science and Technology
Indian Institute of Technology Kharagpur
November 2017**

©2017 Sri Phani Krishna Karri. All rights reserved.

CERTIFICATE OF APPROVAL

Certified that the thesis entitled **LEARNING REPRESENTATIONS FOR RETINAL OPTICAL COHERENCE TOMOGRAPHY IMAGE ANALYSIS** submitted by **Sri Phani Krishna Karri** to the Indian Institute of Technology, Kharagpur, for the award of the degree Doctor of Philosophy has been accepted by the external examiners and that the student has successfully defended the thesis in the viva-voce examination held today.

(Member of DSC)

(Member of DSC)

(Member of DSC)

(Supervisor)

(Supervisor)

(External Examiner)

(Chairman)

CERTIFICATE

This is to certify that the thesis entitled **Learning Representations for Retinal Optical Coherence Tomography Image Analysis**, submitted by **Sri Phani Krishna Karri** to Indian Institute of Technology Kharagpur, is a record of bona fide research work under my (our) supervision and consider it worthy of consideration for the award of the degree of Doctor of Philosophy of the Institute.

(Supervisor)

(Supervisor)

Date:

DECLARATION

I certify that

- a. The work contained in the thesis is original and has been done by myself under the general supervision of my supervisor.
- b. The work has not been submitted to any other Institute for any degree or diploma.
- c. I have followed the guidelines provided by the Institute in writing the thesis.
- d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.
- e. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.
- f. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Sri Phani Krishna Karri

Dedicated to My Family

ACKNOWLEDGMENT

I would like to thank all people who have helped and inspired me in life. First of all, I would like to thank my PhD supervisors Professors Jyotirmoy Chatterjee and Debjani Chakraborty, not only for their continual trust and support, but also for encouraging me to explore optical coherence tomography (OCT), medical image analysis and machine learning at School of Medical Science and Technology. On an additional note Dr. Sambuddha Ghosh (Regional Institute of Ophthalmology, Kolkata) has patiently shared his in-depth knowledge of retina and its pathologies from the perspective of clinical practices. Professor Sina Farsiu (Duke university) too deserves the special mention for open-sourcing retinal OCT datasets for research purposes. I am deeply thankful to my colleagues and their enthusiasm for our progress starting as the Medical Imaging Laboratory in 2009 and maturing to the Laboratory on Multimodal Imaging and Computing for Theranostics in 2013. To begin with, Hrushikesh Garud, Dr. Debdoott Sheet and Vishnu Bajpai, for mentoring me through my life in research and outside it. My friends Kaushik Basak, Tamogna Ojha, Amit Suveer, Subhamoy Mandal, Debanjan Das, Rusha Patra, Pankaj Wasnik, Biswajoy Ghosh, Sailesh Conjeti, Abhijit Guha Roy, Niladri Garai, Swarnadip Chatterjee, Kaushik Das Jaswanth Manku and Siva Kumar Katragadda, for always being there by my side when needed.

My colleagues and seniors when I started my PhD, Raunak Kumar Das, Roshan Joy Martis, Muthu Rama Krishnan, Anji Anura, Swarnendu Bag, Amrita Chaudhary, Atasi Sarkar, Satarupa Banerjee, Mounika Rajput, Moushumi Mondol, Arpan Guha Mazumder, Vikram Venkatraghavan, Richa Malviya, Puneet Manocha and Rithesh Khare would always have that special place reserved in my heart. My labmates Koushik Mallik, Debnath Das, Aniruddha Das, Krishnabrata Panda, also bear this special mention for being there by my side. Swadeep, Rajendra, Biswajoy, Sailesh, Jyoti and Abhinav from the M. Tech batch of 2012-14 for helping me improve my perception about medical imaging. Colleagues across the campus, Kausik Basak, Rusha Patra, Tripti Swarnakar, Sripad Bhatlawande, Chandrasekhar Shendkar, Subhranil Koley, Subir Ghosh and not to forget Phani for supporting my research endeavors and with the Machine Learning Reading Group, IEEE Engineering in Medicine and Biology Student Club and the Student Branch. Professors Pranab Kumar Dutta, Manjunatha Mahadevappa, Analava Mitra, Pabitra Mitra deserve the special mention for closely scrutinizing my research and for supporting my dreams to explore territories. I specially thank Prof. Saina

Farsiu, Dr. Stephanie Chiu, P. P. Srinivasan and Dr. L. Fang for answering numerous questions regarding the Duke's data sets.

Sri Phani Krishna Karri

Contents

Dedication	vii
List of Figures	xv
List of Tables	xix
List of Symbols and Abbreviations	xxi
Artwork Licenses	xxv
Graphical Abstract	xxix
Abstract	xxxi
1 Introduction	1
1.1 Objectives and Scope	3
1.2 Literature Review	5
1.2.1 Retina	5
1.2.2 Retinal imaging modalities	8
1.2.3 Image analysis	11
1.2.4 Machine learning	15
1.2.5 Learning representations	16
1.3 Related Work	19
1.3.1 State-of-the-art for retinal layer segmentation	20
1.3.2 State-of-the-art for retinal pathology classification (AMD and DME) .	22
1.3.3 State-of-the-art feature descriptors for retinal pathologies (AMD and DME)	24
1.3.4 State-of-the-art for restoring and reconstructing retinal OCT	26
1.4 Targeted Applications	28

Contents

1.5	Organization of the Thesis	34
2	Learning Layer-Specific Edge Representations for Retinal Layer Segmentation	39
2.1	Introduction	39
2.1.1	Need for structured prediction	40
2.1.2	Structured prediction with classical learning machines	41
2.1.3	Structured random forests for edge	41
2.1.4	Identified approach	43
2.2	Preamble to Solution	44
2.3	Materials and Methods	45
2.3.1	Experimental bed and baselines	45
2.3.2	Curating data for SRE	45
2.3.3	Data preparation	46
2.3.4	Model training and prediction	47
2.4	Results and Discussion	51
2.4.1	Governing parameters for training and testing of SRE_n	51
2.4.2	Delineating a given test image	51
2.4.3	Delineating under various scenarios	53
2.4.4	Quantitative performance	53
2.4.5	Alternate approaches	55
2.5	Summary	56
3	Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach	59
3.1	Introduction	59
3.1.1	Need for modeling a manifold representation	61
3.1.2	Manifold representation modeling	62
3.1.3	Identified approach	62
3.2	Preamble to Solution	63
3.3	Materials and Methods	64
3.3.1	Experimental bed	64
3.3.2	Retinal flattening	65
3.3.3	Feature quantification	66
3.3.4	Proposed framework	67
3.4	Results and Discussion	70
3.4.1	Spiral example	70

3.4.2	Synthetic toy example	70
3.4.3	Modeling characteristics and predicting capability of proposed algorithm for retinal pathologies	72
3.4.4	K-fold validation performance in comparison to baselines	75
3.4.5	Regulating parameters	75
3.5	Summary	76
4	Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies	79
4.1	Introduction	79
4.1.1	Need for learning feature quantifiers as representations in input space	80
4.1.2	Automated representation learning	81
4.1.3	GoogleNet	82
4.1.4	Visualizing learned representations	82
4.1.5	Identified approach	84
4.2	Preamble to Solution	84
4.3	Materials and Methods	86
4.3.1	Experimental setup	86
4.3.2	Data preparation	86
4.3.3	Modifications to trained GoogLeNet on ImageNet	87
4.3.4	Fine-tuning	87
4.3.5	Identifying potential response	88
4.3.6	Visualizing abstract patterns	89
4.4	Results and Discussion	89
4.4.1	Repeatability	89
4.4.2	Image-level performance	90
4.4.3	Visualization	91
4.4.4	Baseline	91
4.4.5	Crossvalidation	94
4.5	Summary	94
5	Low-Rank CNN Filters for Simultaneous Reconstruction and Restoration of Sparsely Sampled Retinal OCT Image	97
5.1	Introduction	97
5.1.1	Need for sparse representations	98
5.1.2	Extending sparse representations for OCT denoising	99

Contents

5.1.3	Identified approach	101
5.2	Preamble to Solution	103
5.3	Materials and Methods	103
5.3.1	Experimental setup	103
5.3.2	Alterations considered while adapting CNN architectures	104
5.3.3	Factorizing filters	105
5.4	Results and Discussion	108
5.4.1	Training parameters	109
5.4.2	Reconstructing a given test image	109
5.4.3	Training and testing characteristics	113
5.4.4	Benchmark	113
5.5	Summary	115
6	Summary and Conclusion	117
6.1	Contribution of this Work	118
6.2	Re-usability of Work	120
6.3	Future Scope of Work	121
Appendices		123
A	Statistical representation	125
B	Multiple layer-specific edge modeling	125
C	Transcendence of means-set based classifier into neural network	125
Dissemination of the Work		129
Resume		131
Bibliography		135

List of Figures

1.1	Illustrates the evolution of machine learning from the perspective of medical image analysis and employing modeled representations in various spaces with different blocks of IAB for retinal image analysis in this thesis.	4
1.2	Illustrating the motivation and scope of considered aim.	5
1.3	A sagittal plane of eye illustrating the ray diagram of eye lenses and the need of retina to be on the focal plane.	6
1.4	Illustrating the increase of geographical atrophy in retina with grade (increasing) of AMD and corresponding vision loss.	7
1.5	The schematic of retinal SD-OCT in respect to optical path of photons illustrating the employment of grating and spectrometer for time-space encoding which is crucial for rapid imaging.	10
1.6	Depicting the cross-sectional information (non-invasive) of normal subject's retina through SD-OCT B-scan.	11
1.7	Illustrating various layers of retina imaged with OCT and comparing with retina histology.	11
1.8	Depicting the drusen deposition between RPE and BM in OCT B-scan image of an AMD subject.	12
1.9	Illustrating the detachment and deformation in OCT B-scan image of a DME subject.	12
1.10	Representing an environment through skeleton structure with curves and lines instead of duplicating the information.	17
1.11	Illustration the role of <i>representation</i> for converting information from various domains into corresponding engineering knowledge.	18
1.12	Display of a typical clinical OCT screen with automated retinal nerve fiber layer (RNFL) thickness computation which depicts the impact of segmentation algorithm in ophthalmology for semi-quantitative analysis of retina through retinal profiling.	29

List of Figures

1.13	Anticipated delineated layers of retina in DME subject's OCT B-scan depicting both segmentation (left side) and upper contour of the layers (right side).	31
1.14	Illustrating capability of a classifier to segregate given images into various pathological states (nomral, AMD and DME i.e., circles, diamonds and plus respectively) in two dimensional feature space. The X and Y axis represents the retinal OCT HOG feature space projected onto 2D space.	32
1.15	Illustrating the capability of feature quantifier depicting deformations induced due to drusen deposition in an OCT image of an AMD subject.	33
1.16	Illustrating simultaneous restoration and reconstruction of sparsely sampled OCT image with reduced noise and improved information.	34
2.1	Distinction between conventional random forest prediction (only label) and structured random forests prediction (structure) given a feature map.	41
2.2	Training and prediction process of structured forest for modeling and predicting edge representations along with functionality of decision node.	43
2.3	Proposed approach empowers SRE to predict layer subjective edge.	45
2.4	Images required for preparing training dataset. (a) Input image, (b) label image and (c) contour image.	46
2.5	Flow of retinal image processing blocks involved in the (a) training phase and (b) testing phase.	48
2.6	Processed outputs of layer 'n' selection block for retinal layer 3 from top. (a) Label image, (b) contour image, (c) processed label binary mask and (d) processed contour binary mask	49
2.7	Block diagram illustrating image inputs during training and resulting image at each block during testing. The output of each model (SRE for $layer_n$) depicts the edge of subjective layer fulfilling the anticipated objective.	52
2.8	Predictions of the proposed approach on test images with various deformations illustrating its robustness to varied deformation levels and noise. The predicted layer are same as considered in subsection 2.3.2 and color coding is same as considered in Figure. 1.13.	54
2.9	Illustration of layer segmentation performance in the presence of deformation (red), noise (yellow), shadows (orange) and low gradients (blue) by the proposed method in comparison to other baselines.	55

List of Figures

3.1	Illustrating improved resolution of subspace in feature space on addition of new control point.	63
3.2	Image standardization through retinal flattening i.e., shifting the A-scan lines so lower boundary of RPE (including drusen in case of AMD) forms a straight line.	66
3.3	Illustrates the capacity of proposed approach to model complex structures (spiral example) through identifying control points (asterisk and rectangle). The modeling capacity of the approach can be depicted through color allocated to each spiral point during prediction and consistency within each spiral is achieved in a few iterations. X and Y axis are spatial coordinates of sampled spirals.	71
3.4	The proposed approach is capable of identifying control points which autonomously constructs the boundaries subjective to the scenarios ((b) linearly separable and (c) non-linearly separable) generated from toy example.	73
3.5	Illustrating the modeling capability of proposed approach for separating retinal pathologies (circle, diamond and plus implies normal, AMD and DME) along each iteration. The zoomed sub-images illustrates the capability of resolving ambiguous samples (along the boundary) with each iteration. The X and Y axis represents the retinal OCT HOG feature space projected onto 2D space.	74
4.1	Illustrating the variations of gradient orientations along geographical atrophy (due to drusen deposition) depicted by a HOG descriptor.	80
4.2	Commonly employed GoogLeNet architecture with illustration of inception block for classification applications.	83
4.3	Illustrating an example of neural style application for adapting Van Gogh's starry night style to dog image.	85
4.4	Illustrating the repeatability (in terms of test set image level accuracy) of the transfer learning experiment with varied random seed point for initializing fully connected layers.	90
4.5	Faster convergence and higher accuracy for fine-tuning pre-trained model (transfer learning) in comparison to training a model with random initialization.	90

List of Figures

4.6	Identified potential response for normal, AMD and DME cases with low responses in case of normal image, peak responses at drusen depositions in case of AMD and high responses along deformation side (right) for DME case. Such information is more enriched in comparison to HOG descriptor's depiction. The color bar represents convolution response of the identified filter with input image or previous filter response.	92
4.7	Visualizing the maximum response of GoogLeNet's 21 st layer for distinction between normal, AMD and DME cases where the responses are clearly distinct from each other which reduces the burden on a classifier. The color bar represents convolution response of the identified filter with input image or previous filter response.	93
5.1	Illustrating the neural network which functionally equivalent to sparse coding framework for image super resolution.	100
5.2	Illustrating the impact of separable filters. The result of image convolved with a 15×15 filter is equivalent to image convolved with separable filters decomposed from corresponding 15×15 filter.	102
5.3	Illustrating the existence of common separable filters between distinct 2D filters.	102
5.4	Adapting SCCNN for super-resolution to restore high definition noise free image.	106
5.5	Proposed approach fusing edge aware filters and learning filters through geometrical constraints.	107
5.6	Training and testing process of conventional (adapted) SCCNN and modified SCCNN for restoring sparsely sampled retinal OCT images.	110
5.7	Illustrates the capability of reconstructed image through proposed and base line (Sparse coding) from a given interpolated image.	111
5.8	Illustrates the absolute error between expected noise free image and estimated noise free images based on various approaches.	112
5.9	Learning characteristics of conventional and modified architecture on training and testing set at each epoch.	114
A.1	Block diagram for modeling and predicting various retinal layers through statistical representations	126
B.1	Capability of segmenting pathological image based on multiple layer-specific edge model.	127

List of Tables

2.1	Mean absolute difference evaluation across the classical graph approach (AN), kernel guided-graph approach (AD) and layer-specific edge guided-graph approach (proposed) in comparison to Expert 1 and Expert 2 annotations.	56
2.2	F-score evaluation across the classical graph approach (AN), kernel guided-graph approach (AD) and layer-specific edge guided-graph approach (proposed) in comparison to Expert 1 and Expert 2 annotations. The lower boundary of the choroid layer is unavailable, the F-score is computed for only the first seven layers.	56
3.1	Decision pooling performance across proposed and baseline algorithms . .	75
4.1	Cross-validation accuracy of transfer learning with various class weights and loss weights settings	94
5.1	Comparison of adapted SCCNN, modified SCCNN, sparse coding and structured random forests across various metrics	115

List of Symbols and Abbreviations

List of Symbols

μ	Mean
A	Magnitude map of HOG
BC	Binary contour mask
BL	Binary label mask
C	Contour image
D	Subset of data matrix where data points of same class are grouped
d	Feature vector or feature space dimension
DBC	Dilated BC
F	Feature tuple
\mathbf{f}	Vectorized feature patch
H	Histogram map of HOG
I	Image
i, j, q, o	Iterators
K	Computed distance of a sample to each cluster centroid
k	Number of classes
L	Label image
l	Class labels

List of Symbols and Abbreviations

\hat{l}	Predicted class labels
m	Membership value of FCM
M	Set of mean vectors
n	Retinal layer number
P	Feature patch
R	Responses of ANN hidden units
r	Quantization bins of a continuous value
u	Computed membership of a sample to each cluster centroid
x, y	Spatial coordinates of image
Y, N	Spatial locations in image with and with out edge
z	Low dimensional binary vector in SRE

List of Abbreviations

AMD	Age-related macular degeneration
ANN	Artificial neural network
CNN	Convolutional neural networks
DME	Diabetic macular edema
HOG	Histogram of oriented gradients
IAB	Image analysis blocks
LR	Learning Representations
LRN	Local response normalization
ML	Machine learning
OCT	Optical coherence tomography
PSNR	Peak signal to noise ratio
RF	Random forests
RPE	Retinal pigment epithelium

List of Symbols and Abbreviations

SC	Sparse coding
SCCNN	Sparse coding based CNN
SD-OCT	Spectral domain optical coherence tomography
SGD	Stochastic gradient descent
SRF	Structured random forests
SRE	Structured random forests for edges
SSIM	Structured similarity index measure
SVM	Support vector machine
TD-OCT	Time domain OCT
tSNE	t-distributed stochastic neighborhood estimation

Artwork Licenses

Graphical abstract source images are adapted from

http://people.duke.edu/~sf59/Chiu_BOE_2014_dataset.htm

http://people.duke.edu/~sf59/Srinivasan_BOE_2014_dataset.htm

http://people.duke.edu/~sf59/Fang_TMI_2013.htm

Figure 1.3 is adapted from

<https://www.flickr.com/photos/nationaleyeinstitute/7544655864>

Figure 1.4 is adapted from

http://media.indiatimes.in/media/content/2015/Mar/1113805_1437743576.jpg

Figure 1.5 is adapted from

http://obel.ee.uwa.edu.au/wp-content/uploads/2015/01/Fig_2_SDOCT.png

Figure 1.6 is adapted from

http://people.duke.edu/~sf59/Srinivasan_BOE_2014_dataset.htm

Figure 1.7 is adapted from

http://vsri.ucdavis.edu/gallery/retinal_images

Figure 1.8 is adapted from

http://people.duke.edu/~sf59/RPEDC_Ophth_2013_dataset.htm

Figure 1.9 is adapted from

http://people.duke.edu/~sf59/Chiu_BOE_2014_dataset.htm

Figure 1.10 is adapted from

Artwork Licenses

<http://csunplugged.org/image-representation/>

Figure 1.11 is adapted from

https://commons.wikimedia.org/wiki/File:Signal_processing.png

<http://maxpixel.freecreativity.com/Image-Painting-Painted-Butterflies-Two-Paint-1635747>

https://commons.wikimedia.org/wiki/File:Korean_text_test.svg

<https://commons.wikimedia.org/wiki/File:SchrodingerEquation.svg>

https://upload.wikimedia.org/wikipedia/commons/thumb/d/d9/Gaussian_2d_1.svg/640px-Gaussian_2d_1.svg.png?1496364664341

https://en.wikipedia.org/wiki/Benoit_Mandelbrot#/media/File:Mandel_zoom_08_satellite_antenna.jpg

https://en.wikipedia.org/wiki/Random_geometric_graph#/media/File:Random_geometric_graph.svg

Figure 1.12 is adapted from

<https://www.slideshare.net/tushyaop/optical-coherence-tomography-38710702>

Figure 1.13 is adapted from

http://people.duke.edu/_sf59/Chiu_BOE_2014_dataset.htm

Figure 1.15a is adapted from

http://people.duke.edu/_sf59/Srinivasan_BOE_2014_dataset.htm

Figure 1.16 is adapted from

http://people.duke.edu/_sf59/Fang_TMI_2013.htm

Figures 2.3a, 2.4a, 2.5, 2.7, 2.8 and 2.9 are adapted from

http://people.duke.edu/_sf59/Chiu_BOE_2014_dataset.htm

Figures 2.3a, 4.1, 4.6 and 4.7 are adapted from

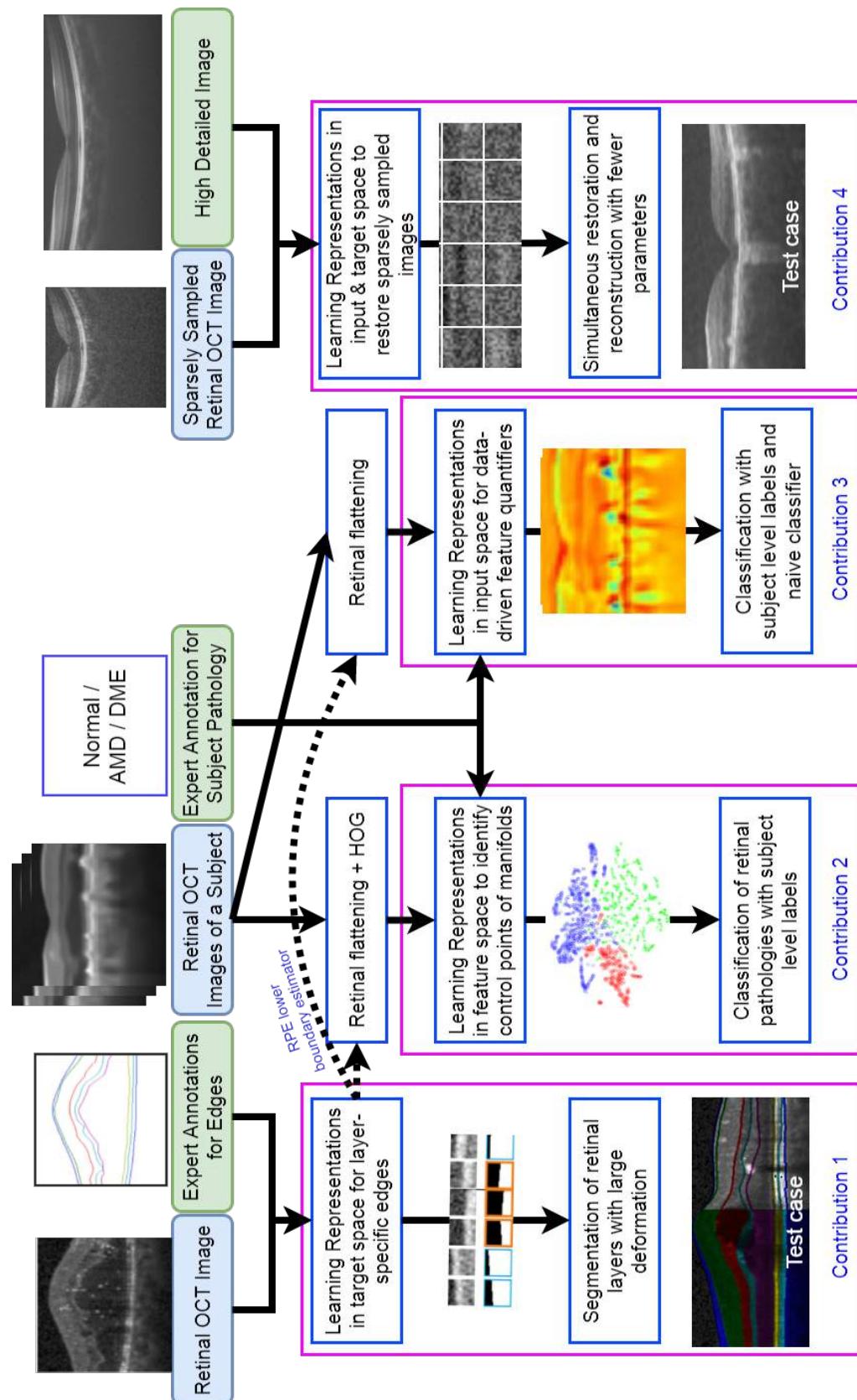
http://people.duke.edu/_sf59/Srinivasan_BOE_2014_dataset.htm

Figure 4.3 is adapted from

<http://khabaroff.dropmark.com/291878/7114916>

Figures 5.2 and 5.7 are adapted from
http://people.duke.edu/~sf59/Fang_TMI_2013.htm

Graphical Abstract



Abstract

Vision is an integral part of human sensory network and retinal pathologies can lead to irreversible vision loss. In early precise diagnosis of such pathologies, noninvasive methods like optical coherence tomography (OCT) is employed to study disease progression in retina. Automated retinal screening by OCT imaging is augmented through machine learning based image analysis blocks (IAB). However, such IAB suffers with limitations in analyzing OCT images having artifacts, poor signal-to-noise ratio and structural deformation in different retinal pathologies. Recent advances in learning representations (LR) including structures can address such medical image analysis problem appropriately through transforming information to engineering knowledge. Present work addresses various aspects of IAB for retinal image analysis in segmenting retinal layers, classifying retinal pathologies, quantifying multiple pathologies through data-driven descriptors and restoring sparsely sampled images. Improvement of individual block in IAB is achieved independently through learning representations in various spaces (input, feature and target) and associating each representation-space combination with functionality of an IAB block. This is achieved through equipping structured random forests with the capability of predicting layer-specific edges for segmentation. Guiding Fuzzy C-Means to model manifolds rather boundaries between classes for classification. Transfer learning feature quantifiers for effective quantification of pathologies. Finally, image restoration is performed by inducing low-rank filters in deep learning architectures. The efficacy of LR is validated by considering Duke's retinal OCT benchmark datasets. The retinal layers of pathological subjects are delineated with an average F-score of 0.86 and 0.85 in comparison to experts' annotations. Feature space representation modeling and data-driven feature quantifiers based approaches have elucidated retinal pathologies with a decision pooling accuracy of 0.94 and 0.91 with only subject level labels. The restoration block had recovered sparsely sampled (50%) retinal images with 28.15 peak signal to noise ratio with fewer parameters. The classification and restoration blocks augments remote clinics through identification of pathological subjects and reducing the data to be transferred to main clinics. Hence the present study depicts the efficacy of improved retinal layer delineation, classification and feature quantification of pathologies in addition to restoration of sparsely sampled OCT images. The future scope involves fusion of data-driven feature descriptors and guided Fuzzy C-Means for end-to-end training of retinal image analysis framework.

Keywords: Optical coherence tomography, retinal image analysis, machine learning, knowledge representation, fuzzy membership, diabetic macular edema, age-related macular degeneration.

Chapter 1

Introduction

“The aim of art (science) is to represent not the outward appearance of things, but their inward significance.”

—Aristotle

“Spoon feeding in the long run teaches us nothing but the shape of the spoon.”

—E.M. Forster

Automated modeling of patterns for interpretation and recognition has been a common practice for reducing expert’s workload and augmenting performance. Machine learning (ML) is widely employed for such automated learning and has been used in various streams of the medical community. ML based image analysis algorithms are adapted for automating visual inspection based screening and they have demonstrated exemplary performance in radiology (Wang & Summers 2012), cytology (Cruz & Wishart 2006) and ophthalmology (Abramoff et al. 2010). This work considers improving various integral blocks to automatically screen retinal pathologies, insofar as the retina is the most vital sensory organ and retinal pathologies have surged in the current generation (Bressler 2004, Bourne et al. 2014). Ophthalmologists employ optical coherence tomography (OCT) to screen retinal pathologies, because it is the only device to image

cross-sectional information with micrometer resolution in real-time. Current work targets ML based image analysis blocks to screen retinal pathologies based on OCT images. The employment of block combinations is subjective to the screening approach i.e., semi-automated or automated. In case of semi-automated screening (Chiu et al. 2012) only a segmentation block is employed as it can delineate retinal layers and depict layer thickness profiles so that an ophthalmologist can attribute such profiles to various pathologies (Chauhan & Marshall 1999, Fuller et al. 2007, Farsiu et al. 2014). However, automated screening framework (Chen et al. 2013) includes classification along with segmentation to detect pathologies. Irrespective of application ML based blocks have failed in multiple cases (Rathke et al. 2014), because experts identify features and interpret underlying representations, whereas traditional ML approaches are anchored around mapping features to a probability space (Bishop 2006). This has led to unsettling results (Lafferty et al. 2001, Kortschieder et al. 2011) and raised the need for modeling representations not only in medical domains but also in intricate domains ranging from natural language processing to decision making (Joachims et al. 2009). Thus, advanced ML algorithms are capable of modeling structures and representations in various spaces (Kortschieder et al. 2011, Bengio et al. 2013, Domingos 2007). Even though such algorithms are recently popular in the machine vision community, adaptation of such algorithms for the medical community is still an open problem (Rathke et al. 2014, Tajbakhsh et al. 2016).

Reducing an ophthalmologist's workload through image analysis has motivated biomedical researchers to automate blocks of retinal image analysis (Abramoff et al. 2010, Marleen 2016), namely image segmentation, classification, restoration, etc. An individual block of image analysis operates on different spaces such as the input space (image), transformed space, feature space and target space (or domain) (Domingos 2007, 2012, Bengio et al. 2013). Thus, this thesis is anchored around the modeling of representations in various spaces of retinal OCT images and in improving attributed blocks of retinal image analysis (see Figure 1.1) in health and pathologies. The considered pathologies

are age-related macular degeneration (AMD) and diabetic macular edema (DME). The flow of objectives have been set based on the hierarchy of spaces in which representations are learned i.e., target, feature and input. Additionally, the classification and feature quantification processes are dependent on layer delineation.

The autonomy of blocks in IAB is maintained to evaluate each image analysis block independently. The other reason being employed datasets are acquired from different OCT machines. The explored spaces and anticipated representations include the following: predicting an image edge structure in the target space for semantic segmentation. Identification of control points representing class manifolds in the feature space for effective classification. Heuristics free hierarchical quantification of abstract representations of retinal pathological aberrations in the input image (space). The decomposition of representative quantifiers (feature quantifiers) for input space, with application in image restoration. The last application identifies representations in input space to restore representative target image and hence involves crafting representations in both the input and target spaces.

Duke’s OCT databases (Chiu et al. 2015, Srinivasan et al. 2014, Fang et al. 2013) are considered for evaluation of each objective as they are publicly available for research and are commonly employed in retinal image analysis community as benchmark datasets. The datasets are stored in either .mat files or loss-less image formats and packed as zip file, so there is no loss of information due to web transfer.

1.1 Objectives and Scope

The aim of this thesis is to adapt and modify conventional machine learning algorithms to learn representations in various (input, feature and target) spaces in enhancing the performance of various image analysis blocks for retinal OCT. The following objectives have been set to achieve this aim:

1.1. Objectives and Scope

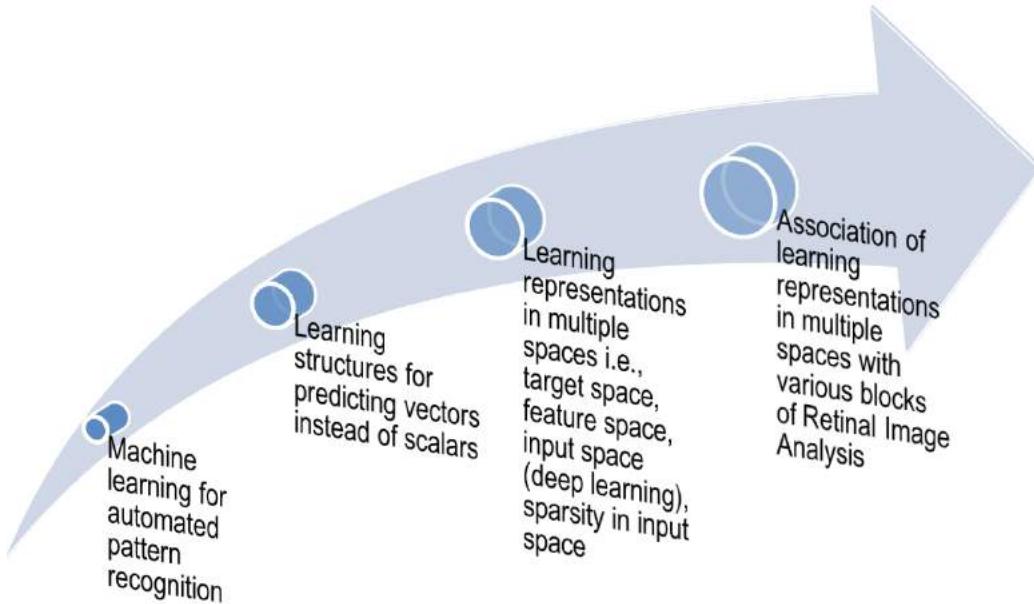


Figure 1.1: Illustrates the evolution of machine learning from the perspective of medical image analysis and employing modeled representations in various spaces with different blocks of IAB for retinal image analysis in this thesis.

- Modeling the edges of various retinal layers in an OCT image for accurate delineation of layers in health and pathology.
- Developing a classifier capable of representing the manifolds (in feature space) belonging to various pathological conditions (AMD, DME and Normal) of retina.
- Transfer-learning data-driven feature descriptors to effectively represent of multiple retinal pathologies in feature space and evaluate them with a naive classifier.
- Reducing the number of parameters to simultaneously restore and reconstruct a noise-free image from a sparsely sampled OCT image with a limited trade-off of information.

The motivation for choosing prior mentioned aim is illustrated in Figure 1.2.

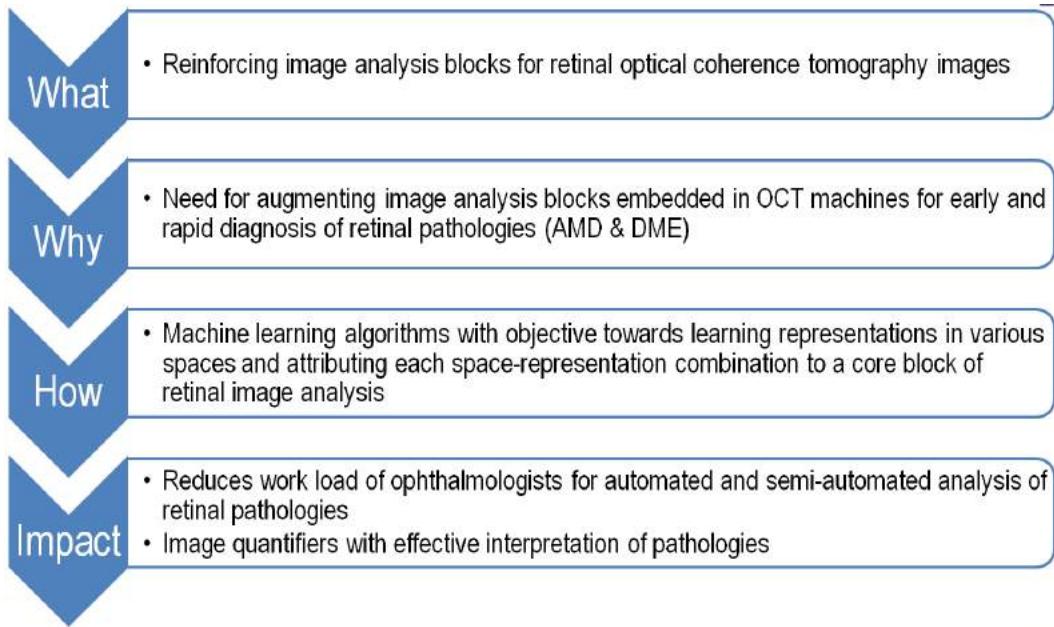


Figure 1.2: Illustrating the motivation and scope of considered aim.

1.2 Literature Review

This section initially depicts the need of retina which is followed by imaging modalities for early diagnosis of retinal pathologies. It also includes various blocks of image analysis and their basic functionality. Finally, the section illustrated the evolution of learning representations from machine learning.

1.2.1 Retina

The origin of vision is credited to the big bang of evolution which triggered the concept of competition for food and resulted in the inception of intelligence (Parker 2003). The level of intelligence increased on account of numerous aspects, one of which is the accumulation and accommodation of information from eye (Kobayashi & Kohshima 2001, Roorda & Williams 1999). The eyes aids to interpret the perturbing environment and help the system to adapt accordingly (Bourne et al. 2014). Structural complexity has been an integral part of the eye, embedding multi-functionality in the organ (Roorda & Williams

1.2. Literature Review

1999). The eye is an important sensory organ due to its contribution of information to the brain for leading a perturbation free life (Bourne et al. 2014). An illustration of its parts is shown in Figure 1.3 depicting light rays focused on the retina. In abnormal cases like myopia, there is shift of the focal plane from retina, leading to blurred vision. Many people from current generation suffers with eye abnormalities due to the lifestyle and source of all abnormalities need not be eye. These range from reversible vision loss (refractive errors, cornea crystallization) to irreversible vision loss (glaucoma, age-related macular degeneration). Major irreversible vision loss is attributed to retinal pathologies (Bourne et al. 2014, Bressler 2004).

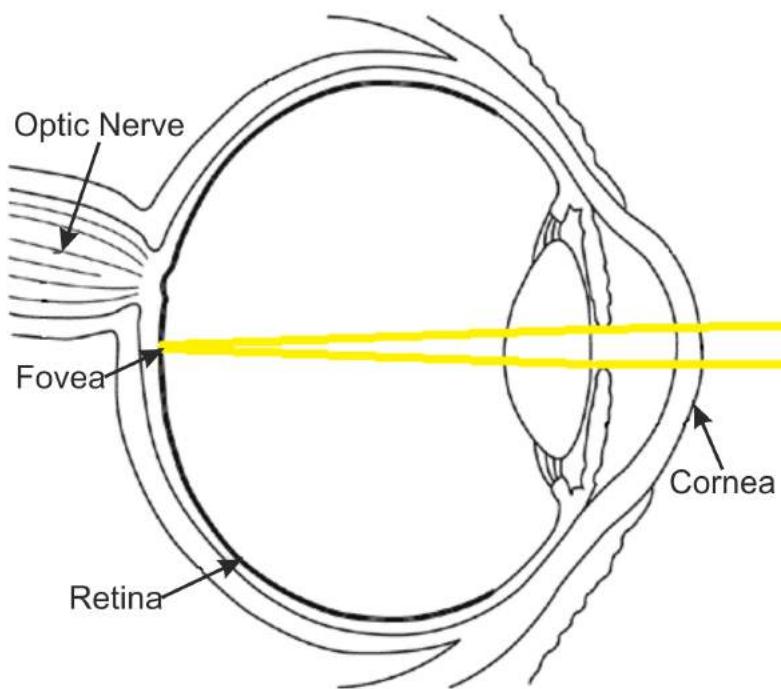


Figure 1.3: A sagittal plane of eye illustrating the ray diagram of eye lenses and the need of retina to be on the focal plane.

The retina is a crucial multi-layered structure comprised of light sensitive tissue with functionality analogous to a smart charge coupling device in a camera i.e., transforming incident photons into encrypted electrical signals (or nerve impulses, in the case of the retina) (Baylor & Fuortes 1970, Polyak 1941). A sensor transfers raw signals to the

1. Introduction

imaging pipeline, but the retina itself process the information before relaying it to the visual cortex in the brain through the optic nerve. One such process is the differentiation of chroma through a different set of cones cells in the retina (Roorda & Williams 1999). Any disruption of retina's functionality due to pathology or trauma, leads to a disruption in the conversion of photons to signals, or information processing, or transferring information to the brain. This leads to misguided or incomplete visual information (see Figure 1.4), which affects various aspects of life from knowledge acquisition to daily activities. The retina also acts as a surrogate for diagnosing various systemic diseases (Klein et al. 1992) and organ localized diseases (Klein et al. 1992, Frohman et al. 2008). Retina abnormalities need not originate from retinal pathology; they can be an effect of systemic or central nervous system pathologies. Thus, the retina also acts as a surrogate to the pathologies pertaining to the central nervous system, insofar as it is the only part of central nervous system that can be visualized without intervention (London et al. 2013).

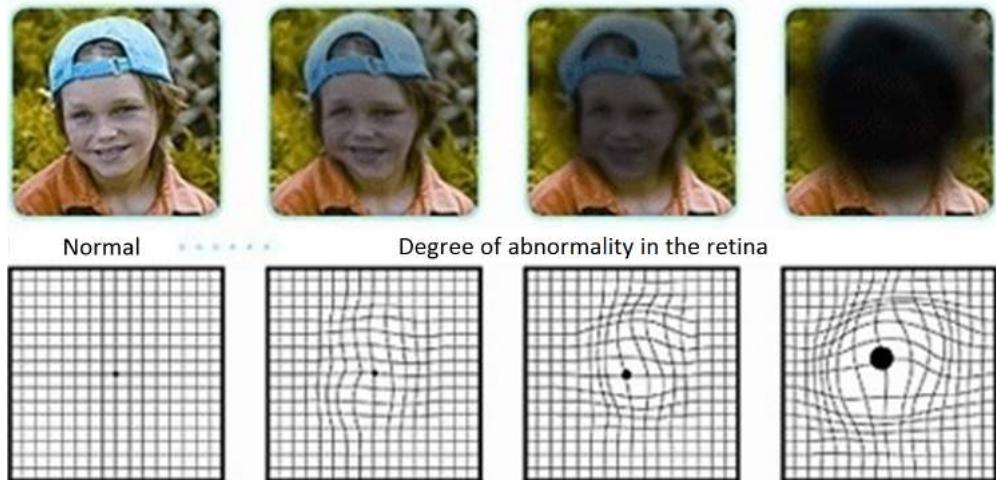


Figure 1.4: Illustrating the increase of geographical atrophy in retina with grade (increasing) of AMD and corresponding vision loss.

1.2. Literature Review

1.2.2 Retinal imaging modalities

To recede irreversible vision loss and regress reversible vision loss regular vision checks are suggested by ophthalmologists (Ciulla et al. 2003). A fundus camera is one of the primary imaging devices for the retina in ophthalmology, owing to its ability to visualize macula and blood vessels non-invasively (Yannuzzi et al. 2004). These features provide salient information for diagnosing various irregularities in the retina and other organs, for which the retina acts as a surrogate. Functional dimensionality has been included through fluorescence angiography for early diagnosis of a few diseases like age-related macular degeneration (AMD) through Lipofuscin or diabetic macular edema (DME) through fluid leakage (Cunha-Vaz et al. 1975) in the retina. Autofluorescence properties of extracellular deposition (Lipofuscin) (Bird et al. 1995) and fluid leakage have added a functional dimension to information in ophthalmic imaging which has boosted early diagnosis of retinal disorders. However, information acquired through all such imaging instruments is merely a surface projection of the actual information. This makes early diagnosis difficult as it needs cross-sectional information. Profiling layers deformation and volumetric deformations is crucial for early diagnosis and accurate prognosis and this is possible only with cross-sectional information. Ophthalmologists employed low coherence interferometer based subsurface imaging device called as optical coherence tomography (OCT) (Adhi & Duker 2013) in order to image cross-sectional information of the retina non-invasively.

Existing *in-vivo* imaging techniques such as MRI could not image retinal layers, owing to limitation in resolution (Cheng et al. 2006). OCT is a subsurface imaging technique with a penetration depth of few millimeters, depending on the refractive index of the sample and with an axial resolution of few microns (Fercher et al. 2003, Schmitt 1999). OCT was invented in an optical communication lab where experiments were performed to identify the distance of the damaged optical fiber from the source using light (specifically, lasers). The same experimental setup has been extended to image a

1. Introduction

biological sample, but the reflected light response is rapid, such that no sensor could previously capture the reflected signals with small time resolution. This lead to the inclusion of the interferometer to capture constructive interference patterns and translate them to A-scans, which are then concatenated to form a B-scans (Huang et al. 1991, Hee et al. 1995). The laser source is limited to the near infrared range for better penetration without loosing resolution.

The optical path of photons in OCT begins with the generation of a laser split using a dichromatic mirror (i.e., a beam splitter), where one is incident onto the sample and the other onto a reference mirror. Both signals are reflected back where a constructive interference pattern is obtained from which time-delay is estimated. This is translated into a single pixel in the A-scan line, whose location corresponds to the placement of the movable reference mirror. Moving the reference mirror and imaging the sample fills up the information in an A-scan line (Fercher et al. 2003). This is the imaging protocol employed by time domain OCT. The setup comprises moving parts and requires considerable acquisition time. Thus frequency-domain based approaches have been introduced to improve the imaging speed and quality in the past decade. One of the refined approaches, spectral domain OCT (SD-OCT) is extensively used for retinal imaging (De Boer et al. 2003, Fercher et al. 2010). Time domain OCT is dependent on a time delay, whereas SD-OCT operates based on the wavelength to determine the spatial location of the reflected photons. Rather than moving a reference mirror, the A-scan line information is acquired in one instance using SD-OCT through evaluating the frequency spectrum of interference between the reflected signal from a fixed reflective mirror sample and the sample. An illustration of the photon path from the source to A-scan generation is provided in Figure 1.5. A scan lines (axial scans) are acquired in an instance, which allows the SD-OCT to acquire multiple B-scans (see Figure 1.6) in same imaging time as TD-OCT. Such multiple B-scans at successive neighboring locations allows for a 3D imaging of the retina (Tomlins & Wang 2005). SD-OCT machines

1.2. Literature Review

such as SPECTRALIS from Heidelberg Engineering are currently being used in clinical practice for rapid and volumetric imaging of retina with high signal strength. The methods proposed for image analysis are based on B-scans, but not 3D data. B-scan based approaches are computationally inexpensive and can be expandable to 3D data if required.

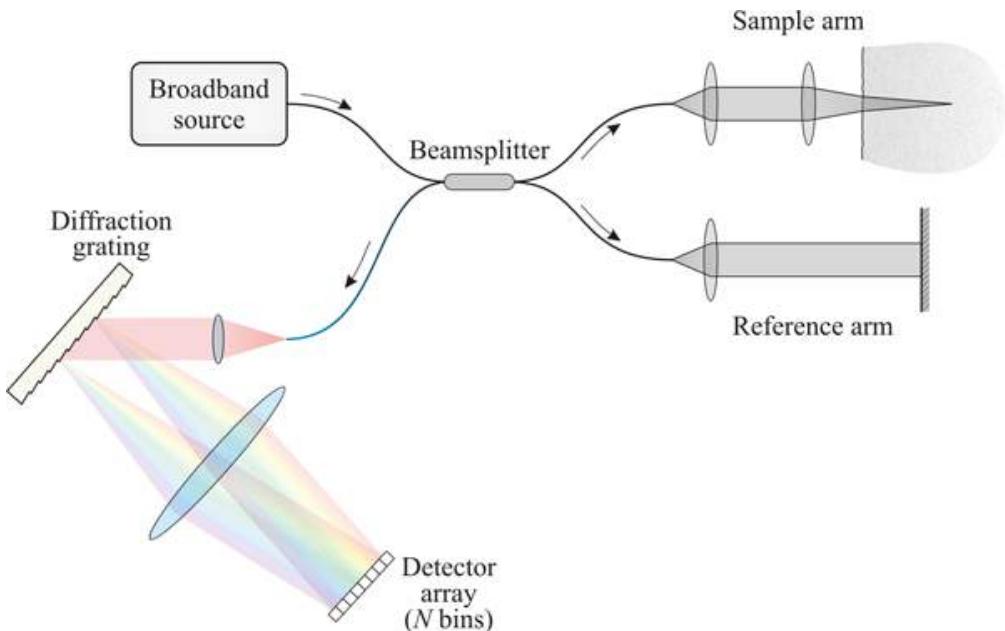


Figure 1.5: The schematic of retinal SD-OCT in respect to optical path of photons illustrating the employment of grating and spectrometer for time-space encoding which is crucial for rapid imaging.

Figure 1.7 shows a normal subjects's retina imaged with SD-OCT and its comparison with histology. This illustration clearly depicts various layers including inner plexiform layer (IPL) and outer plexiform layer (OPL). Multiple experiments have validated the anatomical information imaged through OCT (Lu et al. 2012, Spaide & Curcio 2011, Glocsmann et al. 2003). To enhance the OCT as a diagnostic tool various benchmarks has been established with comparisons to fundus and florescence angiography and microperimetry (Mujat et al. 2005, Brar et al. 2010, Vujosevic et al. 2006) for grading pathologies related to visual acuity (Wong et al. 2015). Studies have been carried out

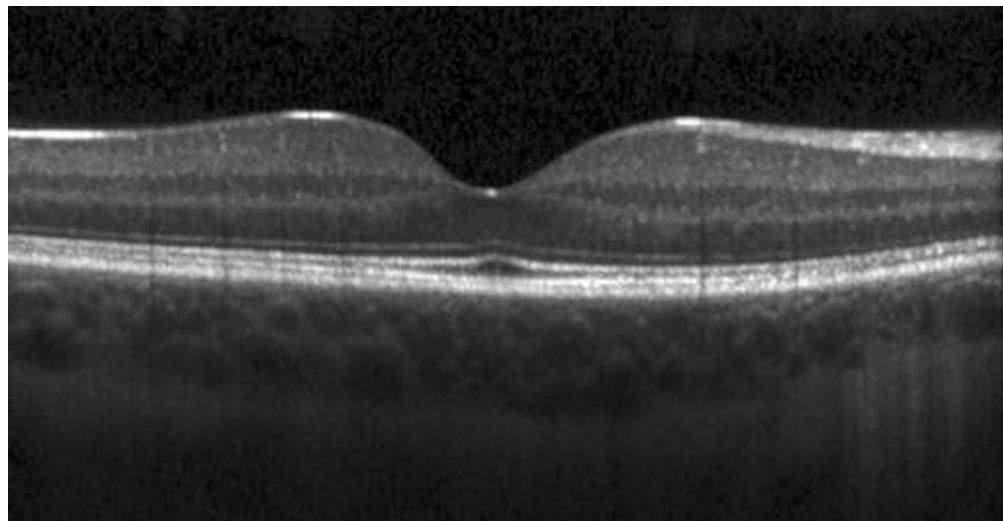


Figure 1.6: Depicting the cross-sectional information (non-invasive) of normal subject's retina through SD-OCT B-scan.

to map retinal layer profiles to various pathologies (Kim et al. 2008, Koozekanani et al. 2001, Kromer et al. 2014, Gaudric et al. 2015, Zangwill et al. 2001) including dry AMD (see Figure 1.8) and DME (see Figure 1.9).

1.2.3 Image analysis

Analyzing such retinal OCT images (namely, B-scans) is crucial for early diagnosis. Automation has improved and speed up the screening process (Marleen 2016). An

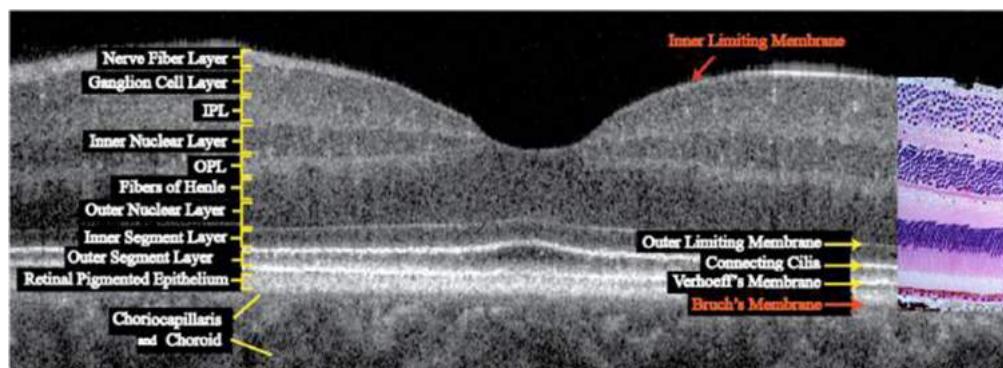


Figure 1.7: Illustrating various layers of retina imaged with OCT and comparing with retina histology.

1.2. Literature Review

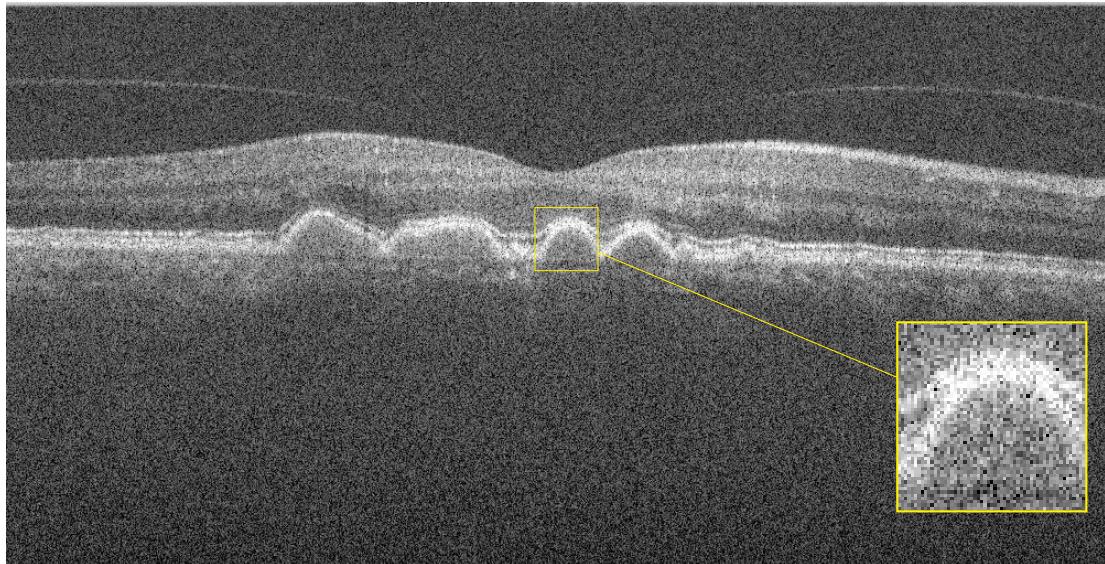


Figure 1.8: Depicting the drusen deposition between RPE and BM in OCT B-scan image of an AMD subject.

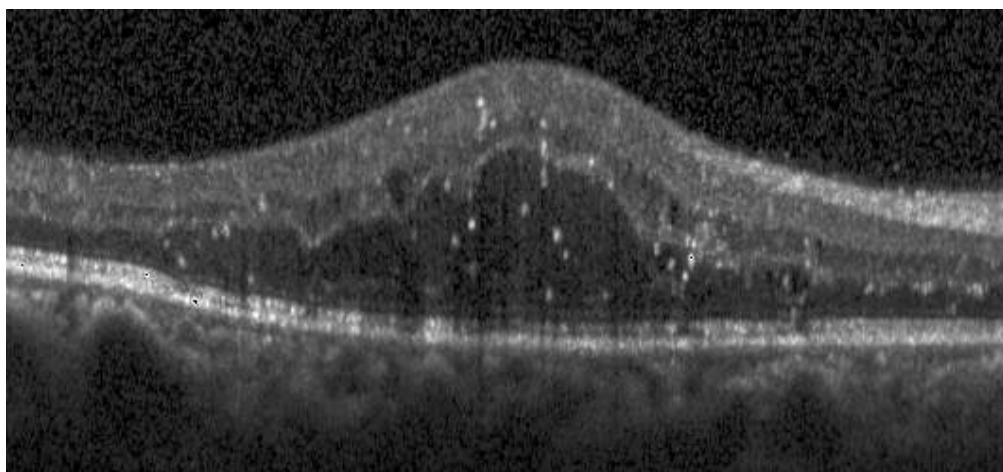


Figure 1.9: Illustrating the detachment and deformation in OCT B-scan image of a DME subject.

overview of the four basic blocks of image analysis is given below.

Segmentation: The characterization of the object should start with a delineation of the object and is the job of the segmentation block. The common hypothesis employed is discontinuity, i.e., two objects in the image can be delineated only if there is some

1. Introduction

distinction in pixel intensity or texture (Pratt 2000). Segmentation algorithms are thus a combination of abstract information from images (e.g., intensity distribution, edges, etc.), a metric to identify the difference in information between neighboring pixels (e.g., entropy, variance, distance, etc.) and optimizer (e.g., line searching, dynamic programming) to identify boundary pixels between objects. Segmentation algorithms range from thresholding (Delon et al. 2007, Tobias & Seara 2002) to graph theory (Boykov & Jolly 2001, Grady 2006, Shi & Malik 2000) including active contours (Vese & Chan 2002). There exists multiple combinations and there is no rule of thumb for the best algorithm. The performance of any algorithm is subjective to application and image information. However, active contours and graph-based approaches are commonly employed. Machine learning has improved the segmentation block but resulted in discontinuous predictions (Lafferty et al. 2001, Kortschieder et al. 2011). Thus recent successful segmentation algorithms include refinement block as a post-processing step (graph based approaches like conditional random fields).

Feature Quantification: Upon delineating the object, it needs to be described quantitatively to generate a set of numbers. This way, a set of rules can be constructed to recognize the object. Clinically significant qualitative descriptors are translated as quantitative descriptors of the object (Fernández 2005). Frequently an existing quantitative feature descriptor (for other applications) corresponding to the clinical descriptor is identified rather than constructing an application specific descriptor (Fujiyoshi 2007). This leads to low accuracy, however, programmers include potential feature descriptors (also from different application) (Quellec et al. 2010) and iteratively remove ambiguous features based on their accuracy, through feature selection approaches. An alternate approach is to adopt feature extraction, where the complete feature set (including ambiguous features) is transformed to a potential feature set (Chen & Lin 2006, Kohavi & John 1997). The dependency level of feature quantification on the segmentation block changes based on the amount of information considered and the kind of features to be

1.2. Literature Review

quantified. Recent algorithms for feature extraction are potential enough to construct data subjective feature quantifiers from standardized image information (Srinivasan et al. 2014).

Classification: Quantified or extracted information from images are termed features or feature vectors and in a feature space, each feature will be depicted as a point. So each image is transformed into a point in feature space. A supervised classifier considers a training set of feature vectors with corresponding class labels and a classifier is constructed with the objective of identifying the class label if given a test feature vector. The classical practice involves covering the feature space with data points by transforming images from various classes. This is referred to as data augmentation. During prediction given an image, the corresponding data point location in feature space is computed. Thus, the label of nearby data points is considered as an image label or class (K-nearest neighbors). Such practices need ample data and storage. Therefore, classifiers are aimed at modeling the decision boundary between datapoints from various classes (Domingos 2012). The datapoints need not be linearly separable, so classifiers are able to transform the feature space autonomously (Yang & Jin 2006) or heuristically (Gaussian or Radial Basis Function kernels in SVM) where data points are linearly separable. The autonomous approach is computationally expensive in large data sets and heuristics are prone to subjectivity. Therefore, nonlinear boundary modeling is performed through ensemble learning (Breiman 2001), where multiple classifiers are bagged together. This way, even a linear classifier can model non-linear boundaries through a piece-wise linear approach.

Reconstruction: Imaging involves a transformation of real-world information to an arrangement of pixels with a facet of complete information, i.e., attenuation in the case of computed tomography scans, water molecules concentration in the case of MRI and lucidity in case of OCT. All information cannot be imaged or stored, however, due to limitations in the imaging protocol or sample. This, leads to corrupt or missing in-

formation. The current block is employed as preprocessing or postprocessing (image transmission for telemedicine) in order to restore missing information (restoration) and remove the noise (filtering) by incorporating the knowledge of distortion functions (e.g., Gaussian or signal to noise ratio) and noise (Bow 2002). In the case of medical applications, pinpointing such information is difficult. Thus, approaches with a hypothesis of homogeneity are successful. Such algorithms exploit the characteristics that image pixels or patches should be similar to neighbouring pixels or patches. However, the exclusive consideration of pixel information or all neighbouring patches has led to blurring at the boundaries (Deng & Cahill 1993). Consequently, considering a subset of neighboring pixels or patches has improved the performance. This is known as adaptive filtering (Paris et al. 2008, Tomasi & Manduchi 1998, Westin et al. 2008, He et al. 2013) or block matching (Dabov et al. 2007). BM3D (Dabov et al. 2007) is one the most successful preprocessing approaches for medical and non-medical applications. Yet, sparse coding is more successful in the case of image restoration. Recently, machine learning approaches have been employed (Ryan Fanello et al. 2014, Burger et al. 2012) for restoration and filtering to avoid any prior knowledge and has been successful in both medical and non-medical applications.

1.2.4 Machine learning

Machine learning (ML) has been an integral part of multiple predictive frameworks (Bishop 2006) in the past few decades and it has been infused into new territories (Esposito & Malerba 2001), such as the medical sciences (Wang & Summers 2012, Marleen 2016). In the inception, it was a minor part of artificial intelligence, but its rate of growth and impact has made it a distinguished field (Bishop 2006). This is because predictive frameworks involve creating a model and estimating the target given test data. Modeling involves an infusion of domain knowledge by hard coding of rules, functions, transformations, etc. Machine learning automates modeling without explicit coding for

1.2. Literature Review

‘how to learn?’ as it involves ‘how to learn?’ based on data (Mitchell 1997). This thesis is exclusively concerned with supervised approaches (Kotsiantis 2007), such that the data employed for constructing a model should include a set of features and their corresponding target values (integers or real numbers) for each feature. Each feature is extracted information from each image for the current scenario and the target value is an integer or real number based on the functionality of the ML algorithm (whether classification or regression) (Bishop 2006). Thus, the model represents a surface used to separate or fit data points. An algorithm is called an ML algorithm if the predictive error of the constructed model reduces with each iteration (i.e., with more “experience”) (Mitchell 1997). Supervised ML algorithm involves training and prediction phases, during training, the model is constructed and during testing, the model is employed to predict target values of the same nature as those employed during training. This constrains the ML to applications with single value predictions like signal classification, image classification, voice recognition, etc.

1.2.5 Learning representations

Complex applications like semantic segmentation or voice translation require predicting an array of numbers, rather than single number (Joachims et al. 2009, Lafferty et al. 2001, Nowozin 2010, Kortschieder et al. 2011). For semantic image segmentation, the model should be capable of transforming every image into a segmented image where each pixel represents a target class value. Classical ML can be adapted for such complex application by predicting each pixel label independently. This offers a partial solution to the problem. Such an approach resulted in an inconsistency between neighboring predictions (Lafferty et al. 2001, Kortschieder et al. 2011). Thus, structured learning has been introduced to predict structured array of numbers (including neighboring predictions). This has led to modifications of loss functions optimized in classic ML like support vector machines (SVM) to predict structured output (Tsochantaridis et al. 2004, Joachims et al. 2009).

Structured ML has different variants, from Bayesian networks to deep learning including the graph-theory. Recently, algorithms modeling, predicting and recognizing structures have been grouped under structured ML (Domingos 2007). In multiple cases, such structures do have a meaning or at least symbolize something such as edges, objects, etc as shown in Figure 1.10. In computer vision or machine vision, these are termed as representations (Marr 1982, Fischler & Elschlager 1973, Porat & Zeevi 1988).

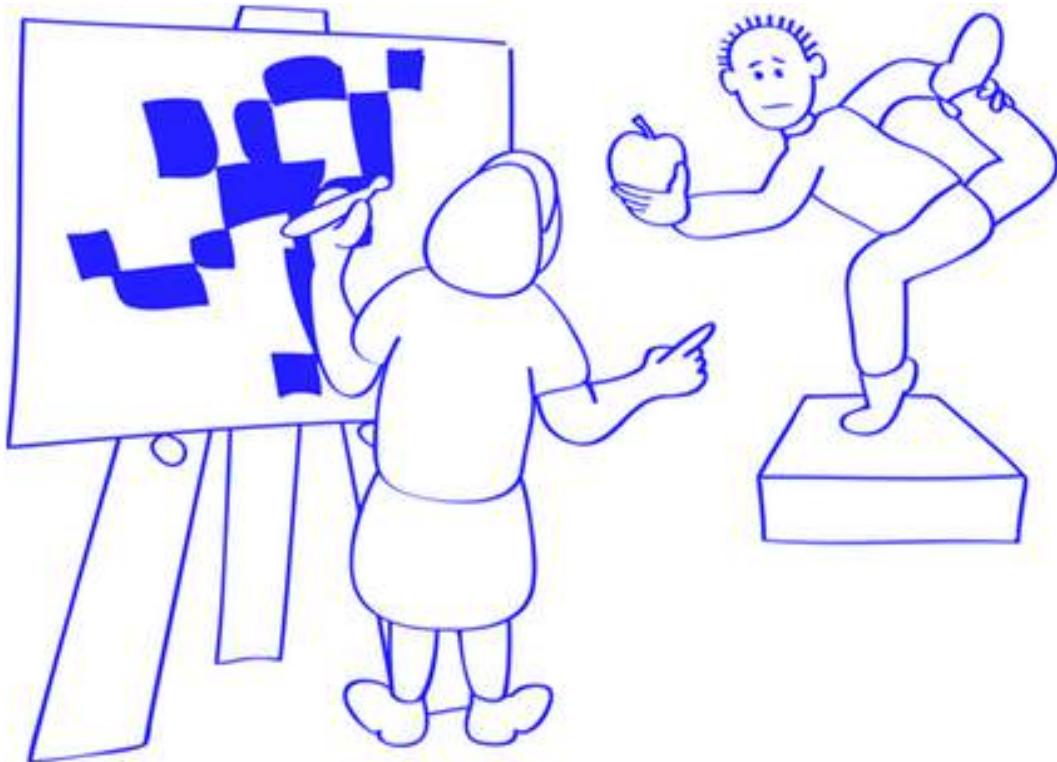


Figure 1.10: Representing an environment through skeleton structure with curves and lines instead of duplicating the information.

A representation algorithm has various facets (Domingos 2007, Bengio et al. 2013), but conventional definition in image processing community is “an algorithm employed for information processing with a target towards explicit entities” (Marr 1982). With the level of ML intervention into multiple applications the *Representation* converts a raw or preprocessed information into array of numbers encapsulating core knowledge for affective task performance (classification or regression) as shown in Figure 1.11. In Figure

1.2. Literature Review

1.11 the subfigures in information column depicts a signal, an image with objects from multiple scales and Korean text. Effective learning algorithms represent the information in the form of multi-scale sinusoidal, fractals and graph which depicts the underlying knowledge for augmenting subjective applications.

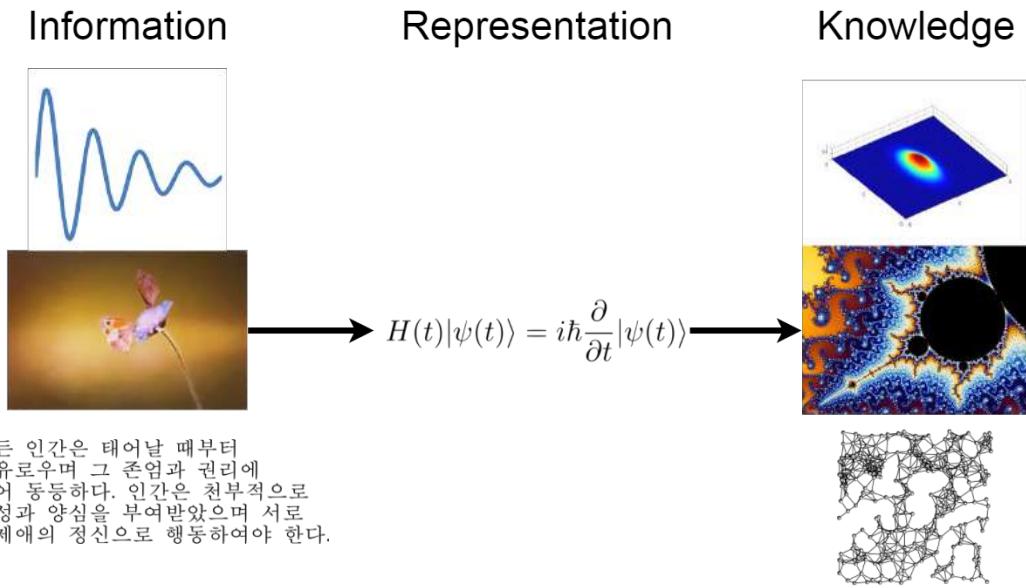


Figure 1.11: Illustration the role of *representation* for converting information from various domains into corresponding engineering knowledge.

The majority of employed algorithms for frameworks in the thesis are structured learning (structures in an input image or predicting in target space) and are aimed at explicit objectives in various spaces (e.g., the input space and target space). Consequently, they are treated as algorithms for learning representations (LR). The remaining part employs unsupervised learning approaches for pattern recognition, intrinsic loss functions in such algorithms are explicitly designed to explore the representations of underlying patterns, so the remaining part also falls under LR.

Recently, deep learning is a popular approach in the artificial intelligence community (Hinton 2007, Bengio et al. 2013), convolutional neural networks (Krizhevsky et al. 2012) are employed because they established state-of-the-art benchmarks in image recognition.

As LR was introduced in the context of deep learning, it was initially constrained to feature learning. Yet, it has the potential to be expanded to various functions other to hierarchical quantification of structures (i.e., hierarchical abstract patterns in deep learning) (Domingos 2007). There are multiple salient cases that need not model the structure, yet still encompass representations in the form of instances in case of k-nearest neighbors, hyperplane in case of SVM and set of rules in case of decision trees (Domingos 2012).

Currently, LR has many facets in machine learning community, owing to its impact on various sub-branches of machine learning like variational learning (Chen et al. 2017), feature learning (Bengio et al. 2013), structured learning, deep learning, etc. Because of its vastness only structured random forests (Dollar & Zitnick 2013), unsupervised learning and deep learning techniques have been considered in this work. The majority of the work in structured learning is developed around structured prediction. Nevertheless, structured prediction is a subset of structured learning; that is, algorithms with structure modeling but without structured prediction can still be considered structured learning (Domingos 2007). For example learning structures (lines, curves, etc.) in the input space (i.e., the image) for classification (Szegedy et al. 2015) is also structured learning. LR is majorly being driven by the computer science community and the data considered involve non-medical images (e.g., cats, dogs, cars, etc.). They need to be adapted to the medical applications (Tajbakhsh et al. 2016). As stated above, machine learning is being infused into multiple predictive applications. Yet, this thesis is more precisely focused on retinal optical coherence tomography analysis.

1.3 Related Work

In this section, a survey of literature related to the targetted applications made in this dissertation is reported.

1.3. Related Work

1.3.1 State-of-the-art for retinal layer segmentation

Retinal layer delineation or segmentation is an evolving and challenging problem in medical image analysis. A wide range of methods has been proposed, varying from graph-based segmentation Kafieh et al. (2013), Chiu et al. (2010), Mayer et al. (2008) to pattern recognition based prediction, in addition to classical intensity based approaches. Few works have articulated all such methods and categorized them based on the information employed (A-scans, images, 3D scans, edges, low- level features, etc.) and algorithm (active contours, graph approach, machine learning, etc.). Graph-based approaches (Chiu et al. 2010, 2012, 2015, Dufour et al. 2013), especially dynamic programming approaches, offer superior performance and have been widely employed owing to their limited computational complexity even under the presence of noise. Graph-based approaches are powerful enough for exhaustive layer delineation in resource-limited computational facilities through iterative approaches (Chiu et al. 2015). Graph-based approaches involve constructing a graph based on the edge and nodes, where edges are computed based on image pixel information. Conventional graph-based segmentation involves searching for optimal label assignment for each node that corresponds to each pixel in the image, so the label of each node is treated as the label of the pixel that translates to retinal delineation algorithms. The approaches considered for edge computation vary from method to method and are commonly based on the image’s low-level features, heuristics, machine learning predictions and regularization (Dufour et al. 2013, Chiu et al. 2015, Rathke et al. 2014, Kafieh et al. 2013). In OCT images, edges are computed and a dynamic program is employed to identify the shortest path from the top left node to the bottom right node on the image graph, which is then treated as a layer boundary. Graph-based segmentation approaches are concerned more with solving the graph rather than with constructing the graph. Indeed, the majority of graph approaches consider trivial gradient operators, resulting in high responses at all prominent layers. Heuristics (viz., the average layer width and arrangement) are considered in order to avoid ambiguity and

1. Introduction

constraints on the dynamic program’s search space.

Alternatively, approaches have induced the fuzzy search space through a shape regularizer (Dufour et al. 2013) to resolve ambiguity in information particularly due to shadow artifacts. Heuristic constraints on the search space have resulted in failed delineation in cases of rare patterns (or an emergence of new patterns). Machine learning approaches are favored for recognizing a wide variety of patterns and regularizers are constructed through probabilistic graphical models (Rathke et al. 2014). These approaches have demonstrated the feasibility of constructing a regularizer that comprehends wide patterns. Alas, they fail in case of pathologies, owing to large layer deformations like diabetic macular edema (DME). Methods targeting layer segmentation in DME subjects have thus employed models to predict edges or layer information, rather constructing the regularizers for the search space. Such methods have surpassed conventional algorithms. Kernel regression surpassed previous methods in terms of layer segmentation in DME subjects. Learning-based graph edge construction has enhanced the segmentation algorithm’s performance and established the state-of-the-art in multiple instances (Chiu et al. 2015, Rathke et al. 2014, Kafieh et al. 2013). Machine learning approaches to medical image segmentation are anchored around off-the-shelf learning methods (e.g., random forests, support vector machines, etc.). Recent developments in machine learning have led to extending structured prediction or learning representations to image processing applications and led to the introduction of a structured SVM and structured random forests (SRF) (Kontschieder et al. 2011). SRF as it accomodates advantages of random forests, which offer optimal risk minimization, robustness to noise in labels, etc. Indeed, an SRF has been modified to predict object edges rather than objects (Dollar & Zitnick 2013). Most recent approaches have employed adaptive weighted filters (de Sisternes et al. 2017) or feature representation (Montuoro et al. 2017) for better quantification of edge features rather relying on heuristics (Laplacian of Gaussian) completely. Finally graph based approaches delineate layers based on such edge features. However, it

1.3. Related Work

depends on heuristics (de Sisternes et al. 2017) at one point to remove the redundant feature responses.

Collectively, graph based segmentation algorithms have established state-of-the-art performance for delineating retinal layers and machine learning based predictions are used as edge during solving graphs. The limitations of conventional retinal segmentation algorithms are:

- Fails in case of large deformation in retinal layers due to pathologies.
- Fails to discriminate feeble layers as edges from different layers come closer.
- ML based approaches have inconsistency between predictions with in a neighborhood of pixels.

1.3.2 State-of-the-art for retinal pathology classification (AMD and DME)

Any image- analysis pipeline is incomplete without the identification or recognition of the target and current applications are designed to identify (or screen) the pathological subjects. Initial approaches for screening involved mimicking clinical practices, namely layer delineation and classification, insofar as retinal-width profiles can be indexed to various pathologies (Comander et al. 2011, Dai & Sun 2011, Farsiu et al. 2014, Fuller et al. 2007, Shahidi et al. 2005). Such classifiers are a curated list of decision rules practiced in ophthalmology and effectiveness of such rules depends on the accurate performance of the preceding block (Ehnes et al. 2014). That is, segmentation remains an evolving and challenging field. Such rules are limited to morphological variations or changes in width profiles. Recent descriptors include gradient-based features. Moreover, low-level feature descriptors with classifiers capable of modeling the decision boundaries in the feature space have been introduced in the screening pipeline. This reduces the level of dependency on the segmentation block (Quellec et al. 2010, Srinivasan et al. 2014,

Somfai et al. 2014).

Classifiers are supervised, unsupervised, or semi-supervised, based on the availability of data. However, supervised classifiers are commonly employed for image-based screening systems. This is because ample data is required to identify the representative population of data samples in the feature space and supervised approaches need only a limited amount of data (compared to unsupervised classifiers), because they model the decision boundary between the data samples from various classes in the feature space (Kotsiantis 2007). Support vector machines (Burges 1998), decision trees (Breiman 2001), etc. initially extrapolate the representations or statistics of the population based on the training dataset, which is sampled from the population. Then, the decision boundary (or boundaries, in the case of ensemble learning) is modeled as linear or nonlinear (piecewise linear). There are no rules of thumb for selecting a classifier to guarantee performance, because there are multiple tuning knobs (or hyperparameters) in the ML-based modeling process (Domingos 2012). This means that specialized ML expert knowledge is needed regarding model characteristics in various scenarios, in order to tweak the parameters for the required application. The majority of retinal pathology screening methods include an SVM (Srinivasan et al. 2014, Massich et al. 2016, Syed et al. 2016, Anantrasirichai et al. 2013, Khalid et al. 2017) or an ANN (Somfai et al. 2014, Devarakonda et al. 2016) out of which SVM is widely employed. The majority of such approaches aim to identify only a single pathology, but the proposed approach aims at identifying multiple pathologies (e.g., both AMD and DME) in a manner similar to (Srinivasan et al. 2014).

Transforming the current application for anomaly detection involves the adaptation of an unsupervised approach for detecting pathological cases (Sankar et al. 2016). Such approaches are successful provided that all variations of the normal subjects have been effectively modeled. This is not possible with sampled data. Semi-supervised approaches are not yet employed for retinal OCT screening. The primary reason for this could be

1.3. Related Work

that the majority of such approaches are graph-based (e.g., in random walks) which are computationally expensive and require large space when storing complex graphs (e.g., a Laplacian matrix is commonly stored). Dictionary learning can also be extended for pathology detection (Sun et al. 2017) but it needs over-complete dictionaries.

Collectively, classifying low-level (HOG) features quantified from pre-proceed image (retinal flattening and filtering) have established state-of-the-art performance for screening of retinal pathologies through reducing level of dependency over segmentation block. The limitations of classic classifiers for identification of pathologies are:

- Selection of wrong kernels or hyper-parameter can plunge the accuracies.
- Modeling the boundaries between classes rather the skeleton of manifolds in feature space.
- Imbalance in number of samples per class.

1.3.3 State-of-the-art feature descriptors for retinal pathologies (AMD and DME)

The previous section discussed classical algorithms for retinal screening. These algorithms employed width profiles as features (Zangwill et al. 2001, Comander et al. 2011, Farsiu et al. 2014, Shahidi et al. 2005) and translated them to representative features by considering the order statistics of width profiles (viz., the mean and standard deviation). In the case of AMD, the deposition of drusen beneath retinal pigment epithelium (RPE) can be reflected in the variation of the width profile from the RPE to Bruch's membrane (the upper boundary of choroid) across each A-scan and this can be quantified with the standard deviation (Farsiu et al. 2014). The amount deposited can be quantified through the difference in the means of subjects with and without a pathology, which is translated according to the severity of the pathology for diagnosis. The screening process involves only identifying the pathology, but not its severity. Thus, we limit the state-of-the-art

1. Introduction

to identification. In the case of DME, fluid deposition commonly occurs above the RPE, such that if new clinical features arise in the target pathology or need to quantify a new pathology, the entire target quantifiers changes (Medeiros et al. 2005). Constructing common features (in a pathology independent manner) is helpful when identifying various pathologies, rather than pushing each OCT image through a pathology-subjective screening system. Current applications target both AMD and DME pathologies. Hence, common feature descriptors capable of delineating subjects into three classes (normal, AMD and DME) must be considered. Some approaches have implemented simultaneous identification and delineation of drusen (Chen et al. 2013) and fluid deposition (Chiu et al. 2015, Fernández 2005) to classify AMD and DME cases, but they require a robust segmentation approach.

To reduce the degree of dependency on segmented layers, low-level image descriptors are employed. Such approaches depend only on retinal flattening, which involves identifying the lower boundary of the retina, which is one of the hyperreflective bands (Spaide & Curcio 2011) in retinal OCT images. Low-level image descriptors (e.g., a HOG i.e., histogram of oriented gradients) (Quellec et al. 2010, Srinivasan et al. 2014, Somfai et al. 2014) are employed, because they act as a partial representation of the pathology's features and collections of such descriptors can represent various pathologies with distinguishability in the feature space. From the perspective of linear algebra, they can act as a governing variable in an equation. The classifier then identifies the weights for the equation to separate the pathologies. The questionable aspect is the *number of features that need to be incorporated*, since not all existing feature descriptors can be employed. Thus, feature descriptors where each quantifies a partial aspect of the pathology are incorporated and filtered through feature extraction approaches such as principal component analysis (Albarak et al. 2013). Recently it has been established that deep learning has major contribution towards identification of retinal pathologies (Lee et al. 2017), which is inspired from a form of sparse coding (Sun et al. 2017). How-

1.3. Related Work

ever, majority of such approaches are designed for identification of single pathology but not for multiple pathologies.

Collectively, low-level features (HOG) quantifiers with capability of quantifying geographical atrophies in retina are selected based on clinical knowledge and they have established state-of-the-art performance for screening of retinal pathologies. Recently feature extraction techniques (deep learning) are capable of constructing feature quantifiers from raw informations. This reduces the dependency over clinical expert knowledge to some extent i.e., the region of interest parameters. The limitations for adapting deep learning algorithms for quantification of pathologies are:

- Need of large data for constructing data-driven feature quantifiers.
- Data-driven feature quantifiers doesn't guarantee better performance to conventional features in all medical applications (Tajbakhsh et al. 2016).
- What did those learned feature quantifiers were meant to be?

1.3.4 State-of-the-art for restoring and reconstructing retinal OCT

Ophthalmologists in clinical practice prefer a highly detailed image with good peak signal-to-noise ratio (PSNR). This can be achieved by dynamically modifying the shape of the reflective mirrors in OCT (known as adaptive optics) (Jian et al. 2013), or by restoring the actual information via software (known as computational imaging) (Fang et al. 2013), rather than the governing instrumentation. Computational imaging techniques employed for noise removal in OCT image restoration can be categorized into multi-image-based (Alonso-Caneiro et al. 2011, Mayer et al. 2012) and single-image-based techniques (Adler et al. 2004), insofar as SD-OCT is capable of providing 3D scans. Single-image-based techniques are preferred, because they can be extended to multiple images and they are less computationally complex. Classic techniques for reconstructing highly detailed images (with a high PSNR) from a sparsely sampled image

1. Introduction

employ a single- or two-stage approach. In the absence of noise, the two-stage approach involves interpolation and deblurring, which is inapplicable to OCT. In the presence of noise, the two - stage approach involves interpolating and denoising (Mujat et al. 2005) or vice-versa. Such approaches employ conventional interpolation techniques like the bicubic technique (Hou & Andrews 1978) and they are more concerned with denoising approaches. Such denoising algorithms include anisotropic (Mujat et al. 2005, Gregori & Knighton 2004, Garvin et al. 2008, Fernandez et al. 2004), median (Koozekanani et al. 2001, Shahidi et al. 2005, Srinivasan et al. 2008, Lee et al. 2010, Boyer et al. 2006), 2D linear smoothing (Huang et al. 1991), low pass filter (Hee et al. 1995), mean filter (Ishikawa et al. 2005, Mayer et al. 2008), directional (Bagci et al. 2008), adaptive kernel (Mishra et al. 2009), SVM (Fuller et al. 2007) and wavelet (Adler et al. 2004, Zlokolica et al. 2007, Gupta et al. 2008, Quellec et al. 2010) algorithms. Some offer a combination of filtering techniques (Dai & Sun 2011, Baroni et al. 2007) and the employment of ML has demonstrated encouraging results (Fuller et al. 2007).

Most modern successful approaches to denoising OCT images are dictionary- based (Kafieh & Rabbani 2013) and many employ sparse representation (Fang et al. 2012, 2013) as their core algorithm. Such algorithms employ either raw information from spectroscopy (i.e., the frequency domain) or from the constructed image (i.e., the spatial domain). The proposed approach employs the latter, because the former approach requires tapping into the imaging pipeline and this is not an option in commercially available OCT. Notably, (Fang et al. 2013) verified through an empirical experiment that a sparse-coding based approach offers superior performance compared to standard denoising algorithms on a benchmark OCT dataset.

The sparse-coding cost function and linear mapping are hybridized by various approaches with subtle changes. Kernel regression (Takeda et al. 2007) based approaches are also explored. BM3D and BM4D have been employed to denoise based on single and multiple images, respectively. ML-based approaches have been used to preprocess med-

1.4. Targeted Applications

ical images and these have been successful by employing random forests (Ryan Fanello et al. 2014). Empirical experiments on non-medical data have also established that a neural- network- based architecture can perform similarly to BM3D (Burger et al. 2012). Moreover, dictionary based learning approaches (Esmaeili et al. 2017) have been successful and these are advantageous insofar as they free the pipeline from knowledge of the noise model, which for OCT is stochastic in nature. Finally, principal component analysis can also filter noise, because it helps to retain repetitive patterns in the image patches to construct the dictionaries that are processed through principal component analysis (Zhang et al. 2015). This effectively reduces the load on dictionaries to remove noise. Recently multi-step image restoration approaches are considered (Fang et al. 2017) which involves segmentation process. The aim of the thesis is learning representations in a single space but not multiple spaces for a single application.

Collectively, sparse coding based frameworks are effective in reconstructing sparsely sampled OCT images and established state-of-the-art on a standard dataset in comparison to conventional image filtering based reconstruction frameworks. The limitations of sparse coding based frameworks are:

- Only patch level prediction which needs an additional steps to mosaicing during testing.
- Large number of model parameters.
- No end-to-end training.

1.4 Targeted Applications

Impact of image analysis has driven medical community to embed medical image analysis algorithms into clinical OCTs (see Figure 1.12).

In the current work, multiple frameworks are proposed aiming at improving the performance of retinal image analysis blocks independently. Such contributions impact

1. Introduction

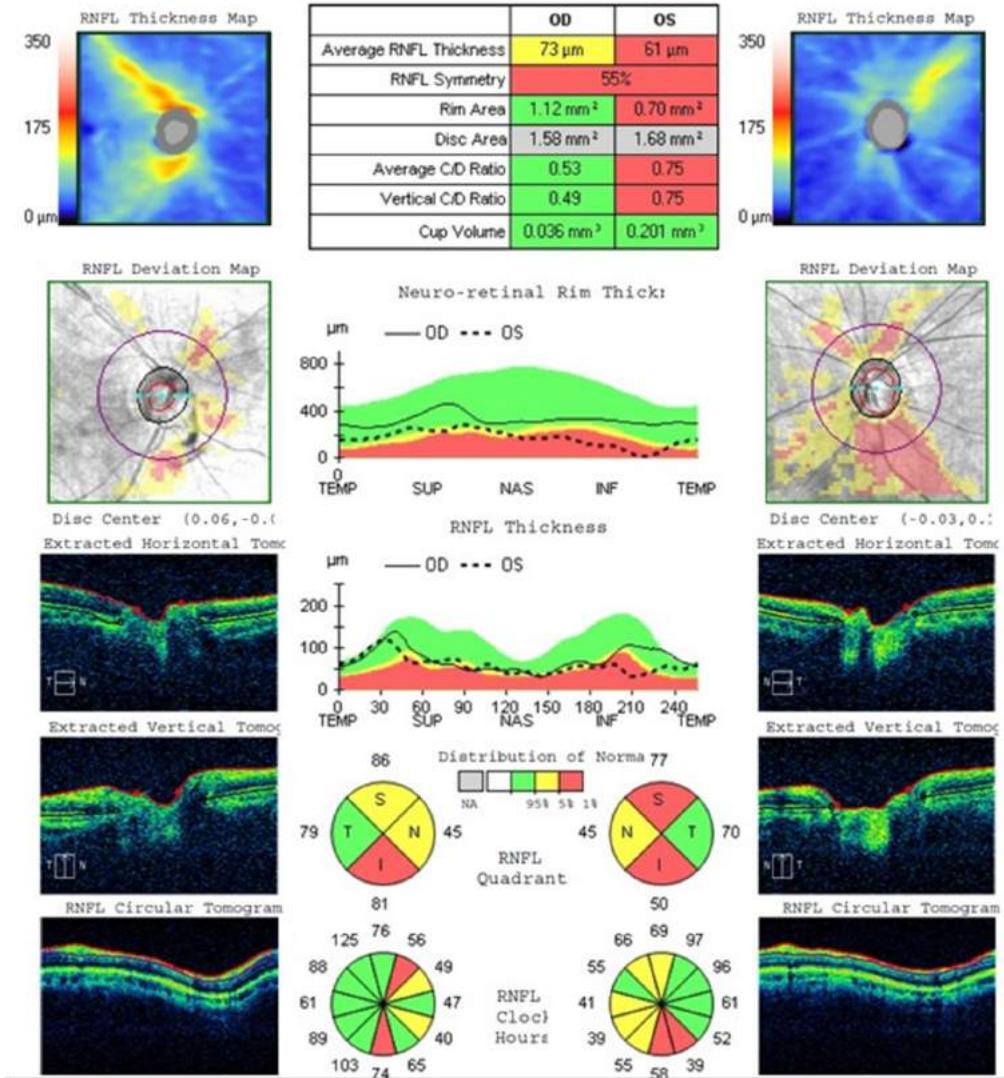


Figure 1.12: Display of a typical clinical OCT screen with automated retinal nerve fiber layer (RNFL) thickness computation which depicts the impact of segmentation algorithm in ophthalmology for semi-quantitative analysis of retina through retinal profiling.

various applications namely retinal layer delineation, retinal pathology classification, feature quantifiers subjective to pathology and image restoration. Retinal layer delineation is commonly employed to profile retinal layer thickness due to its ability for early diagnosis of retinal pathologies and system level pathologies like neuronal degeneration. Delineation also impacts the image analysis pipeline through predicting and flattening

1.4. Targeted Applications

the lower boundary of RPE, which is a crucial image preprocessing step for retinal pathology identification frameworks. Intra-retinal layer segmentation is one of the challenging problems and the field is still progressing, owing to challenges such as feeble layers, noise, low gradients, etc. However, precise early diagnosis through retinal layer profiling requires an accurate delineation of all retinal layers. Alternatively, approaches have been proposed to minimize the dependency of screening over layer profiles by including low-level features in automated screening. Recent classifier based screening approaches are dependent on only features quantified from RPE lower-boundary segmentation and model the boundaries between classes in feature space. Boundaries in feature space for practical applications are non-linear and samples near boundaries are non-separable. Boundaries in case of non-linearity or complex manifolds can be modeled through ensemble approaches with carefully identified hyper-parameters (type and parameters of kernel in-case of SVM). However, non-separable cases indicated the failed representation of the feature quantifiers. So, either the feature space needs to be projected to different feature space or identification of new feature quantifiers. Data-driven feature quantifiers have received major attention in recent times as they autonomously construct feature quantifiers which can project images onto a feature space where samples from different classes can be effectively separated by a naive classifier. Alternatively from imaging perspective, sparsely sampled OCT images are commonly used to reduce imaging time and storage space that can be retrieved through a restoration block. Computational imaging has received much attention in the form of image restoration blocks, because it aids with restoring sparsely sampled scans. Along with restoration, such approaches are capable of simultaneous filtering.

Initial applications explore the semantic segmentation of OCT images with diabetic macular edema (see Figure 1.13), because they exhibit large deformation, a challenge for retinal layer segmentation algorithms. In case of graph based segmentation this is attributed to inconsistency of edge predictions between neighboring pixels, overlap of

edges, missing edges, etc. Such barriers can be overcome by employing structured prediction of edges which can be treated as learning representations in the target space. The proposed approach additionally predicts a class conditional edge, such that a graph-based solver will not be confused regarding edges from different layers. Inner limiting membrane (ILM), nerve fiber layer (NFL), inner plexiform layer (IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer, inner segment myeloid (ISM), retinal pigment epithelium (RPE) and Bruch's membrane (BM) layers are considered for current application.

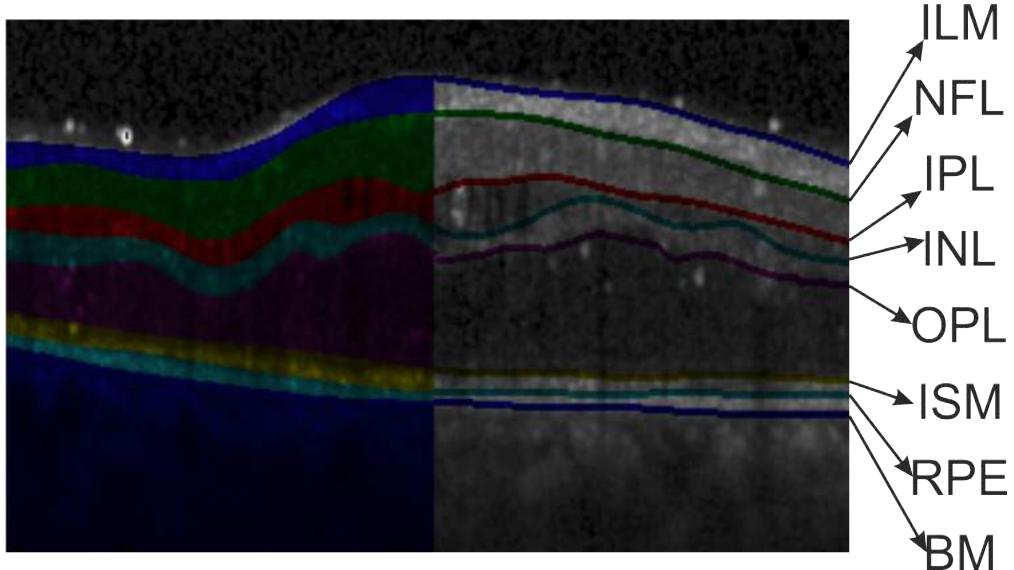


Figure 1.13: Anticipated delineated layers of retina in DME subject's OCT B-scan depicting both segmentation (left side) and upper contour of the layers (right side).

The second application explores the task to identify subjects with various pathology states of retina (see Figure 1.14) based on a classifier. Conventionally classifiers model the boundaries between classes and employ ensemble approach. In case of complex manifolds in feature space such approaches need carefully crafted hyper-parameters and a wrong choice can reduce the performance. The proposed approach represent the complex manifolds in the feature space with control points. So, such control points can be translated to boundaries in L2 space.

1.4. Targeted Applications

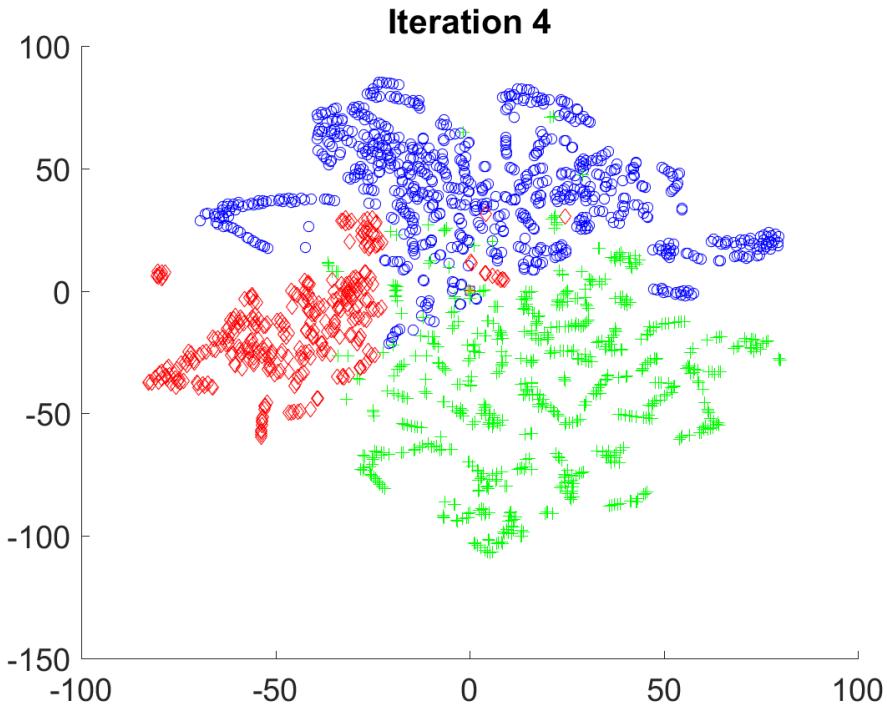


Figure 1.14: Illustrating capability of a classifier to segregate given images into various pathological states (nomral, AMD and DME i.e., circles, diamonds and plus respectively) in two dimensional feature space. The X and Y axis represents the retinal OCT HOG feature space projected onto 2D space.

The third application emphasizes the construction of application-specific feature descriptors (see Figure 1.15) that reduce the burden on classifiers. Features are deemed to be the weakest link in low-level information based detection problems (Parikh & Zitnick 2011). Data-driven feature quantifiers are successful for various applications as they construct hierarchical representations based on input space. Such frameworks need ample amount of data which is a bottle neck in case of medical applications which can be taken care by transfer learning. A major criticism over such frameworks are they being black-box features. A attempt has been made to interpret that the feature quantifiers are identifying the geographical atrophies induced due to pathology (see Figure 1.15b) in a given image (see Figure 1.15a).

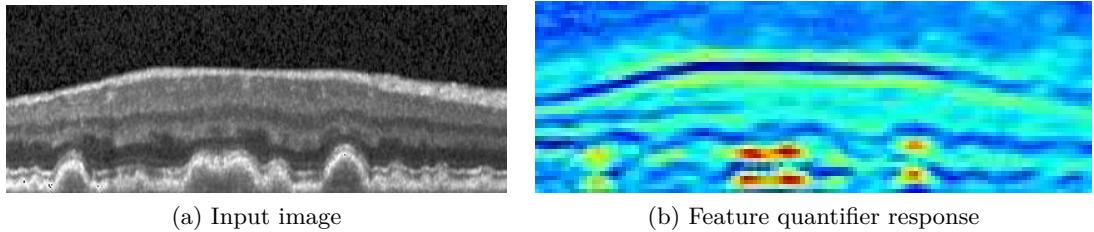


Figure 1.15: Illustrating the capability of feature quantifier depicting deformations induced due to drusen deposition in an OCT image of an AMD subject.

The fourth application explores the possibility of reducing the parameters required to reconstruct and restore a high definition image (see Figure 1.16b) from a sparsely sampled images (see Figure 1.16a). Sparse representation learning based deblurring and filtering can be molded as a neural network learning process and recently a functionally equivalent convolutional neural network (CNN) is proposed to speed up the prediction. This approach allows for end-to-end learning, fast predictions and fewer parameters, insofar as an sparse representation based approach learns dictionaries in a mutually exclusive format based on patches rather than images. The CNN learns representations in the form of filters and the dependency in each filter can be exploited to reduce the weights required. Signal processing community has termed these approaches to be separable filters, matrix decomposition, etc. The impact of ML has impeached filter separability and led to learning separable filters.

Clinical impacts of such applications are:

- Accurate retinal layer profiles for effective clock images and retinal flattening.
- Accurate identification of retinal pathologies for automated screening in remote clinics.
- Interpretation of constructed feature descriptors.
- Real-time reconstruction and restoration of sparsely sampled images.

1.5. Organization of the Thesis

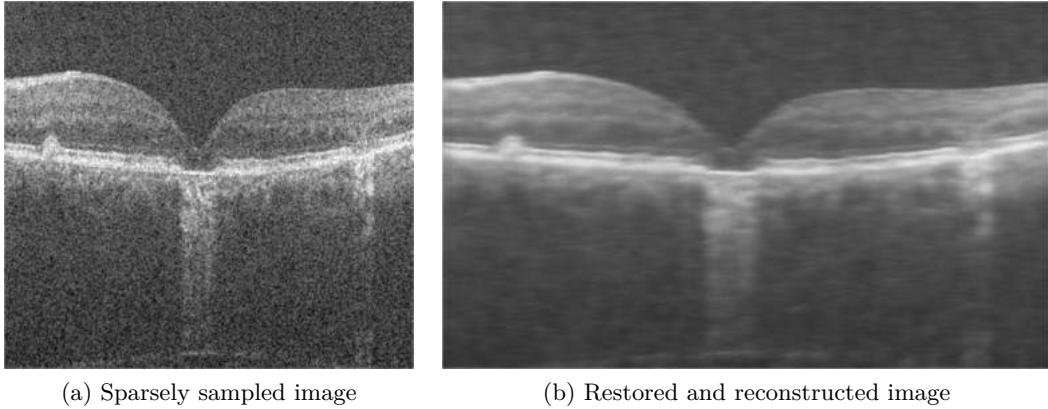


Figure 1.16: Illustrating simultaneous restoration and reconstruction of sparsely sampled OCT image with reduced noise and improved information.

1.5 Organization of the Thesis

The thesis is application driven. Consequently, the impact of learning representations on each image analysis block has been validated via a comparing with various baselines. The flow of the contributory chapters is arranged according to the space in which representations need to be learned, viz., the target, feature space and finally the input image. The thesis is accordingly organized as follows:

Chapter 1: Introduction: In the current chapter, the motivation for aim and required objectives are identified. The introduction chapter also emphasizes the role of an eye as a sensory organ and the need for automated tools to accelerate the screening pipeline for retinal pathologies. It renders the need for OCT imaging for accurate diagnosis and the challenges of adapting conventional image analysis algorithms as retinal image analysis blocks. It also describes the requirement of learning representations to improve the performance of the blocks. The chapter includes a brief description of the association between the representation space and an image analysis block. It also illustrates related work and effective baselines in various blocks of retinal image analysis. Finally, the target applications are discriminated inorder to grossly illustrate the need and path taken for improving the image analysis blocks.

Chapter 2: Learning Layer-Specific Edge Representations for Retinal

Layer Segmentation: The second chapter presents the employment of *one versus all* approach for equipping structured random forests (for edges) to incubate retinal layer information along with the edge to accurately delineate retinal layers. It also describes the need for structured prediction of edges and the need for accurate edges for graph-based segmentation approaches. A basic understanding of structured random forests for edges (SRE) during training and prediction is included. The chapter discusses the training and testing frameworks to predict layer-specific edges for a test image. A description of each process during training and testing is included. To illustrate the output characteristics for testing (prediction) framework image outputs at each process during testing are included. The chapter also contains the details of the public dataset considered and the experimental setup to validate the impact of the proposed approach in comparison to existing baselines on Duke’s OCT diabetic macular edema dataset.

Chapter 3: Modeling the Manifolds with Guided Fuzzy C-Means for

AMD and DME Subjects Elucidation: A Supervised Approach: This chapter presents a classifier that aims to represent the samples in feature space, rather than modeling the decision boundary like a conventional classifier. This is done to identify retinal pathologies like age-related macular degeneration (AMD) and diabetic macular edema (DME). The chapter describes the transformation of OCT images to feature vectors from Chapter 2 for retinal flattening and histogram of oriented gradients (HOG) for feature quantification. It includes additional toy examples for better understanding of training characteristics of the proposed approach. It also depicts the modeling characteristics of the proposed approach in the case of retinal feature vectors by transforming high vectors to two-dimensional space. The chapter also contains details of the experimental setup and considered baselines. To validate the impact of proposed approach Duke’s OCT classification dataset comprising normal and pathological subjects with subject level labels. The performance of the proposed classifier is compared with

1.5. Organization of the Thesis

support vector machines (Burges 1998) with various kernels on a Duke OCT pathology classification dataset. It also includes the preprocessing and feature quantification pipeline employed.

Chapter 4: Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies: This chapter presents an adaptation of a standard convolutional neural networks (CNN) architecture to quantify retinal pathologies and identify potential filter for understanding of information processing in the architecture. The repeatability of experimentation is also illustrated as neural network architectures (i.e., CNN) are subjective to initialization. This chapter includes a comparison of the modeling capability of CNN architectures trained with random initialization to that of transfer learning of existing architectures. For classification applications, the effectiveness of feature quantifier can be illustrated by its capacity to empower a naive classifier. Thus, the chapter also illustrates the impact of such features for augmenting a naive classifier's performance in comparison to a standard feature quantifier. It describes an algorithm to identify the potential response at each block of a CNN autonomously. HOG features are employed for a comparison using Duke's OCT pathology classification dataset, where in the preprocessing and standardization pipeline is free of heuristics.

Chapter 5: Low-Rank CNN Filters for Simultaneous Reconstruction and Restoration of Sparsely Sampled Retinal OCT Image: This chapter presents a modified sparse coding inspired CNN architecture for simultaneous restoration and reconstruction of a sparsely sampled B-scan, with fewer parameters. The chapter illustrates the adaptation of sparse coding for restoration and explains functionally equivalent neural network. It describes the need for extension of learnable separable filter technique for a CNN architecture. It adapts and describes a CNN architecture that is functionally equivalent to sparse coding for super resolution tasks. Finally, the chapter contains a brief overview of separable filters. Thus, the chapter connects the dots (sparse coding, CNN and separable filters) and validates on the performance on

1. Introduction

OCT denoising and interpolation dataset with various metrics.

Chapter 6: Summary and Conclusion: Finally, this chapter initially illustrates considered challenges and how learning representations have overcome such challenges. It depicts an ensemble of contributions with respect to each targeted application. It also discusses the possible future directions of the algorithms for each retinal image analysis block.

Chapter 2

Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

A graph optimizer performance for retinal layer delineation can be augmented through providing layer (or class) information along with edge and such layer-specific edges can be achieved through ‘one versus all’ approach.

—The Learner

2.1 Introduction

Autonomous retinal layer segmentation approaches for OCT machines have been developed to reduce ophthalmologists' burden by computing width profiles (Mayer et al. 2008, Kafieh et al. 2013). It attributes retinal layer-width profiles to various pathologies (Kanamori et al. 2003, Drexler & Fujimoto 2008, Farsiu et al. 2014). Conventional heuristics (image analysis expert rules) based algorithms are successful in case of small morphological deviations in layer structures. DME is one of the leading cause of irreversible vision loss and induce large deformations in retinal layers. Human intervention

2.1. Introduction

is required due to the failure of conventional algorithms in case of large deformations in retinal structures. The performance is further reduced in case of subsequent retinal layers having similar intensities (in OCT), noise with speckled nature, lack of information due to shadow artifacts etc. The current chapter targets to delineate the retinal layers accurately irrespective of deformation level even under the presence of noise and artifacts. This can be achieved through employing structured ML based edge prediction (modeling edge representations in target space) rather classical ML. Additionally, only the edges of anticipated layers are identified unlike conventional edge prediction where edges of all layers are predicted. Such layer-specific edge information is crucial for delineating feeble or largely deformed layers.

2.1.1 Need for structured prediction

Graph-based approaches are effective for OCT based retinal layer segmentation and edge predicted based on ML approaches have established remarkable baselines (Kafieh et al. 2013). Random forests (RF) are considered for current application because they can model complex decision boundaries during training and make rapid predictions during testing. A common problem in extending ML for segmentation is the inconsistency in prediction between neighboring pixels, resulting in unappealing semantic segmentation. This can be resolved with a post-processing step involving heuristics and handcrafted parameters. Another approach is to append conditional random fields to refine current pixel prediction (unary potential), by considering neighborhood pixel predictions (pair-wise potential). Such practices avoid end-to-end training, including RF, owing to architectural incompatibility. Structured random forests (Kontschieder et al. 2011), equip the RF with the capability of predicting an array, instead of a number or label, based on a feature vector that can be reshaped as a 2D patch (i.e., the label patch) for image processing, as shown in Figure 2.1.

2. Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

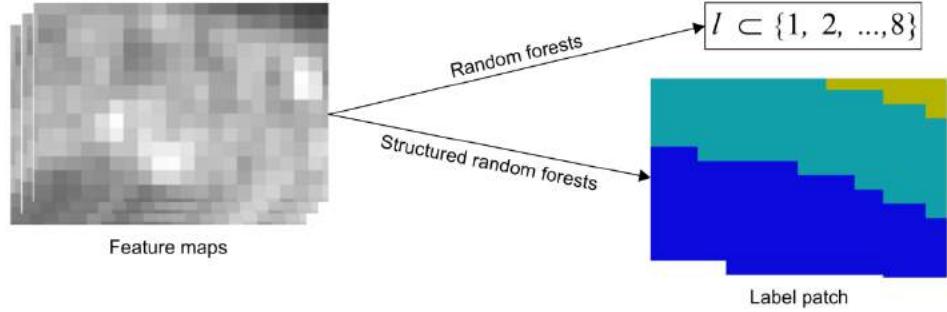


Figure 2.1: Distinction between conventional random forest prediction (only label) and structured random forests prediction (structure) given a feature map.

2.1.2 Structured prediction with classical learning machines

A RF is an ensemble of decision trees. However, it is computationally infeasible to construct rules at the decision nodes of a decision tree based on label patches. This issue has been addressed by indexing every possible label patch pattern. Training set for a RF comprises feature and mapped indexes. During prediction, the estimated index is mapped back to the corresponding pattern. A pitfall to this approach is that the number of unique patterns grow exponentially with label patch size. For example, a 16×16 2D label patch with two label states can theoretically result in $2^{16 \times 16}$ patterns or indexes and the consequent need for extreme machine learning. Clustering based distinct pattern identification has been explored to exploit the similarity within the label patterns. This limits the number of indexes and can employ any off-the-shelf RF. The same can be extended to label patches consisting of edges instead of semantic labels.

2.1.3 Structured random forests for edge

Initial development phases of SRF for image analysis were aimed at avoiding expert-intervened clustering and equipping decision nodes to handle label patches. This was achieved by reducing the vectorized label patch dimension through random sampling and evaluating for semantic segmentation applications. Such sampling also induces noise into the modelling pipeline which avoids overfitting. Structured forests for edge (SRE) (Dollar

2.1. Introduction

& Zitnick 2013) integrates clustering and random sampling at decision nodes, with clustering that explores the structural similarity between patches and dimension-reduction through sampling avoids the model from overfitting. Training an SRE (basically, an ensemble of decision trees) involves employing feature vectors, edge patches and label patches, as shown in Figure 2.2a (feature vectors are reshaped for visualization), where each tree is trained with different data. Training a tree involves constructing a decision node (the circles in Figure 2.2b) rules and a leaf node (the rectangles in Figure 2.2c) decisions. An SRE decision node is equipped to map label patches to a low-dimensional binary vector, z , such that clustering can be performed in low-dimensional space. Training a decision tree is initialized from the root (top) node where all labels are assigned to a binary vector (z) through intermediate mapping. Then, clustering (two classes) assigns a provisional class (blue and orange) to each label patch and corresponding feature vector (see Figure 2.2b) (Dollar & Zitnick 2013). Decision-node rules are constructed based on the feature vectors and class information through an impurity measure. Feature vectors are diverted towards the left or right branch based on the decision-node rule. In each branch, label patches corresponding to feature vectors are retrieved from the input data and the provisional class information is removed. Subsequently, each branch is appended with another node. The z mapping, clustering, node rule construction, splitting, class information removal and node appending are repeated for individual branches. This recursion at a branch is stopped when the dissimilarity between label patches falls below a tolerance threshold (left branch in Figure 2.2b). The last node of a branch is treated as the leaf node. In each leaf node, the least distant label patch to other label patches is stored as the representative label patch, as shown in Figure 2.2c. Edge patches corresponding to representative label patches are retrieved from input data and stored. Given a test feature vector, each tree in the SRE directs the vector towards a leaf node based on decision-node rules, as shown in Figure 2.2d. The edge patch stored in the leaf node is considered as the tree decision. The mean of tree decisions is treated as the

2. Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

SRE decision (i.e., mean of edge patches). All patches in Figure 2.2 are kept the same only for illustration purposes. The feature vector and label patch need not be the same dimension, as it involves z mapping. However, the label patch and edge patch need to be of the same dimension. During prediction, SRE estimates the edge patch with the same size of training edge patches.

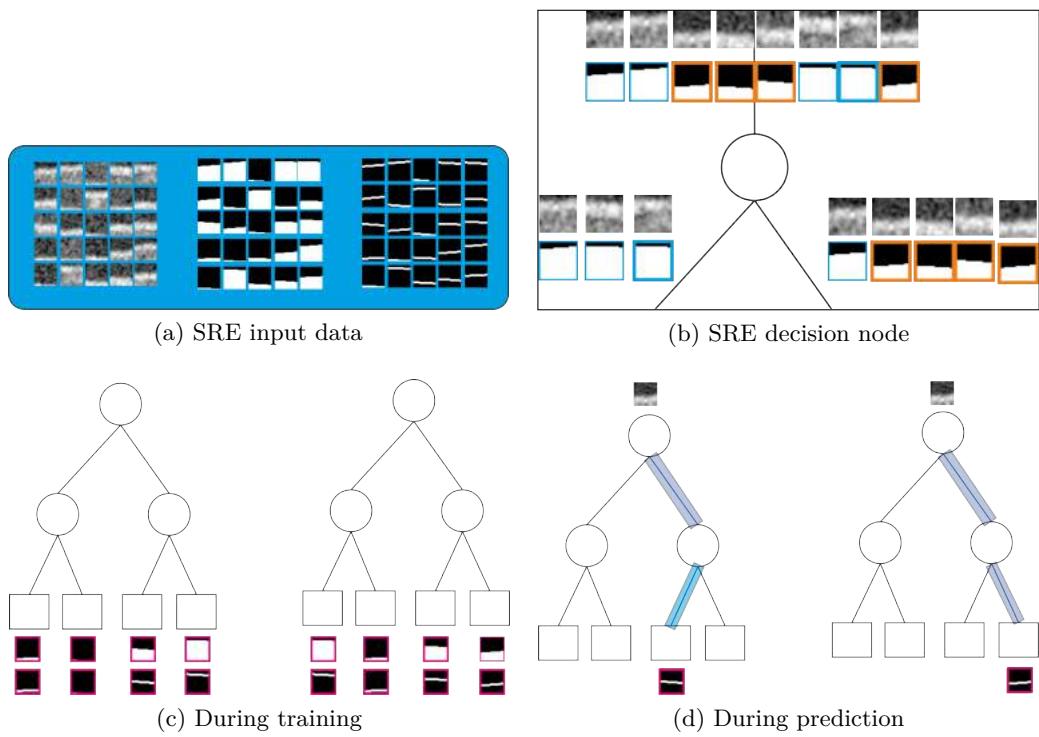


Figure 2.2: Training and prediction process of structured forest for modeling and predicting edge representations along with functionality of decision node.

2.1.4 Identified approach

All machine learning approaches (including structured prediction) employed for graph-edge construction are aimed at predicting edges, but not the object or layer class contributing to the edge. It is hypothesized that the inclusion of layer-class information along with edge can augment the dynamic program's path prediction. Such information suppresses artifact induced edges, avoids the fusion of edges from different layers, shows

2.2. Preamble to Solution

resilience to shadow artifacts, prevents dynamic programming from jumping across paths (between layers) in case of feeble layers and latches onto the edge even under large deformation. The learned edge is compatible not only with the dynamic program but also any off-the-shelf graph optimization. In other words, the proposed approach reinforces existing graph optimizers and it should not be considered as a competitive alternative.

2.2 Preamble to Solution

For the application considered the edges should be layer-specific. That is, the edges of all layers should not be modeled as single class. However, SRE are exclusively designed to consider binary labels and predict binary labels. SRE can be equipped to predict layer-specific edges by employing a *one versus all* approach (multiclass classification). Such an approach avoids any intervention or modification of learning algorithm, so it can be extended to any structured learning algorithm. The *one versus all* approach can be achieved by appending a layer ‘n’ selection block for target images (labels and edges) and training a model (i.e., an SRE for each layer). The layer ‘n’ selection block binarizes a given label image (L) using equation (2.1), where x and y are the spatial coordinates of the image. The current application involve eight layers. Thus, eight SRE models need to trained. To show that the proposed modification is not subjective to the feature considered, the same features as those in (Dollar & Zitnick 2013) are employed. Given an image (Figure 2.3a), the difference between the predictions of the SRE and those of the proposed method are illustrated in Figure 2.3b and 2.3c, respectively.

$$O(x, y) = \begin{cases} 1 & \text{if } L(x, y) = n \\ 0 & \text{else} \end{cases} \quad (2.1)$$

2. Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

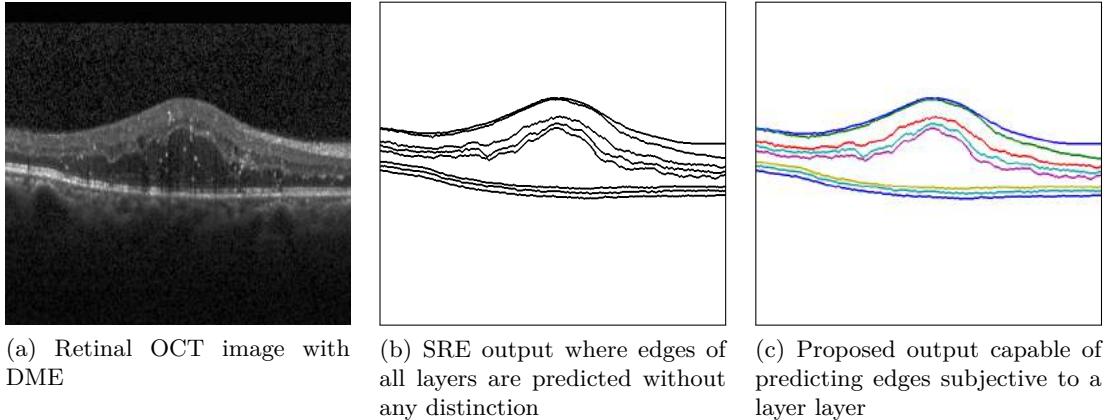


Figure 2.3: Proposed approach empowers SRE to predict layer subjective edge.

2.3 Materials and Methods

2.3.1 Experimental bed and baselines

Duke’s OCT DME dataset (Chiu et al. 2015) comprises 110 images from 10 subjects along with the upper boundary of eight layers annotated by two clinical experts. The dataset was released by Saina Farsiu Group. MatLab 2014b was used for the experimentation. The Piotr’s image (and video) toolbox and the edges toolbox were used for the feature maps, guided sampling and the structured random forest for the edges. The codes for feature tuple construction was modified for the current application. An Ubuntu workstation with 112 GB of RAM is used during training. However, edge prediction, refinement and smoothening require only eight GB of RAM. Kernel-regression based segmentation (Chiu et al. 2015) established a benchmark (AD) performance by guiding the classical approach (AN) for robust segmentation. Therefore, these two methods are considered as the baselines for the quantitative and qualitative comparisons.

2.3.2 Curating data for SRE

The proposed approach is aimed at identifying the upper edge of the inner limiting membrane, nerve fibre layer, inner plexiform layer, inner nuclear layer, outer plexiform

2.3. Materials and Methods

layer, inner segment myeloid, retinal pigment epithelium (RPE) and choroid (i.e., Bruch's membrane). These layers are here after termed as Layer 1 (inner limiting membrane) to 8 (Bruch's membrane), respectively (see Figure 1.13). The proposed approach is supervised so it requires a ground-truth along with SD-OCT images. The contour image (C) with the upper edge of the layers annotated by experts for a given an image is treated as the ground-truth. A standard SRE requires an SD-OCT image, label image and edge image (contour image). The label image ($L \forall L(x, y) \subset \{0, 1, \dots, 8\}$) is a semantic segmented map constructed from the contour image. An example of SD-OCT image (Figure 2.4a) with its corresponding label (Figure 2.4b) and contour images (Figure 2.4c) with pseudo-colors is provided in Figure 2.4

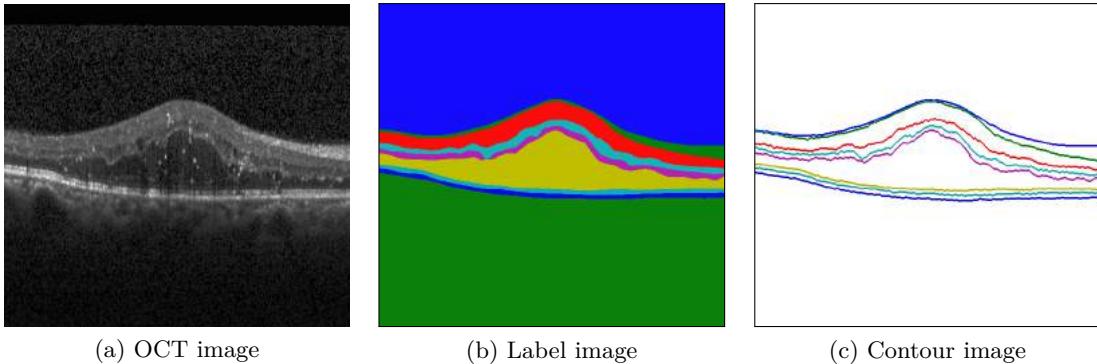


Figure 2.4: Images required for preparing training dataset. (a) Input image, (b) label image and (c) contour image.

2.3.3 Data preparation

The provided data is prepared to match the proposed method's data requirements. The proposed method considers only a single ground-truth (i.e., contour) image corresponding to an OCT image. So for each layer, along each column, experts annotated locations of contour are averaged. The 'nan' (not a number) values along a contour are estimated through 1D interpolation. Averaged locations (including interpolated) are considered as a row values and a contour image was constructed by allotting the layer index 'n' (n

2. Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

$\subset \{1,..,8\}$) to the pixel. Along each column, the pixels between ‘n’ contour and ‘n+1’ contour are replaced with ‘n’ value to construct the label image. For eighth layer, pixels between eighth contour and last row are replaced with ‘8’. In contour and label images, only columns from 120 to 650 are considered, owing to lack of expert annotations.

2.3.4 Model training and prediction

Similar to conventional supervised approach, the proposed framework involves two phases (training and testing): training a single SRE for the upper edge (contour) of each layer (Figure 2.5a) and predicting the upper edge of each layer (Figure 2.5b). Any off-the-shelf graph-based segmentation technique can be employed for this framework. Each block employed is described below.

Image correction

Initially, the dynamic range of the input SD-OCT image is standardised to [0,1]. The OCT image-acquisition protocol chisels the acquired A-scans to match a predefined B-scans and replace the missing values with maximum values. This results in saturated values along the boundaries of the OCT image. This is resolved by replacing pixel values equal to 1 with 0.01 (minimum value of information) and the resultant image is treated as I .

Image intensity and gradient features

Image intensity values and a histogram of oriented gradients (HOG) are commonly employed by various approaches for edge detection in computer vision. At each pixel of I , the gradient magnitudes are computed (in a 9×9 window), normalized and treated as a magnitude map (A). HOG features are extracted (using four orientation bins and a 9×9 window) with a stride of one resulting in four histogram maps (H) where each map is the result of each orientation. It is commonly suggested to incorporate scale-space, so

2.3. Materials and Methods

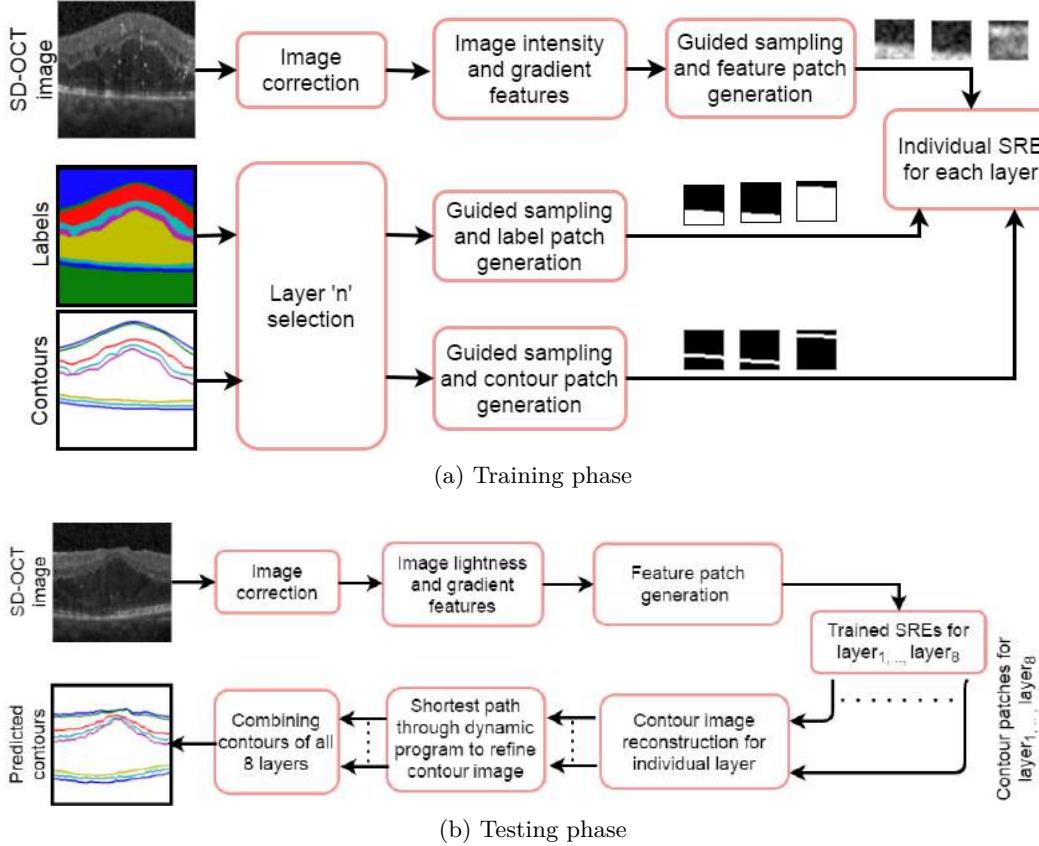


Figure 2.5: Flow of retinal image processing blocks involved in the (a) training phase and (b) testing phase.

an image re-sampling is performed over I at scale 0.5 to generate $I_{0.5}$. The magnitude map and HOG feature maps are computed on $I_{0.5}$ and the maps are resampled to match the dimensions of I , resulting in $A_{0.5}$ and $H_{0.5}$. A feature tuple (F) is generated by concatenating all feature maps ($I, A, H, A_{0.5}$ and $H_{0.5}$) along the third dimension.

Layer ‘n’ selection

This is the crucial block as it expands the capability of a classical SRE to predict layer specific edges using *one-versus-all* approach. Given C (contour image), the block creates a binary contour mask (BC) using equation (2.1), with only the upper edge (contour) of layer ‘n’ ($n \subset \{1, 2, \dots, 8\}$) as the foreground. In other words, if spatial location (x, y) in

2. Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

C is ‘n’ then corresponding BC location is ‘1’ else the location of BC is ‘0’. Given L (i.e., label image), the block creates a binary label mask (BL) through equation (2.1), where the posterior region of BC is the foreground and its anterior region is the background. Given L (see Figure 2.6a) and C (see Figure 2.6b) the constructed BL (Figure 2.6c) and BC (Figure 2.6d) for Layer 3 is illustrated in Figure 2.6. Here, BC is also used as a guide for sampling the feature, label and contour patches.

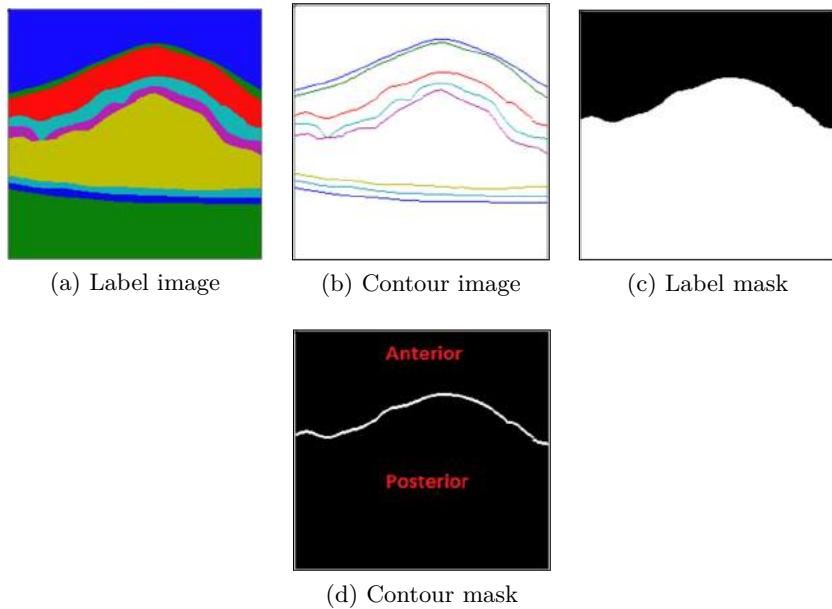


Figure 2.6: Processed outputs of layer ‘n’ selection block for retinal layer 3 from top. (a) Label image, (b) contour image, (c) processed label binary mask and (d) processed contour binary mask

Guided sampling

The BC is morphologically dilated with a disk of radius 16, which results in DBC . All DBC locations with value ‘1’ are collectively termed as the positive spatial locations Y ($Y=(x_1,y_1), (x_2,y_2), \dots \forall DBC(x,y)=1$). The remaining locations other than Y are treated as negative spatial locations N . As each image can result in around quarter million locations only partial set of Y and N are retained.

2.3. Materials and Methods

Patch generation

A patch size of $32 \times 32 \times 11$ is considered to extract from F (11 dimensions) based on Y and N . This results in a set of feature patches (P) with each patch size being $32 \times 32 \times 11$. Each feature patch is flattened through a lexicographical arrangement resulting in a feature vector \mathbf{f} (1D array).

The Label patches and contour patches each with 16×16 dimensions are extracted from BL and BC based on Y and N . The correspondence of the central pixel between feature patches, labels patches and contour patches is retained.

Training individual SRE

Each SRE block is an ensemble of six decision trees, where each tree is constructed with decision nodes and leaf nodes based on flattened feature patches, label patches and contour patches. A trained SRE stores representative label patches and corresponding edge (contour) patches for layer ‘n’ and denoted by SRE_n . The training process is repeated for all eight layers.

Contour image reconstruction

Given a set of feature vectors extracted from a test image a trained SRE_n predicts edge patch (layer ‘n’) corresponding to each feature vector. All such patches are overlapped or mosaiced to construct layer ‘n’ edge map with the same dimensions as the test image. Insofar as feature vectors are extracted with a stride, the overlapping regions of predicted patches are averaged.

Shortest path

Given an edge map, shortest path involves graph edge computation (Chiu et al. 2010) and concatenation with a column vector ($1e^{-6}$) on the left and the right sides. The shortest path from the top-left node to the bottom-right node of the graph is computed

2. Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

based on the dynamic program. This path is smoothed using the Savitzky Golay filter (Savitzky & Golay 1964). The smoothed path is projected back to the image space and considered as the layer ‘n’ contour (the upper edge).

Combining contours

Upon repeating SRE prediction, shortest path and smoothing for all layers, eight layer contours are attained. These contours are fused by allotting each layer contour with a pseudo-colour and treated as the layer delineation result.

2.4 Results and Discussion

2.4.1 Governing parameters for training and testing of SRE_n

A total of 55 images from the first five subjects were considered for training. Piotr’s image and video toolbox was used to construct feature tuple. The upper-limit of Y and N set size was 10^6 and 1.5×10^6 , respectively. The dimension of z for the low-dimension projection was set to 256. Piotr’s edges toolbox was used for training an SRE with six trees. Each tree employed 25% of data. Because training different trees in an SRE_n is mutually exclusive the model training process can be parallelized.

During testing, not all six trees need to be used for the prediction if the user interested in speed. Owing to mutual independence, each thread of the machine can be employed for each tree prediction. Dijkstra’s algorithm was employed in the form of a dynamic program for the shortest path detection, since it offers balanced performance with less complexity.

2.4.2 Delineating a given test image

Upon completing the training phase, eight SRE models are constructed. Each SRE has the capability to predict upper edge (contour) of the corresponding layer. Figure 2.7

2.4. Results and Discussion

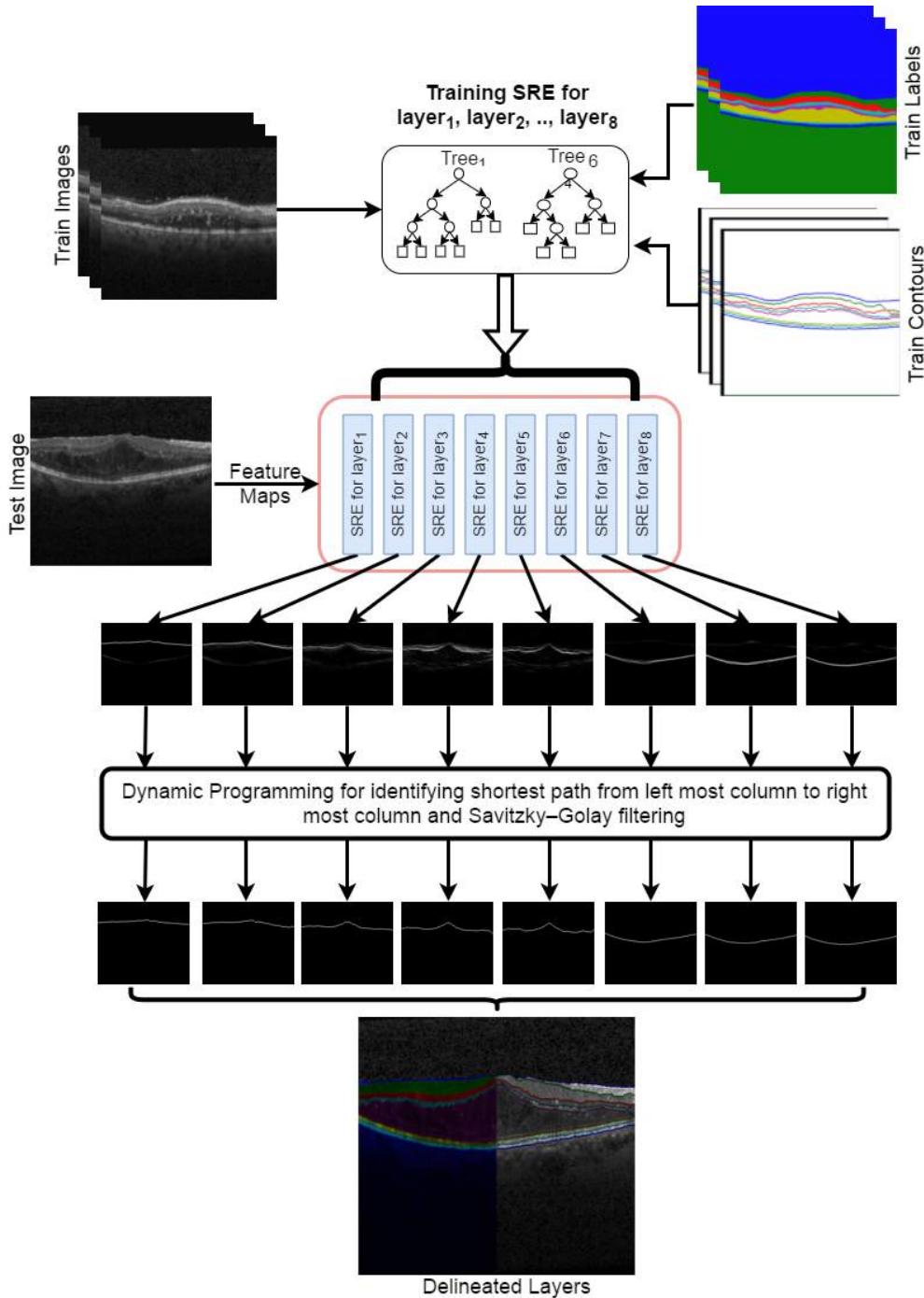


Figure 2.7: Block diagram illustrating image inputs during training and resulting image at each block during testing. The output of each model (SRE for $layer_n$) depicts the edge of subjective layer fulfilling the anticipated objective.

2. Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

illustrates the predictions of each model given a test image and it also shows the image results of major blocks in the prediction pipeline. It can be observed that the edges are indeed predicted, although they are not crisp. So dynamic program refines the edge to a single pixel width. Dijkstra's algorithm results in zig-zag pattern along the layer due to the non convexity in image information, which is smoothed. The resulting contours are considered as layer delineations of the retina. Both the layer-width profiles and edge profiles can be illustrated in Figure 2.7.

2.4.3 Delineating under various scenarios

The robustness of the proposed framework to the influence of deformations, artifacts, noise and shadow is shown in Figure 2.8. The magnified representations are illustrated in Figure 2.8a to depict the efficacy of the proposed approach in the presence of noise-induced due to imaging-protocol, layer deformation due to the pathology, low gradients due to signal attenuation in the posterior retina and shadows due to blood vessels are illustrated in figure 2.8c and figure 2.8d respectively.

The resulting prediction of the proposed approach agrees more with expert annotations than kernel-regression features based segmentation, both visually (see Figure 2.9) and statistically (see Table 2.2).

2.4.4 Quantitative performance

Testing was performed on the remaining 55 images. The mean of the absolute difference between the predicted contour and the expert contour along the column was employed as a metric for comparing the proposed measure with baselines (see Table 2.1). This metric illustrates the deviation of predicted contours from expert annotations. This is crucial for computing retinal-width profiles. The comparison was performed by considering two expert annotations independently.

The F-score was employed for a quantitative comparison with the established base-

2.4. Results and Discussion

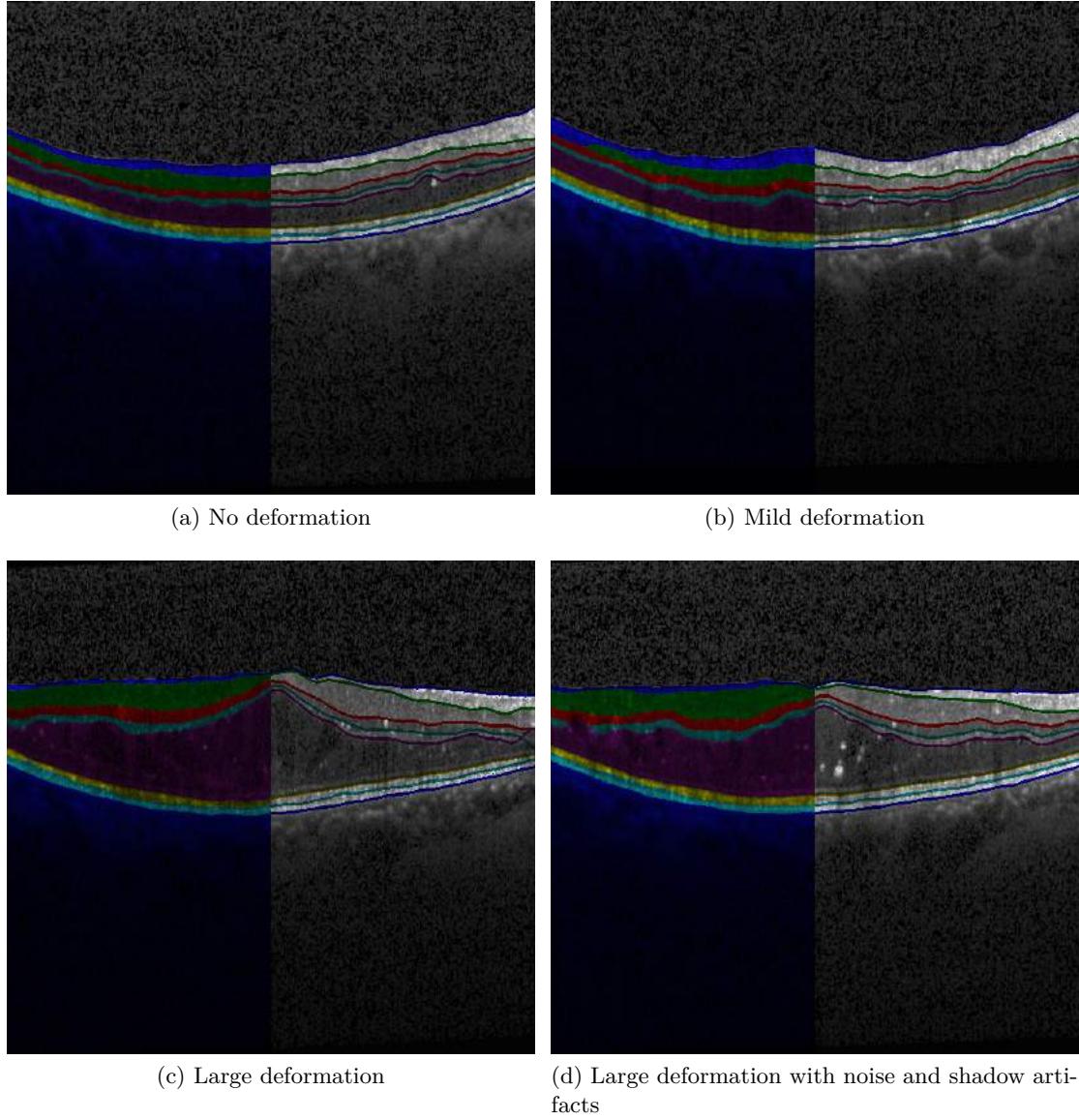


Figure 2.8: Predictions of the proposed approach on test images with various deformations illustrating its robustness to varied deformation levels and noise. The predicted layer are same as considered in subsection 2.3.2 and color coding is same as considered in Figure. 1.13.

lines in Table 2.2 on the considered dataset. From the perspective of the image processing, the F-score is employed to measure the segmentation accuracy. From the perspective of statistical evaluation, it is a combination of three statistics (viz., true positive, false

2. Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

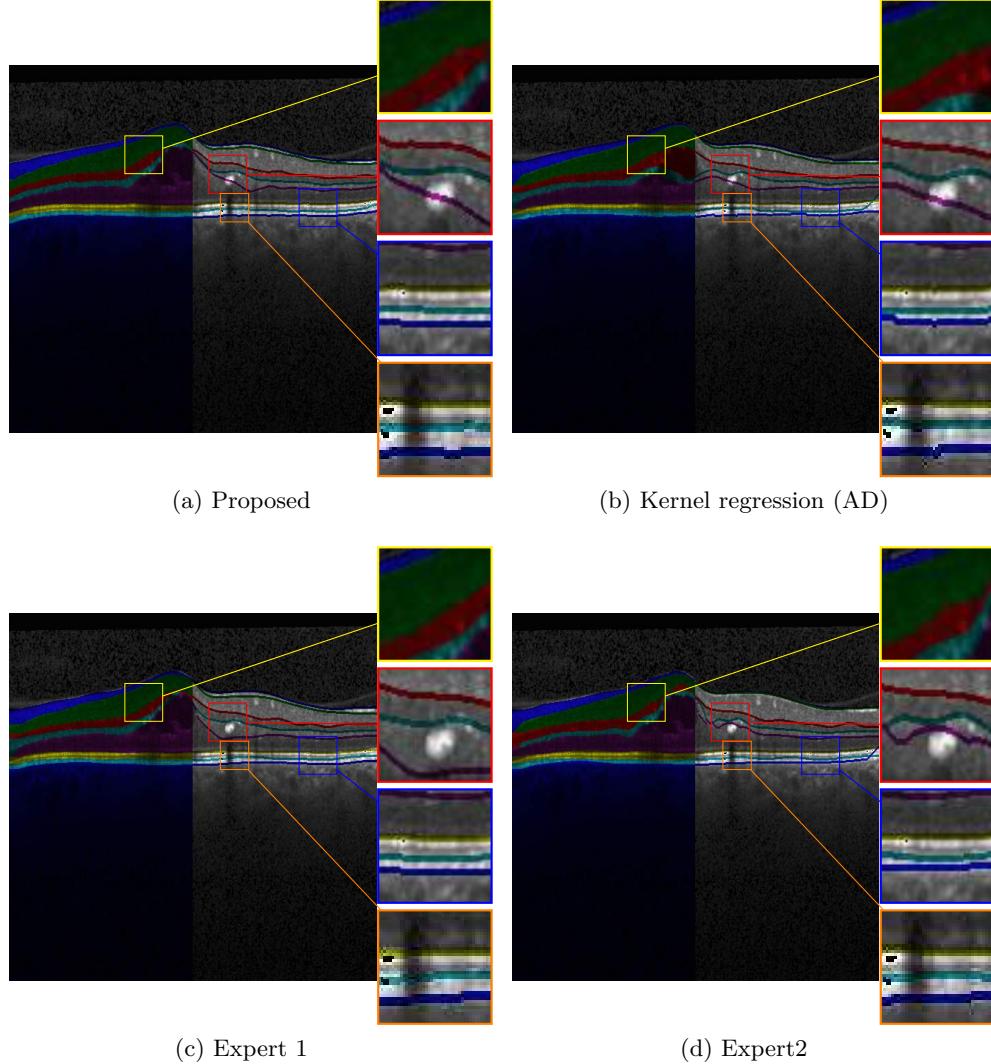


Figure 2.9: Illustration of layer segmentation performance in the presence of deformation (red), noise (yellow), shadows (orange) and low gradients (blue) by the proposed method in comparison to other baselines.

positives and false negatives).

2.4.5 Alternate approaches

Employment of statistical representations rather to structural representation (see Appendix A) further reduces the performance. In case of SRE the construction of z with

2.5. Summary

Table 2.1: Mean absolute difference evaluation across the classical graph approach (AN), kernel guided-graph approach (AD) and layer-specific edge guided-graph approach (proposed) in comparison to Expert 1 and Expert 2 annotations.

	Expert 1 vs				Expert 2 vs		
	Expert 2	AN	AD	Proposed	AN	AD	Proposed
Layer 1	1.14	1.099	1.319	0.969	1.024	1.189	0.906
Layer 2	1.683	3.962	1.708	1.625	4.25	1.91	1.826
Layer 3	1.681	5.936	2.013	1.698	6.03	2.092	1.853
Layer 4	1.752	5.598	2.168	1.704	5.816	2.431	1.753
Layer 5	1.959	5.311	2.486	2.146	5.317	2.406	2.125
Layer 6	1.103	1.125	1.064	0.863	1.037	1.043	0.901
Layer 7	1.273	1.042	1.185	1.086	1.271	1.328	1.229
Layer 8	1.191	1.348	1.182	0.863	1.623	1.399	1.112

Table 2.2: F-score evaluation across the classical graph approach (AN), kernel guided-graph approach (AD) and layer-specific edge guided-graph approach (proposed) in comparison to Expert 1 and Expert 2 annotations. The lower boundary of the choroid layer is unavailable, the F-score is computed for only the first seven layers.

	Expert 1 vs				Expert 2 vs		
	Expert 2	AN	AD	Proposed	AN	AD	Proposed
Layer 1	0.864	0.778	0.853	0.874	0.768	0.85	0.868
Layer 2	0.903	0.772	0.895	0.909	0.764	0.887	0.9
Layer 3	0.797	0.652	0.757	0.807	0.649	0.751	0.802
Layer 4	0.747	0.67	0.747	0.77	0.65	0.728	0.756
Layer 5	0.941	0.868	0.931	0.944	0.869	0.933	0.944
Layer 6	0.862	0.878	0.872	0.889	0.868	0.864	0.878
Layer 7	0.829	0.823	0.824	0.868	0.791	0.803	0.845

binary states was identified as the reason for the binary edge predictions. A modified z equips the SRE to models all layer edges in a single model. Such an approach results in less training time but fails to match the performance in respect to all layers (see Appendix B), insofar as a single model cannot handle complex patterns.

2.5 Summary

The proposed approach extended the applicability of structured edge prediction for retinal OCT segmentation. The edge prediction models reduced the need for the heuristics

2. Learning Layer-Specific Edge Representations for Retinal Layer Segmentation

e.g., that Layer 1 upper edge can have a moderate positive gradient, that Layer 6 upper edge can have a steep positive gradient, that Layer 8 upper edge can have a steep negative gradient, etc. Such reduction was achieved through autonomous modeling of patterns through machine learning. This was needed because pathologies can arise multiple patterns and development of expert based heuristics for all such patterns burdens the expert. The proposed approach added a new dimension of information to any edge predicting algorithm. It was empirically validated that edge-predicting algorithms can encapsulate such class-/layer- subjective information. It was also validated that such information reinforces graph-based approaches. Because a complete framework has been proposed for layer delineation, the SRE can be replaced with any existing structured prediction. For simplicity, the proposed approach considers the mutual independence between layers in ‘combining the contours of all eight layers’ block during testing. Depending on the interest of the programmer, any empirical rule can be imposed through this block. The prediction process estimates edge maps with low computational complexity, such that they can be integrated into the clinical pipeline. The pathology classification and quantification chapters (Chapter 3 and 4) also shows that SRE₈ can be employed for retinal flattening in images acquired on different machines to diagnose various pathologies (AMD and DME). This broadens the applicability of the proposed approach as an image preprocessing step for retinal OCT analysis.

HOG features are subjective to gradients and structured predictions are robust to missing gradient to a certain extent (specifically, half the predicted edge patch size). However, this cannot handle large missing edges. Extending the proposed approach to volumetric data can overcome the hurdle to a certain extent, insofar as neighboring images can provide additional or missing information. Upon arrival of new data, current models should be able to fine-tune the rules in decision nodes. This can be achieved with adapting gradient boosting trees optimization to SRE.

Chapter 3

Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

Modeling of manifolds rather to boundaries in feature space can improve the classification performance. Unsupervised algorithms are effective for such purposes as they are intrinsically designed to identify underlying manifolds. However, unsupervised algorithms needs to be guided for improved performance in comparison to supervised approaches.

—The Learner

3.1 Introduction

Computer-aided screening tools have been developed to identify retinopathy in subjects to reduce the workload on ophthalmologists. There is a wide range of algorithms to do so. Some involve retinal layer profiles prediction (as described in Chapter 2) that are

3.1. Introduction

attributed to various pathologies. Alternate approaches have been proposed to reduce the dependency of retinal layer profiles by using low-level features and machine learning. These approaches have grown rapidly by constructing pathology subjective feature descriptors and classifiers (machine learning algorithms) capable of delineating complex structures. The current application is more concerned with the classifier and the Chapter 4 deals with the construction of a feature descriptor. The increased availability of computational power and ample data has boosted the machine learning community to design a wide variety of classifiers. Classifiers dominate the pattern recognition community so strongly that the employment of a classifier is commonplace in the case of automated analysis. Conventional structured learning like Structured forests (see Chapter 2) aims to predict patterns in a target space. It is identified that understanding the representation of manifolds in feature space is crucial for effective classification. Considering SVM kernels as an example, where employing linear kernel to classify non-linear manifolds results in plunged performance. Carefully crafting such kernels entails an understanding of the manifold representation, but image analysis commonly involves more than two features, so rendering human interpretations is difficult. This has been overcome through dimensionality-reduction techniques for visualization. These techniques are used to projects high dimensional features onto two-dimensional space. This allows for a visual interpretation of the manifold representations and identifies transformation or kernels to make them linearly separable in actual feature space. Large-dimensional feature space has hampered methods seeking to identify the optimal projection with less complexity. Metric learning is another approach for constructing distance functions with an appropriate measure of the distance between data points (Bellet et al. 2013) or for transforming the computed features to a new feature space where samples from different classes are linearly separable. Classifiers are equipped to handle such manifolds by including kernel transformations along with decision rules construction. However, this is computationally expensive, insofar as the search space has increased. Ensemble learning

3. Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

is employed to autonomously model the boundaries of complex manifolds through a base classifier without modeling the actual manifolds. Alternatively, unsupervised approach explores the underlying manifolds but fails if manifolds are close. Current work targets to fuse the properties of ensemble learning and supervised learning for effective elucidation of pathological subjects based on retinal OCT images. The difference between existing classifiers and the proposed classifier is that the former aims to identify the boundary but the proposed classifier models the manifold representation of each class independently.

3.1.1 Need for modeling a manifold representation

In image analysis feature quantification involves transforming the image into a D -dimensional feature vector that resembles a point in a D -dimensional feature space. It is the classifier's job to shatter the space into pieces through decision rules, where each piece represents all or the majority of samples from a single class. A D -dimensional space with r quantization bins along each dimension requires r^D samples to fill the space and construct global decision where r is capable of accommodating millions of variants (thanks to double precision) and D ranges in the 1000's to accommodate complete image information. Any amount of existing training data will be insufficient to make global decisions. In practice, global decisions are constructed based on limited data. Because there can be multiple decision boundaries, SVMs introduced an optimal decision boundary by constructing support vectors that represent the boundaries of individual classes. This is the reason that SVM is still preferred over random forests in scenarios where training data is limited. However, support vectors can never represent the manifold effectively because they represent only the boundary condition of the manifolds of each class. This remains unaltered in the case of SVM and various classifiers because classifiers aim at separating samples rather than cumulative representation. This makes support-vector construction prone to outliers or noise. Thus a

3.1. Introduction

cumulative representation is needed instead of a decision representation (hyperplane in case of SVM).

3.1.2 Manifold representation modeling

Conventional learning machines are concerned with constructing complex decision boundaries through a basic model (i.e., with ensemble learning). As an example, a linear classifier can be used for the non-linear problems through a piecewise linear approach. As such this is an indirect approach whereby the algorithms are designed to separate the manifolds from different classes by shattering the space into subspace and further shattering the subspace with data points from multiple classes. The decision rules reflect the boundaries of the manifolds more effectively than in SVM, resulting in an improved representation of the manifold. Alternatively, kernel machines are deployed to transform features for linear separability but insofar as separability is a governing factor, these approaches do not actually model the manifold representing a class.

Graph-based approaches are a backup choice for representing manifolds by exploring the local neighbor adjacency. Graph-based approaches have dominated unsupervised (spectral clustering) and semi-supervised (random walks) approaches. In the case of supervised algorithms, non-graph algorithms are effective for storage and rapid predictions upon the arrival of new test data. Algorithms inspired by Voronoi diagrams, like the K-nearest neighbours were undermined after the rise of statistical learning theory with effective representations of a data population.

3.1.3 Identified approach

Voronoi diagram have indulged algorithms that are computationally efficient due to their non-parametric nature. Further, they provide for easy storage, requiring only control points and a testing phase involves a Euclidean distance measure which is translated to decision boundary on L2 space. Each generated subspace resembles the shattered space

3. Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

pieces by a classifier and the boundaries between subspaces are classified by piece-wise linear decision boundaries. To represent complex patterns, the subspaces are further divided with new control points (i.e., the fractals in a Voronoi diagram). This resembles the Voronoi fractals used in computer graphics. Such an approach is extended to modeling manifold representations through an ensemble of unsupervised modeling. Unsupervised approaches are explicitly designed to understand the patterns in feature space given the unavailability of class labels. The approach relies on an unsupervised approach, but one governed by label information.

3.2 Preamble to Solution

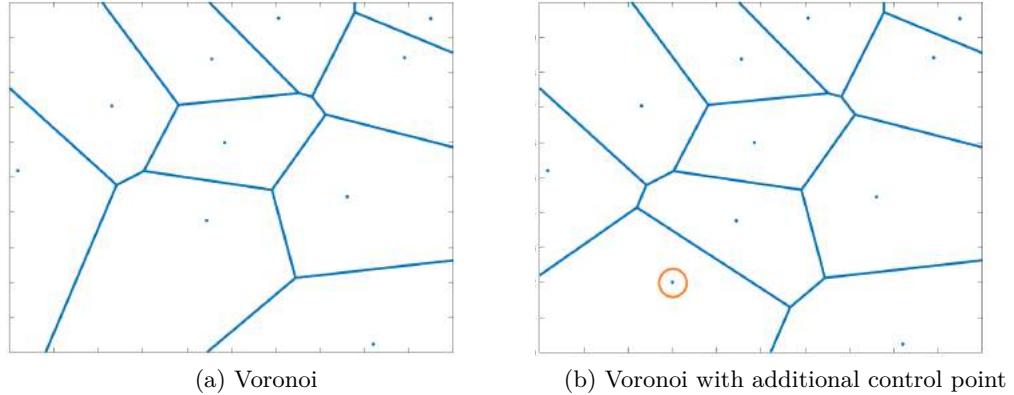


Figure 3.1: Illustrating improved resolution of subspace in feature space on addition of new control point.

A base model needs to identify control points and each control point results in subspace. A collection of such subspaces of a class resembles the manifold. From algorithm's perspective, control points can be considered as mean vectors of K-means clustering where the label of each centroid is known. In the test phase, the cluster (or subspace in a Voronoi diagram) of each test sample is identified based on a topology that resembles a Voronoi diagram (see Figure 3.1a). However, it is evident that K-means clustering fails

3.3. Materials and Methods

in cases of nonlinear patterns, because K-means follows a Euclidean metric. Thus, an identified cluster with samples from multiple classes can be further divided (see Figure 3.1b) resulting in new clusters with increased resolution. This resembles the fractals in the subspace, as illustrated in Figure 3.1. This reduces the proposed approach to a boosting approach with class conditional K-means. Fuzzy C-Means (FCM) is employed because it has the capability of resolving this ambiguity to a certain extent. It also acts a node to incorporate expert knowledge into the system in order to attract wider applicability. During prediction, K-means clustering strategy is considered. That is, a hard boundary is established to keep the computational complexity in mind. In the case of a lack of information, replacing the membership function with the impulse function transforms a FCM into classical K-means clustering algorithm.

3.3 Materials and Methods

3.3.1 Experimental bed

3.3.1.1 Toy example

Spiral dataset

A toy spiral example is commonly used to validate the modeling capability of proposed approach. A two-class spiral example (of 1000 samples) was considered, where diamonds represent Class ‘1’ and circles represents Class‘2’.

Synthetic dataset

A simple toy example (of 16 samples) was used to explore two scenarios, illustrating the characteristics of the algorithm and its boundaries. Eight samples were considered for training and the remainder for testing. The toy example comprised three groups of points (a, b and c in Figure 3.4a). The scenarios involved:

3. Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

i Grouping a and b to form Class 2 and considering c as Class 1.

ii Grouping a and c to form Class 2 and considering b as Class 1.

The first scenario simulates class imbalance, whereas second scenario simulates non-linearly separable scenario.

3.3.1.2 Retinal pathologies dataset

Duke's OCT classification dataset (Srinivasan et al. 2014) with normal, Diabetic macular edema (DME) and Age-related macular degeneration (AMD) subjects is used for experimentation. The datasets consists of 15 subjects per class totaling 45 subjects and 3231 images. A 15 fold validation is performed where each fold in the training set included eight subjects from each class. The remaining subjects are used as the test set. Thus, the first fold contains Subjects 1-8 for training and Subjects 9-15 for testing. The fifteenth fold has Subject 1-7 and subject 15 (following circular indexing) for training and Subjects 8-14 for testing. Ambiguous images in each training set are removed and test-set images retained.

3.3.2 Retinal flattening

As shown in Figure 3.2a images are processed through various blocks before feature extraction, in order to standardize and enhance feature presence. The retinal flattening in Figure 3.2 is a common practice in retinal image analysis, requiring a lower boundary for the retina. Thus, the SRE₈ model from Chapter 2 is employed to predict the lower boundary of the retina. The predicted boundary is smoothened using a morphological closing operation. Each column in the actual image is shifted circularly, so that the retina's lower boundary (including the drusen complex) forms a straight line. Each image is smoothened without loosing the edge information and this is done using the BM3D filter. To reduce the complexity, each image is scaled down by half, using bicubic interpolation (see Figure 3.2b).

3.3. Materials and Methods

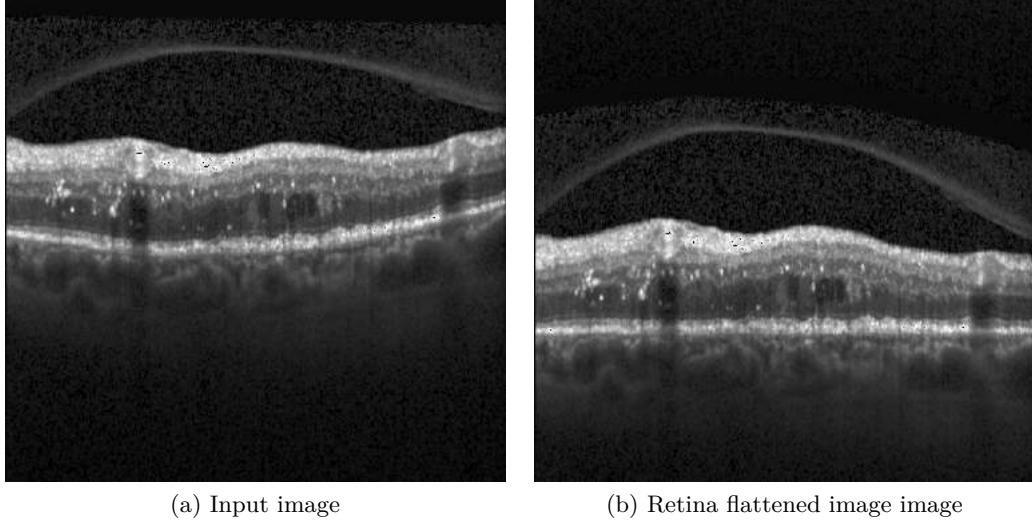


Figure 3.2: Image standardization through retinal flattening i.e., shifting the A-scan lines so lower boundary of RPE (including drusen in case of AMD) forms a straight line.

3.3.3 Feature quantification

The feature quantification process employed by (Srinivasan et al. 2014) was used. This involves cropping of the image to dimensions of 45×150 , such that the retina's lower boundary is in five rows from the bottom of the image and the central column remains in the center. HOG features are then acquired on the cropped image patch. The image patch is scaled down by 0.5 and 0.25 followed by HOG feature computation for each scaled image patch. Each SD-OCT image is quantified to a feature vector with 15K dimensionality. The feature vector corresponding to each image is concatenated to form a data matrix.

The reason for using HOG features over data-driven feature quantifiers is that the former are less effective quantifiers resulting in complex manifolds in the feature space and thus burdening the classifier which illustrates the true potentiality of the classifier.

3. Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

3.3.4 Proposed framework

Generally a classifier requires a data matrix (formed by concatenating features vectors training samples) and corresponding labels l ($l \in \mathbb{R}^k$, k is number of classes). The proposed approach constructs k means-sets, where each set consists of centroids (or control points) for the corresponding class. Framework initialization involves partitioning the data matrix (X) into k number of datasets (D_1, D_2, \dots, D_k where the second set has samples from the second class) and k null sets are considered as the initialization of the means-sets.

The j^{th} dataset is divided into two clusters using Fuzzy C-means clustering (Bezdek et al. 1984) and identified means are concatenated to the j^{th} means-set. Upon repeating the above step for all k classes, the Euclidean distance between each sample in the data matrix and all means in each means-set is computed. For each sample, the closest mean is identified and the corresponding means-set index (or class) is treated as the estimated class. Wrongly classified samples are identified and partitioned into k datasets (D_1, \dots, D_k) based on the corresponding true labels such that samples from each set belong to the same class. This process is repeated until any partition has no more than two samples. To identify wrongly classified samples, a classic K-means strategy is employed as a hard-decision approach. It will be interesting in future research to consider computing membership along with the decision in order to rank the importance (or severity) of the sample.

An algorithmic flow for the proposed approach is shown in Algorithm 1. A data matrix X and label vector l with k classes is taken as input and returns a set of mean vectors (M), where each M matrix is a collection of control points representing the respective class. Functionaries in the algorithm are illustrated below:

- FCM refers to a standard fuzzy C-means clustering algorithm.
- Pairwise distance (pDist) is the standard function in Matlab with the Euclidean measure.

3.3. Materials and Methods

```

Data:  $X$  (dataset) and  $l$  (label vector with  $k$  classes)
Result:  $M_1, M_2, \dots, M_k$  (mean-sets)

1 Initialization:  $M_i \leftarrow \{\emptyset\}$ 
2 ;  $D_i \leftarrow X_j; \forall l_j = i$ 
3 ; for  $iter = 1$  to  $10$  do
4   for  $i = 1$  to  $k$  do
5     if  $\#samples$  in  $D_i > c$  then
6        $\{\mu_1, \mu_2, \dots, \mu_c\} \leftarrow FCM(D_i, clusters=c)$ 
7        $M_i \leftarrow \{M_i | \{\mu_1, \mu_2, \dots, \mu_c\}\}$ 
8     end
9   end
10  for  $i = 1$  to  $k$  do
11     $dist_i \leftarrow min\_means(pDist(X, M_i))$ 
12  end
13   $\hat{l} \leftarrow min\_Idx(dist_1, \dots, dist_k)$ 
14  Err  $\leftarrow \hat{L} \neq l$ 
15  for  $i = 1$  to  $k$  do
16     $| D_i \leftarrow \{ \forall X_n \in \text{class } i \text{ AND Err}(c) \equiv \text{TRUE} \}$ 
17  end
18 end

```

Algorithm 1: Proposed algorithm for identifying means-sets (or control points) of manifolds for classification

- min_means is the minimum value along the axis of the means (i.e., the row).
- min_Idx gives the index of minimum value along each row.

Step-by-step flow of proposed approach is as follows:

1. Partitioning data matrix into c number of data sets.
2. Initializing mean sets (control points) with Null sets.
3. Each partitioned data set has samples from the respective classes.
4. Dividing each data set into n clusters with Fuzzy C-Means clustering (FCM (Bezdek et al. 1984)).
5. Concatenating identified centroids (means) for i^{th} data set with i^{th} mean set.

3. Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

6. Estimating euclidean distance between each sample of the data matrix and each mean set.
7. Estimating label of the sample based on closest mean set index.
8. Partitioned Wrongly classified samples into c number of data sets following the condition of Step 3.
9. Steps 4-8 are repeated if any partition has more than n samples.

During prediction, given a test sample, each means-set is considered and Euclidean distance is computed. To reduce the metadata, only the least distance is stored along with the label of the means-set. This results in k distances whose inversion reflects the possibility of a given test sample falling into each class. Thus, the index of the minimum distance is retrieved and associated as a predicted class for test samples. Considering a standard FCM the equivalence between the index of maximum membership and minimum distance for a data sample is given in equation (3.1).

$$\begin{aligned}
 u_j &= \frac{1}{\sum_{i=1}^k \left(\frac{|X_s - \mu_j|}{|X_s - \mu_i|} \right)^{\frac{2}{m-1}}} \\
 K_j &= \sum_{i=1}^k \left(\frac{|X_s - \mu_j|}{|X_s - \mu_i|} \right)^{\frac{2}{m-1}} \\
 K_j &= \sum_{i=1}^k \left(\frac{|X_s - \mu_j|^{\frac{2}{m-1}}}{|X_s - \mu_i|^{\frac{2}{m-1}}} \right) \\
 K_j &= |X_s - \mu_j|^{\frac{2}{m-1}} \sum_{i=1}^k \left(\frac{1}{|X_s - \mu_i|^{\frac{2}{m-1}}} \right)
 \end{aligned} \tag{3.1}$$

(Lemma : second part is constant for each sample)

$$K_j \propto |X_j - \mu_j|^{\frac{2}{m-1}}$$

⟨Lemma : if $a < b$ & $m > 1$ then $a^{\frac{1}{m-1}} < b^{\frac{1}{m-1}}$ ⟩

$$K_j \propto |X_s - \mu_j|^2$$

3.4. Results and Discussion

Thus the index for maximum u_j is the same as the minimum K_j where X_s , μ_i , m , k , u_j , U , K and j denote the random data sample, mean of i^{th} cluster, membership, classes, membership of sample with class j , membership values of a sample with all classes, distance values of a sample and centroid index respectively.

3.4 Results and Discussion

3.4.1 Spiral example

The proposed algorithm with four clusters was considered. With each iteration, the means-set vectors of classes ‘1’ and ‘2’ are represented in Figure 3.3 by the red asterisk and blue squares, respectively. At the end of each iteration, all samples are considered as the test set and predictions ‘1’ and ‘2’ are respectively illustrated as red and blue in Figure 3.3. Thus, a red diamond and blue circle show correct predictions and the remaining combinations are incorrect predictions. It is evident from the evaluation that the proposed approach is capable of modeling complex structures. The example also serves to illustrate the self-correcting nature between Iterations 1 and 2. A red asterisk control point at 0.1 along X-axis and 1.1 along the Y-axis correctly classifies diamonds but wrongly classifies circles. Thus, a blue rectangle was automatically introduced in the second iteration to counter it at 0.1 along the X-axis and 1 along the Y-axis. The same trend can be observed in Iteration 2 with the red asterisk at approximately -0.4 along the X-axis and 0.8 along the Y-axis.

3.4.2 Synthetic toy example

In both scenarios After training the proposed approach (with 2 clusters), it generated two centroids (control points) per class, as illustrated by the red asterisk for Class ‘1’ and the blue square for Class ‘2’. The feature space was divided based on these centroids, with blue denoting Class ‘1’ and yellow denoting class ‘2’ (note: using the same color for

3. Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

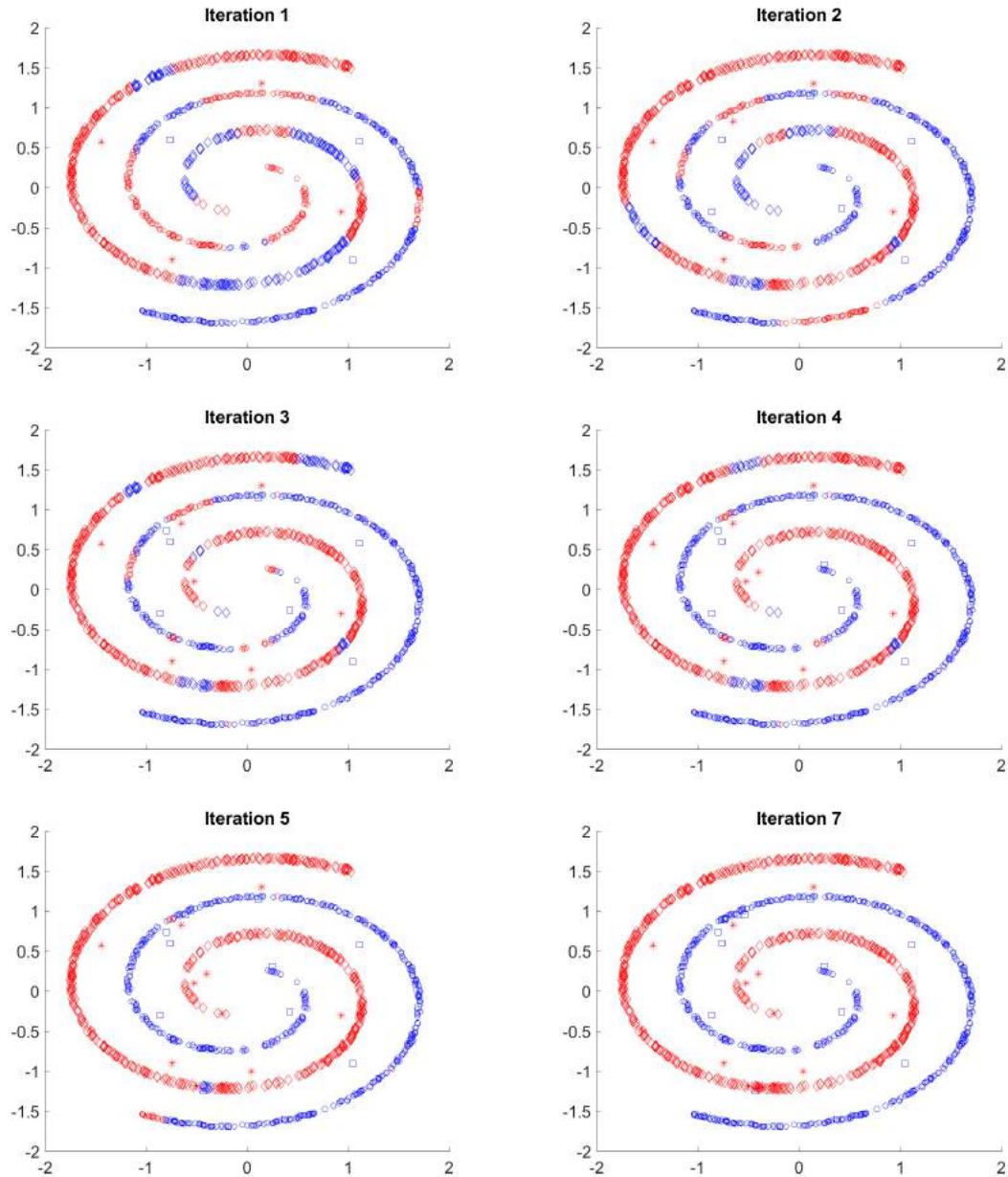


Figure 3.3: Illustrates the capacity of proposed approach to model complex structures (spiral example) through identifying control points (asterisk and rectangle). The modeling capacity of the approach can be depicted through color allocated to each spiral point during prediction and consistency within each spiral is achieved in a few iterations. X and Y axis are spatial coordinates of sampled spirals.

3.4. Results and Discussion

centroids affects the contrast). The contours reflect the boundary conditions of different confidence levels in Figure 3.4. The black line shows the hard decision boundary. The predictions of the test dataset for the scenarios are illustrated in Figures 3.4b and 3.4c. The circles and diamonds represent the ground-truth classes and colors blue and red represent the predicted classes of the test set. Thus, blue circles and red diamonds show correct predictions and other combinations of shape and color are false predictions. Figures 3.4b and 3.4c also demonstrate that the algorithm is capable of constructing linear and piecewise linear (i.e., nonlinear) boundaries simultaneously without the need for kernel knowledge.

3.4.3 Modeling characteristics and predicting capability of proposed algorithm for retinal pathologies

Following the training pipeline proposed by (Srinivasan et al. 2014), each image in a training set is translated to features, as defined above and the subject-level label is considered the image level label (weakly supervised). The concatenation of feature vectors and labels (scalar) for all images in the training set result in a data matrix and a label vector. Figure 3.5 shows classification training characteristics (of three centroids) for the first fold, along with each iteration. Because the feature space extends to the thousands, each samples is projected onto 2D space through t-SNE (Maaten & Hinton 2008). Following the same notations as above, the shapes are true classes and colors are predicted classes (note: ambiguous images were removed during training). Diamonds, circles and plus signs are true first, second and third classes, respectively. Red, blue and green are predicted first, second and third classes, respectively. Even though the first iteration managed to classify the majority of the samples, there were misclassification between samples from different classes. The same is shown in the magnified versions of each iteration. The bottom-right image in Figure 3.5 illustrates the convergence of the system.

3. Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

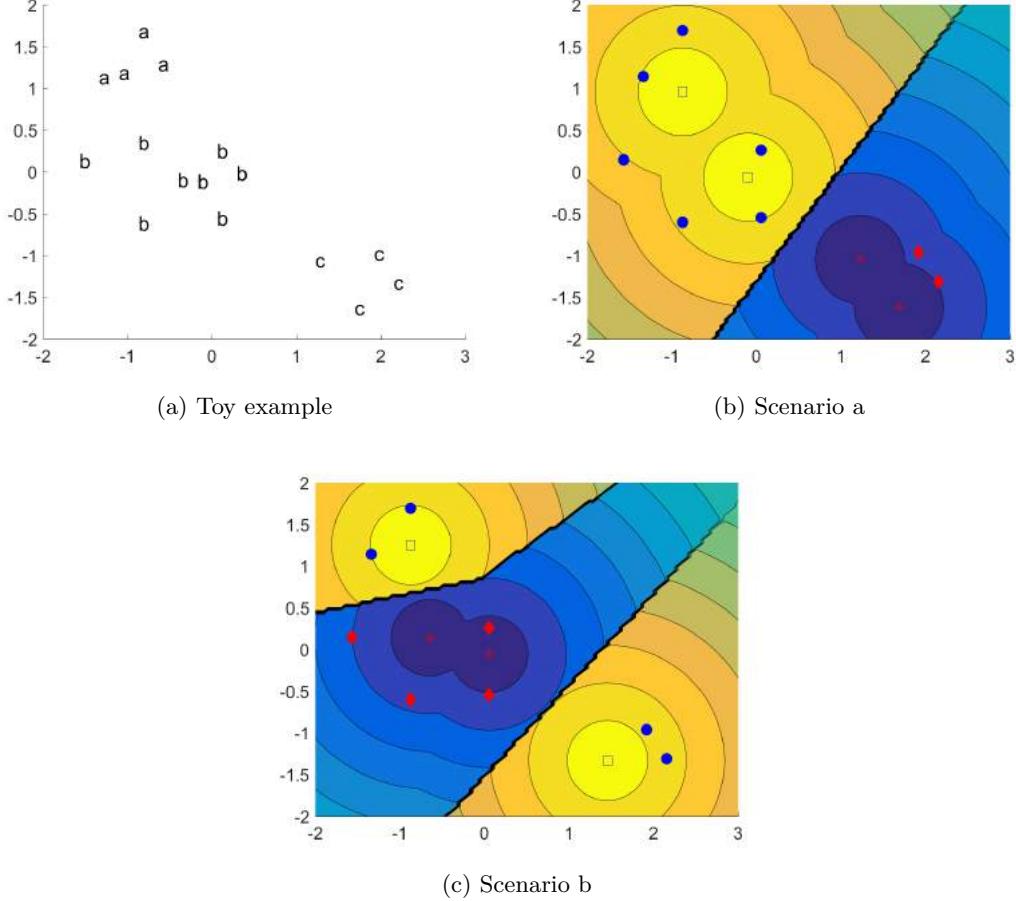


Figure 3.4: The proposed approach is capable of identifying control points which autonomously constructs the boundaries subjective to the scenarios ((b) linearly separable and (c) non-linearly separable) generated from toy example.

The data standardizing pipeline for the training set has been extended to the test dataset and constructed test data matrix and test labels. During testing, the class of the closest mean vector (or control point) is allotted as the prediction for the test sample. The image-level prediction accuracy for the proposed approach, a linear SVM, an SVM with Gaussian kernel and an SVM with radial basis function (RBF) kernels are 0.868, 0.855, 0.452 and 0.452, respectively.

3.4. Results and Discussion

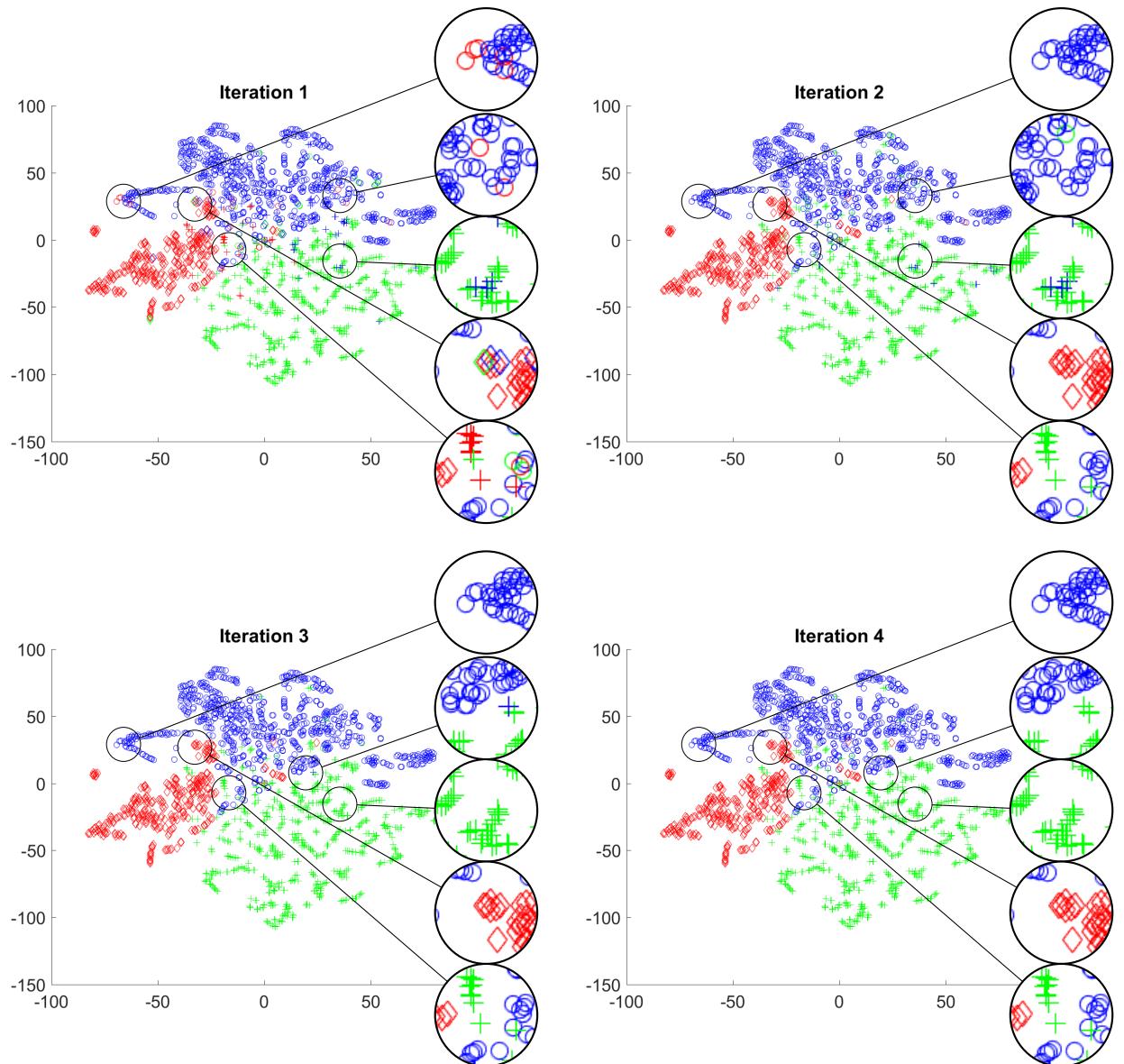


Figure 3.5: Illustrating the modeling capability of proposed approach for separating retinal pathologies (circle, diamond and plus implies normal, AMD and DME) along each iteration. The zoomed sub-images illustrates the capability of resolving ambiguous samples (along the boundary) with each iteration. The X and Y axis represents the retinal OCT HOG feature space projected onto 2D space.

3. Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

Table 3.1: Decision pooling performance across proposed and baseline algorithms

Method (parameters)		Normal	AMD	DME
Proposed	$c=2, m=1.2$	0.9905	0.8476	0.9333
	$c=3, m=1.2$	1.0000	0.8667	0.9143
	$c=3, m=1.1$	1.0000	0.8762	0.9524
Random forest	Trees = 1	0.9714	0.7429	0.9048
	Trees = 2	1.0000	0.7810	0.4000
	Trees = 3	1.0000	0.8286	0.8952
	Trees = 4	1.0000	0.8476	0.8000
	Trees = 5	1.0000	0.8571	0.8857
SVM	Linear	0.9429	0.8857	0.9143
	Gaussian	1	0	0
	RBF	1	0	0

3.4.4 K-fold validation performance in comparison to baselines

The protocol proposed by (Srinivasan et al. 2014) was employed to the test accuracy. Decision pooling accuracy is commonly used with Duke’s dataset for comparison. Every image in a pathological subject need not have a pathological state. Consequently, majority voting (i.e., decision pooling) of all image predictions in a subject was used as the subject level prediction. The average accuracy of decision pooling for test sets across 15 fold validation is depicted in Table 3.1. This illustrates the impact of governing parameters when modeling the performance of the proposed approach with different parameter combinations.

Random forests (with different number of trees) and an SVM with different kernels were considered the baselines, in order to illustrate the impact of proposed approach. Indeed, structured SVM or SRF were not employed insofar as they are designed for structured predictions.

3.4.5 Regulating parameters

There are two regulating factors for the proposed algorithm: the membership function and the resolution of the finest subspace after clustering. *membership function:* FCM

3.5. Summary

clustering is the generalized version of K-means clustering and considers a higher-order membership function whose degree is regulated by a factor. The membership factor (m) is helpful to define the fuzzy overlap between clusters. *Number of clusters*: This is another user defined factor for handling manifolds with more convoluted structures (e.g., spiral). The higher the number of clusters, the more refined the subspaces after clustering. However, this increases computational time and the number of control points that must be stored and it also affects the testing time.

It might be asked "why there are two clusters. That is, why should we not adopt mean per class?". In answer, consider a toy example with two concentric circles, where samples on the first circle belong to first class and so on. Upon the first iteration with one cluster, the means-set has one mean vector per class where both are aligned. At the end of each iteration during error computation, all samples are allotted to one of the classes (with a high number of samples) and all samples from another class are erroneous samples. During the second iteration of proposed algorithm, the means-set of erroneous samples remains the same as it was previous. Thus, the algorithm will be stuck in a void.

3.5 Summary

Classification through the proposed approach can create subspaces with piecewise wise linear boundaries through identifying control points representing the manifolds in feature space. The approach utilizes class-subjective manifold representations and optimizes by considering class labels for auxiliary cost computation. Whereas conventional classifiers aim at a global separation of the manifolds i.e., to compute decision boundaries. The decision tree involves identifying decision rules based on entropy and the SVM involves identifying the hyperplane with maximum separability based on support vectors (i.e., a boundary representation of all samples). In other words, one does not consider the representation of the manifold and the other considers the boundary sample representing

3. Modeling the Manifolds with Guided Fuzzy C-Means for AMD and DME Subjects Elucidation: A Supervised Approach

the boundary but not the manifold itself. As illustrated in various empirical experiments, SVM is opted for small training dataset and random forest is for large training datasets (insofar as sampling statistics aligns with population statistics).

The proposed approach frees the system from knowledge regarding the manifold's structure or spaces and expert information can be incorporated through m if needed. This affects its accuracy in current application as the selection of an incorrect kernel results in reduced accuracies. There is an important difference between human intervention through kernels in an SVM and through m in the proposed approach. In the case of an SVM, the expert must know which kernel to select. In case of the proposed approach, on the other hand, if the expert does not know m , FCM can be replaced with K-means clustering. The control point (centroids) strategy employed is resilient to noise, because the proposed approach can identify a separate control point for the outlier. Voronoi fractals offers a clear advantage by employing global-to-local modeling, rather than simultaneously modeling global and local representations, which is computationally expensive. This strategy also makes the approach free of decision boundary computations, since bounds with respect to centroids on L2 space effectively act as decision boundaries. Collectively, from a naive perspective, if the feature space is considered a stable water surface, then centroids are an introduction of point force at different locations. The ripples represent the contours and colors represent various sources of the forces. The intersection of ripples from various sources, then, is treated as the constructed decision boundary.

Chapter 4

Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies

Transfer learning protocol for CNN architectures is subjective to application and associating learned feature quantifiers responses with clinical land marks can bridge the void between clinical community and image analysis community to some extent.

—The Learner

4.1 Introduction

A Classifier performance depends on low-level information computed by feature quantifiers and any information loss during this quantification process affects the classifier's performance. An effective feature quantifier can empower even a naive classifier to delineate the pathologies. As an example, the HOG feature descriptor has boosted pedestrian detection solutions, even with a standard decades-old classifier (Dalal & Triggs 2005). The HOG features are widely employed for retinal classification, because they can represent geographical atrophy in terms of gradient orientations, as illustrated in Figure 4.1.

4.1. Introduction

Such orientations cannot distinguish between the normal curvature (fovea region) or geographical atrophy (due to drusen deposition in case of AMD or fluid in case of DME). The current work constructs hierarchical representations in input space to effectively quantify the geographical atrophies without heuristics and give high responses only at abnormal locations. Additionally such responses need to be visualized and interpreted as commonly available feature quantifiers (like local binary pattern) only represents partial information.

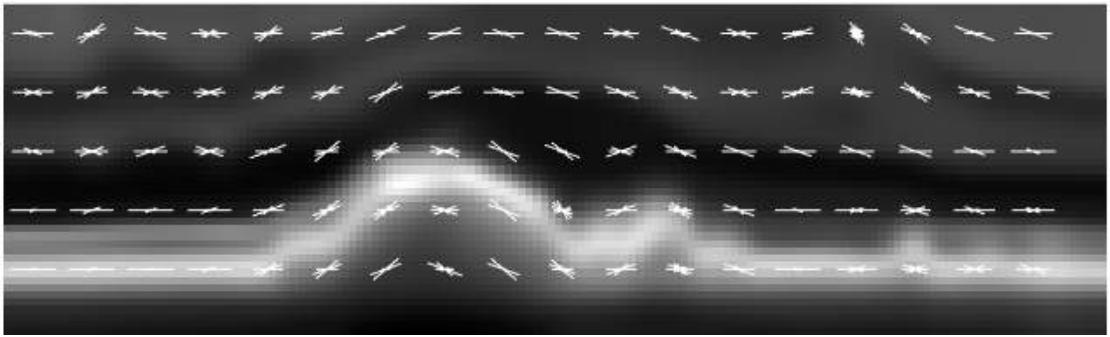


Figure 4.1: Illustrating the variations of gradient orientations along geographical atrophy (due to drusen deposition) depicted by a HOG descriptor.

4.1.1 Need for learning feature quantifiers as representations in input space

Conventional practice involves identifying representative features to delineate target classes and formulate descriptors that quantify subjective features. It is always difficult to construct a mathematical expression to represent an anticipated feature. Thus, multiple mathematical basis quantifying various partial aspects of the anticipated feature are considered. For instance, an orange can be quantified as round (i.e, its eccentricity is 1), medium in size (i.e, a radius of 3 cm), having a bright red color, medium green color, no blue color, etc. Common hurdles with this approach stem from the facts that representative features for low-frequency patterns are difficult to attain, some residual information is always held back irrespective of the pool of features considered, the whole

4. Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies

process is reiterated upon consideration of a new pathology, etc. This burden is reduced by employing existing salient feature descriptors for various applications. This way, the attained features are refined by selecting a feature subset (filter, wrapper, or hybrid approaches) or by feature extraction (principle component analysis, an autoencoder, etc.) to quantify anticipated features. Current algorithms for feature extraction are capable of constructing anticipated features from raw information, thus reducing the dependency on salient features.

4.1.2 Automated representation learning

The feature extraction process has been an integral part of image analysis, transcending from choice to need in the analysis pipeline. Classical techniques involve constructing a transfer function to project computed salient features onto a feature space in which samples from different classes are easily distinguishable by a classifier. There are both supervised and unsupervised parts to feature extraction and the proposed approach has adopted the former due to its effectiveness. The approaches of the past decade have illustrated the potentiality for considering raw information as input and constructing feature descriptor for computing feature quantification and extraction in a single transfer function. The existence of such a transfer function depends on the anticipated features. Even when it exists, identifying a complex transfer function is computationally expensive. To overcome such hurdles, a hierarchical approach is employed. According to this approach, a single transfer function is replaced by a sequence of transfer functions. Given raw information, multi-stage approaches sequentially process raw information through a set of transfer functions. This process is functionally equivalent to an anticipated single transfer function. A convolutional neural networks (CNN), a common architecture employed in deep learning, is one such approach to image analysis. From a signal processing perspective, convolving information with filters in a layer of a CNN is equivalent to imposing a transfer function. A CNN is like a Lego piece, whereby multiple architectures

4.1. Introduction

can be generated using four basic types of building blocks. Not all combinations can construct the anticipated transfer functions set, however, so the selection of architecture is crucial. The GoogLeNet (Szegedy et al. 2015) architecture is considered here, because it established benchmark performance and offers fewer number of filter weights, making it less prone to overfit.

4.1.3 GoogleNet

CNN architectures (LeCun et al. 1998, Krizhevsky et al. 2012, Szegedy et al. 2015) for image classification involve combinations of the convolution block, with learnable filters, an activation block, a cross-map local response normalization (LRN) block, a pooling block, a fully connected (traditional neural network) block with learnable weights, a dropout and a loss block, as illustrated in Figure 4.2. Given an image or the responses, the convolution block convolves with a set of learnable filters. The activation block alters each element in the input based on a user-defined function. The local response normalization acts as regularizer for unbounded activation functions. The pooling block replaces each element based on a statistical operation (maximum or mean) in respect to neighboring elements (image dilation or mean filtering). The fully connected layer is a traditional artificial neural network. The loss block computes the error between the predicted class and the actual class. The error computed in the loss layer is utilized to correct the filter weights of each convolution block with gradients computed through error back-propagation.

4.1.4 Visualizing learned representations

Interpreting learned filters has been a common trend in image analysis. This interpretation is used to determine whether anticipated representations are modeled and to understand the characteristics of the information processing pipeline. There are a wide variety of approaches to depict the feature representations from first layer filters, includ-

4. Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies

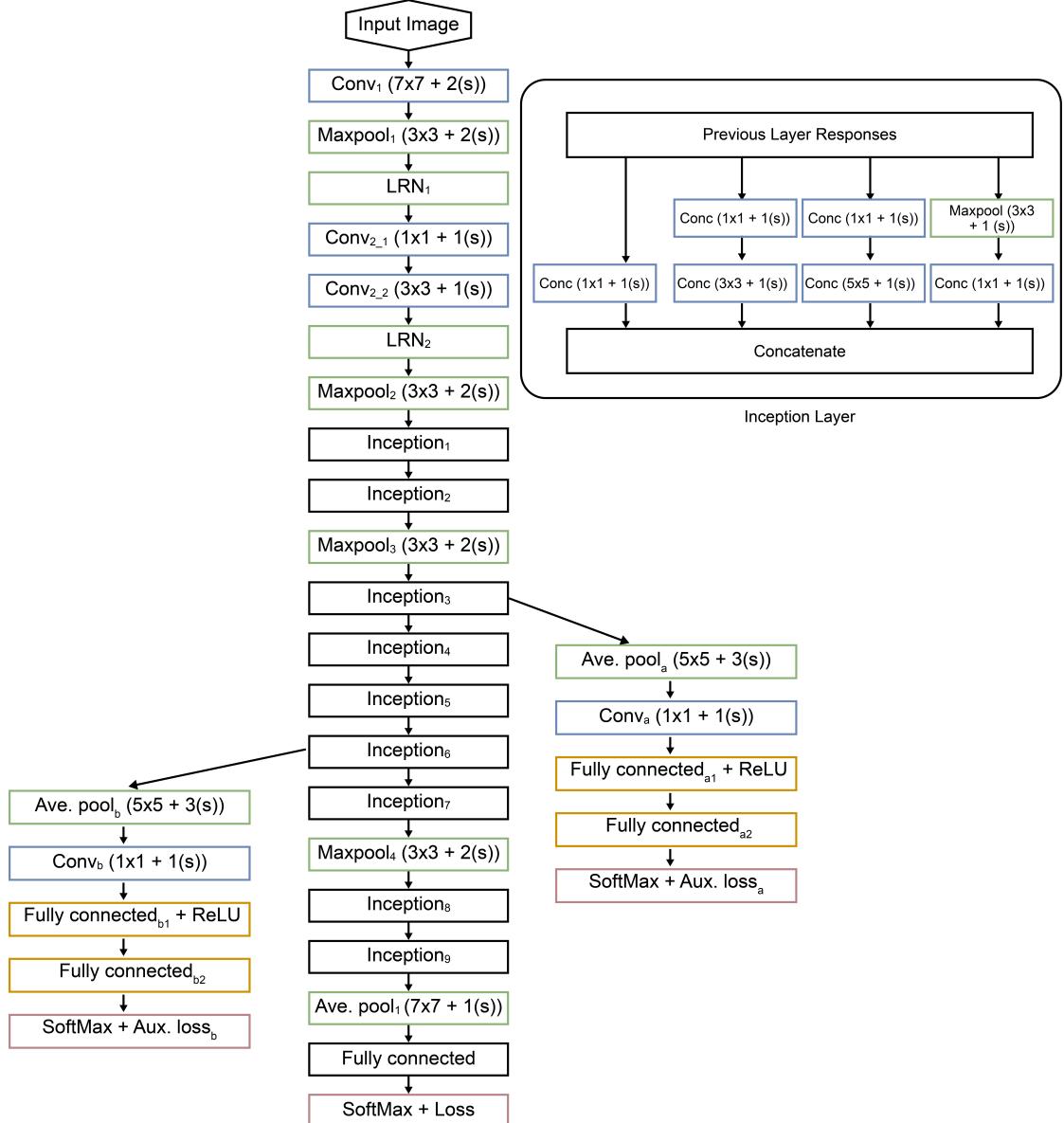


Figure 4.2: Commonly employed GoogLeNet architecture with illustration of inception block for classification applications.

ing hand picking the important filters, illustrating the salient points in the input image, etc. There is also interest in interpreting information from hierarchical layers and this has been a major drive in trusting deep learning. Such information will not always be as elegant as FaceNet (for detecting human faces) but it can reflect distinguishable abstract

4.2. Preamble to Solution

patterns for different classes.

4.1.5 Identified approach

Learning representations from raw information with a CNN requires ample data to identify repetitive patterns or potential patterns. The transfer learning approach is employed in case of crunch in image data which is common in medical domain and retinal OCT in particular. Such an approach requires a pretrained model. For this research, GoogLeNet trained on large image dataset (ImageNet) is available online and has been employed. The reason for choosing GoogLeNet is that it has fewer parameters or weights in comparison to other architectures. Reason being, overfitting is a common problem when transfer learning CNN architectures, particularly in cases with less training data. A common way around is choosing a model with fewer parameters. Thus, GoogLeNet is employed, insofar as it has relatively few parameters, compared to modern deep CNN architectures. Transfer learning offers improved performance in comparison to random initialization, but it need not be better to a classical approach (always). Thus, HOG features were used for a comparison.

Visualizing representations is a two-fold process: firstly, a potential filter in the first layer is identified autonomously and interpreting its capacity to describe salient feature; and second, gross interpretation of final layer's representations to provide a visual validation of the difference in abstract patterns from different classes.

4.2 Preamble to Solution

Transfer learning as a method for learning representations for a specific application is a common practice in various domains and particularly with microscopic and computed tomography images with Gaussian as underlying distribution. However, retinal OCT's speckle distribution is double exponential and its image distribution depends on adaptive compression techniques. The proposed approach explores the expandability of

4. Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies

GoogLeNet to retinal OCT images, with modifications to handle low probabilities at the final layer. It also deviates from conventional training strategies regarding data standardization and optimization for improved performance. Retinal flattening and preprocessing are similar to the process described in the chapter 3, but without a heuristically identified region of interest, i.e., by cropping of the flattened retinal image with reference to the lower boundary of the retina. The multi-scale image is not considered as the input, because GoogLeNet inherently computes at multiple-scales in the inception block, by convolving with filters of multiple receptive fields i.e., filter sizes 3×3 and 5×5 .

The first layer in GoogLeNet has 64 filters and identifying a potential filter autonomously is crucial. The proposed approach is gradient-based and it is inspired from neural paint (Gatys et al. 2015) insofar as it corrects the filters contributing to the style (see Figure 4.3). The initial filters depict feature descriptors similar to low-level information processing and act as a foundation for successive filters. The final responses must also be depicted, because they are the sample points in the feature space and illustrate inter-class variance. The top layer’s feature descriptors aim for sparse representation. Thus, gross representation is illustrated by the statistical combination (max. operand) of all responses.



Figure 4.3: Illustrating an example of neural style application for adapting Van Gogh’s starry night style to dog image.

4.3. Materials and Methods

4.3 Materials and Methods

4.3.1 Experimental setup

Duke's OCT classification dataset is used for the experiment. The Torch API is employed for training and testing the CNN architectures along with the dependencies listed described in Chapter 2. A workstation with 12 GB GPU is required, because CNN architectures can be trained more rapidly on GPU than a CPU.

The train set is a collection of first eight subjects images from each class whether AMD, DME, or normal (24 subjects in total) and the corresponding labels were considered. Training set images is processed to generate mean subtracted images. The test set comprises the remaining seven subjects' images from each class and the corresponding labels are used to compute the accuracy. Training set images were processed to generate mean-subtracted images and divided them into batches. GoogLeNet trained on the ImageNet dataset is employed for fine-tuning on the training set. The test accuracy of the complete test set is computed after each epoch.

4.3.2 Data preparation

Every image's saturated regions (with an intensity value of 255) is replaced with 10 and normalized. Retinal flattening is performed such that the lower contour is shifted to 70% of the height. Each image is resized to 224×224 since employed GoogLeNet was trained on images with such dimensions. Images are filtered with BM3D filtering, because OCT is constructed from speckle information. A three channel image is constructed by replicating the BM3D output, since employed GoogLeNet was trained on color images and each replicated output is treated as channel information. Following the strategy in (Srinivasan et al. 2014), the label for each image is allocated. Approximately 53% of the images are considered for training and remaining images for testing without any overlap of the subjects between training and testing sets. After dividing the dataset in this way,

4. Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies

the ‘mean image’ (3x224x224) from the training set is computed and subtracted from each training and testing image.

4.3.3 Modifications to trained GoogLeNet on ImageNet

- Fully connected layer, fully connected layera2, fully connected layerb2 have three outputs because current classification task has three classes i.e., normal, AMd and DME.
- All the weights for the fully connected layers and convolutional layers in auxiliary branches are randomly initialized because there is a need to learn feature space shifts between ImageNet data and OCT data.
- All final layers have been changed from softmax to log softmax because the softmax probabilities are low and this avoids an underflow of gradients.
- Negative log likelihood is considered as loss because product of logits in case of likelihood are converted to sum of logits which avoids drastic fall due to a single small value.
- The new GoogLeNet is trained with Adagrad (Duchi et al. 2011) optimizer rather than Adam or stochastic gradient descent (SGD) method. Indeed, using Adam results in impulsive step lengths and SGD requires expert introduced learning rates at different epochs.

4.3.4 Fine-tuning

An experiment involved fine-tuning a pre-trained CNN with a mean subtracted training set and simultaneously predicting the accuracy on a mean subtracted test set at each epoch.

The training phase involved breaking the mean subtracted training set into batches of ≤ 64 images. On the forward pass of each batch, the CNN predicts 64 labels from which

4.3. Materials and Methods

the loss is computed through negative log-likelihood. The loss is translated to gradients at each layer using the chain rule in error back-propagation, which corrects the filter weights. The amount of correction incurred is determined with an Adagrad optimizer. Each iteration involves selecting a subsequent batch, a forward pass, a backward pass and weight correction. Under the current setup, the training set is broken into 28 batches. A collection of 28 iterations is treated as an epoch and training is performed on 50 epochs. That is, complete training data has been pushed though the CNN for times.

The Testing phase, at each epoch, splits mean subtracted testing sett into batches of ≤ 65 images. A forward pass of CNN predicts labels for each batch and such labels are concatenated as predicted labels for the test set. After completing the 50 epochs, mode (majority voting) of all test image predictions from a subject is treated as subject level decision (Srinivasan et al. 2014).

4.3.5 Identifying potential response

The first layer of a fine-tuned CNN has numerous filters, resulting in a large number of responses. In a few cases of modifying CNN architectures, experts interpret the filters and responses to identify repetitive filters, missing patterns, etc. Such response visualization also helps non-experts to understand the information processed at each layer. Conventionally, representative responses are hand picked. The current approach is automated, however, inspired more by the basic definition of pathology (i.e., deviation from normalcy). Thus, given an abnormal test image, if the fine-tuned GoogLeNet predicts it as abnormal (i.e., as AMD or DME) then considering a pseudo ground-truth ‘normal’ (irrespective of being AMD or DME) error is back-propagated. Now the filters that are more responsible for predicting the image label are expected to correct more through gradients. In other words, for each layer, the filter with the highest gradient is treated as the potential filter and the corresponding response is treated as the potential response (or representative) response.

4. Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies

4.3.6 Visualizing abstract patterns

Top layer response patterns are abstract and sparse by nature. Thus, all such information must be accumulated for visualization. The ‘maximum response’ has a special place in CNN architectures for increasing the local receptive field of the CNN. The same strategy is employed for gross visualization of abstract patterns. All responses of the 21st layer (Inception₉) are concatenated along the third axis and maximum operation along the third axis results in a 2D map. This 2D map is considered the visualization of abstract patterns.

4.4 Results and Discussion

4.4.1 Repeatability

The experiment was to fine-tune GoogLeNet on a defined training set over 50 epochs and compute the prediction accuracy on a defined test set at every epoch (recall that the training set and test set do not overlap). The prediction (test) accuracy was computed using GoogLeNet’s predicted labels for all test set images and the ratio of the number of correctly predicted labels to the total number of test images. Notably, CNN convergence and predictions are subjective to initialization. Random initialization in fully connected layers results in performance differences from experiment to experiment. Similarly, the dropout block also randomly nullifies (i.e., sets to zero) the percentage of responses. This also results in changes to the training model’s performance from experiment to experiment. Such randomness introduced during model fine-tuning produces variable results for each experiment. In practice, multiple (10) experiments are repeated and the best model (96%) is stored. The test accuracy of each experiment for every epoch is illustrated in Figure 4.4.

4.4. Results and Discussion

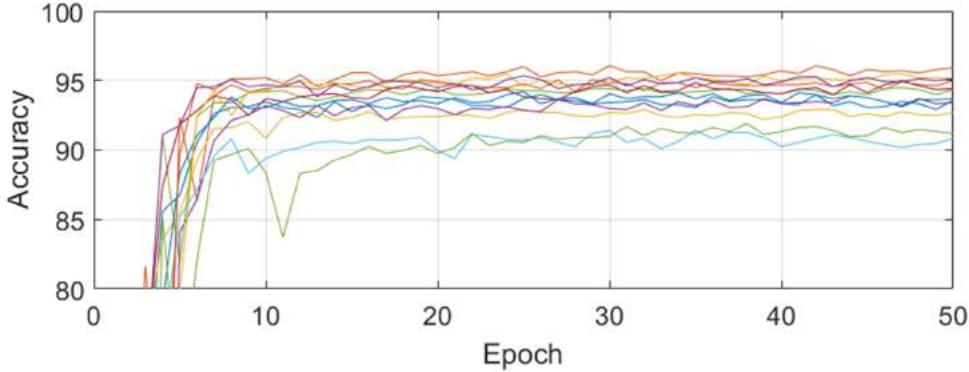


Figure 4.4: Illustrating the repeatability (in terms of test set image level accuracy) of the transfer learning experiment with varied random seed point for initializing fully connected layers.

4.4.2 Image-level performance

Keeping the repeatability in mind, the model with 94% accuracy was considered for evaluations. The impact of transfer-learning was illustrated by a comparison with GoogLeNet trained with random initialization in all filters. The plots in Figure 4.5 present the prediction accuracy on the complete test set (i.e, the remaining seven subjects) images and illustrates faster convergence with improved accuracy by using the transfer-learning approach.

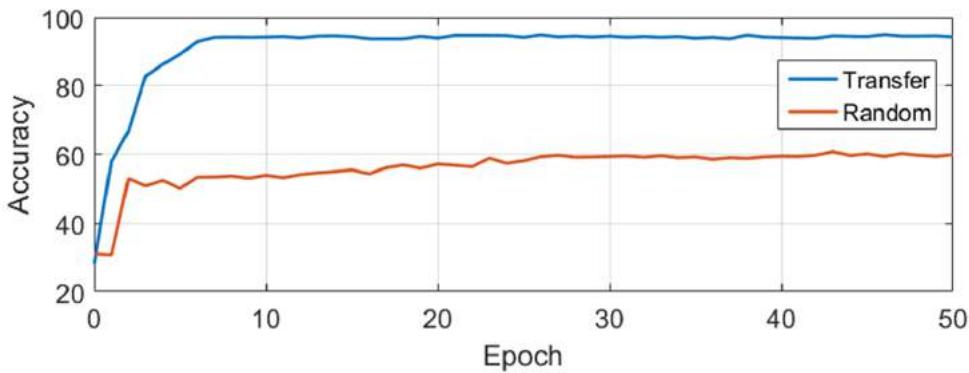


Figure 4.5: Faster convergence and higher accuracy for fine-tuning pre-trained model (transfer learning) in comparison to training a model with random initialization.

4.4.3 Visualization

Given a $3 \times 224 \times 224$ OCT image with AMD, the potential response at the first layer was identified autonomously, as shown in Figure 4.6. The Conv1 layer includes the filter that quantifies curves along with edges but proposed approach autonomously identified curves as potential filters. Curve filters are crucial for identification of deformations in the RPE layer due to drusen deposition. This also illustrates that the filters constructed appropriate descriptors for target curves (drusen deposited area). To achieve the same conventional practices employ bank of filters (wavelets) with attributes of the curves. Given a $3 \times 224 \times 224$ OCT image with DME, the potential response at the first layer was identified autonomously, as shown in Figure 4.6. The Conv1 potential filter identifies the deformations with respect to total retina. So the responses are high on the side of deformation but not the complete ILM. The normal case contains squashed responses, owing to a lack of deformations subjective to pathology. For a better contrast, each response was normalized and associated with pseudo-colors through Python plot tool where dark blue and dark red denotes lowest and highest response, respectively.

The gross response of GoogLeNet's 21st layer in in normal, AMD and DME cases are illustrated in Figure 4.7. Unlike the first layer, the final layer's abstract patterns are clearly distinct in the case of various classes.

4.4.4 Baseline

The images in the training set were processed to generate the BM3D filtered image. The process in (Srinivasan et al. 2014) was employed for HOG feature computation without cropping. A softmax classifier was employed for classification. This acted as a baseline to illustrate the predictive capability of the constructed feature quantifiers from the fine-tuned GoogLeNet. Thus, this shows grossly HOG features compared to fine-tuned GoogLeNet filters.

4.4. Results and Discussion

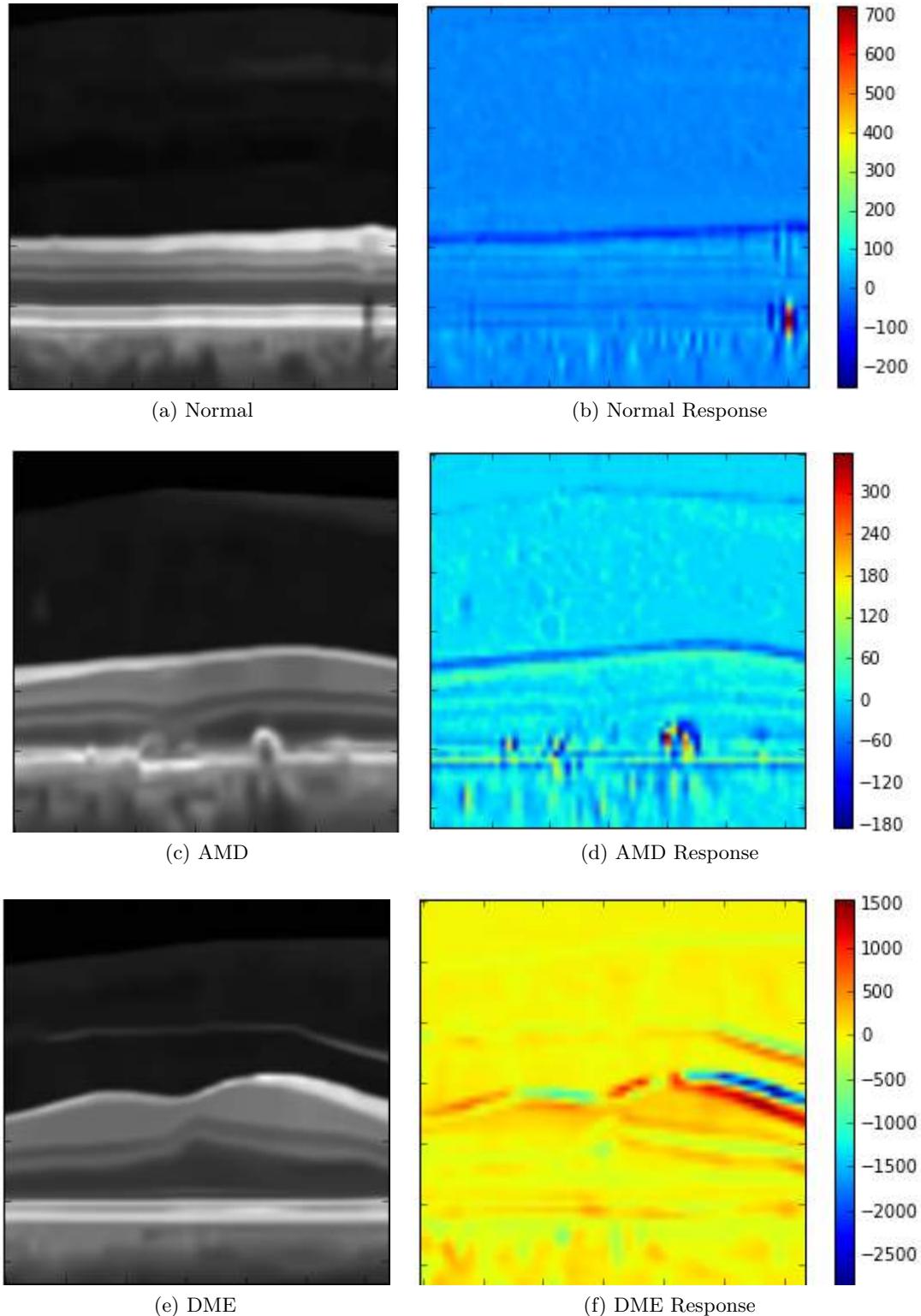


Figure 4.6: Identified potential response for normal, AMD and DME cases with low responses in case of normal image, peak responses at drusen depositions in case of AMD and high responses along deformation side (right) for DME case. Such information is more enriched in comparison to HOG descriptor's depiction. The color bar represents convolution response of the identified filter with input image or previous filter response.

4. Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies

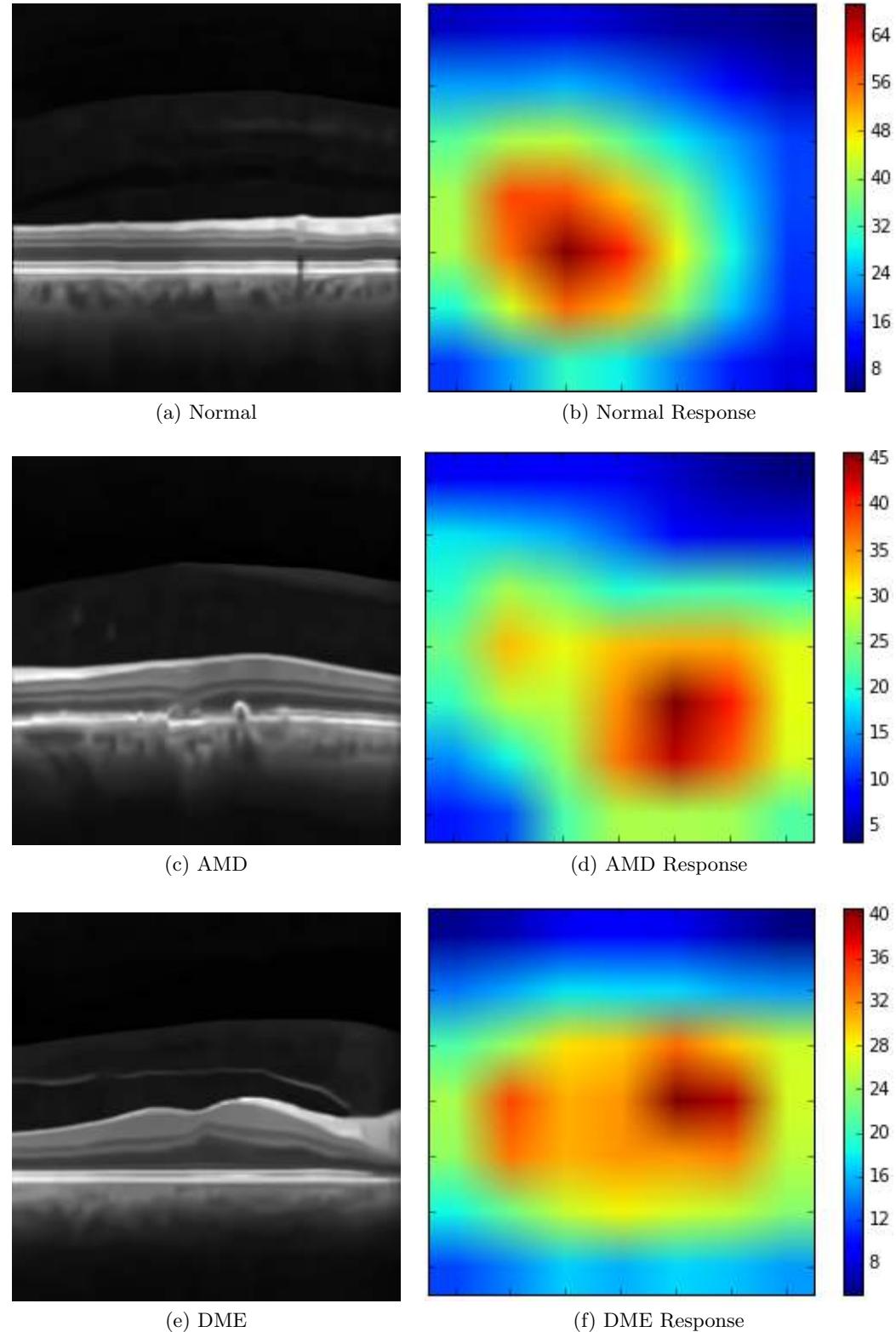


Figure 4.7: Visualizing the maximum response of GoogLeNet's 21st layer for distinction between normal, AMD and DME cases where the responses are clearly distinct from each other which reduces the burden on a classifier. The color bar represents convolution response of the identified filter with input image or previous filter response.

4.5. Summary

Table 4.1: Cross-validation accuracy of transfer learning with various class weights and loss weights settings

Class Weights			Loss Weights			Normal	AMD	DME
0.3	0.3	0.4	1	1	1	0.99	0.89	0.84
0.3	0.3	0.4	1	0.1	1.00E-04	0.99	0.89	0.86
0.2	0.2	0.6	1	1.00E-07	1.00E-14	0.98	0.85	0.81
0.33	0.33	0.33	1	1	1	1	0.88	0.81
0.33	0.33	0.33	1	0	0	0.98	0.85	0.85

4.4.5 Crossvalidation

Following the 15 fold validation described above, training and test sets for each fold were constructed. The average accuracy of decision pooling across all folds for the proposed approach for normal, AMD and DME cases was 0.99, 0.89 and 0.86, respectively. The average accuracy of decision pooling across all folds for the baseline was 1, 0.9 and 0.78. In comparison to carefully crafted feature descriptors (HOG), the proposed approach is comparable in cases of normal and AMD and demonstrated considerable superiority in the case of DME.

The above mentioned performance is the best performance of the CNN. This was achieved by tweaking the governing factors of the training process such as class weights and loss weights. Class weights are helpful to compensate for unbalanced data, i.e., imbalance in the number of data points per class. A GoogLeNet has multiple losses (main and auxiliary losses), loss weights regulate the contribution of each loss. The variation in performance with various combinations of class imbalance and loss weights is illustrated in Table 4.1.

4.5 Summary

CNN has been successful at automatically crafting feature quantifiers for particular image analysis applications. CNN based algorithms require a large amount of data and a large number of epochs for convergence. To avoid such learning barriers models trained

4. Transfer Learning based Data-driven Feature Quantifiers for AMD and DME Pathologies

on similar datasets must be fine-tuned for various applications (Mesnil et al. 2012, Yosinski et al. 2014). Transfer-learning based classification or segmentation of medical images mostly pertains to microscopic and CT images. The underlying image distributions follow a Gaussian distribution similar to non-medical images. The current application involves fine-tuning models for OCT images whose imaging physics incorporate speckles of both information and noise.

In contrast to the widely employed VGG architecture, an effective architecture for image classification was identified and extended to the classification of retinal pathologies. A standard softmax classifier was employed, but to overcome underflow in softmax a ‘log’ was appended in case of CNN output. The level of correcting weights based on loss was regulated by the learning rate and this should be finer with the progression of each epoch. AdaGrad or Adam optimizers are used to avoid human intervention in weight decay. However, Adam was prone to blowing learning rates. Hence, the AdaGrad approach was used. A common criticism of neural networks is ‘non-deterministic models’, i.e., upon the change of initialization, the final model characteristics change. The fully connected layers need to be randomly initialized for fine-tuning so statistical converge of the model is illustrated through repeating the experiment on a fold. Conventionally, the test accuracy is a surrogate for learning capabilities of the model. However, depicting layer-wise signal processing characteristics generate more confidence to the medical and biomedical communities. Conventionally, the salient location in the input image is illustrated, but identifying the potential filter or gross representation autonomously has not been well explored. The proposed block partially solves this by autonomously identifying potential filters and corresponding responses at each layer. The proposed automated selection is a ranking process. Hence, all filter responses can be arranged based on their potentiality. The initial layer filters are crucial, because CNN being a sequential (hierarchical) process any information lost in the initial layer cannot be retrieved by any top layer. The top layer filters have significance from the perspective

4.5. Summary

of constructing the feature space and distinction between the abstract patterns from different classes is reflected in the distinction of data points in feature space. So, both initial layer filter responses and 21st layer (Inception₉) responses were illustrated. As future scope the initial filter's response can be visually validated by the expert, whether or not the potential filter (i.e., feature descriptor) has been constructed.

Chapter 5

Low-Rank CNN Filters for Simultaneous Reconstruction and Restoration of Sparsely Sampled Retinal OCT Image

Sparsely sampled OCT images can be effectively restored with fewer parameters through adapting CNN (functionally equivalent to sparse coding) for super resolution application. The required parameters can be further reduced through learning low-rank filters with geometrical constraints.

—The Learner

5.1 Introduction

Low coherence interferometry avails SD-OCT to image the subsurface morphology of a sample along the transverse plane with a lateral and axial resolution of few micrometers. However, the images acquired suffers from low peak signal to noise ratio (PSNR). As SD-OCT is capable of acquiring a multiple images in few seconds, imaging the same

5.1. Introduction

region for multiple instances and temporal averaging results in image with higher PSNR (noise-free image). Thus, clinical SD-OCTs are equipped to image a region of the retina for multiple instances through flying-spot scanning technique. The noise-free image (HD image) is constructed through affine registration of multiple images and averaging such registered images (instances). However, patients exhibits involuntary movements (McNabb et al. 2012) and imaging the same location results in pixel displacement (non-rigid i.e., not affine) between instances which induces blur in reconstructed noise free image. This restrains the number of B-scans (instances) and A-scans while imaging to avoid motion blurs (Yun et al. 2004). The need for noise free image has resulted in the need for single-image-based reconstruction and restoration algorithms. This can be achieved through ANN which models the representations in input space and predicts the representations in target space (i.e., high detailed image) of retinal OCT image. The required parameters in neural network can be reduced by inducing geometrical constrains inspired from separable filters.

5.1.1 Need for sparse representations

Computational imaging includes multiple approaches such as initially denoising the images (Chitchian et al. 2012, Avanaki et al. 2014) and then estimating the high detailed noise-free image (I_H) from a sparsely sampled noisy image (I_S) (Bian et al. 2015) or vice-versa. Alternative algorithms employ sparse dictionary learning or sparse coding. Such approaches are multi-image based and require neighbouring location images that must intervene with the imaging protocol for standard OCT. Therefore, the current approach aims at a single-image-based restoration. A comparison of all algorithms including image restoration, filtering and sparse coding (SC) approaches on a benchmark OCT dataset shows that a sparse coding based approach offers superior performance for OCT images (Fang et al. 2013). Given a set of image patches, sparse coding iteratively identifies a dictionary through *orthogonal matching pursuit* with the capability of trans-

forming image patches to sparse representations (vectors). Sparse coding is designed to identify sparse representations through introducing sparsity measure as regularizer in cost function during construction of dictionaries, which is major reason for its superior performance. Such dictionaries are capable of transforming image patches into a sparse vector and projecting back sparse representations to image patches. This is employed by computational imaging for super-resolution applications.

5.1.2 Extending sparse representations for OCT denoising

For OCT applications, sparse coding is wrapped as supervised learning where image patches from sparsely sampled noisy images (I_S) and corresponding reconstructed noise-free images (I_H) are required. Noise-free patches (P_H) and noisy patches (P_S) are generated based on I_H and I_S . Noise-free patches are separated using K-means clustering and noisy patches are assigned to the same cluster as corresponding noise-free patches. Then, the centroids for each cluster are identified. In each cluster, dictionaries are learned independently for noisy patches and noise-free patches. Then, each patch is transformed into a sparse vector employing a noisy patch based dictionary (B_S) or a noise free patch based dictionary (B_H). A mutual matrix (E) is computed through singular value decomposition in order to transform the sparse vector of the noisy patch (α_S) into a sparse vector of the noise free patch (α_H). Given a test noisy patch, then, the closest noise-free patch cluster and corresponding dictionary (B_S) is identified. The identified B_S transforms the noisy patch into a sparse vector and a mutual matrix transforms this sparse vector into a sparse vector of the noise free patch. Finally, the noise free patch based dictionary estimates the noise free patch from the sparse vector of noise free patch. This approach learns dictionaries and the mutual matrix in a mutually exclusive manner (i.e., independently). This raises the need for multiple clusters, each with dictionaries containing thousands of elements and the consequent need to learn a few 100k parameters.

5.1. Introduction

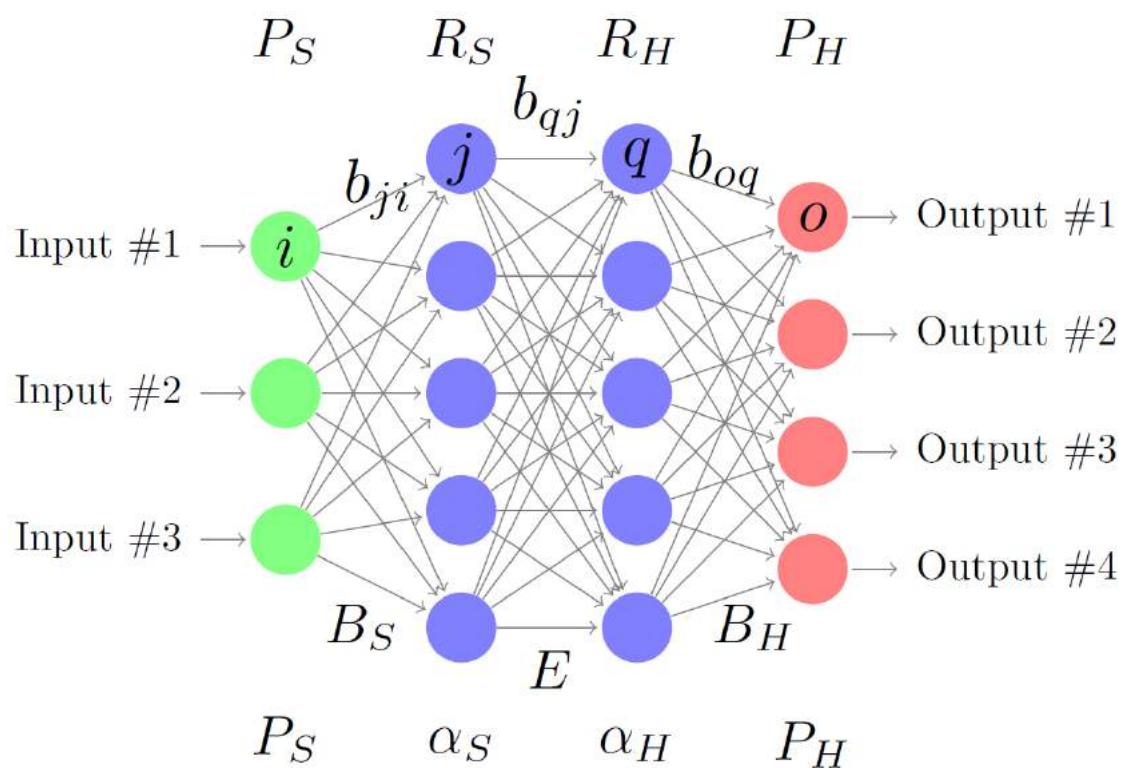


Figure 5.1: Illustrating the neural network which functionally equivalent to sparse coding framework for image super resolution.

5. Low-Rank CNN Filters for Simultaneous Reconstruction and Restoration of Sparsely Sampled Retinal OCT Image

Figure 5.1 illustrates the neural representation of dictionaries for OCT applications. Given sparsely sampled noisy image patches (P_S) and their corresponding high-dimensional (P_H) patches, from a sparse coding perspective, B_S , α_S , E , α_H and B_H represent the dictionary for noisy patches, sparse vectors of noisy image patches, the mutual matrix, sparse vectors of noise free image patches and a dictionary for noise free patches, respectively. From the perspective of a neural network, it is a multilayered perceptron architecture with two hidden layers where b_{ji} , b_{qj} and b_{oq} are expected to act as surrogates for B_S , E and B_H , respectively. These are identified through gradient-descent approach.

5.1.3 Identified approach

Sparse coding has been widely employed to reconstruct and denoise applications independently, so the current approach aims to combine both. Any ANN architecture can be revamped as a CNN and recent frameworks have extended sparse coding equivalent neural network architecture to a CNN architecture for rapid operations in super-resolution applications. The current approach aims to extend the capability of such architectures for simultaneous restoration and reconstruction. In the case of a CNN, intra-dependency within filters is a common and this can be exploited to further reduce learnable parameters through signal processing.

In signal processing, separability is an approach for factorization which reduces the total number of filter elements (or weights) by identifying low-ranked filters. Figure 5.2 shows separable filters, where the response of the image on the left-hand side and the right-hand side are the same. This results in a reduction of total filter weights and the number of multiplications (convolution is a combination of multiplication and summation). For example, a 3×3 filter requires nine weights and multiplications on the left side but only six weights and multiplications on the right side. The only draw back to the use of a separable filter is its sequential nature.

5.1. Introduction

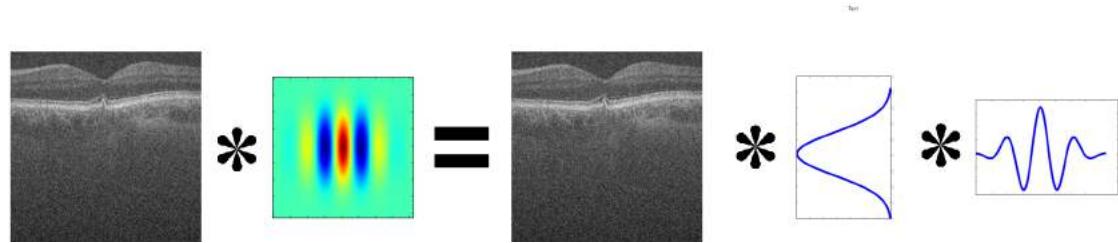


Figure 5.2: Illustrating the impact of separable filters. The result of image convolved with a 15×15 filter is equivalent to image convolved with separable filters decomposed from corresponding 15×15 filter.

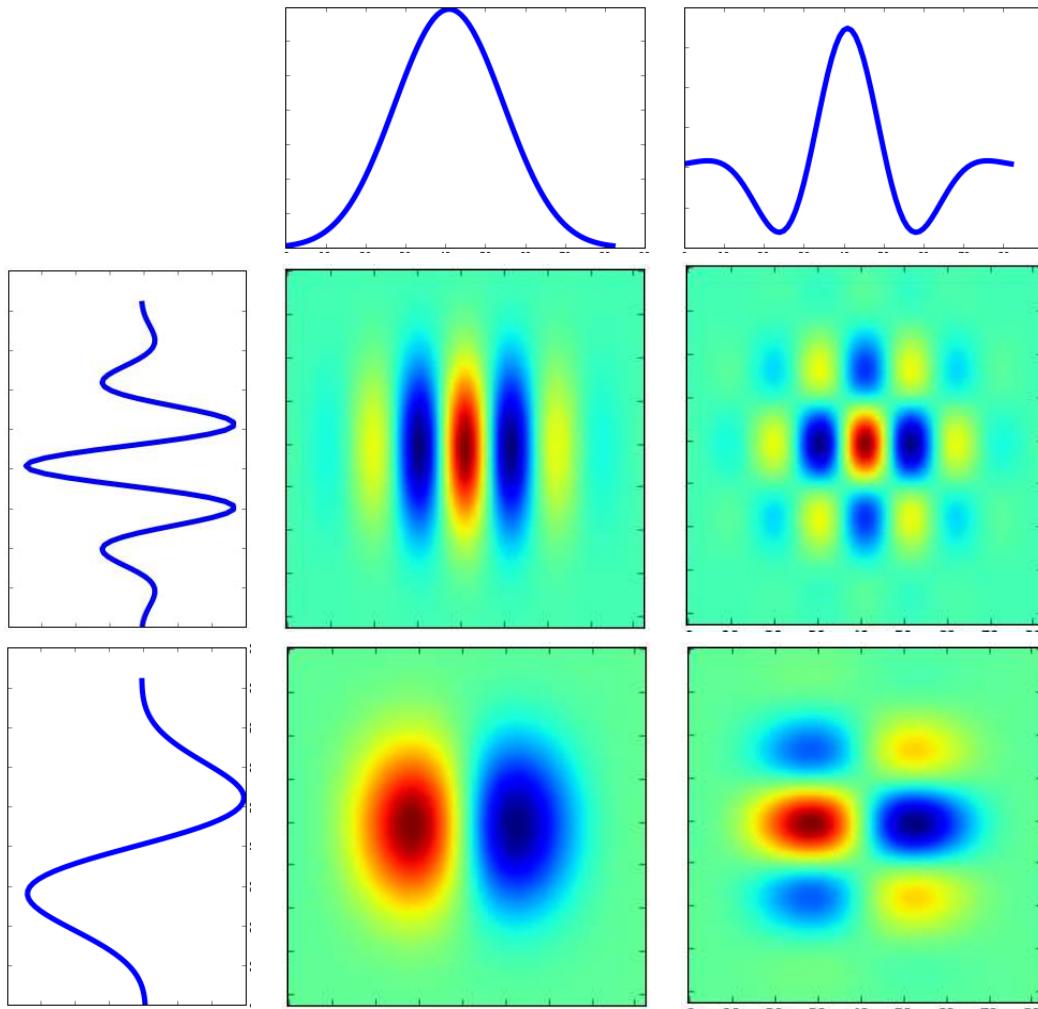


Figure 5.3: Illustrating the existence of common separable filters between distinct 2D filters.

5.2 Preamble to Solution

An existing architecture for super-resolution can be extended to denoising application with modification to the receptive fields (i.e., filter sizes) of CNN. The number of weights (or parameters) can be further reduced by exploring filters with interdependency because conventional CNN employs convolution operation with 2D or 3D filters. The adaptation of separable filter technique for the current application can further reduce the parameters required. Learnable separable filters is an alternate approach with the hypothesis that common separable filters can exist and can be learned. Figure 5.3 illustrates common 1D filters for various 2D filters. For learning separable filters, imposing geometrical constraints are sufficient in case of deep learning (Rigamonti et al. 2013), but such an approach considers filter with a rank of two (to resolve instability during training) and the approach is sequential or parallel in nature. Thus, the approach is aimed at single rank filters without any trade-off in accuracy. One common draw back of such a constraint during training is the impulsive error-rate, which affects the gradients for correction. This can be resolved through the selection of an appropriate optimizer to track gradients from previous iterations during error back-propagation, rather than adopting a simple stochastic gradient descent method.

5.3 Materials and Methods

5.3.1 Experimental setup

Duke’s SD-OCT restoration dataset is used (Fang et al. 2013). This dataset comprises 10 training samples and 17 testing samples where each sample has a noisy image and noise free image. The experimentation setup proposed in Fang et al. (2013) is considered involving reducing A-scans in noisy image by 50% and sought to estimate the corresponding noise free image. The computational complexity of the experimentation requires 8 GB of RAM and a 2 GB GPU with dependency over Python and Torch.

5.3. Materials and Methods

5.3.2 Alterations considered while adapting CNN architectures

The sparse coding based CNN (SCCNN) (Dong et al. 2014) architecture is a supervised approach. A set of interpolated low-resolution image patches and their corresponding true high-resolution image patches are thus required. A CNN can be deployed over an image to generate a super-resolved image as output. For the current application, the entire image is considered as the input, such that the output is also an entire image. This avoids predicted patches based image mosaicing and speeds up the prediction process. For the current approach, a set of noisy sparsely sampled images I_S and their corresponding true HD images I_H are required. Initially I_S is interpolated (I_{LH}) to match the spatial resolution of I_H . A basic convolutional block of the CNN involves the convolution of the input with filters whose dimensions $h \times w \times a$ (height, width and depth) are user-defined and a non-linear transformation (i.e., a rectified linear unit) of each filter response (\mathbf{r}). The unification of both processes is represented as follows:

$$\text{output} = \max(0, \text{input} * \mathbf{r}_{h \times w \times a}) \quad (5.1)$$

It is to be noted that pooling in conventional CNN architectures is excluded for super-resolution images. A CNN is a cascade of convolution blocks where the output of each block is treated as the input for the subsequent block. Indeed, Dong et al. (2014) did not employ an upsampling layer and compensated for this by interpolating the input patches to match the size of output patches. Thus, bicubic interpolation was employed as an initial estimate of the missing A-scans in a sparsely sampled image. The filters size was increased in order to increase the receptive field of the filters. The architecture's input is an interpolated noisy image and the output image is a restored noise-free image. For the experiment, the input image was a gray scale image and the input filter depth was set to '1'.

A SC equivalent CNN (SCCNN) architecture (see Figure 5.4) involves learning a

cascade of three convolution blocks, where U is equivalent to dictionary for the noisy patch, V is equivalent to a mutual matrix and W is equivalent to a dictionary for noise free patches. Thus, U transforms each pixel image in I_{LH} to LH sparse coefficients α_{LH} (equivalent to α_S), V maps α_{LH} to sparse coefficients of the noise free image α_H and W estimates each pixel in the noise free image I_H . Extending the equation 5.1 to the above blocks results in the following equations 5.2, 5.3 and 5.4, where h, w and d are user-defined and can change from block to block:

$$O_U = \max(0, I_{LH} * U^{h \times w \times a}) \quad (5.2)$$

$$O_V = \max(0, O_U * V^{h \times w \times a}) \quad (5.3)$$

$$\hat{I}_H = \max(0, O_V * W^{h \times w \times a}) \quad (5.4)$$

The number of filters and filter sizes are user defined but to match the mutual matrix functionality V , the filter size can only be 1×1 and number of W filters must be one, because the predicted image has only one channel. A forward pass of an interpolated noisy image through the SCCNN is expected to predict a noise free image. The mean square error between the estimated image and true noise free image is treated as an error. In pursuit of minimizing the error, the U, V and W filters are corrected through error back-propagation as shown in Figure 5.4. The core distinction between Dong et al. (2014) and the proposed approach is in terms of application and inclusion of separability.

5.3.3 Factorizing filters

The separable filter approach utilizes the dependency in filter representations and aims to decompose a 2D filter into two unique vectors (Uy, Ux). The response attained through

5.3. Materials and Methods

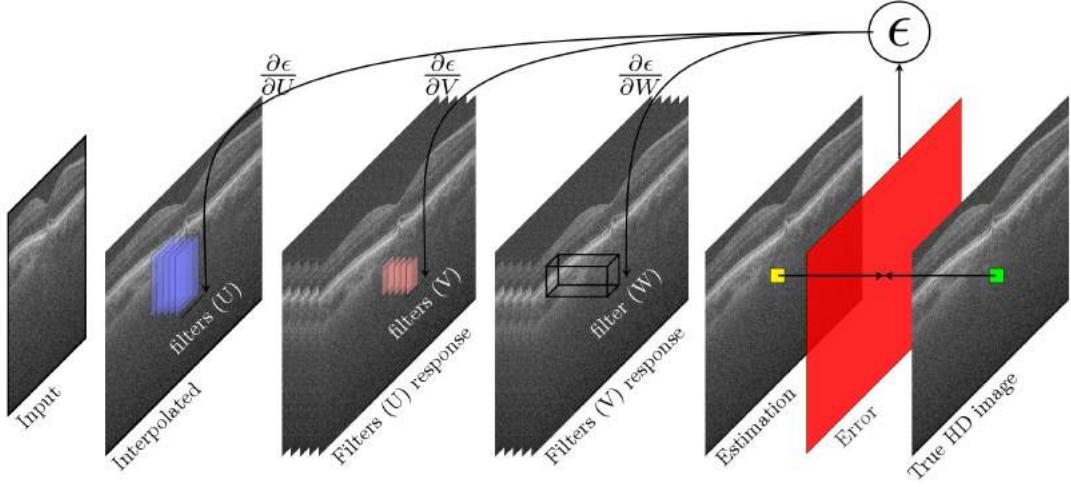


Figure 5.4: Adapting SCCNN for super-resolution to restore high definition noise free image.

sequence of identified 1D convolutions is same as convolving with 2D filter.

$$O_{Uy} = I_{LH} * Uy^{h \times 1 \times a} \quad (5.5)$$

$$O_U = \max(0, O_{Uy} * Ux^{1 \times w \times a}) \quad (5.6)$$

Rather than decomposing an identified 2D filter, a different branch learns separable filters with the hypothesis that separable filters can be shared (see Figure 5.3) between multiple 2D core filters (Rigamonti et al. 2013). These architectures learn separable filters with geometrical constraints, showing that core filters can be reconstructed with weighted combination of separable filters. In the process of reducing parameters the sequential steps have been increased. Thus, alternate approaches have been explored to compute both 1D filters in a parallel manner (Xu et al. 2015) and to combine them at the end of the CNN before their reconstruction in deeper architectures(similar to equation 5.13). Thus, the latter approach reduces sequential processing. Such approaches employ inference blocks to reconstruct image information from gradients.

5. Low-Rank CNN Filters for Simultaneous Reconstruction and Restoration of Sparsely Sampled Retinal OCT Image

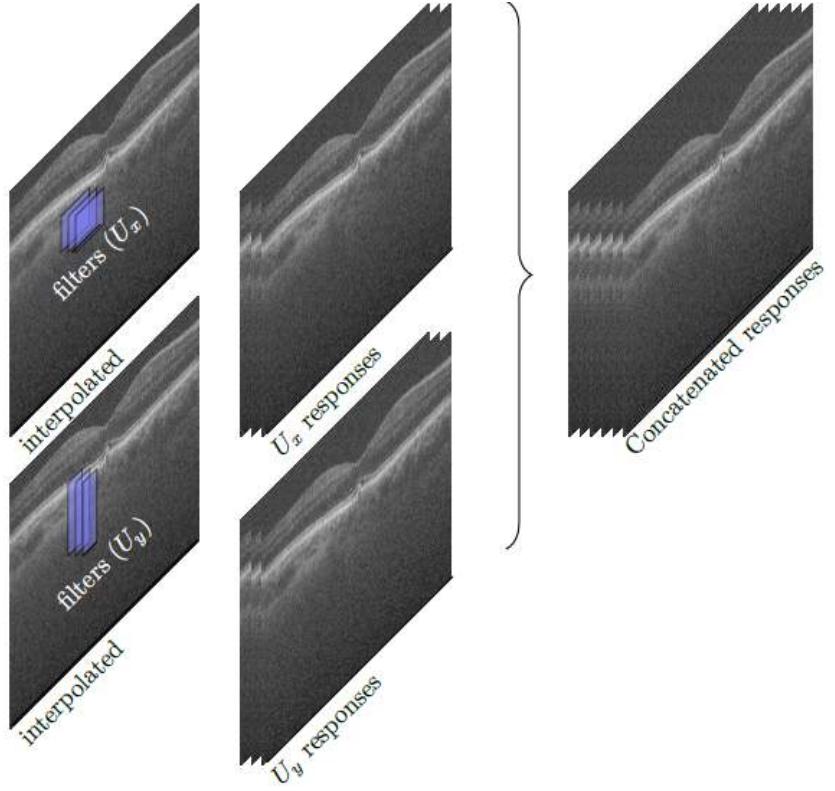


Figure 5.5: Proposed approach fusing edge aware filters and learning filters through geometrical constraints.

$$O_{Uy} = \max(0, I_{LH} * Uy^{h \times 1 \times a}) \quad (5.7)$$

$$O_{Vy} = \max(0, O_{Uy} * Vy^{h \times 1 \times a}) \quad (5.8)$$

$$O_{Wy} = \max(0, O_{Vy} * Wy^{h \times 1 \times a}) \quad (5.9)$$

$$O_{Ux} = \max(0, I_{LH} * Ux^{1 \times w \times a}) \quad (5.10)$$

$$O_{Vx} = \max(0, O_{Ux} * Vx^{1 \times w \times a}) \quad (5.11)$$

$$O_{Wx} = \max(0, O_{Vx} * Wx^{1 \times w \times a}) \quad (5.12)$$

$$\hat{I}_H = \text{inference}(O_{Wx}, O_{Wy}) \quad (5.13)$$

The proposed approach fuses both approaches by only factorizing the U filter of SC-CNN (i.e., SCCNN with 5.7 and 5.10). The U filters (see Figure 5.4) are replaced with

5.4. Results and Discussion

low-rank filters (see Figure 5.5) through a geometrical constraint. Effectively equation 5.2 is replaced with ensemble of 5.14, 5.15 and 5.16. The V filters, W filters and corresponding responses are unaltered (i.e., 5.3 and 5.4). The V filters 5.3 of the CNN can be accountable for the functionality of both the weighted combination of responses and the mutual matrix.

Employing such factorization through geometrical constraint renders the entire system impulsive and standard stochastic gradient descent approaches are prone to local minima. A Conjugate gradient approach is thus employed to handle the impulsive nature of the gradients while learning separable filters.

$$O_{Uy} = \max(0, I_{LH} * Uy^{h \times 1 \times a}) \quad (5.14)$$

$$O_{Ux} = \max(0, I_{LH} * Ux^{1 \times w \times a}) \quad (5.15)$$

$$output_{block1}^i = (O_{Uy}, O_{Ux}) \quad (5.16)$$

where h , w , a and number of filters are user-defined.

5.4 Results and Discussion

The current approach has two objectives i.e., adaptation of SCCNN architecture designed for super resolution purposes to retinal OCT restoration and reduction of required model parameters. The model trained with SCCNN architecture is expected to perform superior to sparse coding. The model trained with modified SCCNN architecture (with low-rank filters through geometrical constraint) is expected to perform comparatively to sparse coding and requires less number of parameters.

5.4.1 Training parameters

For this experiment, U was 120 filters that were 15×15 in size, V was 80 filters that were 1×1 in size and W was 1 filter that was 15×15 in size. Large filter sizes result in large receptive fields (large neighborhood) which is need for overcoming stochastic noise in the OCT image. Each noisy image was interpolated with bicubic kernel to the size of noise-free image and treated as the input to the SCCNN (see Figure 5.6a). Each image was considered as a minibatch and gradients were corrected after each epoch. In contrast to conventional practices where a single image is discriminated to thousands of patches, recent practices employ the complete image, owing to the convolution property of a CNN. Training was performed for 1200 epochs and a conjugate gradient was employed to correct the weights. The training phase needs a set of sparsely sampled OCT images and noise free image (see Figure 5.6a).

In the case of modified architecture (modified SCCNN) Ux was 60 filters that were 15×1 in size and Uy was 60 filters that were 1×15 in size. The responses were concatenated at the input of V . Thus 120 responses were received as inputs, which is the same as the SC based CNN architecture (i.e., SCCNN). The training and testing (see Figure 5.6b) frameworks for both SCCNN and modified SCCNN (with geometrical constraints) is illustrated in Figure 5.6.

5.4.2 Reconstructing a given test image

Sample test images reconstructed through the bicubic interpolation, sparse coding and modified SCCNN (with interpolated image as input) are shown in Figure 5.7. The approaches have successfully reduced the noise and restored image. Additionally, the approaches are capable of restoring structures (measured with SSIM). The red and green rectangles in Figure 5.7 illustrated the removal of noise and retainment of structures. The orange rectangle illustrates the clear distinction between hyper-reflective bands. The orange rectangle depicts the existence of deformation. The speckle noise above ILM

5.4. Results and Discussion

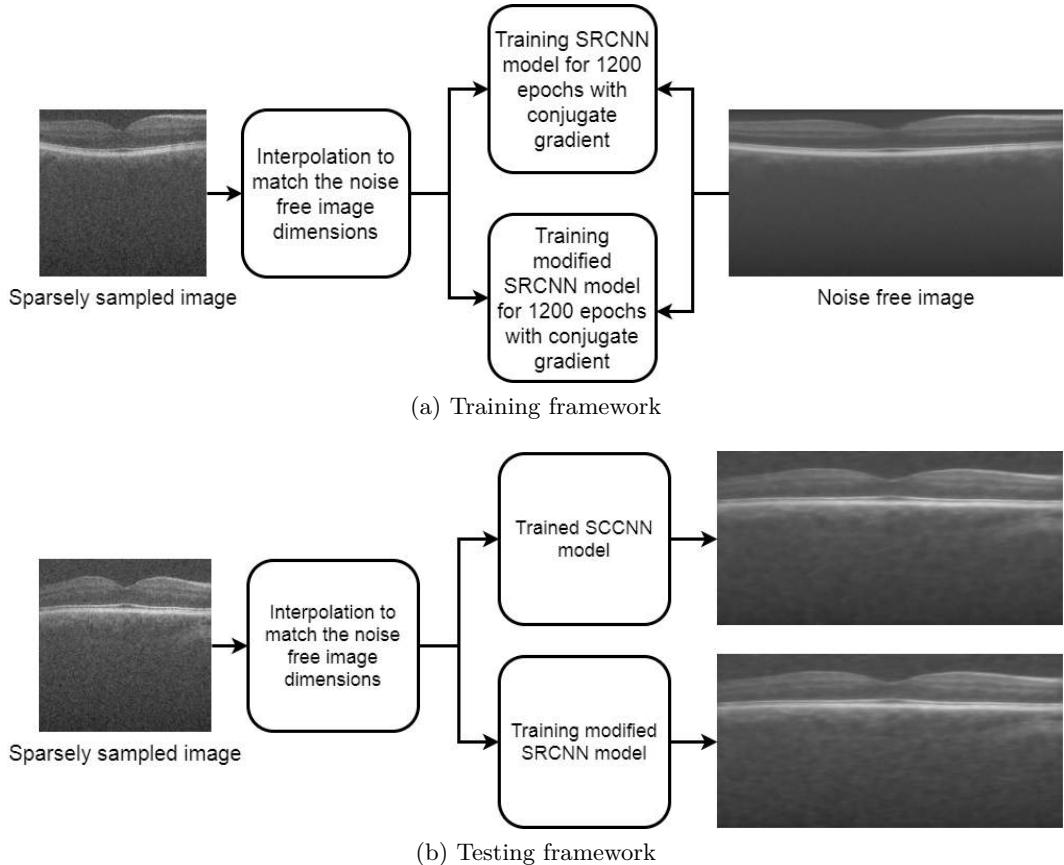
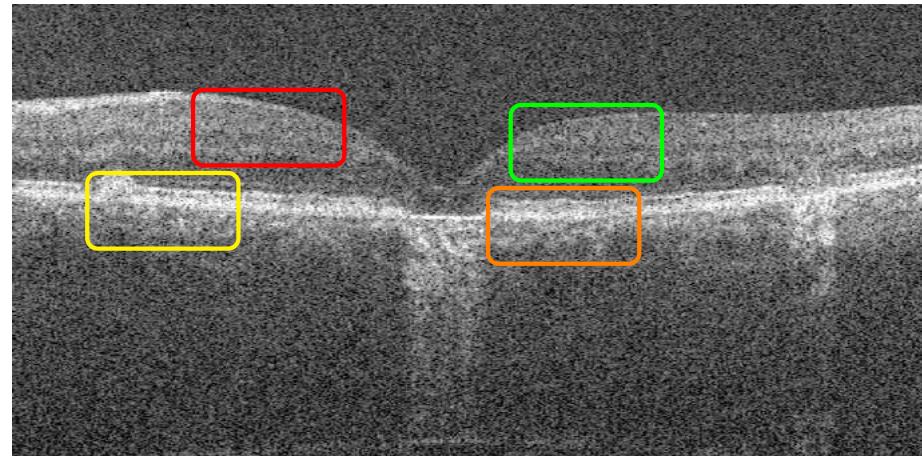


Figure 5.6: Training and testing process of conventional (adapted) SCCNN and modified SCCNN for restoring sparsely sampled retinal OCT images.

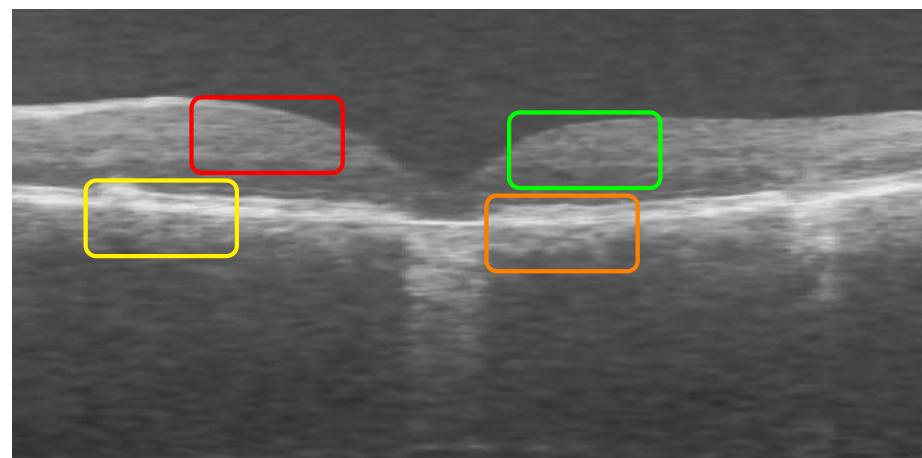
and below BM are completely smoothed illustrating that the algorithms are capable of distinguishing speckles reasonable for information and noise.

The absolute difference values between anticipated image (I_H) and reconstructed images through bicubic interpolation (for reference), SCCNN, modified SCCNN and sparse coding are illustrated in Figure 5.8. The mean absolute difference error for bicubic interpolation, SCCNN, modified SCCNN and sparse coding are 0.1034, 0.037, 0.0374 and 0.0417 respectively.

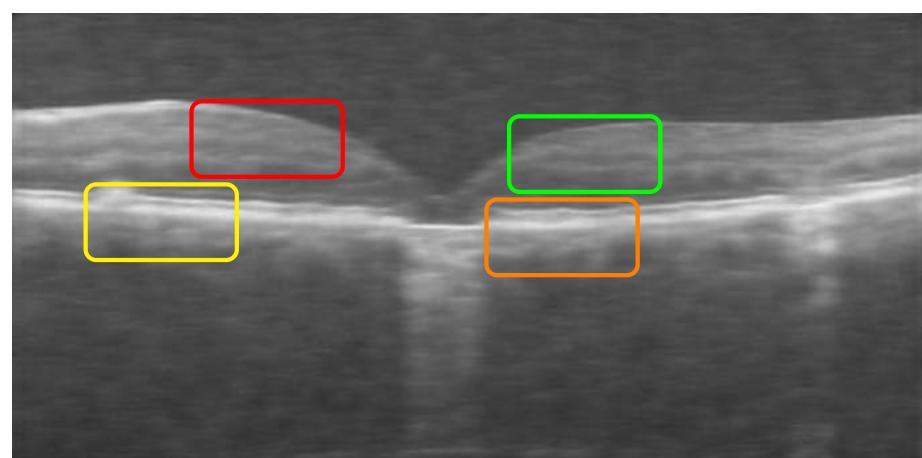
5. Low-Rank CNN Filters for Simultaneous Reconstruction and Restoration of Sparsely Sampled Retinal OCT Image



(a) Bicubic interpolation



(b) Sparse coding based reconstruction (SSIM:0.64)



(c) Modified SCCNN based reconstruction (SSIM:0.68)

Figure 5.7: Illustrates the capability of reconstructed image through proposed and base line (Sparse coding) from a given interpolated image.

5.4. Results and Discussion

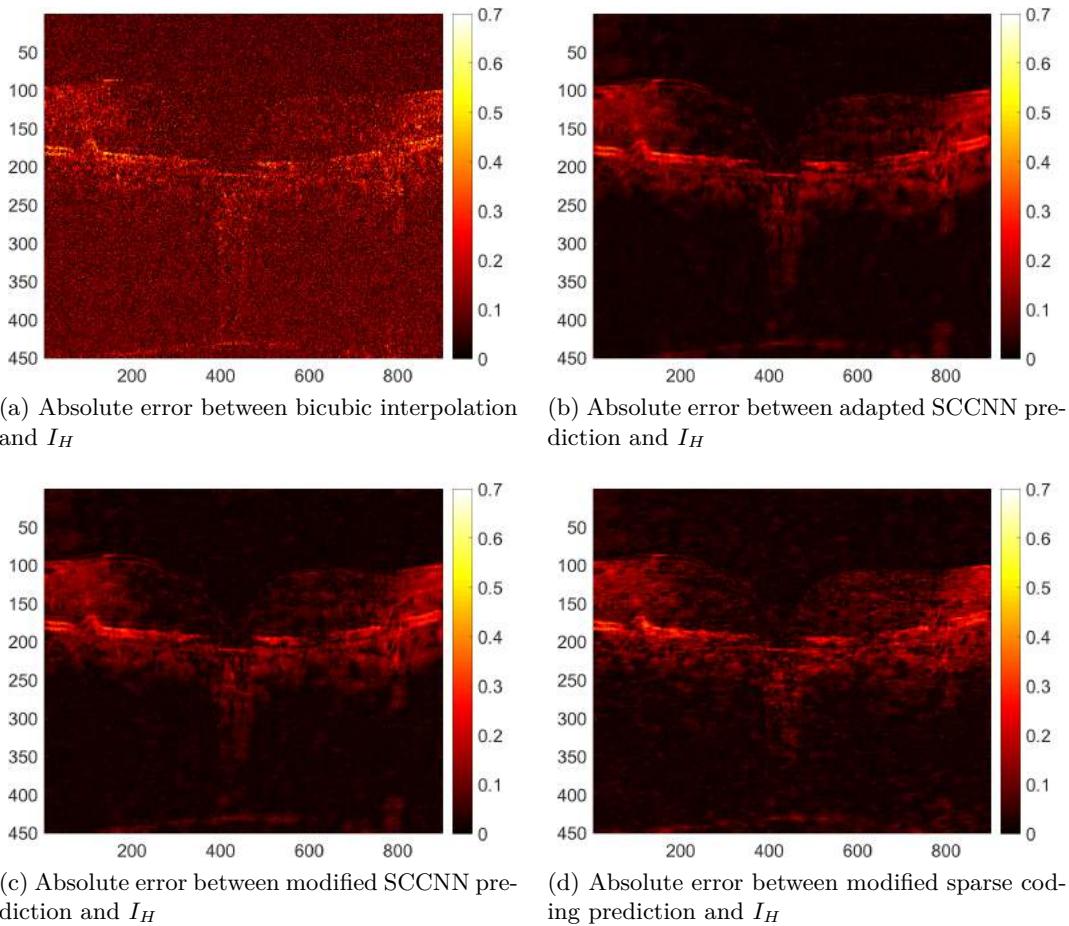


Figure 5.8: Illustrates the absolute error between expected noise free image and estimated noise free images based on various approaches.

5.4.3 Training and testing characteristics

The mean PSNRs of the estimated HD images for both the training and test sets at each epoch from the SCCNN and the modified architectures are shown in Figure 5.9. It can be observed that SCCNN and modified SCCNN (with geometrical constraints) follow similar training and prediction trend with slight trade-off in PSNR. It is to be noted that the models during initial phases of training exhibit impulsive nature and the same can be observed from training performance. The model has achieved maximum possible PSNR by 400 epochs but continued to illustrate the converge of the model.

5.4.4 Benchmark

A test set with 17 samples was used to validate the adapted and modified architecture. The SC ((Fang et al. 2013) without neighboring information) was used as baseline. Each architecture was trained following the protocol for single-image-based restoration. Standard image quality metrics were used for the evaluation. The PSNR defines the ratio between peak signal strength and noise. This metric is a log scaled version of MSE. The Max. error and L2 norm are variants of MSE to measure the difference of mean difference in brightness between images. On a side note SSIM measures the similarity in structures between images.

The Fang et al. (2013) depicts that sparse coding had superior performance in comparison to BM3D and Tikonov regularizer (classical image processing) based reconstruction on the dataset so image processing based baselines are not considered. SRF based super resolution architectures (Schulter et al. 2015) are recently proposed. So the present study extended its applicability for OCT restoration and considered as relevant baseline. The Table 5.1 comprises performance evaluation of proposed approaches (adapted SC-CNN and modified SCCNN with low-rank filters) in comparison to existing baselines for OCT. The table also includes the required parameters for each method. The adapted with large receptive fields resulted in the best qualitative performance and quantita-

5.4. Results and Discussion

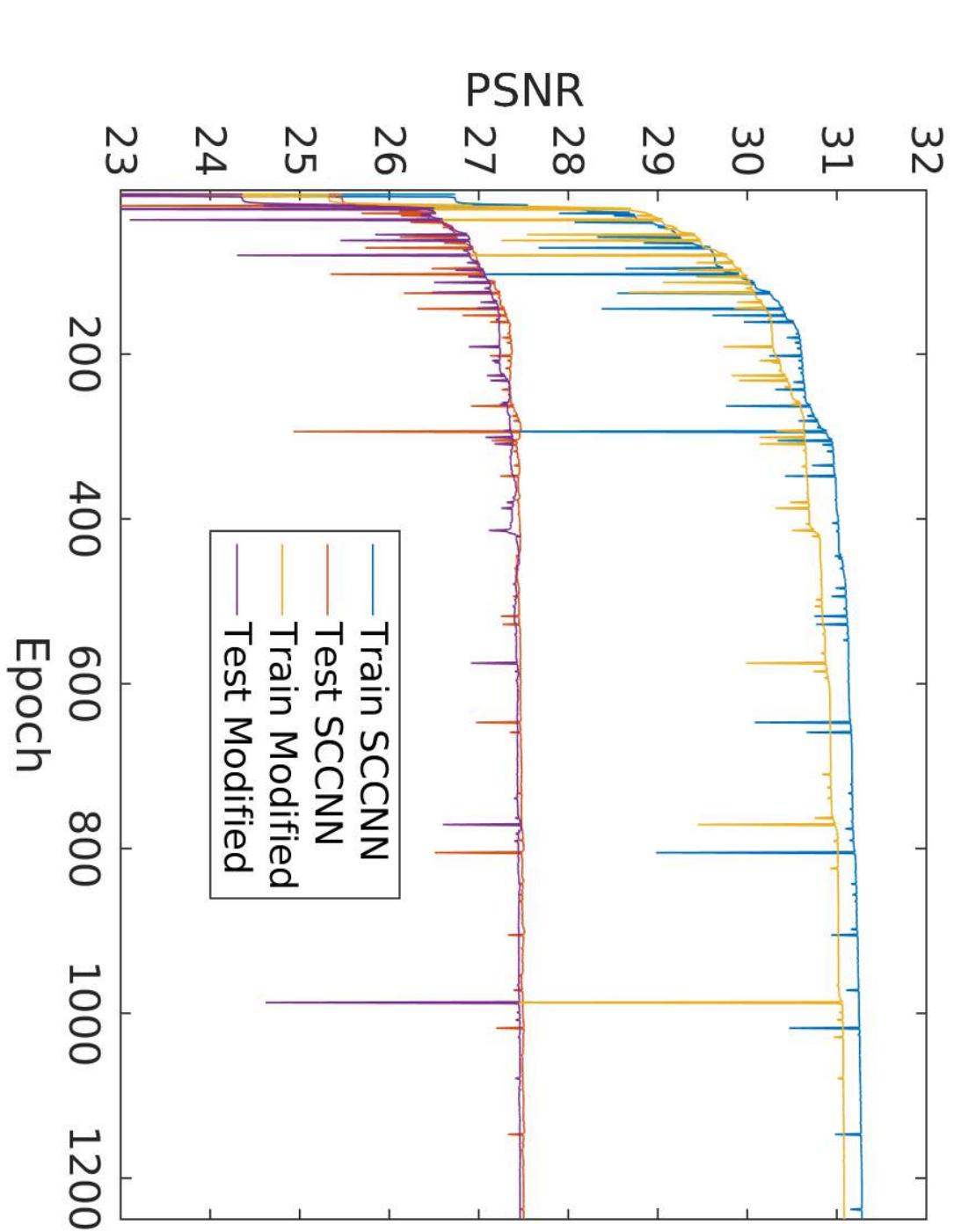


Figure 5.9: Learning characteristics of conventional and modified architecture on training and testing set at each epoch.

5. Low-Rank CNN Filters for Simultaneous Reconstruction and Restoration of Sparsely Sampled Retinal OCT Image

Table 5.1: Comparison of adapted SCCNN, modified SCCNN, sparse coding and structured random forests across various metrics

Approach (parameters)	PSNR	MSE	Max. Error	L2 norm	SSIM
Adapted SCCNN (54K)	28.24	115.58	92.52	0.99	0.69
Modified SCCNN (29K)	28.15	116.68	90.03	0.99	0.68
SC (>100K)	27.64	128.42	95.32	1.02	0.66
SRF (10 trees and 15 depth)	26.17	168.71	198.70	0.98	0.61

tive performance (Table 5.1) in comparison to other approaches. Low-rank filter based SCCNN (modified SCCNN) achieved a comparable performance, with 45% fewer parameters. This depicts thats the core filters for retinal OCT reconstruction can be factorized with limited loss of informations.

5.5 Summary

Computational algorithms have been helpful at handling the constraints from instrumentation or imaging protocols and for estimating true information for effective screening. Sparse coding based approaches have been effective at restoring sparsely sampled images, with an existing CNN architecture whose functionality is identical to sparse coding. However, these have been designed for super-resolution applications. The employed CNN architecture and training procedure has been altered from the size of the receptive field to the optimizer for simultaneous image restoration and reconstruction. Such trained CNN is capable of estimating noise free images with reasonable image quality. The parameters needed are considerably reduced with the use of a sparse-coding pipeline. Additionally prediction of entire image through employment of GPU has reduced the testing time. Sparse coding offers superior performance in comparison to other baselines in terms of various metrics, ranging from signal strength to structure. Inspired by separable filters and edge-aware filtering, a geometrical constraint was imposed over the first layer’s filters. This lead to the construction of low-rank filter factorization, which effectively reduced the number of parameters required. This imposition had a

5.5. Summary

negative effect on system stability, however, consideration of gradients from previous epochs equipped the optimizer to handle such impulsiveness. This demonstrates that a redefined architecture has the capability of predicting noise free and restored images without any substantial trading off in accuracy.

Chapter 6

Summary and Conclusion

“Once you eliminate the impossible, whatever remains, no matter how improbable, must be the truth.”

—Arthur Conan Doyle

The work proposed in the thesis has aimed at modeling representations in various spaces in order to improve medical image analysis. The performance of each block of the analysis was evaluated for retinal applications in comparison to various baselines. The block specific approach is based on the maxim behind Bellman’s equation: that optimal systems can be achieved by starting at an optimal point and maintaining optimal decisions at each step. Thus, in the future, even if the core algorithm in a block changes, the proposed algorithms for other blocks will retain as optimal decision. The state-of-the-art algorithms employed in each block of medical image analysis were based on ML, but they fail in pathological cases, owing to their incapacity to learn multiple patterns.

Learning representations are essential for effective modeling and this is commonly used by various blocks of image analysis for improved predictions. Representation space depends on the application or block (for segmentation or classification) because a segmentation block is more concerned with representation in a target space and a classifier is more concerned with representative control points in feature space. Thus, multiple algorithms need to be identified and adapted to retinal image analysis, each with the

6.1. Contribution of this Work

capability of modeling representations in different spaces. Because any given proposal might not be effective for all applications, each proposed algorithm must be validated against existing baselines. Accordingly, the approaches proposed here were compared with various baselines for a comprehensive qualitative evaluation.

6.1 Contribution of this Work

The proposed algorithms are capable of segmenting layers even in pathological cases with large deformations. They can also identify pathological subjects with weakly supervised datasets or subject level labels. Data-driven feature quantifiers were constructed to reduce the level of expert intervention and interpreted the feature responses. Sparsely sampled images generated to avoid motion artifacts and to save storage memory which can be restored with fewer parameters through low-rank filters. These algorithms can affect clinical practice insofar as they can provide accurate profiles of the retinal layers width and automated screening of retinal pathologies. Consequently, ophthalmologists can focus on pathological subjects, comparing the responses depicted by a feature quantifier with standard clinical features. They can also acquire images more rapidly and store them in less space. The latter offers benefits to telemedicine. Along with augmenting the performance, the proposed approaches reduce both space and time complexity.

In chapter 2 a training framework was developed to allow standard edge-detection algorithms, such as SRE, to model layer-specific edges in retinal OCT images. Given an image patch, each model is capable of predicting the corresponding edge representations in the target space. This initially validates the hypothesis that edges from different objects can be modeled independently. This improves segmentation performance when delineating layers, which is pivotal for retinal layer profiles and thickness maps. The proposed segmentation algorithms are robust to layer deformations (due to pathology), noise (due to imaging principles) and artifacts (due to imaging protocols). The proposed framework can be employed with any other off-the-shelf edge-prediction algorithms, re-

6. Summary and Conclusion

flecting its flexibility. The model for RPE lower boundary can be employed for retinal flattening, which is a standard step in retinal OCT classification. These flattened images can be quantified using features for the construction of feature space.

In chapter 3 a classifier inspired by Voronoi fractals was developed to identify representative control points in feature space. This approach permits the model to autonomously identify feature-space structures without kernel knowledge. This proposal was validated using a toy example with various scenarios, whereby the same algorithm constructed the appropriate boundaries according to the feature space. The modeling capability of the classifier was verified by modeling a two-spiral example. Accurate identification of retinal pathologies was also demonstrated by the proposed classifier.

In chapter 4 data-driven feature quantifiers exhibited comparative or improved performance over traditional feature descriptors (HOG) for retinal pathologies identification. Existing data-driven feature quantifiers have been adapted for retinal applications through transfer learning. An appropriate CNN architecture with fewer parameters was considered, as it is less prone to overfitting. Based on a neural network algorithm, its consistency was validated in multiple experiments. Such practices reduce the dependency on heuristics (by experts) during the employment of traditional feature descriptors. In addition, such architectures can visualize the step-by-step transformation of the input space's raw information to high-level abstract representations. The abstract patterns allow for even naive classifiers to identify pathological subjects effectively. An automated approach was proposed to identify the potential response at each CNN block, for easy visual interpretation of information processing and abstract patterns.

Finally, in chapter 5 low-rank filters based image restoration was proposed to restore and filter sparsely sampled images simultaneously. Such blocks facilitate faster imaging, reduced image storage and transmit images more quickly for telemedicine applications. As sparse representation based algorithms have been successful for such applications, a CNN architecture (for super-resolution) that is functionally equivalent to sparse coding

6.2. Re-usability of Work

was adapted for retinal OCT image restorations. Subtle changes such as increasing the receptive field of the CNN filters were needed to reduce speckle noise has resulted in improved performance in restoring images. The core change of imposing a geometrical constraint on the filters reduced the number of parameters needed, with a negligible tradeoff in performance on the test set. During training, a complete image was considered, rather than patches (as in conventional practice) and this eliminated the need for image mosaicing of predicted patches during testing which is time-consuming.

Insofar as ML formed the model's back-end, the same algorithms can be employed for various datasets acquired from different OCT image, or with different imaging protocols, without much effort. Upon on arrival of a new dataset, the modeled representations can be further modified through transfer learning. Additionally, representations have the capability of modeling multiple patterns with limited training data. Because the proposed algorithms are supervised in nature, they cannot autonomously attribute a new pattern to a class, as can be done using an unsupervised algorithm. However, this direction can be explored through semi-supervised approaches.

6.2 Re-usability of Work

All the codes written for the thesis has been released on Github.

- Repository https://github.com/ultrai/Chap_1 holds the code for retinal layer segmentation
- Repository https://github.com/ultrai/Chap_4 holds the code for retinal pathologies classification
- Repository https://github.com/ultrai/Chap_3 holds the code for constructing data driven feature for classifying retinal pathologies
- Repository <https://github.com/ultrai/SRCNN> holds the code for simultaneous

restoration and reconstruction of retinal OCT

6.3 Future Scope of Work

A dataset common to all blocks must be acquired on a single machine to execute all of the proposed algorithms in a pipeline. The data-driven feature quantifiers must have a classifier embedded into the architecture. However, in its current state, end-to-end training between the feature quantifier and the proposed classifier is infeasible, owing to architectural incompatibility. An extension of the image- level algorithms to a volumetric analysis that considers neighboring images is trivial, but it is necessary to improve performance.

Individual chapter- specific future directions can also be stated. First, the HOG features for the SRE can be replaced with data-driven feature quantifiers. Gradient boosted tree techniques can be extended to the SRE such that the models can adapt the criteria at each node upon the arrival of new data. Representation learning-based semantic segmentation can be an alternative to the generation of delineated layers without the need for edge prediction and a dynamic program framework.

The proposed classifier considers a hard margin, but the algorithm is capable of generating soft margins as well. The latter are crucial for determining the confidence level of a decision. The unsupervised algorithm for exploring control points can be replaced with more effective algorithms that can determine the number of clusters autonomously (e.g., spectral clustering). A neural network architecture can be designed (see Appendix C) that is functionally equivalent to the proposed classifier for end-to-end training of the data-driven feature quantifier.

Current practices widely employ CNN architectures for the construction of a data- driven feature quantifier. Interest in this has resulted in multiple CNN architectures that can be explored in terms of their adaptability to retinal pathology identification. As the architecture changes, the optimal parameters combination must be found (e.g.,

6.3. Future Scope of Work

class weight, loss weight, etc.). The visualization can also be improved for effective interpretations of the representations.

For the restoration block, the MSE loss should be considered, which aims exclusively at a mean signal match. As OCT images have structures, combinations of SSIM and MSE can improve both the mean signal and structure. The SC based CNN is only two layers deep and it has been empirically shown that deeper architectures can improve performance. Thus, the number of layers in the proposed CNN architecture can be increased to improved the overall performance.

Appendices

A Statistical representation

The speckle distribution in retinal OCT changes from layer to layer and can be modeled through tissue photon interaction. Such interactions are represented through multi-scale statistics and attenuation. A semantic segmentation framework has been constructed for Duke’s AMD dataset by considering multi-scale statistics as features and random forest as classifier (see Figure A.1). The complete details of work flow and experimentation has been given in Karri et al. (2016).

B Multiple layer-specific edge modeling

The layer segmentation chapter employs ‘One vs all’ approach involves training eight models i.e., one for each layer. This is because the SRE cumulates edge from different layers as single edge and the bottleneck is traced to z being binary. On replacing z to be a multivalued vector the SRE is capable of modeling all layer specific-edges simultaneously (see Figure B.1). Such approach only reduces training time complexity. .

C Transcendence of means-set based classifier into neural network

Proposed classifier holds two computational phases i.e., computing the Euclidean distance with mean vectors and identifying the sample with *min* distance. This can be interpreted as neural network with one hidden layers and output layer with neurons same as number of classes. The means-set can be considered as weights for first hidden layer, as Euclidean distance is inversely proportional to a dot product. Considering selective dropout and max operation, the resultant gives the probability of a sample falling into each class. The weights can be modified on arrival of new data through gradient descent. Replacing the softmax classifier in conventional deep learning, architectures

Appendices

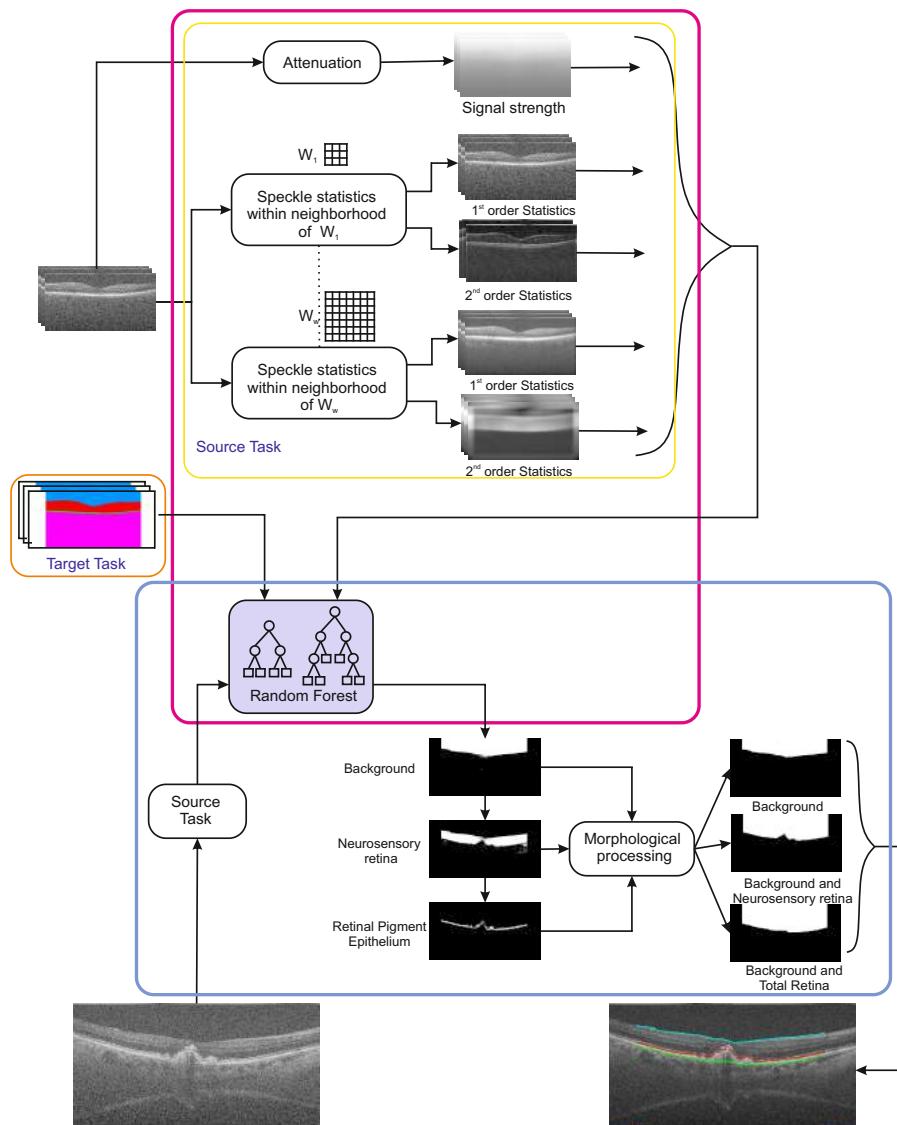


Figure A.1: Block diagram for modeling and predicting various retinal layers through statistical representations

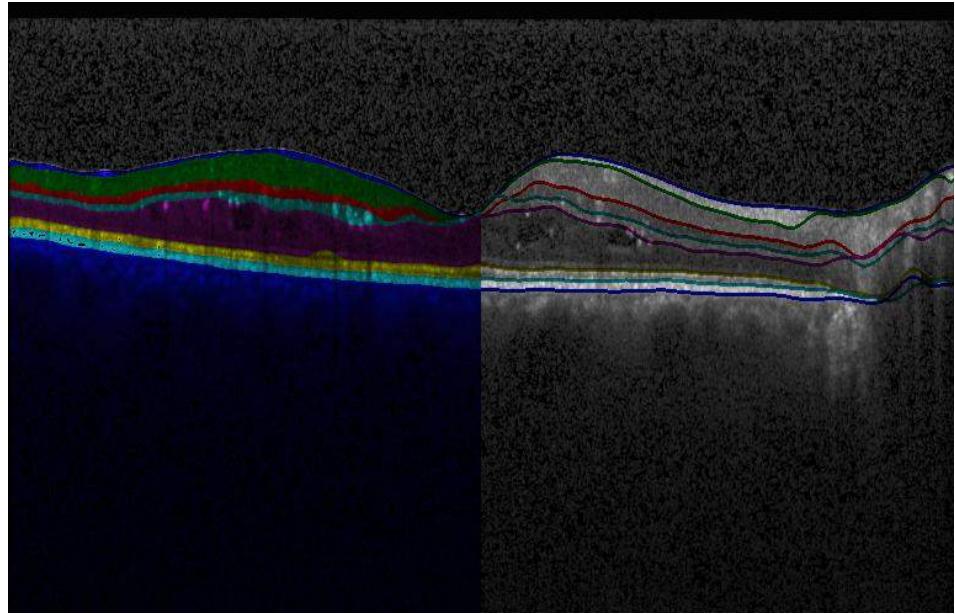


Figure B.1: Capability of segmenting pathological image based on multiple layer-specific edge model.

with proposed classifier can be the immediate future direction.

Dissemination of the Work

Journal

- Karri, S. P. K., Chakraborty, Debjani, and Chatterjee, J. (2017). Guided Fuzzy C-Means: A Supervised Approach.(under preparation).
- Karri, S. P. K., Chakraborty, Debjani, and Chatterjee, J., 2017. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. *Biomedical Optics Express*, 8(2), pp.579-592.
- Karri, S. P. K., Chakraborty, Debjani, and Chatterjee, J., 2016. Learning layer-specific edges for segmenting retinal layers with large deformations. *Biomedical Optics Express*, 7(7), pp.2888-2901.
- Karri, S.P.K., Chakraborty, Debjani, and Chatterjee, J., Rapid training of layer specific edges for segmentation of retinal layers in OCT. *Current Indian Eye Research Journal of Ophthalmic Research Group*, p.73.

Conference

- Karri, S. P. K., Chakraborty, Debjani, Ray, A. K., and Chatterjee, J. (2017, October). Learning Representations through Ensemble of Fuzzy C-Means for Identification of Retinal Pathologies. In ACM Internet of Things and Machine Learning conference (Accepted).
- Karri, S. P. K., Garai, N., Nawn, D., Ghosh, S., Chakraborty, Debjani, and Chatterjee, J., 2016, September. Retinal layer delineation through learning of tissue

Dissemination of the Work

photon interaction in optical coherence tomography. In Technology Symposium (TechSym), 2016 IEEE Students (pp. 46-51). IEEE.

- Karri, S. P. K., Garai, N., Nawn, D., Ghosh, S., Chakraborty, Debjani, and Chatterjee, J., 2016, September. Simultaneous reconstruction and restoration of sparsely sampled optical coherence tomography image through learning separable filters for deep architectures. In Technology Symposium (TechSym), 2016 IEEE Students' (pp. 52-55). IEEE.
- Sheet, D., Karri, S. P. K., Katouzian, A., Navab, N., Ray, A.K. and Chatterjee, J., 2015, April. Deep learning of tissue specific speckle representations in optical coherence tomography and deeper exploration for in situ histology. In Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on (pp. 777-780). IEEE.

Resume

Sri Phani Krishna Karri

RESEARCH AREAS

- **Machine learning:** Transfer, representation, ensemble and deep learning; multiscale pattern matching; feature selection.
- **Computer vision & Image analysis:** Filtering and restoration; semantic segmentation; tracking; registration.
- **Physics of medical imaging:** Optical coherence tomography; microscopy; Fundus imaging.

EDUCATION

2011 - PRESENT	Doctor of Philosophy (Synopsis given) LEARNING REPRESENTATIONS FOR MEDICAL IMAGE ANALYSIS <i>IIT Kharagpur, India</i>
2009 - 2011	Master of Technology MEDICAL IMAGING AND IMAGE ANALYSIS <i>IIT Kharagpur, India</i>
2005 - 2009	Bachelor of Technology ELECTRICAL AND ELECTRONICS <i>Acharya Nagarjuna University, India</i>

RECENT PUBLICATIONS

- Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration, *Biomed. Opt. Exp.* 8 (2), 579-592, 2017
- Learning layer-specific edges for segmenting retinal layers with large deformations, *Biomed. Opt. Exp. (IF:3.3)* 7 (7), 2888-2901, 2016
- Simultaneous reconstruction and restoration of sparsely sampled optical coherence tomography image through learning separable filters for deep architectures, *IEEE Techsym'16*, 2016

Medical Imaging Lab, SMST, IIT Kharagpur, Kharagpur, WB, 721302, India
+91 94753 65053
pkkarri.mm@iitkgp.ac.in
<http://www.linkedin.com/in/spkkarri>

CODING

GITHUB: <https://github.com/ultrai/>
LANGUAGES AND APIs: Python, Matlab, Lua and Torch

ACADEMIC CONTRIBUTIONS

TA	Digital Image Processing (SMST) '12 - '13 Medical Imaging Lab '12 - '14 Digital Image Processing (EE) '14
REVIEWER	IEEE Int. Symp. Biomed. Img.'14 - '17 IEEE Tech. Sym.'14 Biomed. Opt. Exp. OSA

SCIENTIFIC DISSEMINATION

JOURNALS	Biomed. Optics. Expr. J. Biomed. Optics.
CONFERENCES	IEEE Int. Symp. Biomed. Imaging; IEEE Int. Con. Sys. Med. Biomed.; IEEE Int. Con. Biomed. Health Inf.;
CO-AUTHORED	6 Journals; 9 Conferences; 6 Short abstracts
GOOGLE SCHOLAR	sri phani krishna karri

EXPERIENCE

FEB. 2014 - PRESENT
SkinCurate Research Private Limited
Founder

MAY. 2013 - SEPT. 2013

Resume

Zeiss India

Research intern @ Center for Advanced Research division for 3D OCT registration

JAN 2010 - JULY 2012

Indian Institute of Technology Kharagpur

Core member for image analysis @ DIT Project

AWARDS & ACHIEVEMENTS

- 2014 **Young Scientist Travel Grant**
Dept. of Human Resource Dev., India
- 2013 **GE Edison Challenge**
GE John F. Welch Tech. Centre, India
- 2012 **Topper**
Computer Vision and Advanced Digital Image Processing Course '12
- 2009 **Gate (EE) '09**

GRANT WRITING

- Skin lesion diagnosis with *in situ* imaging and characterization through learning of multispectral signatures for augmentation of clinical insight in effective rural healthcare delivery. *Sponsor:* GE Edison Challenge 2013, *Award:* INR 1,000,000 (\$15,000).
- Multispectral optical imaging for real-time functional characterization and monitoring of wound healing. *Sponsor:* BIG BIRAC 2015, *Award:* \$66,000.

ONGOING PROJECTS

- A guided framework for unsupervised algorithms to incorporate supervised data.
- Medical data compression.
- Tracking of cells with event detection.

COMPLETED PROJECTS

- Semantic segmentation of retinal layers and fluid through learning representations. (Collaborated and published in Biomed. Opt. Exp. 2017)

- Registration of 3D OCT volumes. (Work at Zeiss)
- Color deconvolution of stain for cervical screening. (Mentored and published in IEEE INDICON 2016)
- Localization and segmentation of PAP smeared nuclei for screening cervical cancer. (IEEE ICBHI 2014)
- Tracking of HaCat cells under phase contrast microscopy. (Poster Presentation ISBI 2013)

PERSONAL DETAILS

- Born on 2 May 1987 (30 years)
- Father Occupation: Farmer
- Married
- One sibling
- Permanent residence is: D.no 16-42-11/1, NGO colony, Tanuku (534211), AP.

REFERENCES (HYPERLINKED)

- [Dr. Jyotirmoy Chatterjee \(IIT Kharagpur\)](#)
- [Dr. Debjani Chakraborty \(IIT Kharagpur\)](#)
- [Dr. Pabitra Mitra \(IIT Kharagpur\)](#)

Bibliography

- Abramoff, M., Garvin, M. & Sonka, M. (2010), ‘Retinal Imaging and Image Analysis’, *IEEE Reviews in Biomedical Engineering* **3**, 169–208.
- Adhi, M. & Duker, J. S. (2013), ‘Optical coherence tomography—current and future applications.’, *Current opinion in ophthalmology* **24**(3), 213–21.
- Adler, D. C., Ko, T. H. & Fujimoto, J. G. (2004), ‘Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter’, *Optics letters* **29**(24), 2878–2880.
- Albarak, A., Coenen, F. & Zheng, Y. (2013), Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction, in ‘The 17th Annual Conference in Medical Image Understanding and Analysis (MIUA)’, pp. 59–64.
- Alonso-Caneiro, D., Read, S. A. & Collins, M. J. (2011), ‘Speckle reduction in optical coherence tomography imaging by affine-motion image registration’, *Journal of biomedical optics* **16**(11), 116027–1160275.
- Anantrasirichai, N., Achim, A., Morgan, J. E., Erchova, I. & Nicholson, L. (2013), ‘Svm-Based Texture Classification in Optical Coherence Tomography’, *IEEE 10th International Symposium on Biomedical Imaging* (APRIL), 1320–1323.
- Avanaki, M. R., Marques, M. J., Bradu, A., Hojjatoleslami, A. & Podoleanu, A. G.

Bibliography

- (2014), A new algorithm for speckle reduction of optical coherence tomography images, in ‘SPIE BiOS’, International Society for Optics and Photonics, pp. 893437–893437.
- Bagci, A. M., Shahidi, M., Ansari, R., Blair, M., Blair, N. P. & Zelkha, R. (2008), ‘Thickness Profiles of Retinal Layers by Optical Coherence Tomography Image Segmentation’, *American Journal of Ophthalmology* **146**(5).
- Baroni, M., Fortunato, P. & La Torre, A. (2007), ‘Towards quantitative analysis of retinal features in optical coherence tomography.’, *Medical engineering & physics* **29**(4), 432–41.
- Baylor, D. & Fuortes, M. (1970), ‘Electrical responses of single cones in the retina of the turtle’, *The Journal of physiology* **207**(1), 77.
- Bellet, A., Habrard, A. & Sebban, M. (2013), ‘A survey on metric learning for feature vectors and structured data’, *arXiv preprint arXiv:1306.6709* .
- Bengio, Y., Courville, A. & Vincent, P. (2013), ‘Representation learning: A review and new perspectives’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(8), 1798–1828.
- Bezdek, J. C., Ehrlich, R. & Full, W. (1984), ‘Fcm: The fuzzy c-means clustering algorithm’, *Computers & Geosciences* **10**(2-3), 191–203.
- Bian, L., Suo, J., Chen, F. & Dai, Q. (2015), ‘Multiframe denoising of high-speed optical coherence tomography data using interframe and intraframe priors’, *Journal of biomedical optics* **20**(3), 036006–036006.
- Bird, A., Bressler, N., Bressler, S., Chisholm, I., Coscas, G., Davis, M., De Jong, P., Klaver, C., Klein, B., Klein, R. et al. (1995), ‘An international classification and grading system for age-related maculopathy and age-related macular degeneration’, *Survey of ophthalmology* **39**(5), 367–374.

Bibliography

- Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Vol. 4.
- Bourne, R. R. A., Jonas, J. B., Flaxman, S. R., Keeffe, J., Leasher, J., Naidoo, K., Parodi, M. B., Pesudovs, K., Price, H., White, R. A., Wong, T. Y., Resnikoff, S. & Taylor, H. R. (2014), ‘Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe: 1990-2010.’, *The British journal of ophthalmology* **98**(5), 629–38.
- Bow, S. T. (2002), *Pattern recognition and image preprocessing*, CRC Press.
- Boyer, K. L., Herzog, A. & Roberts, C. (2006), ‘Automatic recovery of the optic nerve-head geometry in optical coherence tomography’, *IEEE Transactions on Medical Imaging* **25**(5), 553–570.
- Boykov, Y. Y. & Jolly, M.-P. (2001), Interactive graph cuts for optimal boundary & region segmentation of objects in ND images, in ‘Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001’, Vol. 1, pp. 105—112.
- Brar, M., Yuson, R., Kozak, I., Mojana, F., Cheng, L., Bartsch, D.-U., Oster, S. F. & Freeman, W. R. (2010), ‘Correlation between morphologic features on spectral-domain optical coherence tomography and angiographic leakage patterns in macular edema.’, *Retina (Philadelphia, Pa.)* **30**(3), 383–9.
- Breiman, L. (2001), ‘Random forests’, *Machine Learning* **45**(1), 5–32.
- Bressler, N. M. M. (2004), ‘Age-Related Macular Degeneration Is the Leading Cause of Blindness’, *Jama* **291**(15), 1900–1901.
- Burger, H. C., Schuler, C. J. & Harmeling, S. (2012), Image denoising: Can plain neural networks compete with bm3d?, in ‘Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on’, IEEE, pp. 2392–2399.

Bibliography

- Burges, C. J. (1998), ‘A tutorial on support vector machines for pattern recognition’, *Data mining and knowledge discovery* **2**(2), 121–167.
- Chauhan, D. S. & Marshall, J. (1999), ‘The interpretation of optical coherence tomography images of the retina.’, *Investigative ophthalmology & visual science* **40**(10), 2332–2342.
- Chen, Q., Leng, T., Zheng, L., Kutzscher, L., Ma, J., De Sisterne, L. & Rubin, D. L. (2013), ‘Automated drusen segmentation and quantification in SD-OCT images’, *Medical Image Analysis* **17**(8), 1058–1072.
- Chen, X., Kingma, D. P., Salimans, T., Duan, Y., Dhariwal, P., Schulman, J., Sutskever, I., Abbeel, P., Science, C. & Science, C. (2017), Variational Lossy Autoencoder, in ‘International Conference on Learning Representation’, pp. 1–14.
- Chen, Y.-w. & Lin, C.-j. (2006), Combining SVMs with Various Feature Selection Strategies, in ‘Feature Extraction’, Vol. 324, pp. 315–324.
- Cheng, H., Nair, G., Walker, T. A., Kim, M. K., Pardue, M. T., Thulé, P. M., Olson, D. E. & Duong, T. Q. (2006), ‘Structural and functional mri reveals multiple retinal layers’, *Proceedings of the National Academy of Sciences* **103**(46), 17525–17530.
- Chitchian, S., Mayer, M. A., Boretsky, A. R., van Kuijk, F. J. & Motamedi, M. (2012), ‘Retinal optical coherence tomography image enhancement via shrinkage denoising using double-density dual-tree complex wavelet transform’, *Journal of biomedical optics* **17**(11), 116009–116009.
- Chiu, S. J., Allingham, M. J., Mettu, P. S., Cousins, S. W., Izatt, J. A. & Farsiu, S. (2015), ‘Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema’, *Biomedical optics express* **6**(4), 1172–1194.
- Chiu, S. J., Izatt, J. A., O’Connell, R. V., Winter, K. P., Toth, C. A. & Farsiu, S. (2012), ‘Validated automatic segmentation of AMD pathology including drusen and

Bibliography

- geographic atrophy in SD-OCT images', *Investigative Ophthalmology and Visual Science* **53**(1), 53–61.
- Chiu, S. J., Li, X. T., Nicholas, P., Toth, C. A., Izatt, J. A. & Farsiu, S. (2010), 'Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation.', *Optics express* **18**(18), 19413–28.
- Ciulla, T. A., Amador, A. G. & Zinman, B. (2003), 'Diabetic Retinopathy and Diabetic Macular Edema: Pathophysiology, screening, and novel therapies', *Diabetes Care* **26**(9), 2653–2664.
- Comander, J., Gardiner, M. & Loewenstein, J. (2011), 'High-resolution optical coherence tomography findings in solar maculopathy and the differential diagnosis of outer retinal holes', *American Journal of Ophthalmology* **152**(3), 413–419.
- Cruz, J. A. & Wishart, D. S. (2006), 'Applications of machine learning in cancer prediction and prognosis', *Cancer informatics* **2**, 59.
- Cunha-Vaz, J., De Abreu, J. F. & Campos, A. (1975), 'Early breakdown of the blood-retinal barrier in diabetes.', *British Journal of Ophthalmology* **59**(11), 649–656.
- Dabov, K., Foi, A. & Katkovnik, V. (2007), 'Image denoising by sparse 3D transformation-domain collaborative filtering', *IEEE Transactions on Image Processing* **16**(8), 1–16.
- Dai, Q. & Sun, Y. (2011), Automated layer segmentation of optical coherence tomography images, in 'Proceedings - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011', Vol. 1, pp. 142–146.
- Dalal, N. & Triggs, B. (2005), Histogram of Oriented Gradients for Human Detection, in 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005', Vol. 1, pp. 886–893.

Bibliography

- De Boer, J. F., Cense, B., Park, B. H., Pierce, M. C., Tearney, G. J. & Bouma, B. E. (2003), ‘Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography’, *Optics letters* **28**(21), 2067–2069.
- de Sisternes, L., Jonna, G., Moss, J., Marmor, M. F., Leng, T. & Rubin, D. L. (2017), ‘Automated intraretinal segmentation of sd-oct images in normal and age-related macular degeneration eyes’, *Biomedical Optics Express* **8**(3), 1926–1949.
- Delon, J., Desolneux, A., Lisani, J. L. & Petro, A. B. (2007), ‘A nonparametric approach for histogram segmentation’, *IEEE Transactions on Image Processing* **16**(1), 253–261.
- Deng, G. & Cahill, L. (1993), An adaptive gaussian filter for noise reduction and edge detection, in ‘Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.’, IEEE, pp. 1615–1619.
- Devarakonda, S. T., Vupparaboina, K. K., Richhariya, A., Chhablani, J. & Jana, S. (2016), ‘Automated Detection of Retinal Disorders from OCT Images using Artificial Neural Network’, *2016 IEEE Annual India Conference (INDICON)* .
- Dollar, P. & Zitnick, C. L. (2013), Structured forests for fast edge detection, in ‘Proceedings of the IEEE International Conference on Computer Vision’, pp. 1841–1848.
- Domingos, P. (2007), ‘Structured machine learning: Ten problems for the next ten years’, *Proceedings of 17th International* pp. 1–4.
- Domingos, P. (2012), ‘A few useful things to know about machine learning’, *Communications of the ACM* **55**(10), 78–87.
- Dong, C., Loy, C. C., He, K. & Tang, X. (2014), Learning a deep convolutional network for image super-resolution, in ‘Computer Vision–ECCV 2014’, Springer, pp. 184–199.
- Drexler, W. & Fujimoto, J. G. (2008), ‘State-of-the-art retinal optical coherence tomography’.

Bibliography

- Duchi, J., Hazan, E. & Singer, Y. (2011), ‘Adaptive subgradient methods for online learning and stochastic optimization’, *Journal of Machine Learning Research* **12**(Jul), 2121–2159.
- Dufour, P. A., Ceklic, L., Abdillahi, H., Schroder, S., De Dzanet, S., Wolf-Schnurrbusch, U. & Kowal, J. (2013), ‘Graph-based multi-surface segmentation of oct data using trained hard and soft constraints’, *IEEE transactions on medical imaging* **32**(3), 531–543.
- Ehnes, A., Wenner, Y., Friedburg, C., Preising, M. N., Bowl, W., Sekundo, W., Zu Bexten, E. M., Stieger, K. & Lorenz, B. (2014), ‘Optical Coherence Tomography (OCT) Device Independent Intraretinal Layer Segmentation.’, *Translational vision science & technology* **3**(1), 1.
- Esmaeili, M., Dehnavi, A. M., Rabbani, H. & Hajizadeh, F. (2017), ‘Speckle noise reduction in optical coherence tomography using two-dimensional curvelet-based dictionary learning’, *Journal of medical signals and sensors* **7**(2), 86.
- Esposito, F. & Malerba, D. (2001), ‘Machine learning in computer vision’, *Applied Artificial Intelligence* **15**, 693–705.
- Fang, L., Li, S., Cunefare, D. & Farsiu, S. (2017), ‘Segmentation based sparse reconstruction of optical coherence tomography images’, *IEEE transactions on medical imaging* **36**(2), 407–421.
- Fang, L., Li, S., McNabb, R. P., Nie, Q., Kuo, A. N., Toth, C. A., Izatt, J. A. & Farsiu, S. (2013), ‘Fast acquisition and reconstruction of optical coherence tomography images via sparse representation’, *IEEE Transactions on Medical Imaging* **32**(11), 2034–2049.
- Fang, L., Li, S., Nie, Q., Izatt, J. A., Toth, C. A. & Farsiu, S. (2012), ‘Sparsity based denoising of spectral domain optical coherence tomography images’, *Biomedical optics express* **3**(5), 927–942.

Bibliography

- Farsiu, S., Chiu, S. J., O'Connell, R. V., Folgar, F. A., Yuan, E., Izatt, J. A., Toth, C. A., Group, A.-R. E. D. S. . A. S. D. O. C. T. S. et al. (2014), Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography, Vol. 121, Elsevier, pp. 162–172.
- Fercher, A. F., Drexler, W., Hitzenberger, C. K. & Lasser, T. (2003), ‘Optical coherence tomography - principles and applications’, *Reports on Progress in Physics* **66**(2), 239–303.
- Fercher, A. F., Drexler, W., Hitzenberger, C. K. & Lasser, T. (2010), ‘Optical coherence tomography – development, principles, applications’, *Zeitschrift für Medizinische Physik* **20**(4), 251–276.
- Fernández, D. C. (2005), ‘Delineating fluid-filled region boundaries in optical coherence tomography images of the retina’, *IEEE Transactions on Medical Imaging* **24**(8), 939–945.
- Fernandez, D. C., Villate, N., Puliafito, C. A. & Rosenfeld, P. J. (2004), ‘Comparing total macular volume changes measured by optical coherence tomography with retinal lesion volume estimated by active contours’, *Iovs* **45**, U61.
- Fischler, M. A. & Elschlager, R. A. (1973), ‘The Representation and Matching of Pictorial Structures Representation’, *IEEE Transactions on Computers* **C-22**(1), 67–92.
- Frohman, E. M., Fujimoto, J. G., Frohman, T. C., Calabresi, P. A., Cutter, G. & Balcer, L. J. (2008), ‘Optical coherence tomography: a window into the mechanisms of multiple sclerosis’, *Nature Clinical Practice Neurology* **4**(12), 664–675.
- Fujiyoshi, H. (2007), ‘Gradient-Based Feature Extraction -SIFT and HOG’, *Computer* **107**(206), 211–224.
- Fuller, A. R., Zawadzki, R. J., Choi, S., Wiley, D. F., Werner, J. S. & Hamann, B.

Bibliography

- (2007), ‘Segmentation of three-dimensional retinal image data’, *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 1719–1726.
- Garvin, M. K., Abràmoff, M. D., Kardon, R., Russell, S. R., Wu, X. & Sonka, M. (2008), ‘Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search’, *IEEE Transactions on Medical Imaging* **27**(10), 1495–1505.
- Gatys, L. A., Ecker, A. S. & Bethge, M. (2015), ‘A neural algorithm of artistic style’, *arXiv preprint arXiv:1508.06576*.
- Gaudric, A., Krivosic, V. & Tadayoni, R. (2015), ‘Outer retina capillary invasion and ellipsoid zone loss in macular telangiectasia type 2 imaged by optical coherence tomography angiography’, *Retina* **35**(11), 2300–2306.
- Gloesmann, M., Hermann, B., Schubert, C., Sattmann, H., Ahnelt, P. K. & Drexler, W. (2003), ‘Histologic correlation of pig retina radial stratification with ultrahigh-resolution optical coherence tomography’, *Investigative Ophthalmology and Visual Science* **44**(4), 1696–1703.
- Grady, L. (2006), ‘Random walks for image segmentation’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(11), 1768–1783.
- Gregori, G. & Knighton, R. (2004), ‘A robust algorithm for retinal thickness measurements using optical coherence tomography (stratus oct)’, *Investigative Ophthalmology & Visual Science* **45**(13), 3007–3007.
- Gupta, V., Chan, C. C., Poh, C.-L., Chow, T. H., Meng, T. C. & Koon, N. B. (2008), Computerized automation of wavelet based denoising method to reduce speckle noise in oct images, in ‘Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on’, IEEE, pp. 120–123.
- He, K., Sun, J. & Tang, X. (2013), ‘Guided image filtering’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(6), 1397–1409.

Bibliography

- Hee, M. R., Izatt, J. A., Swanson, E. A., Huang, D., Schuman, J. S., Lin, C. P., Puliafito, C. A. & Fujimoto, J. G. (1995), ‘Optical coherence tomography of the human retina.’, *Archives of ophthalmology* **113**(3), 325–32.
- Hinton, G. E. (2007), ‘Learning multiple layers of representation’.
- Hou, H. S. & Andrews, H. C. (1978), ‘Cubic Splines for Image Interpolation and Digital Filtering’, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **26**(6), 508–517.
- Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., Hee, M. R., Flotte, T., Gregory, K., Puliafito, C. A., Et, A. & al., E. (1991), ‘Optical coherence tomography.’.
- Ishikawa, H., Stein, D. M., Wollstein, G., Beaton, S., Fujimoto, J. G. & Schuman, J. S. (2005), ‘Macular segmentation with optical coherence tomography’, *Investigative Ophthalmology and Visual Science* **46**(6), 2012–2017.
- Jian, Y., Zawadzki, R. J. & Sarunic, M. V. (2013), ‘Adaptive optics optical coherence tomography for in vivo mouse retinal imaging.’, *Journal of biomedical optics* **18**(5), 56007.
- Joachims, T., Hofmann, T., Yue, Y. & Yu, C.-N. (2009), ‘Predicting structured objects with support vector machines’, *Communications of the ACM* **52**(11), 97.
- Kafieh, R. & Rabbani, H. (2013), Optical coherence tomography noise reduction over learned dictionaries with introduction of complex wavelet for start dictionary, in ‘SPIE Optical Engineering+ Applications’, International Society for Optics and Photonics, pp. 885826–885826.
- Kafieh, R., Rabbani, H., Abramoff, M. D. & Sonka, M. (2013), ‘Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map’, *Medical Image Analysis* **17**(8), 907–928.

Bibliography

- Kanamori, A., Nakamura, M., Escano, M. F., Seya, R., Maeda, H. & Negi, A. (2003), ‘Evaluation of the glaucomatous damage on retinal nerve fiber layer thickness measured by optical coherence tomography’, *American journal of ophthalmology* **135**(4), 513–520.
- Karri, S., Garai, N., Nawn, D., Ghosh, S., Chakraborty, D. & Chatterjee, J. (2016), Retinal layer delineation through learning of tissue photon interaction in optical coherence tomography, in ‘Technology Symposium (TechSym), 2016 IEEE Students”, IEEE, pp. 46–51.
- Khalid, S., Akram, M. U., Hassan, T., Nasim, A. & Jameel, A. (2017), ‘Fully automated robust system to detect retinal edema, central serous chorioretinopathy, and age related macular degeneration from optical coherence tomography images’, *BioMed research international* **2017**.
- Kim, K. H., Maguluri, G. N., Cusato, K., Barlow, R. B. & Boer, J. F. D. (2008), ‘Monitoring mouse retinal degeneration with high-resolution spectral-domain optical coherence tomography’, *Journal of Vision* **8**(1), 1–11.
- Klein, R., Klein, B. E., Moss, S. E. & Linton, K. L. (1992), ‘The beaver dam eye study: retinopathy in adults with newly discovered and previously diagnosed diabetes mellitus’, *Ophthalmology* **99**(1), 58–62.
- Kobayashi, H. & Kohshima, S. (2001), ‘Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye’, *Journal of human evolution* **40**(5), 419–435.
- Kohavi, R. & John, G. (1997), ‘Wrappers for feature subset selection’, *Artificial intelligence* **97**(1), 273—324.
- Kontschieder, P., Bulo, S. R., Bischof, H. & Pelillo, M. (2011), Structured class-labels in

Bibliography

- random forests for semantic image labelling, in ‘Computer Vision (ICCV), 2011 IEEE International Conference on’, IEEE, pp. 2190–2197.
- Koozekanani, D., Boyer, K. & Roberts, C. (2001), ‘Retinal thickness measurements from optical coherence tomography using a Markov boundary model’, *IEEE Transactions on Medical Imaging* **20**(9), 900–916.
- Kotsiantis, S. B. (2007), ‘Supervised machine learning: A review of classification techniques’, *Informatica* **31**, 249–268.
- Krizhevsky, A., Sutskever, I. & Geoffrey E., H. (2012), ‘ImageNet Classification with Deep Convolutional Neural Networks’, *Advances in Neural Information Processing Systems 25 (NIPS2012)* pp. 1–9.
- Kromer, R., Serbecic, N., Hausner, L., Froelich, L., Aboul-Enein, F. & Beutelspacher, S. C. (2014), ‘Detection of retinal nerve fiber layer defects in Alzheimer’s disease using SD-OCT’, *Frontiers in Psychiatry* **5**(FEB).
- Lafferty, J., McCallum, A. & Pereira, F. C. N. (2001), ‘Conditional random fields: Probabilistic models for segmenting and labeling sequence data’, *ICML ’01 Proceedings of the Eighteenth International Conference on Machine Learning* **8**(June), 282–289.
- LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998), ‘Gradient-based learning applied to document recognition’, *Proceedings of the IEEE* **86**(11), 2278–2324.
- Lee, C. S., Baughman, D. M. & Lee, A. Y. (2017), ‘Deep learning is effective for classifying normal versus age-related macular degeneration oct images’, *Ophthalmology Retina* **1**(4), 322–327.
- Lee, K., Abràmoff, M. D., Niemeijer, M., Garvin, M. K. & Sonka, M. (2010), ‘3-D segmentation of retinal blood vessels in spectral-domain OCT volumes of the optic nerve head’, *Proceedings of SPIE Medical Imaging 2010 Biomedical Applications in Molecular Structural and Functional Imaging* **7626**, 76260V–76260V–8.

Bibliography

- London, A., Benhar, I. & Schwartz, M. (2013), ‘The retina as a window to the brain—from eye research to cns disorders’, *Nature Reviews Neurology* **9**(1), 44–53.
- Lu, R.-W., Curcio, C. a., Zhang, Y., Zhang, Q.-X., Pittler, S. J., Deretic, D. & Yao, X.-C. (2012), ‘Investigation of the hyper-reflective inner/outer segment band in optical coherence tomography of living frog retina’, *Journal of Biomedical Optics* **17**(6), 060504.
- Maaten, L.-v.-d. & Hinton, G. (2008), ‘Visualizing data using t-sne’, *Journal of Machine Learning Research* **9**(Nov), 2579–2605.
- Marleen, D. B. (2016), ‘Machine learning approaches in medical image analysis: From detection to diagnosis’, *Medical Image Analysis* **33**, 94–97.
- Marr, D. (1982), ‘Vision: A Computational Investigation of Visual Representation in Man’, *Phenomenology and the Cognitive Sciences* **8**(4), 397.
- Massich, J., Cheung, C. Y., Alsaih, K. & Lema^, G. (2016), ‘Classification of SD-OCT volumes with multi- pyramids , LBP and HoG descriptors : Application to DME detection Classification of SD-OCT volumes with multi pyramids , LBP and HOG’, *IEEE Engineering in Medicine and Biology Society (EMBC)* (August), 1344–1347.
- Mayer, M. A., Borsdorf, A., Wagner, M., Hornegger, J., Mardin, C. Y. & Tornow, R. P. (2012), ‘Wavelet denoising of multiframe optical coherence tomography data’, *Biomedical optics express* **3**(3), 572–589.
- Mayer, M. A., Tornow, R. P., Bock, R., Hornegger, J., Kruse, F. E., Mayer, MA., Tornow, RP., Bock, R., Hornegger, J. & Kruse, F., Mayer, M. A., Tornow, R. P., Hornegger, J. & Kruse, F. E. (2008), ‘Automatic Nerve Fiber Layer Segmentation and Geometry Correction on Spectral Domain OCT Images Using Fuzzy C-Means Clustering’, *Invest. Ophthalmol. Vis. Sci.* **49**(23), E–abstract 1880.
- McNabb, R. P., LaRocca, F., Farsiu, S., Kuo, A. N. & Izatt, J. A. (2012), ‘Distributed

Bibliography

- scanning volumetric sdoct for motion corrected corneal biometry', *Biomedical optics express* **3**(9), 2050–2065.
- Medeiros, F. A., Zangwill, L. M., Bowd, C., Vessani, R. M., Susanna Jr., R. & Weinreb, R. N. (2005), 'Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography', *Am J Ophthalmol* **139**(1), 44–55.
- Mesnil, G., Dauphin, Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I. J., Lavoie, E., Muller, X., Desjardins, G., Warde-Farley, D. et al. (2012), 'Unsupervised and transfer learning challenge: a deep learning approach.', *ICML Unsupervised and Transfer Learning* **27**, 97–110.
- Mishra, A., Wong, A., Bizheva, K. & Clausi, D. A. (2009), 'Intra-retinal layer segmentation in optical coherence tomography images.', *Optics Express* **17**(26), 23719–28.
- Mitchell, T. M. (1997), *Machine Learning*, Vol. 4.
- Montuoro, A., Waldstein, S. M., Gerendas, B. S., Schmidt-Erfurth, U. & Bogunović, H. (2017), 'Joint retinal layer and fluid segmentation in oct scans of eyes with severe macular edema using unsupervised representation and auto-context', *Biomedical Optics Express* **8**(3), 1874–1888.
- Mujat, M., Chan, R. C., Cense, B., Hyle Park, B., Joo, C., Akkin, T., Chen, T. C. & de Boer, J. F. (2005), 'Retinal nerve fiber layer thickness map determined from optical coherence tomography images', *Optics Express* **13**(23), 9480–9491.
- Nowozin, S. (2010), 'Structured Learning and Prediction in Computer Vision', *Foundations and Trends® in Computer Graphics and Vision* **6**(3-4), 185–365.
- Parikh, D. & Zitnick, C. L. (2011), Finding the weakest link in person detectors, in 'Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on', IEEE, pp. 1425–1432.

Bibliography

- Paris, S., Kornprobst, P., Tumblin, J. & Durand, F. (2008), ‘Bilateral Filtering: Theory and Applications’, *Foundations and Trends® in Computer Graphics and Vision* **4**(1), 1–75.
- Parker, A. (2003), In the blink of an eye: How vision sparked the Big Bang of evolution, in ‘In the blink of an eye: How vision sparked the Big Bang of evolution’, p. i.
- Polyak, S. L. (1941), ‘The retina.’
- Porat, M. & Zeevi, Y. Y. (1988), ‘The Generalized Gabor Scheme of Image Representation in Biological Machine Vision’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**(4), 452–468.
- Pratt, W. K. (2000), ‘Image Segmentation’, *Computer Vision* pp. 579–622.
- Quellec, G., Lee, K., Dolejsi, M., Garvin, M. K., Abràmoff, M. D. & Sonka, M. (2010), ‘Three-dimensional analysis of retinal layer texture: Identification of fluid-filled regions in SD-OCT of the macula’, *IEEE Transactions on Medical Imaging* **29**(6), 1321–1330.
- Rathke, F., Schmidt, S. & Schnörr, C. (2014), ‘Probabilistic intra-retinal layer segmentation in 3-d oct images using global shape regularization’, *Medical Image Analysis* **18**(5), 781–794.
- Rigamonti, R., Sironi, A., Lepetit, V. & Fua, P. (2013), Learning separable filters, in ‘Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on’, IEEE, pp. 2754–2761.
- Roorda, A. & Williams, D. R. (1999), ‘The arrangement of the three cone classes in the living human eye’, *Nature* **397**(6719), 520–522.
- Ryan Fanello, S., Keskin, C., Kohli, P., Izadi, S., Shotton, J., Criminisi, A., Pattacini, U. & Paek, T. (2014), Filter forests for learning data-dependent convolutional kernels, in

Bibliography

- ‘Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition’, pp. 1709–1716.
- Sankar, S., Sidibé, D., Cheung, Y., Wong, T. Y., Lamoureux, E., Milea, D. & Meriaudeau, F. (2016), ‘Classification of SD-OCT volumes for DME detection: an anomaly detection approach’, *Medical Imaging - Proceedings of SPIE* **9785**, 97852O.
- Savitzky, A. & Golay, M. J. (1964), ‘Smoothing and differentiation of data by simplified least squares procedures.’, *Analytical chemistry* **36**(8), 1627–1639.
- Schmitt, J. (1999), ‘Optical Coherence Tomography (OCT): A Review’, *IEEE Journal of Selected Topics in Quantum Electronics* **5**(4), 1205–1215.
- Schulter, S., Leistner, C. & Bischof, H. (2015), Fast and accurate image upscaling with super-resolution forests, in ‘Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition’, pp. 3791–3799.
- Shahidi, M., Wang, Z. & Zelkha, R. (2005), ‘Quantitative thickness measurement of retinal layers imaged by optical coherence tomography’, *American Journal of Ophthalmology* **139**(6).
- Shi, J. & Malik, J. (2000), ‘Normalized cuts and image segmentation’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(8), 888–905.
- Somfai, G. M., Tátrai, E., Laurik, L., Varga, B., Ölvedy, V., Jiang, H., Wang, J., Smiddy, W. E., Somogyi, A. & DeBuc, D. C. (2014), ‘Automated classifiers for early detection and diagnosis of retinopathy in diabetic eyes.’, *BMC bioinformatics* **15**(1), 106.
- Spaide, R. F. & Curcio, C. A. (2011), ‘Anatomical Correlates to The Bands Seen in The Outer Retina by Optical Coherence Tomography: Literature Review and Model’, *Retina* **31**(8), 1609–1619.

Bibliography

- Srinivasan, P. P., Kim, L. a., Mettu, P. S., Cousins, S. W., Comer, G. M., Izatt, J. a. & Farsiu, S. (2014), ‘Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images.’, *Biomedical optics express* **5**(10), 3568–77.
- Srinivasan, V. J., Monson, B. K., Wojtkowski, M., Bilonick, R. A., Gorczynska, I., Chen, R., Duker, J. S., Schuman, J. S. & Fujimoto, J. G. (2008), ‘Characterization of outer retinal morphology with high-speed, ultrahigh-resolution optical coherence tomography’, *Investigative Ophthalmology and Visual Science* **49**(4), 1571–1579.
- Sun, Y., Li, S. & Sun, Z. (2017), ‘Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning’, *Journal of Biomedical Optics* **22**(1), 016012.
- Syed, A. M., Hassan, T., Akram, M. U., Naz, S. & Khalid, S. (2016), ‘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.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2015), Going deeper with convolutions, in ‘Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition’, pp. 1–9.
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B. & Liang, J. (2016), ‘Convolutional neural networks for medical image analysis: Full training or fine tuning?’, *IEEE transactions on medical imaging* **35**(5), 1299–1312.
- Takeda, H., Farsiu, S. & Milanfar, P. (2007), ‘Kernel regression for image processing and reconstruction’, *IEEE Transactions on image processing* **16**(2), 349–366.
- Tobias, O. J. & Seara, R. (2002), ‘Image segmentation by histogram thresholding using fuzzy sets’, *IEEE Transactions on Image Processing* **11**(12), 1457–1465.

Bibliography

- Tomasi, C. & Manduchi, R. (1998), ‘Bilateral Filtering for Gray and Color Images’, *International Conference on Computer Vision* pp. 839–846.
- Tomlins, P. H. & Wang, R. K. (2005), ‘Theory, developments and applications of optical coherence tomography’, *Journal of Physics D: Applied Physics* **38**(15), 2519.
- Tsochantaridis, I., Hofmann, T., Joachims, T. & Altun, Y. (2004), ‘Support vector machine learning for interdependent and structured output spaces’, *International Conference on Machine Learning* p. 104.
- Vese, L. A. & Chan, T. F. (2002), ‘A multiphase level set framework for image segmentation using the Mumford and Shah model’, *International Journal of Computer Vision* **50**(3), 271–293.
- Vujosevic, S., Midena, E., Pilotto, E., Radin, P. P., Chiesa, L. & Cavarzeran, F. (2006), ‘Diabetic macular edema: correlation between microperimetry and optical coherence tomography findings.’, *Investigative ophthalmology & visual science* **47**(7), 3044–3051.
- Wang, S. & Summers, R. M. (2012), ‘Machine learning and radiology’.
- Westin, C.-F., Knutsson, H. & Kikinis, R. (2008), ‘Adaptive Image Filtering’, *Handbook of Medical Image Processing and Analysis* pp. 19–33.
- Wong, R. L. M., Lee, J. W. Y., Yau, G. S. K. & Wong, I. Y. H. (2015), ‘Relationship between Outer Retinal Layers Thickness and Visual Acuity in Diabetic Macular Edema’, *BioMed Research International* **2015**, 1–5.
- Xu, L., Ren, J. S., Yan, Q., Liao, R. & Jia, J. (2015), ‘Deep Edge-Aware Filters’, *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)* **37**, 1669–1678.
- Yang, L. & Jin, R. (2006), ‘Distance metric learning: A comprehensive survey’, *Michigan State Universiy* **2**(2).

Bibliography

- Yannuzzi, L. A., Ober, M. D., Slakter, J. S., Spaide, R. F., Fisher, Y. L., Flower, R. W. & Rosen, R. (2004), ‘Ophthalmic fundus imaging: today and beyond’, *American journal of ophthalmology* **137**(3), 511–524.
- Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. (2014), How transferable are features in deep neural networks?, in ‘Advances in neural information processing systems’, pp. 3320–3328.
- Yun, S., Tearney, G., De Boer, J. & Bouma, B. (2004), ‘Motion artifacts in optical coherence tomography with frequency-domain ranging’, *Optics Express* **12**(13), 2977–2998.
- Zangwill, L. M., Bowd, C., Berry, C. C., Williams, J., Blumenthal, E. Z., Sánchez-Galeana, C. a., Vasile, C. & Weinreb, R. N. (2001), ‘Discriminating between normal and glaucomatous eyes using the Heidelberg Retina Tomograph, GDx Nerve Fiber Analyzer, and Optical Coherence Tomograph’, *Arch Ophthalmol* **119**(July), 985–993.
- Zhang, K., Tao, D., Gao, X., Li, X. & Xiong, Z. (2015), ‘Learning multiple linear mappings for efficient single image super-resolution’, *IEEE Transactions on Image Processing* **24**(3), 846–861.
- Zlokolica, V., Jovanov, L., Pižurica, A., De Keyser, P., Dhaenens, F. & Philips, W. (2007), Wavelet-based denoising for 3d oct images, in ‘Optical Engineering+ Applications’, International Society for Optics and Photonics, pp. 66960P–66960P.

