

Abstract

Adapting learning-based approaches for various building blocks of image analysis have established state of the art performance. Medical experts have employed such blocks for automation of repetitive phases in clinical practices and current work targets ophthalmology. The retina is an integral part of human evolution for an effective interpretation of the world through visual sensing. Ophthalmologists currently employ optical coherence tomography (OCT) to interpret retinal layer profiles for early diagnosis of retinal and system level pathologies. The majority of employed machine learning algorithms for retinal image analysis are ‘predictive learning’ whose sole aim is reducing prediction error. Learning representations is an alternative approach whose mathematical model targets the representation (for major cases structure) in the anticipated space along with minimising classification or regression error (predictive learning). Anticipated space for representations can be structured output, manifolds in feature space, hierarchical feature quantifiers in an image, untangling dependencies in sampled image.

The work explores the impact of learning representations in different spaces through modifying existing structure or representation learning approaches and attribute to various blocks of retinal image analysis. The modifications involve equipping structured random forests with the capability of predicting layer specific edges, guiding an unsupervised algorithm to model manifolds rather boundaries between classes, transfer learning feature quantifiers followed by representation visualisation and finally inducing separability for image restoration with less number of parameters.

The impact of representation learning is validated by considering DUKE’s retinal OCT benchmark datasets. The retinal layers are delineated with an average F-score of 0.86. Feature space representation approach and data driven feature quantifiers (with softmax classifier) can identify retinal pathologies with a decision pooling accuracy of 0.94 and 0.91. The future scope involves fusion of data-driven feature descriptors and proposed classifier for end-to-end training. The restoration block is capable of recovering sparsely sampled images with 28.15 PSNR. The proposed approaches have illustrated comparative or superior performance to established baselines in various blocks of retinal image analysis namely layer segmentation with large deformations through modeling representations in target space, pathology classification through exploring feature space representations, data driven feature quantifier construction through hierarchical rep-

resentations in input space and restoration of sparsely sampled OCT images through sparse representations.

Keywords: Optical coherence tomography, image segmentation, feature quantification machine learning.

Chapter 1

Introduction

Automated modelling of patterns for interpretation and recognition has been a common practice for reduction of human workload and augmentation of performance. Machine learning (ML) is widely employed for pattern recognition and are applied in various streams of the medical community. ML based image analysis algorithms are adapted for automation of visual inspection based screening, but automated analysis of retinal OCT has been challenging. Current work considers improving various integral blocks for automated screening of retinal pathologies as the retina is the most vital sensory organ and retinal pathologies have surged in the current generation. Ophthalmologists employ optical coherence tomography (OCT) for screening retinal pathologies as it is the only device to image cross-sectional information with micrometre resolution in real time. Current application targets ML based image analysis blocks for screening retinal pathologies. Employment of block combination is subjective to screening approach, for example, a segmentation block can delineate retinal layers and depict layer thickness profiles so an ophthalmologist can attribute such profiles to various pathologies. Such ML based blocks have failed in multiple cases because experts identify features and interpret underlying representations, but traditional ML approaches are anchored around mapping features to a probability space. This resulted in unsettling results for

1. Introduction

intricate domains ranging from natural language processing to decision making. So advanced ML algorithms are capable of modelling structures and representations in various spaces. Even though such algorithms are recently popular in machine vision community adaptation of such algorithms for medical community is still an open problem.

Reduction of an ophthalmologist's workload through image analysis has intrigued biomedical researchers to automate blocks of retinal image analysis [1, 2] namely image segmentation, classification, restoration, etc. An individual block of image analysis is operated on different spaces such as input space, transformed space, feature space and target space. So current thesis is anchored around modelling of representations in various spaces of retinal OCT image and improving attributed blocks of retinal image analysis. The explored spaces and anticipated representation are: a structure in target space with application towards semantic segmentation, nonparametric modeling of data representations in feature space with application towards effective classification, hierarchical representations in input space with application towards heuristics free feature quantifiers and decomposition of representative quantifiers (feature quantifiers) for input space with application towards image restoration.

1.1 Objectives and Scope

The core aim of the thesis is to improve the performance of image analysis blocks employed in a computer aided diagnostic pipeline for retinal OCT applications. In the due process, it is identified that representations exist in various domains ranging from an input image to target space including feature space and learning such representations is inevitable for improved performance. The following objectives have been set to achieve this aim:

- Modelling edges of various retinal layers in OCT image for accurate delineation even under large deformations induced due to pathology.

1.2. Organization of the Thesis

- Developing a classifier capable of representing the feature space of samples from various classes rather than modelling the boundary between classes for identification of retinal pathologies.
- Transfer learning of data-driven feature descriptors for identification of retinal pathologies and reinforcing naive classifier.
- Reducing number of parameters required for simultaneous restoration and reconstruction of a noise-free image from sparsely sampled OCT image with a limited tradeoff of information.

In objectives, the autonomy of blocks is due to utilisation of dataset acquired from different OCT machines and is needed as they are benchmark datasets. The Dukes OCT database is considered as benchmark datasets and is publicly available for research.

1.2 Organization of the Thesis

The thesis is application driven, and impact of learning representations on each block has been validated through comparing with various baselines. The flow of the contributory chapters is arranged according to the space in which representation need to be learned, i.e., target, feature space and finally the input image. The thesis is accordingly organised as:

Chapter 1: Introduction: the current chapter emphasises the role of an eye as a sensory organ and need of automated tools to accelerate the screening pipeline for retinal pathologies. In the due process, it acknowledges the need for OCT imaging for accurate diagnosis and challenges of adapting conventional image analysis tools as retinal image analysis blocks. It also depicts the requirement of representation learning for improving the performance of the blocks. The chapter also includes a brief description regarding the association between representation space and an image analysis block.

1. Introduction

Chapter 2: Related Work: presents an overview of state of the art in various blocks of retinal image analysis like retinal layer segmentation, classifiers for identifying retinal pathologies, potential feature descriptors for effective quantification of retinal pathologies and computational imaging techniques to estimate a noise free OCT image. The chapter also discusses the limitations of techniques to lay down the need for proposed methods in the thesis.

Chapter 3: Layer-Specific Edges for Retinal Layer Segmentation: presents employment of *onevsall* approach for equipping structured random forests (for edges) to incubate retinal layer information along with edge for accurate delineation of retinal layers. It also describes the need for structured prediction of edges and need of accurate edges for graph-based segmentation approaches. A basic understanding of structured random forests for edges (SRE) during training and prediction has been included. It depicts the training and testing frameworks to predict layer-specific edges for a test image. A description of each block during training and image outputs at each block during testing. The chapter also contains the details of considered public dataset and experimental setup to validate the impact of proposed approach in comparison to existing baselines on Duke OCT DME dataset.

Chapter 4: Guided Fuzzy C-Means: A supervised approach: presents a classifier which aims to represent the samples in feature space rather modelling the decision boundary like a conventional classifier for identification of retinal pathologies like age-related macular degeneration (AMD) and diabetic macular edema (DME). It describes the transformation of OCT images to feature vectors through chapter 3 for retinal flattening and Histogram of oriented gradients (HOG) for feature quantification. It includes additional toy examples for better understanding of training and modelling characteristics of the proposed approach. It also depicts the modelling characteristics in the case of retinal feature vectors through transforming the high vectors to two-dimensional space. The chapter also contains the details of the experimental setup and

1.2. Organization of the Thesis

considered a public dataset to validate the impact of proposed approach in comparison to existing baselines on Duke OCT diabetic macular edema (DME) dataset. The performance of the proposed classifier is compared with support vector machines [3] with various kernels on a Duke OCT pathology classification dataset. It also includes the preprocessing and feature quantification pipeline employed.

Chapter 5: Construction and visualisation of data-driven feature quantifiers: presents an adaptation of a standard convolutional neural networks (CNN) architecture for identification of retinal pathologies and identify potential filter for the understanding of information processing. As neural network architectures are subjective to initialization repeatability of the experimentation is also illustrated. It includes the comparison of modelling capability of CNN architectures trained with random initialization and transfer learning of existing architectures. For a classification application effectiveness of a feature quantifier can be illustrated by its capability of empowering a naive classifier. So it also illustrates the impact of such features in augmenting a naive classifier performance in comparison to standard feature quantifier. It also describes an algorithm to identify potential response at each block of CNN autonomously. The HOG features are employed for comparison on Duke OCT pathology classification dataset and the preprocessing and standardisation pipeline is free of heuristics.

Chapter 6: Low-Rank CNN Filters for Simultaneous Image Reconstruction and Restoration: presents a modified sparse coding inspired CNN architecture for simultaneous restoration and reconstruction of sparsely sampled B-scan with less number of parameters and slight loss of information. It also illustrates the impact of space coding for restoration and functionally equivalent neural network. It describes the need and extension of learnable separable filter technique to a CNN architecture. It also adapts and describes a CNN architecture which is functionally equivalent to sparse coding for super resolution. It contains a brief overview of separable filters. The chapter connects the dots (sparse coding, CNN and separable filters) and validates the

performance on OCT denoising and interpolation dataset with various metrics.

1.3 Targeted Applications

Each application aims at improving the block performance independently, i.e., retinal layer delineation, retinal pathology classification, pathology subjective feature quantifiers and image restoration. Retinal layer delineation is commonly employed for profiling retinal layer thickness due to its implication for early diagnosis of retinal pathologies and system level pathologies like neuronal degeneration. Delineation also impacts the image analysis pipeline as the lower boundary of the retina is crucial for ‘retinal flattening’ which is an image preprocessing step for retinal pathology identification frameworks. Intra-retinal layer segmentation is one of the challenging problems and still in progressing phase due to feeble layers, noise, low gradients, etc. however, precise early diagnosis requires accurate delineation of all retinal layers. Alternatively, approaches have been proposed to minimise the dependency of screening over layer profiles through inclusion of low-level features in automated screening. Recent classification based screening approaches limited the dependency to lower boundary segmentation as it is one of the trivial problems due to its hyperreflective nature. These classification approaches initially consider the manifestations of pathologies and design or pick up feature quantifiers that depict the pathology manifestations. An ad-hoc approach of trial and error method is employed with various classifiers to identify the suitable classifier for the current application. Recent successful classification approaches construct the data driven feature quantifiers subjective to the application to avoid heuristic dependencies. Sparse sampled OCT images is a common practice in recent times to reduce imaging time and storage space which can be retrieved through restoration block. Computational imaging has received much attention in the form of image restoration block as it aids restoring the sparsely sampled scans. Along with restoration, such approaches are capable of simultaneous filtering.

1.3. Targeted Applications

Initial application explores semantic segmentation of OCT images with diabetic macular edema as they exhibit large deformation and remains challenging for retinal layer segmentation algorithms. Conventional image segmentation algorithms range from image processing algorithms (employing A-scan or image information for active contours) to pattern recognition approaches particularly machine learning based. The successful and dominant algorithms are more of a framework which hybridises gradients or gradient derivatives (Hessian matrix) computation and refinement through graphs or active contours. To avoid subjectivity of performance machine learning approaches are employed for layer modelling or layer edge modelling. The work is concerned with improving the modelling for better prediction which reinforces the graphs performance. This can be achieved through accurate edge detection for graphs. Conventional approaches employ a standard ML algorithm (support vector machines, random forests, etc.) to construct a model which predicts a probability of edge presence given an image patch. Advanced algorithms employ Structured Random forests (SRF) which predicts an edge patch given an image and feature patches, but an SRF predicts all edges irrespective of the layer. Proposed approach predicts class conditional edge, so a graph based solver will not be confused between edges from different layers.

The second application explores the identification of subjects with retinal pathology (age-related macular degeneration and diabetic macular edema) based on a histogram of oriented gradient features. A standard pipeline is considered for transforming an image into a feature vector and semantic segmentation model from previous application has been employed for performing retinal flattening. The pipeline additionally involves filtering, cropping the region of interest based on heuristically identified parameters and HOG feature quantification. Conventional off the shelf classifiers aims at identification and modelling of the boundary between data points from different classes. Graph-based approaches are good at understanding the structures, but an increase of data or dimensions is computationally expensive for training and prediction. So ensemble

1. Introduction

of models approach is commonly employed. Parametric or nonparametric approaches are commonly employed for modelling a structure and proposed approach considered nonparametric approach to avoid a multi-constraint problem. Intrinsic cost functions of unsupervised are aimed to explore the structure in data and labels help to guide the modelling. Proposed approach identifies control points representing various clusters in feature space so during prediction K-nearest neighbours (KNN) strategy is employed. It is commonly difficult for visual evaluation of the learned structures as feature space are usually high dimensional. So t-distributed stochastic neighbourhood estimation (t-SNE) is employed to project the high dimensional space to two-dimensional space to illustrate the structures in feature space.

The third application emphasis on the construction of application-specific feature descriptors which reduce the burden on classifiers. Features are termed to be the weakest link in low-level information based detection problem [4] as potential features descriptors can effectively project data onto feature space where samples from different classes are separable with low complex models. The current application is centred around detection of retinal pathologies which are attributed to geographical atrophy and are translated to retinal layer deformation. HOG descriptors quantify such atrophy but heuristic parameters are required to point the region of interest with reference to RPE location. Upon on interest of including new pathologies, the feature descriptors and heuristic parameters alter. To avoid such scenarios feature learning has emerged, and such approaches are expected to learn representations. Deep learning is part of representation learning which constructs feature descriptors with hierarchical nature. Hierarchical feature descriptors involve multiple sequential feature transformations where raw information is transformed to low-level feature descriptors through initial descriptors which are further transformed to refined features by the second descriptor to enrich the abstraction. The process is continued till the abstract pattern for samples from different classes are different in feature space. Visualising such patterns are crucial for defining various aspects of

1.4. Literature Review

deep learning (number of hierarchical layers, the number of weights in each layer, etc.), so an attempt has been made to visualise such abstract patterns.

The fourth application explores the possibility of reducing the parameters required for reconstruction and restoration of sparsely sampled images. Image acquisition involves sparse sampling to speed up the image acquisition process. Algorithms ranging from kernel to sparse representations have been benchmarked and identified that sparse representation based approaches have been effective. Sparse representation learning based deblurring or filtering process can be moulded as neural network learning process and recently as convolutional neural networks (CNN) to speed up the prediction. This approach allows an end to end learning, speed prediction and less number of parameters because SR based approach learns dictionaries in a mutually exclusive format based on patches but not images and require learning bag of dictionaries. The convolutional neural network learns representations in the form of filters and dependency in the each filter can be exploited to reduce the weights required. Signal processing community has termed these approaches to be separable filters, matrix decomposition, e.t.c. The impact of machine learning has impeached filter separability and lead to learning separable filters. The learning protocols followed various approaches, and parallelizable approach is employed to speed up the process as conventional separable filter approach is sequential in nature. The core aim is to reduce the number of parameters required without trading off the performance.

1.4 Literature Review

1.4.1 Machine learning

Machine learning (ML) has been an integral part of multiple predictive frameworks [5] in the past few decades and being infused into new territories [6] like medical sciences [2,7]. During inception, it is a minor part of artificial intelligence, but the rate of growth and

impact has attributed it to a distinguished field [5]. This is because predictive frameworks involve creating a model and estimate the target given a test data and modelling involves an infusion of domain knowledge through hard coding of rules, functions, transformations, etc. Machine learning automates the modelling without explicit coding for ‘what to learn?’ as it involves ‘how to learn?’ based on data [8]. The complete thesis employs supervised approaches [9] so the data employed for constructing model should include a set of feature and corresponding target values (integer or real) for each feature. Each feature is an extracted information from each image for current scenario and the target value is an integer or real based on the functionality of the ML algorithm being classifier or regression [5] so model represents a surface to separate or fit data points figure 1.1. An algorithm is termed to be ML algorithm if the predictive error of the constructed model keeps reducing with each iteration (experience) [8]. An ML (supervised) algorithm involves training and prediction phases so during training the model is constructed, and during testing, the model is employed to predicted target values which are of same nature as target values employed during training. This constrained the employment to applications with single value predictions like signal classification, image classification, voice recognition, etc.

1.4.2 Learning representations

Complex applications require predicting array of numbers instead of single numbers like semantic segmentation or voice translation [10–13]. Considering semantic segmentation as an example given an image 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 which has solved the problem partially. Such approaches resulted in an inconsistency between neighbouring predictions [11, 13] so structured learning has been introduced to predict the structured array of numbers (which include neighbouring predictions). The

1.4. Literature Review

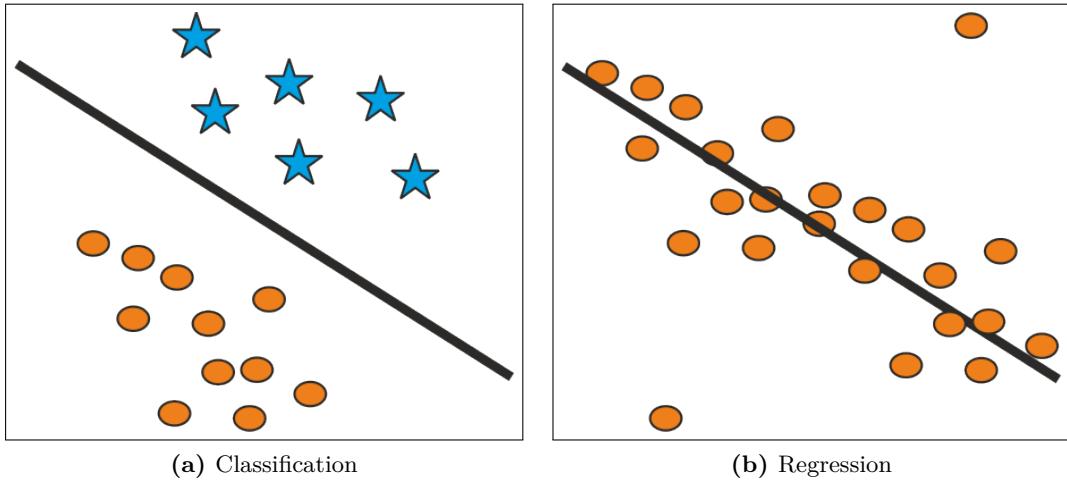


Figure 1.1: Illustrating an example with functional distinction between classification and regression

need has driven to modifications of loss functions to be optimised in classic ML like support vector machines (SVM) and predict structured output [10, 14]. Structured ML has taken different forms from Bayesian networks to deep learning including the graph theory. Recently algorithms with structure modelling, predicting, recognition, etc. capabilities have been grouped under structured ML [15]. In multiple cases, such structures do have a meaning or symbolise something such as edges, objects, etc. In computer vision or machine vision community, they are termed as representations [16–18]. A representation algorithm has various facets [15, 19] but a formal definition considered depicts that an algorithm employed for information processing with a target towards explicit entities [16]. The majority of employed algorithms for proposed frameworks are structured learning (structures in an input image or target space) and are aimed at explicit objectives at various spaces (input space and target space) so are treated to be representation learning (RL) algorithms. The remaining part employs unsupervised learning approaches for pattern recognition, and intrinsic loss function of such algorithms are explicitly designed to explore the representations of underlying patterns, so the remaining part also comes under representation learning.

Deep learning is presently the biggest buzz word in the artificial intelligence community [19,20] and convolutional neural networks [21] are employed as they established state of the art benchmarks in case image recognition. As RL is introduced in the context of deep learning, it was initially constrained as a feature learning but has the potentiality to be expanded to various function other to hierarchical quantification of structures (hierarchical abstract patterns in deep learning) [15]. There are multiple salient cases which need not model the structure yet still encompass representations in the form of instances (k-nearest neighbors), hyperplane (SVM) and set of rules (decision programs) [22].

Currently, representation learning has many facets in machine learning community due to its impact on various sub-branches of machine learning like variational learning [23], feature learning [19], structured learning, deep learning, etc. Due to its vastness only structured random forests [24], unsupervised learning and deep learning techniques have been considered for current work. The majority of the work in structured learning is developed around structured prediction but to be noted that structured prediction is a subset of structured learning i.e., algorithms with structure modelling but no structured prediction can still be structured learning [15]. For example learning structures (lines, curves, etc.) in input space (image) for classification of images [25] is also structure learning. Representation learning is being mainly driven by computer science community and data considered involve non-medical images (cats, dogs, cars, etc.) and need to be adapted to the medical applications [26]. As stated earlier, machine learning is being infused into multiple predictive applications the current work limits the application to retinal optical coherence tomography analysis.

1.4.3 Retina

Origin of vision is credited as the big bang of evolution which triggered the concept of competition for food and resulted in the inception of intelligence [27]. The level of intelligence has been stepped up through various aspects out of which one being

1.4. Literature Review

accumulating and accommodating multiple information from eye [28,29]. Eye interprets perturbing environment to the brain and helps the system to adapt accordingly [30]. Structural complexity has been an integral part of the eye to embed multi-functionality in the organ [29]. The eye is an important sensory organ due to its contribution of information to the brain for leading a perturbation free life [30]. An illustration for functionality of eye is depicted in figure 1.2 where real world light rays are focused onto the retina. In abnormal cases like myopia, there is shift of focal plane from retina leading to blurred vision. A major portion of current generation suffers from eye anomalies due to the lifestyle and source of all anomalies need not be eye. These anomalies range from reversible vision loss (refractive errors, cornea crystallisation) to irreversible vision loss (glaucoma, age-related macular degeneration) and major irreversible vision loss sources are attributed to retinal pathologies [30,31].

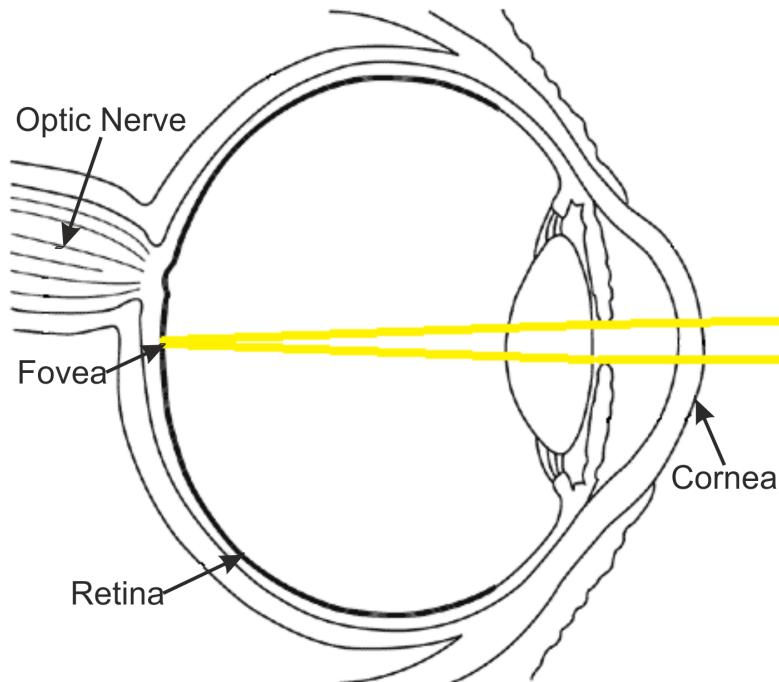


Figure 1.2: Illustrating a sagittal plane of eye

The retina is a crucial multi-layered structure constituting light sensitive tissue with functionality analogous to smart charge coupling device in the camera, i.e., transforming

1. Introduction

incident photons into encrypted electrical signals (nerve impulses in the case of the retina) [32, 33]. A sensor transfers raw signal to the imaging pipeline but retina itself processing the information before relaying to the visual cortex in the brain through the optic nerve. One such processing is differentiation of chroma through a different set of cones cells in the retina [29]. Any disruption of retina functionality due to pathology or trauma leads to disruption in photons to signal conversion, processing of information and relaying it to the brain. This leads to a misguided or incomplete visual information to the brain which affects various aspects of life from knowledge acquisition to daily activities. Retina also acts as a surrogate for diagnosing various systemic diseases [34] and organ localized disease [34,35]. The retina abnormalities need not be originated from retinal pathology and can be an effect of systemic or central nervous system pathologies. So retina also acts as a surrogate to the central nervous system pathologies as it is the only part of central nervous system that can be visualised without intervention [36].

1.4.4 Imaging modalities for retina

To recede the irreversible vision loss and regress the reversible vision loss regular vision checks are suggested by ophthalmologists [37]. A fundus camera is one of the primary imaging devices for retina in ophthalmic institute due to its capability of visualising macula and blood vessels non-invasively [38]. These features provide salient information for diagnosing various irregularities in the retina and other organs for which retina acts as a surrogate. Functional dimensionality has been included through fluorescence angiography for early diagnosis of few diseases like age-related macular degeneration (AMD) through Lipofuscin or diabetic macular edema (DME) through fluid leakage [39] in the retina. Autofluorescence properties of extracellular deposition (Lipofuscin) [40] and fluid leakage has added functional dimension to information in ophthalmic imaging which has boosted early diagnosis of retinal disorders. However, acquired information through all such imaging instruments are a surface projection of actual information and

1.4. Literature Review

made early diagnosis difficult as it needs cross-sectional information. Profiling of layers deformations and volumetric deformations is crucial for early diagnosis and accurate prognosis which is possible only with cross-sectional information. Ophthalmologists employed low coherence interferometer based subsurface imaging device which is termed as Optical Coherence Tomography (OCT) [41] to image cross-sectional information of the retina non-invasively.

Existing *in-vivo* imaging techniques at that time like MRI could not image retinal layers due to the limitation on resolution [42]. OCT is a subsurface imaging technique through optics with a penetration depth of few mm depending on the refractive index of the sample and axial resolution of few microns [43,44]. OCT was invented in an optical communication lab where experiments are performed to identify the distance of the damaged optical fiber from source using light (lasers). The same experimental setup has been extended to image a biological sample, but the reflected light response is rapid that no sensor of that time could 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 concatenated to form a B-scan (image) [45,46]. The laser source is limited to near infrared range to have better penetrability without loosing resolution.

The optical path of photons in OCT starts with the generation of the laser, splitting of the laser using a dichromatic mirror (beam splitter) where one is incident onto the sample and the other onto a reference mirror. Both the signals are reflected back where a constructive interference pattern is attained (time-delay) and translated as a single pixel in 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 [43] is the imaging protocol employed by time domain OCT. Such protocol involves moving parts and long acquisition time, so frequency domain based approaches have been introduced to improve the imaging speed and quality in the past

1. Introduction

decade. One of the refined approaches spectral domain OCT (SD-OCT) is extensively used for retinal imaging [47, 48]. Time domain OCT is dependent on time delay whereas SD-OCT operates based on wavelength to determine the spatial location of the reflected photons. In SD-OCT instead of moving reference mirror information of an A-scan line is acquired in one instance and evaluates the frequency spectrum of interference between reflected signal from fixed reflective mirror sample and the sample. An illustration of the photon path from source to A-scan line is depicted in figure 1.3. As an A scan line (axial scan) is performed in an instance which allows the SD-OCT to acquire multiple B-scans in same imaging time limit of TD-OCT. Such multiple B-scans at successive neighbouring locations leads to 3D imaging of the retina [49]. Proposed methods are based on B-scans but not 3D data as B-scan based approaches are computationally less expensive and can be expandable to 3D data if required.

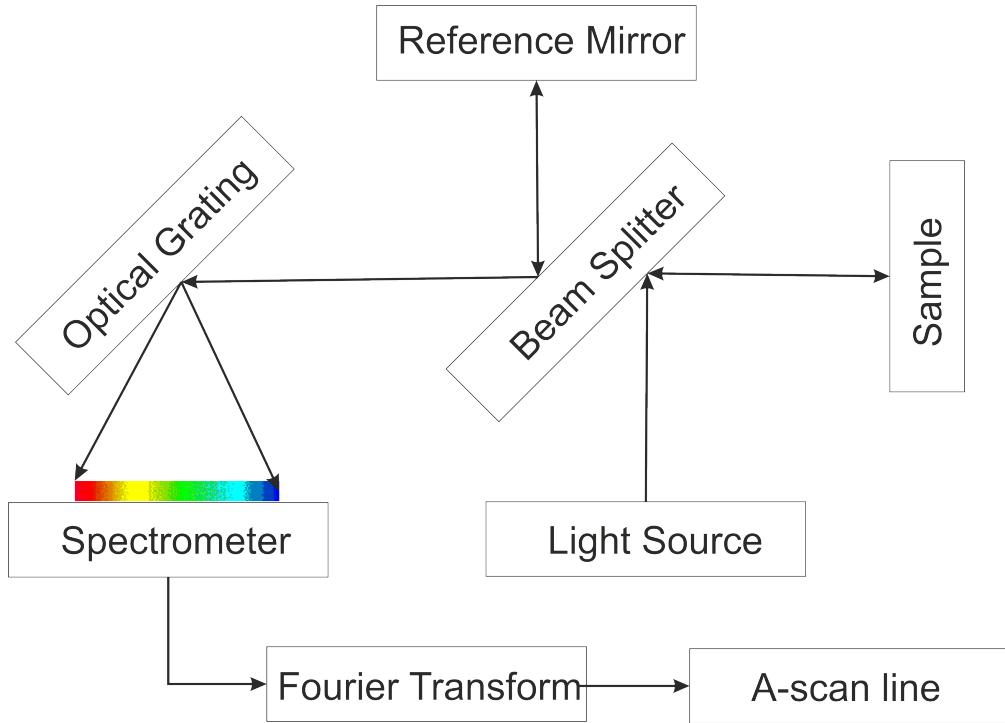


Figure 1.3: Illustrating the schematic of SD-OCT

figure 1.4 illustrates retina of a normal retina imaged with SD-OCT. This illustration

1.4. Literature Review

clearly depicts 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 BruchâŽs membrane (BM) layers. Multiple experimental validations has been performed to validate the anatomical information imaged through OCT with anatomy [50–52]. To enhance the OCT as a diagnostic tool various benchmarks has been established in comparison to fundus and flourescnece angiography and microperimetry [53–55] for grading pathologies related to visual acquity [56]. Studies have been carried out to map retinal layer profiles to various pathologies including drusen depositions and fluid accumulation [57–63].

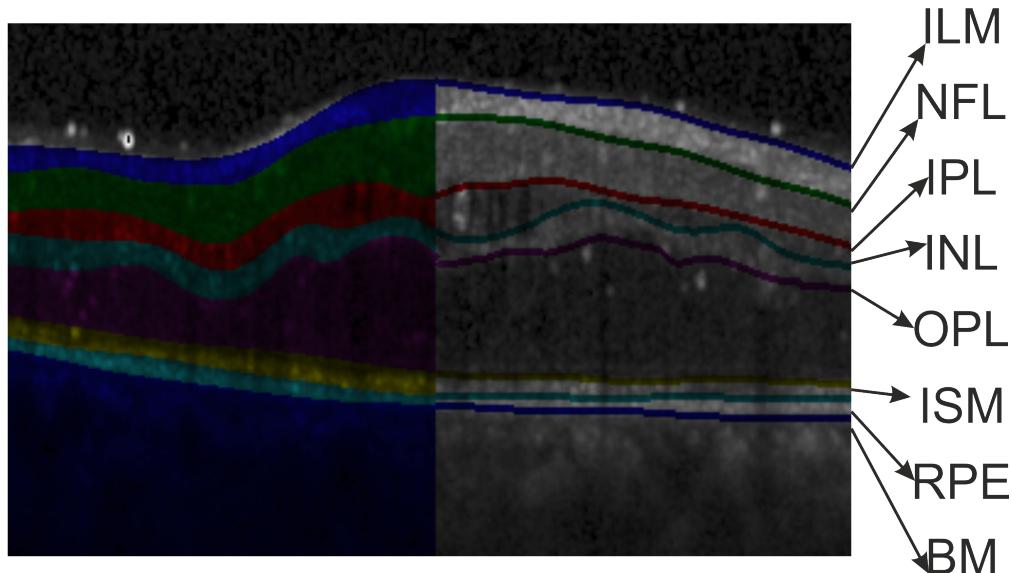


Figure 1.4: Illustrating various layers of retina in OCT B-scan image

1.4.5 Image analysis

Analysing such images (B-scans) is crucial for early diagnosis and automation has improved and speed up the screening process [2]. Overview of targetted four basic blocks of image analysis are given below.

Segmentation: Characterization of the object should start with delineation of the object and is the job of the segmentation block. The common hypothesis employed is

1. Introduction

the discontinuity, i.e., two objects in the image can be delineated if only there is some distinction in pixel intensity or texture [64]. So segmentation algorithms are a combination of abstract information from images (intensity distribution, edges, etc.), a metric to identify the difference of information between neighbouring pixels (entropy, variance, distance) and optimizer (line search, dynamic program) to identify boundary pixels between objects. Segmentation algorithms ranging from thresholding [65, 66] to graph theory [67–69] including active contours [70] employ one such combination. There is no hard thumb-rule for the best algorithm as it is subjective to application and image information. However, commonly employed segmentation algorithms are active contours and graph-based approaches. Machine learning has improved the segmentation block but resulted in discontinuous predictions [11, 13], so recent successful segmentation algorithms involve refinement block (graph based approaches like conditional random fields) as a post-processing step.

Feature Quantification: Upon delineation of the object the object need to be described quantitatively to generate a set of numbers so a set of rules can be constructed to recognise the object. Clinically significant qualitative descriptors are translated as quantitative descriptors to quantify the object [71]. Frequently an existing quantitative feature descriptor (for other application) corresponding to clinical descriptor is identified rather constructing application subjective descriptor [72]. This leads to low accuracies to anticipated classifier accuracy, so programmer includes potential feature descriptors (also from different application) [73] and iteratively removes ambiguous features based on accuracy through feature selection approaches. An alternate approach is feature extraction where the complete feature set (including ambiguous features) are transformed to a potential feature set [74, 75]. To be noted that dependency level of feature quantification on segmentation block is dependent on the amount of information to be considered and the kind of features to be quantified. Available algorithms for feature extraction are potential enough to construct data subjective feature quantifiers from standardised

1.4. Literature Review

image information [76].

Classification: Quantified or extracted information from images are termed as 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. Supervised classifier considers a training set of feature vectors with corresponding class labels and a classifier is constructed with an objective to identify the class label if given a test feature vector. The classical practice involves covering the feature space with data points through transforming images from various classes and termed as data augmentation. During prediction given an image, corresponding data point location in feature space is computed, so the label of nearby data points is considered as image label or class (K-nearest neighbors). Such practices need ample amount of data and storage, so classifiers are aimed to model the decision boundary between datapoints from various classes [22]. The datapoints need not be linearly separable, so classifiers are equipped with the capability to transform the feature space to space autonomously [77] or heuristically (Gaussian or Radial Basis Function kernels in SVM) where data points are linearly separable. The autonomous approach is computationally expensive for large data sets and heuristics are prone to subjectivity so nonlinear boundary modelling is performed through ensemble learning [78] where multiple classifiers are bagged together so even a linear classifier can model non-linear boundary through 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 CT, water molecules concentration in the case of MRI, lucidity in case of OCT, etc. All such information cannot be imaged or stored due to limitations of imaging protocol or sample leading to corrupted or missing information. The current block is employed as preprocessing or postprocessing (image transmission for telemedicine) to restore missing information (restoration) and remove the noise (filtering) through incorporating the knowledge of distortion functions (like Gaussian or Signal to noise

1. Introduction

ratio) and noise [79]. In the case of medical application pinpointing such information is difficult, so approaches with a hypothesis of homogeneity are successful. Such algorithms optimise the characteristics that image pixel or patch should be similar to neighbouring pixels or patches. Considering only pixel information or all neighbouring patches has lead to blurring at boundaries [80] so considering a subset of neighboring pixels or patches has improved performance and commonly called as adaptive filtering [81–84] or block matching approach [85]. BM3D [85] is one the most successful preprocessing approach for medical and non-medical applications, but sparse coding is more successful in the case of image restoration. Recently machine learning approaches are employed [86, 87] for restoration and filtering to avoid any prior knowledge and has been successful for medical and non-medical applications.

Chapter 2

Related Work

In this chapter, a survey of literature related to the contributions made in this dissertation is reported.

The rest of the chapter is organised as follows.

2.1 State of the art for retinal segmentation

Retinal layer delineation or segmentation is an evolving and challenging problem in medical image analysis. A wide range of methods have been proposed, varying from graph-based segmentation [88–90] to pattern recognition based prediction along with classical intensity based approaches. Few works have articulated all such methods and categorised based on information employed (A-scans, images, 3D scans, edges, low-level features) and algorithm (active contours, graph approach, machine learning, etc.). Graph-based approaches [89, 91–94] have superior performance, especially dynamic programming approaches, have been widely employed due to their performance with limited computational complexity even under the presence of noise. Graph approaches are powerful enough for exhaustive layer delineation with low computational facilities through iterative approaches [92], has been proposed. A graph is constructed based on edge and nodes where edges are computed based on image pixel information. Conventional

2. Related Work

graph-based segmentation involves search for optimal label assignment for each node which corresponds to each pixel in the image, so the label of each node is treated as the label of the pixel which translates to retinal delineation algorithms. The approaches considered for edge computation vary from method to method and are commonly based on the image low-level features, heuristics, machine learning predictions, and regularisation [88, 92, 94, 95]. In OCT images edges are computed, and the dynamic program is employed to identify the shortest path from the top left node to the bottom right node on the image graph and treated as layer boundary. Graph-based segmentation approaches are more concerned with solving the graph rather constructing the graph, i.e., edges. So the majority of graph approaches consider trivial gradient operators resulting in high responses at all prominent layers. Heuristics (layer average width and arrangement) are considered to avoid ambiguity and constrain the dynamic program search space. Alternatively, approaches induced fuzzy search space through a shape regularizer [94] to resolve ambiguity in information particularly due to the shadow artifacts. Heuristical constraints on search space have resulted in failed delineation in the case of rare patterns (or an emergence of new patterns). Machine learning approaches are favoured for recognition of a wide variety of patterns and regularizers are constructed through probabilistic graphical models [95]. These approaches have demonstrated the capability of constructing the regularizer comprehending wide patterns but fails particularly in case of pathologies associated to large layer deformations like diabetic macular edema (DME). So methods targeting layer segmentation in DME subjects employed models to predict edges or layer information rather constructing the regularizers for 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 performance and established state of the art in multiple instances [88, 92, 95]. Machine learning approaches for medical image segmentation are anchored around off the shelf learning methods (random forests,

2.2. State of the art for retinal pathology classification

support vector machines, etc.). Recent developments in machine learning have lead to extending structured prediction or learning representations for image processing applications and have lead to the introduction of structured SVM and structured random forests (SRF) [13]. SRF is considered as it holds the properties of random forests which are optimal risk minimization, robust to noise in labels, etc. The SRF is modified to predict object edges instead of objects [24].

2.2 State of the art for retinal pathology classification

Any image analysis pipeline is incomplete without the identification or recognition of target, and current application is for identification (screening) of the pathological subjects. Chronologically initial approaches involve mimicking the clinical practice, i.e., layer delineation and classification of layer profiles as retinal width profiles can be indexed to various pathologies [96–100]. Such classifiers are a curated list of decision rules practised in ophthalmology and potentiality of such rules are dependent on the effective performance of preceding block, i.e., segmentation which is still an evolving and challenging field [101]. Such rules are limited to morphological variations or changes in width profiles but recent descriptors include gradient based features. Low-level feature descriptors with classifiers capable of modelling the decision boundaries in the feature space are introduced in the screening pipeline. This reduces the level of dependency on the segmentation block [73, 76, 102].

Classifiers are categorised to supervised, unsupervised and semi-supervised based on the availability of data out of which supervised are commonly employed for image-based screening systems. This is because ample amount of data is required to identify the representing population of data samples in the feature space and supervised approaches need a limited amount of data (in comparison to unsupervised) as they model the decision boundary between the data samples from various classes in feature space [9]. Support vector machines [3], decision trees [78] etc. initially extrapolates the representations or

2. Related Work

statistics of the population based on the training dataset which is a sampled data from the population. Then the decision boundary or boundaries (in the case of ensemble learning) are modelled as linear, nonlinear (piecewise linear). There are no thumb rules for selection of classifier for guaranteed performance as there are multiple tuning knobs (hyperparameters) in machine learning based modelling process [22]. This demands hefty machine learning expert knowledge regarding model characteristics in various scenarios so the parameters can be tweaked to model better for the required application. Majority of retinal pathology screening methods include SVM [76, 103–105] or ANN [102, 106] out of which SVM is widely employed. The majority of such approach aim to identify only single pathology, but proposed approach aims for multiple pathologies (AMD and DME) similar to [76].

Transforming current application as an anomaly detection involves adaptation of an unsupervised based approach to detect pathological cases [107]. Such approaches are successful as long as all variations of the normal subjects have been effectively modelled which is not possible with a sampled data. Semi-supervised approaches are not yet employed for retinal OCT screening. The primary reason could be that majority of such approaches are graph based (random walks) and computational including storage complexity of graphs (Laplacian matrix is commonly stored) is high.

2.3 State of the art for retinal pathology feature descriptors

It is established in the previous section that classical algorithms for retinal screening employed width profiles as the features [91, 96–100] and translated to representative features by considering order statistics of width profiles (mean and standard deviation). In the case of AMD, the deposition of drusen beneath retinal pigment epithelium (RPE) is can be reflected in the variation of width profile from RPE to Bruch’s membrane

2.3. State of the art for retinal pathology feature descriptors

(upper boundary of choroid) across each A-scan and can be quantified with standard deviation [98]. The amount deposited can be quantified through the difference between means of subjects with pathology and without pathology which is translated to the severity of the pathology for diagnosis. The screening process involves only identification of pathology but not severity we limit the state of the art to identification. In the case of DME, fluid deposition commonly occurs above RPE so if the target pathology changes the whole target features change [98, 108]. Construction of common features (pathology independent) is helpful to identify various pathologies rather pushing each OCT image through a pathology subjective screening system. Current application targets AMD and DME pathologies so common feature descriptors capable of delineation subjects into three classes (normal, AMD and DME) need to be considered. Few approaches have implemented for simultaneous identification and delineation of drusen [109] and fluid deposition [71, 92] to classify AMD and DME cases but needs robust segmentation approaches.

To reduce the degree of dependency on segmented layers low-level image descriptors are employed. Such approaches are dependent on only retinal flattening which involves identification of lower boundary of the retina which is one of the hyperreflective bands [51] in retinal OCT images. Low-level image descriptors (HOG) [73, 76, 102] are employed as they act as a partial representation of the pathology feature and collection of such descriptors can represent various pathologies with distinguishability in the feature space. From linear algebra perspective, they can act a governing variable in an equation and classifier identifies the weights for the equation to separate the pathologies. The questionable aspect is ‘number of features that need to be incorporated’ as all existing feature descriptors can not be employed. So feature descriptors where each quantifies a partial aspect of the pathology are incorporated and are filtered through feature extraction approaches like PCA [110].

2.4 State of the art for restoration and reconstruction of retinal OCT

Ophthalmologists in clinical practice prefer a high detailed image with good peak signal to noise ratio (PSNR). This can be achieved through dynamically modifying the shape of reflective mirrors in OCT optical (adaptive optics) [111] or restoration of actual information through software [112] rather governing instrumentation (computational imaging). Computational imaging techniques employed for noise removal, and restoration of OCT images can be categorised into a multi image based [113, 114] or single image based techniques [93] as SD-OCT is capable of providing 3D scans. Single image based techniques are more favoured as they can be extended to multi-image and computationally less complex. Classic techniques for reconstruction of a high detailed image (with high PSNR) from a sparsely sampled image employ a single-stage or two-stage approach. In the absence of noise two-stage approach involves interpolation and deblurring which is not the case with OCT. In the presence of noise the two stage approach involves interpolating and denoising [53] or vice-versa. Such approaches employ conventional interpolation techniques like bicubic [115] and more concerned on denoising approaches. Such denoising algorithms include anisotropic [53, 116–118], median [58, 100, 119–121], 2D linear smoothing [45], low pass filter [46], mean filter [90, 122], directional [123], adaptive kernel [124], SVM [99] and wavelet [73, 125–127]. Some approaches have combination of filtering techniques [97, 128] and employment of ML has encouraging results [99]

Most modern and successful approaches for denoising OCT images are dictionary based [129] and particularly which employ sparse representation [112, 130] as core algorithm. Such algorithms employ either raw information from spectroscopy (frequency domain) or constructed image (spatial domain), and proposed approach employs later path as former approach needs tapping into imaging pipeline which is not an option availed by commercially available OCT. [112] has performed an empirical experiment to

2.4. State of the art for restoration and reconstruction of retinal OCT

validate that sparse coding based approach has superior performance in comparison 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 [131] based approaches are also explored. BM3D and BM4D approaches have been employed for denoising for single image based and multi-image based denoising techniques respectively. Machine learning based approaches have been impeached into medical image preprocessing techniques and have been successful through random forests [86]. It is also established through empirical experiments on non-medical data that a neural network based architecture can perform similarly to a BM3D [87]. Dictionary based learning approaches have been successful as it makes the pipeline free from knowledge of the noise model which is stochastic in nature for OCT. A PCA can also act as noise filter because it helps to retain repetitive patterns the image patches for constructing dictionaries are processed through PCA [132]. This effectively reduces the load on dictionaries to remove noise.

Chapter 3

Layer Specific Edges for Retinal Layer Segmentation

3.1 Introduction

Autonomous retinal layer segmentation approaches for OCT images have been developed to reduce ophthalmologists burden through computing the width profiles [90]. This transforms an image into an array where each element represents the width of a layer in the corresponding column and a 3D stack to width maps corresponding to each layer. These width maps are clinically referred to as an Early Treatment of Diabetic Retinopathy Study. Such automated width maps manifest a 2D representation of the retina which augments the ophthalmologists screening speed and identify a representative scan effectively for diagnosis [133]. This has intrigued in exploring multiple retinal layers width profiles [98, 134] so mapping of multiple width profiles unveils new relations and augments the diagnosis in comparison to maps generated based on targeted or single layer width. This has raised the need for more effective segmentation algorithms that can deal with layers that have similar intensities as their adjunct layers, morphological deviations in layer structures due to pathology, noise with speckle nature, etc. [88].

3. Layer Specific Edges for Retinal Layer Segmentation

3.1.1 Need of structured prediction

Graph-based approaches are effective for OCT based retinal segmentation, and edge predicted based on machine learning approaches have established remarkable baselines. Random forests (RF) are considered for current application due to the capability of modelling complex decision boundaries during training and rapid predictions during testing. A common problem in extending ML for segmentation is inconsistent prediction between neighbouring pixels resulting unappealing semantic segmentation. This can be resolved with a post-processing step involving heuristics and handcrafted parameters. Another approach is appending conditional random fields for refining current pixel prediction (unary potential) by considering neighbourhood pixels predictions (pair-wise potential). Such practices avoid an end to end training including RF due to architectural incompatibility. Structured random forests [13], equips RF the capability to predict an array (instead of a number or label) based on a feature vector which can be reshaped as a 2D patch (i.e., the label patch) for image processing application, as shown in figure 3.1.

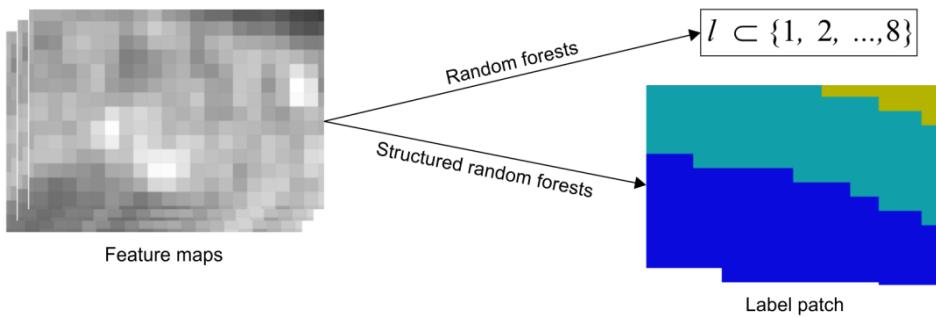


Figure 3.1: Illustrating distinction between conventional prediction and structured random forests prediction

Structured prediction through classical learning machines

Random forests is an ensemble of decision trees and constructing rules at decision node of a decision tree based on label patches is computationally infeasible. This issue has been addressed by indexing every possible label patch pattern. Random forests training set

3.1. Introduction

comprises feature and mapped indexes so during prediction estimated index is mapped back to corresponding pattern. A pitfall of this approach is 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 resulting in a need for extreme machine learning. A 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 edges instead of semantic labels.

3.1.2 Structured random forests for edges

Initial development phases of Structured random forest for image analysis are aimed to avoid expert intervened clustering and equip decision nodes to handle label patches. This is achieved by reducing the vectorized label patch dimension through random sampling and evaluated for semantic segmentation applications. Such sampling also induces noise into the modelling pipeline which avoids overfitting. Structured forests for edge (SRE) [24] integrates clustering and random sampling capability at decision node as clustering explores structural similarity between patches and dimension reduction through sampling refrains the model from overfitting. Training an SRE (an ensemble of decision trees) employs feature vectors, edge patches and label patches, as shown in figure 3.2a (feature vectors are reshaped for visualisation); where each tree is trained with different data. Training a tree involves constructing decision node (circles in figure 3.2b) rules and leaf node (rectangles in figure 3.2c) decisions. A SRE decision node is equipped to map label patch to a low dimensional binary vector (z) so 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 (figure 3.2b) [24]. Decision node rule is constructed

3. Layer Specific Edges for Retinal Layer Segmentation

based on 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 dissimilarity between label patches falls below a tolerant level (left branch in figure 3.2b). Such last node of a branch is treated as the leaf node. In each leaf node, a least distant label patch to other label patches is stored as representative label patch as shown in figure 3.2c. Edge patches corresponding to representative label patches are retrieved from input data and stored. Given a test feature vector, each tree in SRE directs towards a leaf node based on decision node rules, as shown in figure 3.2d. The edge patch stored in the leaf node is considered as tree decision. The mean of tree decisions is treated as SRE decision, i.e., mean of edge patches. For illustration purposes, all patches in figure 3.2 are kept the same but the feature vector and label patch need not be the same dimension as it involves z mapping but 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.

3.1.3 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 hypothesised that inclusion of layer class information along with edge can augment the dynamic program path prediction. Such information suppresses artefact induced edges, avoids fusion of edges from different layers, shows resilience to shadow artefacts, prevents dynamic programming from jumping (between layers) across paths in case of feeble layers, and latches on to the edge even under large

3.2. Preamble to solution

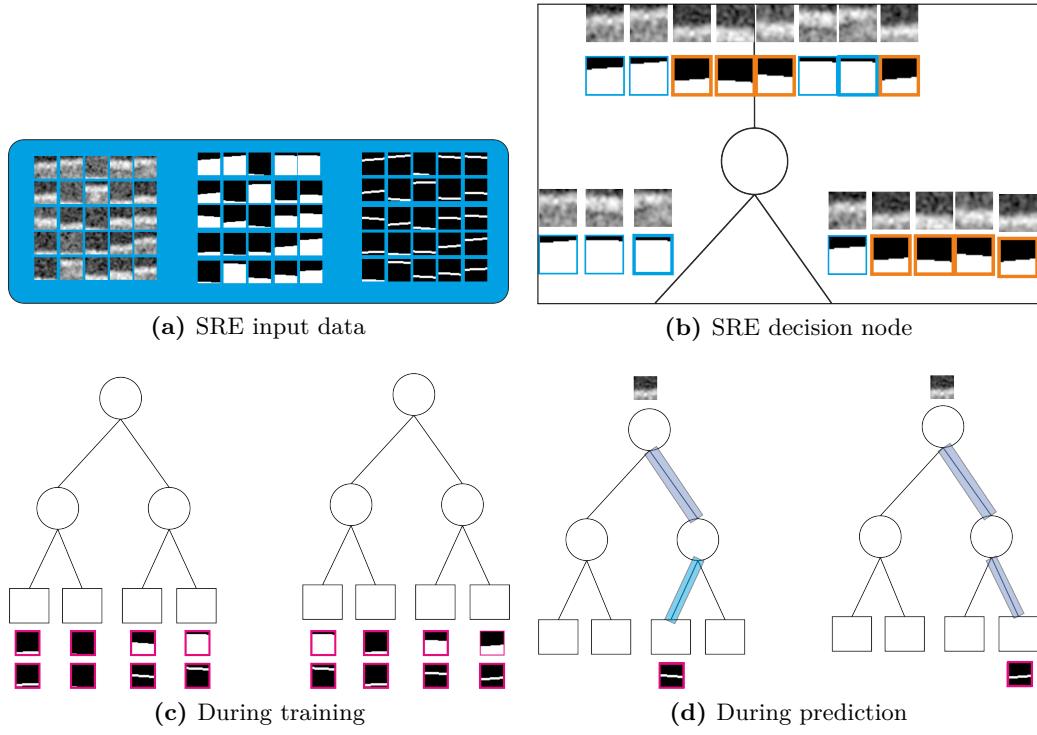


Figure 3.2: Training and prediction process of structured forest for edge

deformation. The learned edge is compatible not only with the dynamic program but also any off-the-shelf graph optimisation. In other words, the proposed approach reinforces existing graph optimizers, and it should not be considered as a competitive alternative.

3.2 Preamble to solution

For the application considered the edges should be layer specific as opposed to modeling the edges of all layers. But as mentioned SRE are designed to consider only binary labels and predict binary labels. SRE can be equipped to predict layer specific edges by employing ‘one vs all’ (multiclass classification) approach. Such approach avoids any intervention into learning algorithm so it can be extended to any structured learning algorithm. The ‘one vs all’ can be achieved through appending layer ‘n’ selection block for target images (labels and edges) and training a model ie. SRE for each layer. The

3. Layer Specific Edges for Retinal Layer Segmentation

layer ‘n’ selection block binarize a given label image (L) through following Equation 3.1 where x and y are spatial coordinates of the image. Current application involve eight layers so eight SRE models need to trained. To depict that the proposed modification is not subjective to feature considered, the same features as in [24] are employed. Given an image (figure 3.3a), figure 3.3 the difference between the predictions of SRE and those of the proposed method are illustrated in figure 3.3b and figure 3.3c.

$$O(x, y) = \begin{cases} 1 & \text{if } L(x, y) = n \\ 0 & \text{else} \end{cases} \quad (3.1)$$

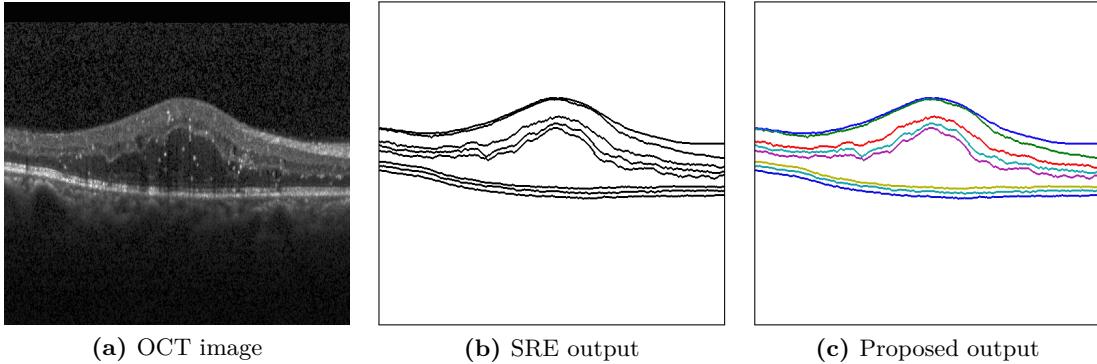


Figure 3.3: Layer information predicted by proposed approach along with predictions of SRE for an image

3.3 Method

3.3.1 Data preparation

The proposed approach is aimed at identifying the upper edge of inner limiting membrane, nerve fibre layer, inner plexiform layer, inner nuclear layer, outer plexiform layer, inner segment myeloid, retinal pigment epithelium (RPE), and choroid (Bruch’s membrane). The above mentioned layers are here after termed as Layer 1 to 8. As proposed approach being supervised, it requires ground-truth along with SD-OCT images. The

3.3. Method

contour image (C) with the upper edge of the layers annotated by experts for a given image is treated as 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 contour image. An example of SD-OCT image (figure 3.4a) with corresponding label (figure 3.4b) and contour images (figure 3.4c) with pseudo-colors are illustrated in figure 3.4

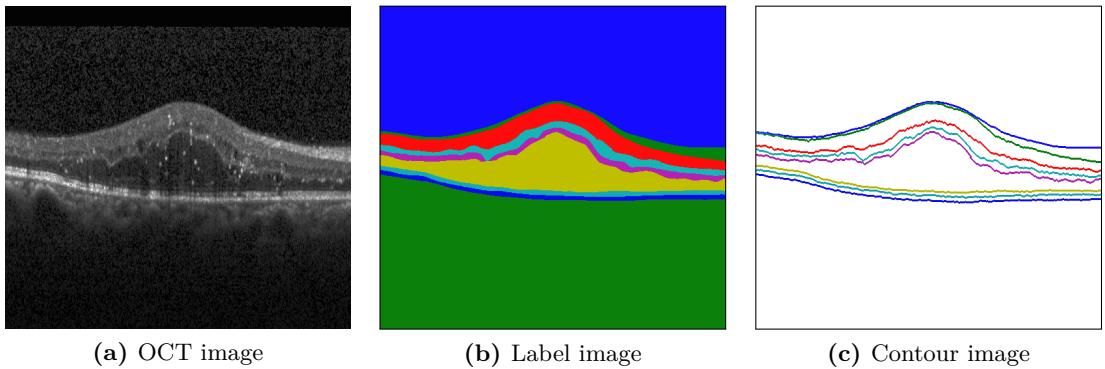


Figure 3.4: Required images for preparing training dataset. (a) Input image, (b) label image, and (c) contour image.

3.3.2 Model training and prediction

As any conventional supervised approach, the proposed framework involves two phases (training and testing): training single SRE for upper edge (contour) of each layer (figure 3.5a) and predicting the upper edge of each layer (figure 3.5b). 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 input SD-OCT image is standardised to [0,1]. The image-acquisition protocol of OCT chisels the acquired A-scans to match a predefined B-scans and replace the missing values with maximum values. This results in saturated values

3. Layer Specific Edges for Retinal Layer Segmentation

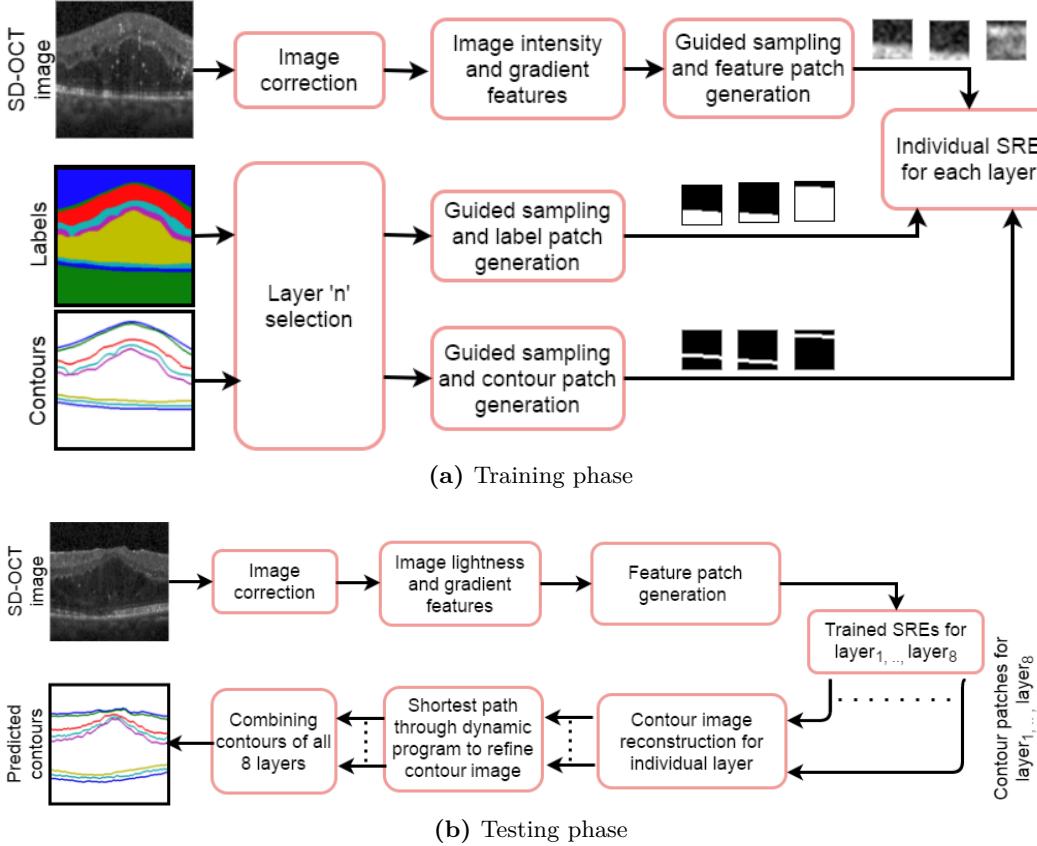


Figure 3.5: Flow of processing blocks involved in (a) training phase and (b) prediction phase

along the boundaries of OCT image. This is resolved by replacing pixel values equal to 1 with 0.01, and the resultant image is treated as I .

Image intensity and gradient features

The image intensity values and the 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 (9×9 window), normalised and treated as a magnitude map (A). HOG features are extracted (four orientation bins and 9×9 window) with a stride of one which results in four histogram maps (H) where each map is the result of each orientation. It is commonly suggested to incorporate scale-space,

3.3. Method

so image resampling is performed over I at scale 0.5 to generate $I_{0.5}$. Magnitude map and HOG feature maps are computed on $I_{0.5}$ and maps are resampled to match the dimensions of I which results in $A_{0.5}$ and $H_{0.5}$. A feature tuple (F) is generated by concatenating all the 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 SRF to predict layer specific edges through one vs all approach. Given a C (contour image), the block creates a binary contour mask (BC) through equation (3.1), with only the upper edge (contour) of layer ‘n’ ($n \in 1, 2, \dots, 8$) as the foreground. In other words, only spatial if the location (x, y) in C is ‘n’ corresponding BC location is ‘1’ else location of BC is ‘0’. Given an L (label image), the block creates a binary label mask (BL) through equation (3.1), where the posterior region of BC is the foreground and its anterior region is the background. Given an L (figure 3.6a) and C (figure 3.6b) the construed BL (figure 3.6c) and BC (figure 3.6d) for Layer 3 is illustrated in figure 3.6. A BC is also used as a guide for sampling the feature, label, and contour patches.

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 ($Y = (x_1, y_1), (x_2, y_2), \dots \forall DBC(x, y) = 1$). 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.

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 resulting in a set of feature patches (P) with each patch size being $32 \times$

3. Layer Specific Edges for Retinal Layer Segmentation

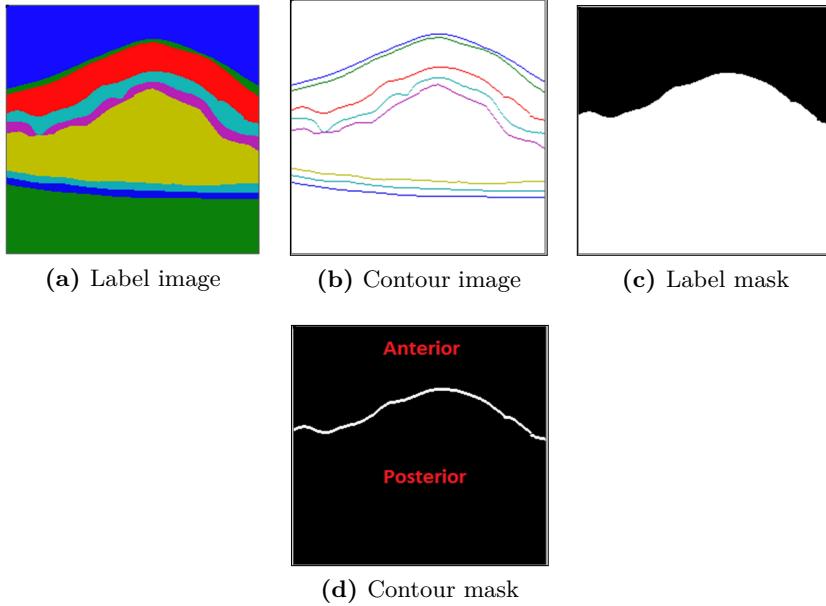


Figure 3.6: Processed outputs of layer $\hat{A} \tilde{Y} n \tilde{A} \tilde{Z}$ selection block for Layer 3. (a) Label image, (b) contour image, (c) processed label binary mask, and (d) processed contour binary mask

32×11 . Each feature patch is flattened through lexicographical arrangement resulting in a feature vector \mathbf{f} (1D array).

Label patches and contour patches each of dimensions 16×16 are extracted from BL and BC based on Y and N . The correspondence of central pixel between feature patches, labels patches, and contour patches are 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 termed as SRE_n . The training process is repeated for all eight layers.

3.4. Results and Discussion

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 same dimensions as the test image. 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 [89] and concatenation with a column vector ($1e^{-6}$) on left and right side. The shortest path from the top-left node to right-bottom node of the graph is computed based on the dynamic program. This path is smoothed using the SavitzkyâŠGolay filter [135]. The smoothed path is projected back to image space and considered as layer ‘n’ contour (the upper edge).

Combining contours

Upon repeating SRE prediction, short 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 layer delineation result.

3.4 Results and Discussion

3.4.1 Experimental bed

A Duke OCT DME dataset comprising 110 images from 10 subjects along with the upper boundary of eight layers annotated by two clinical experts has been released by Saina Farsiu group. MatLab 2014b is considered for experimentation. The Piotr’s image (and video) toolbox and the edges toolbox are used for feature maps, guided sampling, and the structured random forest for edges. The codes for feature tuple construction are modified

3. Layer Specific Edges for Retinal Layer Segmentation

for the current application. Ubuntu installed workstation with 112 GB of RAM is used during training. However, edge prediction, refinement and smoothening require only 8 GB of RAM. Kernel regression based segmentation [92], has established a benchmark (AD) by guiding the classical approach (AN) for robust segmentation. So, these two methods are considered as baselines for quantitative and qualitative comparisons.

3.4.2 Data preparation

The provided data has been prepared to match the proposed method's data requirement. Proposed method considers only one ground truth (contour) image corresponding to an OCT image so for each 'n' experts annotation location of contour along each column is averaged. The 'nan' (not a number) values along a contour are estimated through 1D interpolation. Each averaged location is considered as a row value and a contour image is constructed by allotting the layer index 'n' ($n \in 1,.., 8$) 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 eight layer, along each column pixels between eighth contour and last row are replaced with '8'. In contour and label image only columns from 120 to 650 are considered, owing to lack of expert annotations.

3.4.3 Parameters for training of SRE_n

Images of first five subjects, i.e., 55 are considered for training. PiotrâŽs image and the video toolbox is used for the construction of feature tuple. Upper-limit of Y and N set size is 10^6 and 1.5×10^6 . The dimension of z for low dimension projection is set to 256. PiotrâŽs edges toolbox is used for training SRE with six trees. Each tree employs 25% of data. As the training of different trees in an SRE_n is mutually exclusive the model training process can be parallelized.

3.4. Results and Discussion

3.4.4 Governing parameters of prediction complexity

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

3.4.5 Delineating given test image

Upon training phase completion it is expected to construct eight SRE models each having a capability of predicting upper edge (contour) of the corresponding layer. The figure 3.7 illustrates the predictions of each model given a test image, and it also illustrates the image results of major blocks in the prediction pipeline. It can be observed that the edges are predicted but not crisp. So dynamic program refines the edge to single pixel width. There exists zig-zag pattern due to local optimality approach considered by Dijkstra's algorithm which is smoothed. The resulting contours are considered as layer delineations of the retina. Both layer width profiles and edge profiles can be illustrated with considered illustration figure 3.7.

3.4.6 Delineating capability under various scenarios

The robustness of the proposed framework against the influence of deformations, artefacts, noise and shadow are illustrated in figure 3.8. The magnified representations are illustrated in figure 3.8a to depict the efficacy of proposed approach in the presence of noise-induced due to imaging-protocol, layer deformation due to pathology, low gradients due to signal attenuation in the posterior retina, and shadows due to blood vessels are illustrated in as illustrated in figure 3.8c and figure 3.8d respectively.

The prediction of the proposed approach has more agreeability with expert annotations in comparison to kernel regression features based segmentation, both visually (figure 3.9) and statistically (table 3.2).

3. Layer Specific Edges for Retinal Layer Segmentation

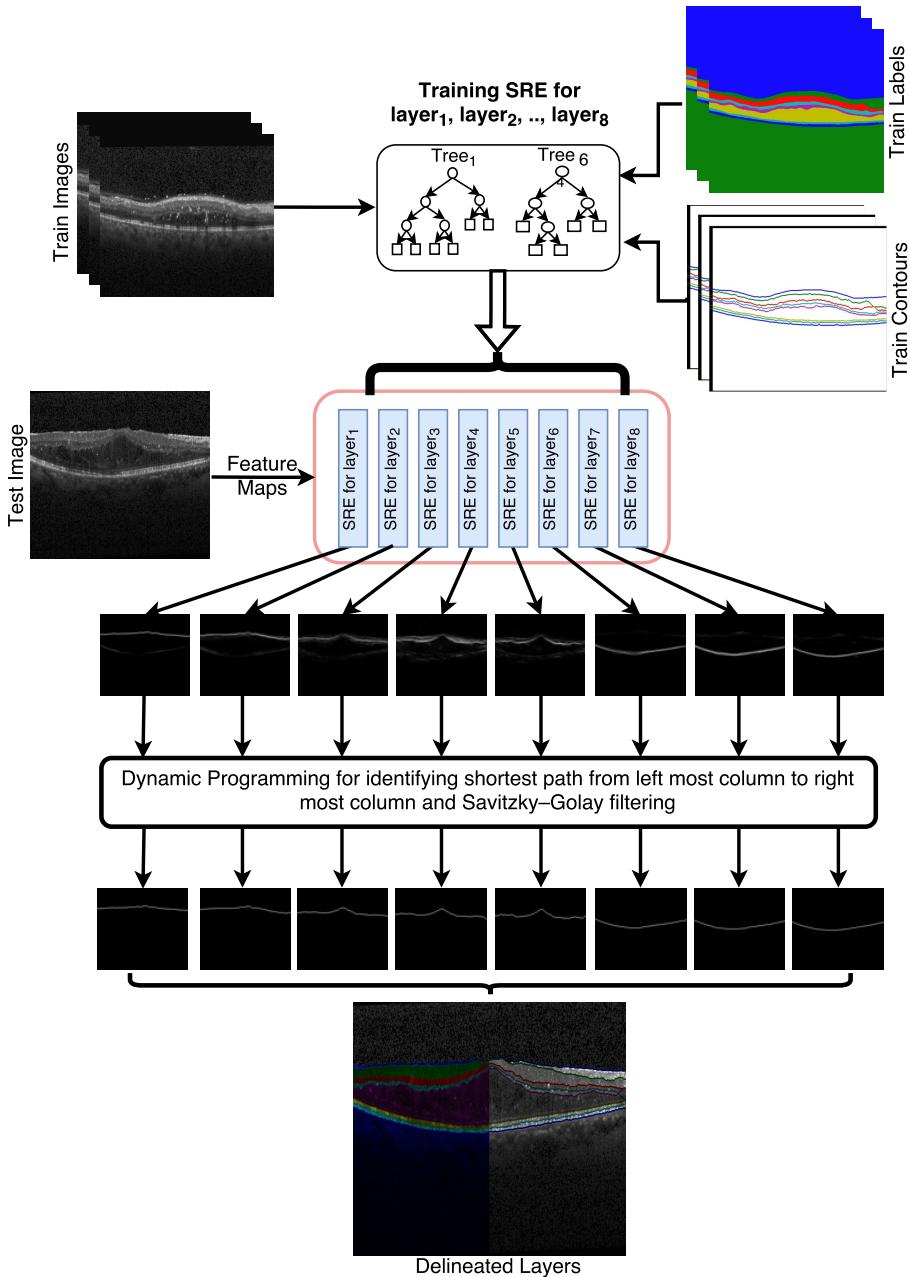


Figure 3.7: Block diagram illustrating image inputs during training and resulting image at each block during testing

3.4. Results and Discussion

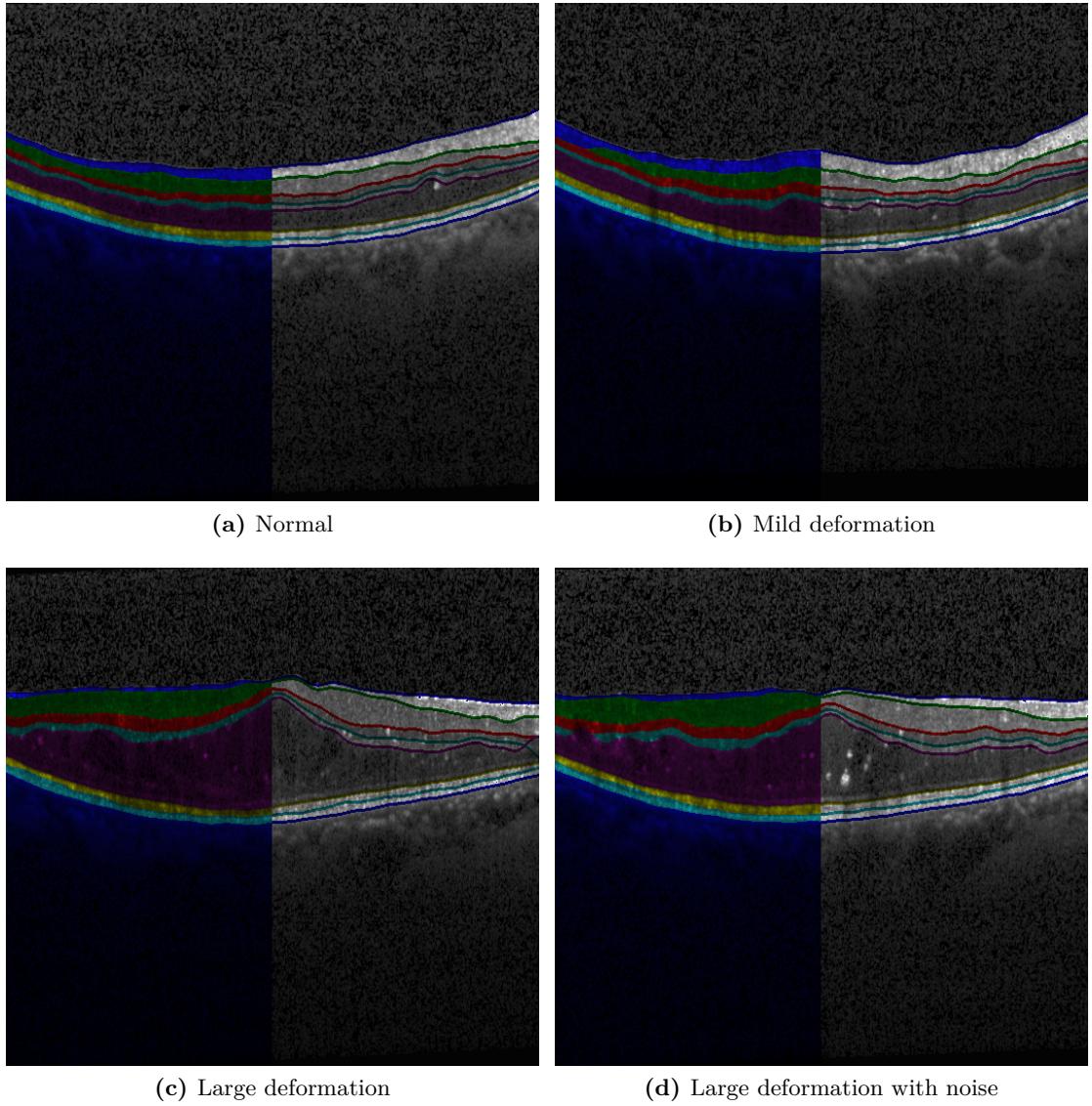


Figure 3.8: Predictions of proposed approach on test images with various deformations

3.4.7 Quantitative performance

Testing was performed on the remaining 55 images. The mean of absolute difference between the predicted contour and the expert contour along the column is employed as a metric for comparing the proposed measure with baselines (table 3.1). This metric illustrates the deviation of predicted contours from expert annotations. This is crucial

3. Layer Specific Edges for Retinal Layer Segmentation

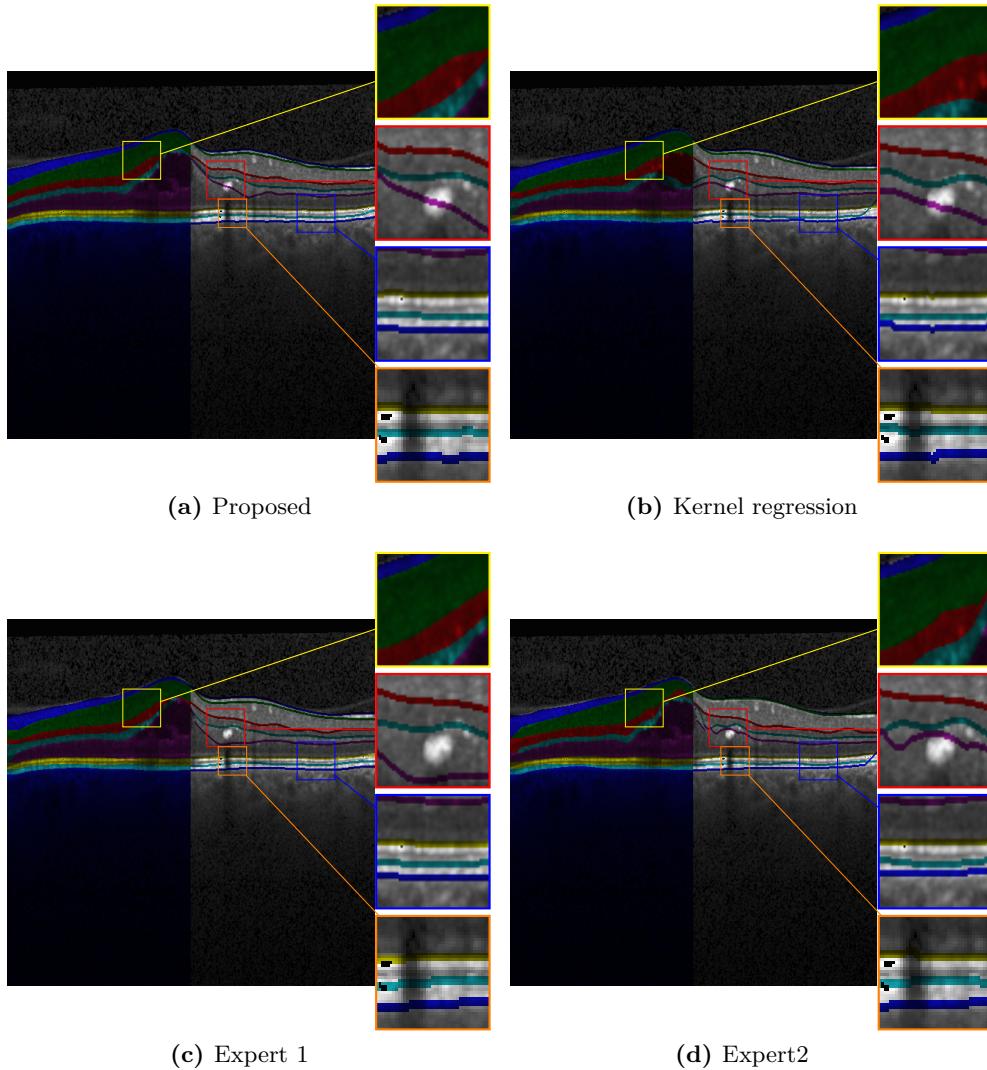


Figure 3.9: Illustration of Layers delineation in the presence of deformation (red), noise (yellow), shadows (orange) and low gradients (blue) by proposed in comparison to other baselines.

3.4. Results and Discussion

Table 3.1: Mean absolute difference evaluation across classical graph approach, kernel guide graph approach and layer specific edge guided graph approach

	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
Layer8	1.191	1.348	1.182	0.863	1.623	1.399	1.112

Table 3.2: F-score evaluation across classical graph approach, kernel guide graph approach and layer specific edge guided graph approach

	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

for computing the retinal width profiles. The comparison is performed by considering two expert annotations independently.

The F-score was employed for quantitative comparison with established baselines table 3.2 on the considered dataset. From image processing perspective F-score is employed to measure the segmentation accuracy, and from statistical evaluation perspective it is a combination of three statistics (true positive, false positives and false negatives). As lower boundary of choroid layer is unavailable so F-score is computed for only seven layers.

3. Layer Specific Edges for Retinal Layer Segmentation

3.4.8 Alternate approach

Construction of z with binary states is identified to be the reason for binary edge predictions and modified z equips the SRE to models all layer edges in a single model. Such approach results in less training time but fails to match the performance as the single model cannot handle complex patterns.

3.5 Summary

To our knowledge, it was the first time that a structured prediction algorithm had been considered for delineating retinal layers. So proposed approach extended the applicability of structured edge prediction for retinal OCT segmentation. This edge prediction models reduced the necessity of the heuristics such as layer 1 upper edge should have a moderate positive gradient, layer 6 upper edge should have a steep positive gradient, layer 8 upper edge should have a steep negative gradient, etc. Proposed approach added a new dimension of information to any edge predicting algorithm. It had been empirically validated that edge predicting algorithms can encapsulate such class (layer) subjective information. It was also validated that such information reinforces graph based approaches. As an end to end framework had been proposed for layer delineation, SRE can be replaced with any existing structured prediction. The proposed approach considers mutual independence between layers in ‘combining contours of all eight layers’ block in the testing phase for simplicity. Upon interest of the programmer, any empirical rule can be imposed through this block. The prediction process estimate edge maps with a low computational complexity so that it can be integrated into the clinical pipeline. It is also illustrated in the following chapter that SRE₈ can be employed for retinal flattening purposes for images acquired on different machines including various pathologies (AMD). This broadens the applicability of proposed approach as an image preprocessing step for retinal OCT analysis.

3.5. Summary

HOG features were very subjective to gradients, and structured predictions were robust to missing gradient to an extent (half the predicted edge patch size), but won't handle large missing edges. Extending the proposed approach to volumetric data can overcome the hurdle to an extent as neighbouring images could provide additional or missing information. Upon arrival of new data, current models should be able to fine-tune the rules in decision node. This can be achieved by adaptive gradient boosted trees capabilities to SRE.

Chapter 4

Guided Fuzzy C Means: A supervised approach

4.1 Introduction

Retinal pathologies are the major cause of irreversible vision loss leading to blindness. Diabetic macular edema (DME) and age-related macular degeneration (AMD) are the highly incident pathologies. Identification of early onset of pathology is crucial for effective therapy. So, computer-aided screening tools have been developed to identify the pathological subjects to reduce the workload on ophthalmologists. There exists a wide range of algorithms, some involve retinal layer profiles prediction (as described in previous chapter chapter 3) and attributed to various pathologies. Alternate approaches have been proposed to reduce the dependency of retinal layer profiles through utilisation of 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 of the classifier, and subsequent chapter deals with the construction of feature descriptor. Availability of computational power and ample data has boosted machine

4. Guided Fuzzy C Means: A supervised approach

learning community to design a vast variety of classifiers. Classifiers dominate pattern recognition community so strongly that employment of classifier is a thumb rule in the case of automated analysis (problem). Structured forests in the previous chapter is an illustration of modelling and predicting patterns in a target space. It is recently identified that understanding the representation of manifolds in feature space is crucial for effective classification. SVM kernels are the classical example where employing linear kernel for classifying non-linear manifolds is like bring a sword to gun fight. Careful crafting of such kernels need an understanding of manifold representation, but image analysis commonly involves more than two features which make human interpretation difficult. This has been overcome through dimensionality reduction techniques for visualisation which projects high dimensional features onto two-dimensional space. This allows 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 such methods to identify the optimal projection with less complexity. Metric learning is another approach to construct distance functions for an appropriate measure of distance between data points [136] or transform the computed features to a new feature space where samples from different classes are linearly separable. Classifiers are equipped to handle such manifolds through including kernel transformations along with decision rule construction. However, it is being computationally expensive as the search space has increased. Ensemble learning is employed to handle complex manifolds through a base classifier.

4.1.1 Need of modelling manifold representation

In image analysis feature quantification transforms the image into a D dimensional feature vector which resembles a point in a D dimensional feature space. It is the classifiers job to shatter the space into pieces through decision rules where each piece represents all or majority samples from a single class. A d dimensional space with r quantization bins

4.1. Introduction

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 1000's to accommodate complete image information. Any amount of existing training data is not sufficient to make global decisions. In practice, global decisions are constructed based on limited data. There can exist multiple decision boundaries, so SVM introduced optimal decision boundary through constructing support vectors which represent the boundaries of individual class. This being the reason in the age of random forests SVM still wins in scenarios where training data is limited. However, support vectors could never represent the manifold effectively because support vectors represent only the boundary condition of the manifolds of each class. This is unaltered in the case of SVM and various classifiers because classifiers aim at separating samples rather than cumulative representing. This makes support vector construction prone to outliers or noise. So a cumulative representation is needed instead of decision representation.

4.1.2 Manifold representation modelling

Conventional learning machines are concerned about constructing complex decision boundaries through a basic model (ensemble learning). As an example, a linear classifier can be used for the non-linear problem through piecewise linear approach. So it is an indirect approach that the algorithms are designed for separating the manifolds from different classes through 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 in comparison to SVM resulting in an improved representation of the manifold. Alternatively, kernel machines are deployed for transforming features for linear separability but separability being a governing factor these approaches do not actually model manifold representing a class.

Graph-based approaches are a backup choice for representing manifolds through exploring local neighbour adjacency. The graph-based approaches have dominated un-

4. Guided Fuzzy C Means: A supervised approach

supervised (spectral clustering) and semi-supervised (random walks). In the case of supervised non-graph algorithms are effective during storing and rapid prediction on arrival of new test data. Voronoi diagram inspired algorithms like K-nearest neighbours were undermined after uprise of statistical learning theory with effective representation of data population statistically.

4.1.3 Identified approach

The idea of Voronoi diagram has indulged algorithms that are computationally efficient due to its non-parametric nature, easy to store as it needs only control points and testing phase involves Euclidean distance measure which can be translated to a dot product. Each subspace generated resembles shattered space pieces by a classifier, and the boundaries between subspaces are classifiers piece-wise linear decision boundaries. To equip the capability of representing complex patterns the subspaces are further divided with new control points (fractals in Voronoi diagram). This resembles fractals in Voronoi used in computer graphics. Such approach is extended to modelling manifold representation through an ensemble of unsupervised modelling. Unsupervised approaches are explicitly designed to understand the patterns in features due to the unavailability of class labels. The approach relies on unsupervised approach but being governed by label information.

4.2 Preamble to Solution

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. Coming to algorithm perspective control points can be considered as mean vectors of K-means clustering (where the label of each centroid is known), and in the test phase, the cluster (subspace in Voronoi) of each test sample is identified based on the topology resembling Voronoi diagram (figure 4.1a). However, it is evident that K-means clustering fails in case of nonlinear patterns because K-means follow Euclidean metric. So an identified cluster

4.3. Method

with samples from multiple classes can be further divided (figure 4.1b) resulting in new clusters with increased resolution and resembles the fractals of the subspace as illustrated in figure 4.1. This boils down the proposed approach to boosting approach with class conditional K-means. Fuzzy C-Means is employed as it has the capability to resolve the ambiguity to an extent. It also acts a node to incorporate expert knowledge into the system to attract wider applicability. During prediction, K-means clustering strategy is considered, i.e., hard boundary keeping the computational complexity in mind. In the case of lack of information replacing membership function with the impulse function makes FCM a classical K-means clustering.

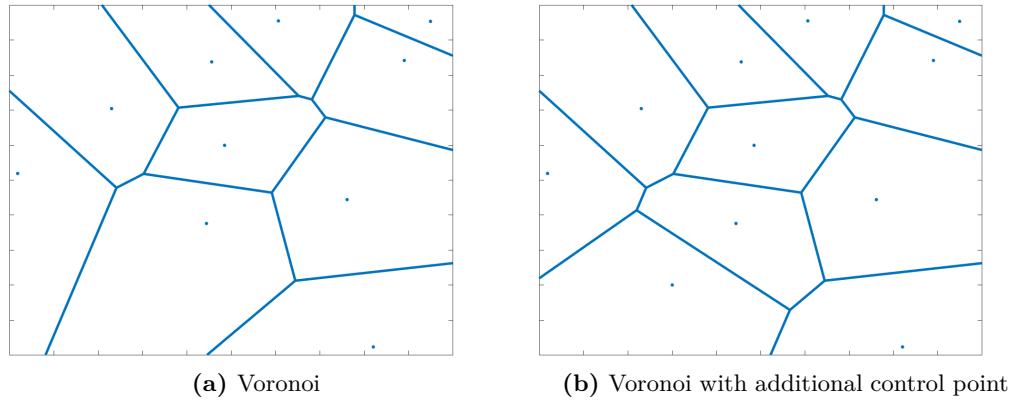


Figure 4.1: Illustrating formation improved resolution of subspace on addition of new control point

4.3 Method

4.3.1 Retinal flattening

Images (figure 4.2a) are processed through various blocks before feature extraction to standardise and enhance feature presence. Retinal flattening figure 4.2 is a common practice in retinal image analysis which requires lower boundary of retina so SRE₈ model from chapter 3 is employed for predicting lower boundary of the retina. The predicted

4. Guided Fuzzy C Means: A supervised approach

boundary is smoothed using a morphological closing operation. Each column in actual image is shifted circularly, so the retina lower boundary (including drusen complex) forms a straight line. Each image is smoothed without loosing the edge information using the BM3D filter. To reduce the complexity, each image is scaled down by half using bicubic interpolation (figure 4.2b).

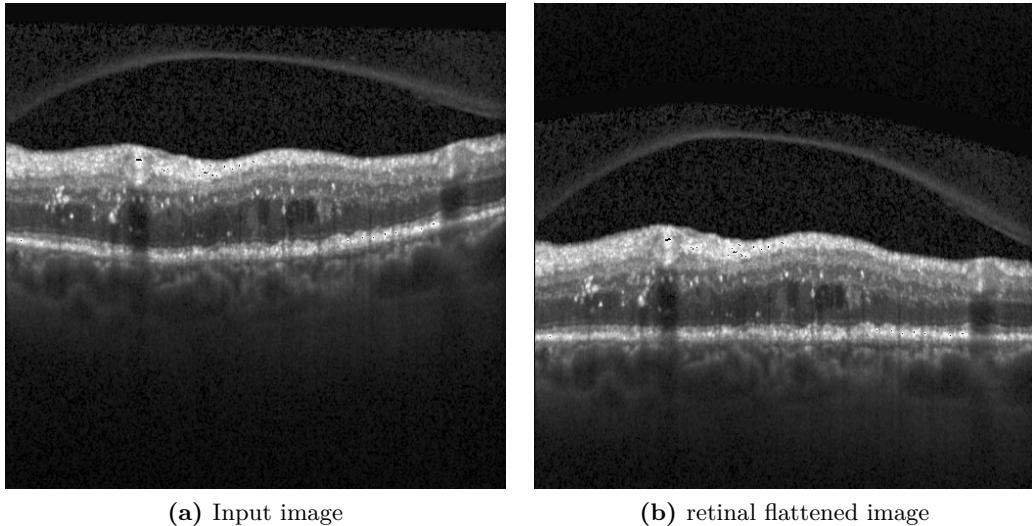


Figure 4.2: Illustrating the data standardization through retinal flattening

4.3.2 Feature quantification

The feature quantification process employed by [76] are used. It involves cropping of the image to dimensions 45×150 such that retina lower boundary is five rows from the bottom of the image and central column stays centre. HOG features are acquired on 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 feature vector of dimensionality 15K. Each feature vector corresponding to each image is concatenated to form a data matrix.

The reason for the choice of HOG features over data driven feature quantifiers is ‘A less effective feature descriptor results in complex manifolds in feature space and burdens

4.3. Method

the classifier which illustrates the true potentiality of the classifier’.

4.3.3 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). Proposed approach constructs k number of means-sets where each set consists centroids (control points) of the corresponding class. Framework initialization involves partitioning of data matrix into k number of datasets (where the second set have samples from second class), and k null sets are considered as initialisation of means-sets.

A j^{th} dataset is divided into two clusters with Fuzzy C-means clustering (FCM [137]) and identified means are concatenated to j^{th} means-set. Upon repetition of the above step for all k classes, 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 corresponding means-set index (class) is treated as estimated class. Wrongly classified samples are identified and partition them into k datasets (D_1, \dots, D_k) based on corresponding true labels such that samples from each set belong to same class. The process is repeated until any partition has no more than two samples. For identification of wrongly classified samples, classic K-means strategy is employed as hard decision approach is considered. Interested work might involve computation of membership along with the decision to include the importance of the sample.

An algorithmic flow of proposed approach is depicted below algorithm 1, which takes a data matrix X and label vector l with k classes as input and return 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 is a standard fuzzy C-means clustering.
- Pairwise distance (pDist) is the standard function in Matlab with the Euclidean measure.

4. Guided Fuzzy C Means: A supervised approach

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

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        $\{\mu_1, \mu_2, \dots, \mu_c\} \leftarrow \text{FCM}(D_i, \text{clusters}=c)$ 
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     $dist_i \leftarrow \text{min\_means}(pDist(X, M_i))$ 
12  end
13   $\hat{l} \leftarrow \text{min\_Idx}(dist_1, \dots, dist_k)$ 
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 identification of control points

During *prediction* given a test sample each mean-set is considered, and Euclidean distance is computed and to reduce the metadata only the least distance is stored along with the label of the mean-set. This results in k number of distances whose inversion reflects the possibility of given test sample falling into each class. So the index of minimum

4.3. Method

distance is retrieved and associated as predicted class for test samples. Considering a standard FCM the equivalence between index of maximum membership and minimum distance for a data sample is given in equation (4.1).

$$\begin{aligned}
 U_j &= \frac{1}{\sum_{i=1}^k \left(\frac{|X_s - \mu_j|}{|X_s - \mu_i|} \right)^{\frac{2}{m-1}}} \\
 K_j &= \sum_{i=1}^k \left(\frac{|X_s - \mu_j|}{|X_s - \mu_i|} \right)^{\frac{2}{m-1}} \\
 K_j &= \sum_{i=1}^k \left(\frac{|X_s - \mu_j|^{\frac{2}{m-1}}}{|X_s - \mu_i|^{\frac{2}{m-1}}} \right) \\
 K_j &= |X_s - \mu_j|^{\frac{2}{m-1}} \sum_{i=1}^k \left(\frac{1}{|X_s - \mu_i|^{\frac{2}{m-1}}} \right)
 \end{aligned} \tag{4.1}$$

(Lemma : second part is constant for each sample)

$$\begin{aligned}
 K_j &\propto |X_s - \mu_j|^{\frac{2}{m-1}} \\
 \left\langle \text{Lemma : if } a < b \quad \& \quad m > 1 \quad \text{then} \quad a^{\frac{1}{m-1}} < b^{\frac{1}{m-1}} \right\rangle \\
 K_j &\propto |X_s - \mu_j|^2
 \end{aligned}$$

So index of maximum U_j is same as minimum K_j where U , K , X_s and j are membership values of a sample, distance values of a sample, a data sample and centroid index respectively.

4.3.4 Spiral example

Two spiral toy set example is commonly considered to validate the modelling capability of any proposed approach. A two class spiral example (1000 samples) is considered where diamonds represent class ‘1’ and circles represents class‘2’. Proposed algorithm with four clusters has is considered. Along each iteration, means-set vectors of class ‘1’ and ‘2’ are represented by the red asterisk and blue squares. At the end of each iteration (i) all samples are considered as the test set and predictions ‘1’ and ‘2’ are

4. Guided Fuzzy C Means: A supervised approach

illustrated as red and blue in figure 4.3. So a red colour diamond and blue colour circle are correct predictions remaining combinations are wrong predictions. It is evident that the proposed approach is capable of modelling complex structures. It also illustrates self-correcting nature between iteration 1 and 2. A red asterisk control point at 0.1 along X-axis and 1.1 along axis correctly classifies diamonds but wrongly classifies circles, so a blue rectangle has been automatically introduced in the second iteration to counter it at 0.1 along X-axis and 1 along Y-axis. The same trend can be observed in iteration 2 with the red asterisk at approximately -0.4 along X-axis and 0.8 along Y-axis.

4.3.5 Regulating parameters

There are two regulating factors for proposed algorithm which are membership function and resolution of finest subspace after clustering. *membership function:* FCM clustering is the generalized version of K-means clustering by considering 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 (spiral). The higher the clusters more refined the subspaces after clustering but increases computational time and number of control points to be stored. This also affects the testing time.

A common question that pops up is "Why 2 clusters? Why not just means per class, i.e., 1 cluster?" and to answer that 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, 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 stay as previous. So the algorithm will be stuck in a void.

4.3. Method

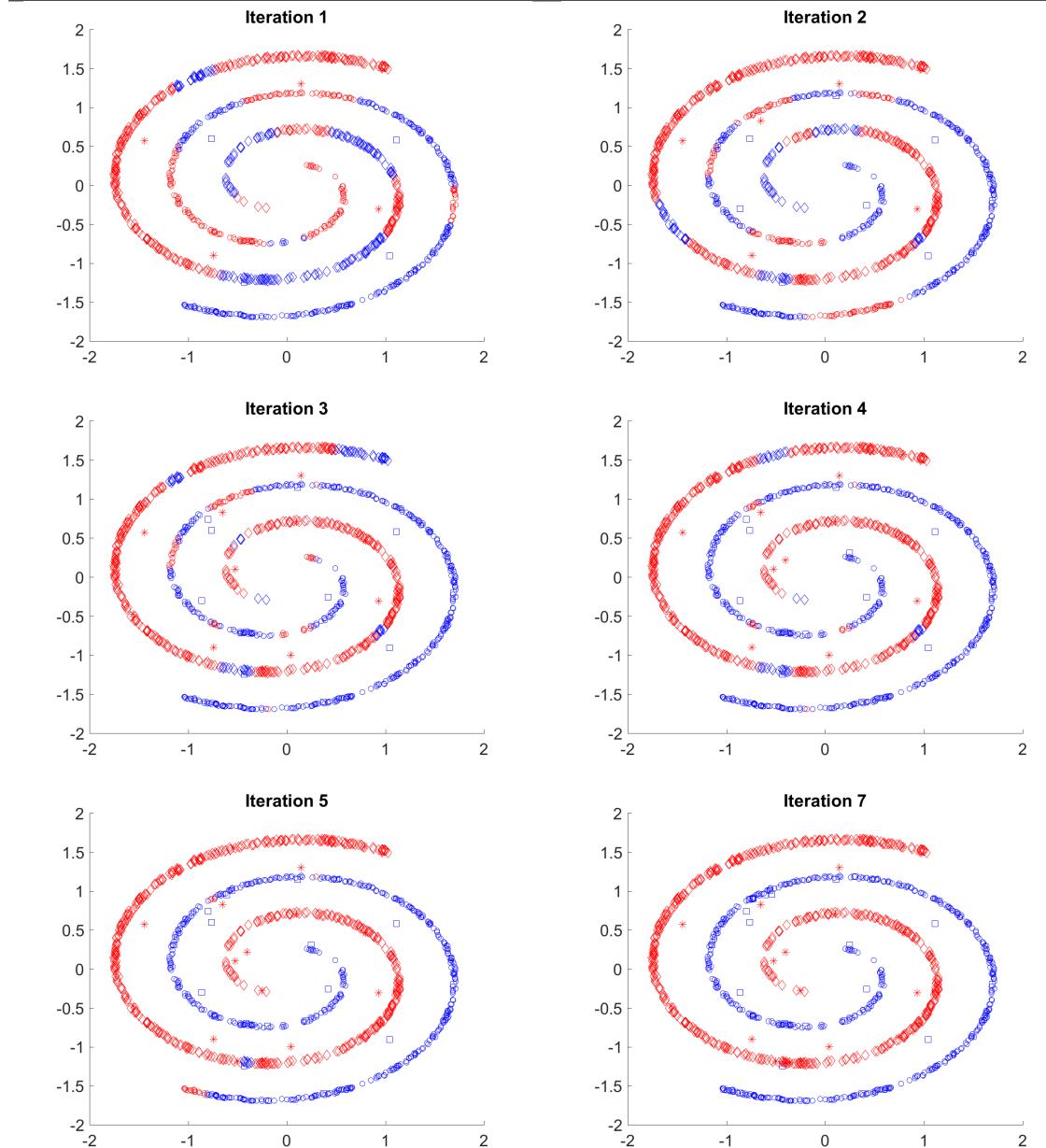


Figure 4.3: Illustrating the capability of delineating complex structureds

4.4 Results and Discussion

The difference between existing classifiers and proposed is the former aims to identify the boundary but the proposed models the representation of each class independently.

4.4.1 Experimental bed

A Duke OCT classification dataset with normal, Diabetic macular edema (DME) and Age-related macular degeneration (AMD) subjects is considered for experimentation. It consists of 15 subjects per class totalling to 45 subjects with 3231 images. A 15 fold validation is performed where each fold in train set have eight subjects from each class and remaining subjects are considered as the test set. So, the first fold has subjects 1-8 as train set and subjects 9-15 as the test set. The fifteenth fold has subject 1-7 including subject 15 (following circular indexing) as train set and subjects 8-14 as the test set. Ambiguous images in each train set are removed, and test set images are retained. The random forest classifier and SVM classifier with various kernels are considered as baselines to evaluate proposed approach. MatLab 2015b (for HOG features) is considered for experimentation. The Piotr's image (and video) toolbox and the edges toolbox and Mahdi Amiri's fuzzy toolbox are used. Along with the retinal classification problem, a toy example is considered to illustrated the potentiality in various scenarios

4.4.2 Toy example

A simple toy example (16 samples) is considered from which two scenarios have been explored to illustrate the characteristics of the algorithm and its boundaries. Eight samples are considered for training, and remaining samples are considered as the test set. The toy example comprises three groups of points a, b and c in figure 4.4a. The scenarios involve:

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

4.4. Results and Discussion

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

The first scenario simulates class imbalance condition and the second scenario simulates non-linearly separable scenario. After training proposed approach (2 clusters) with training data it generates two centroids (control points) per class illustrated as the red asterisk for class ‘1’ and blue square for class ‘2’. The feature space is divided based on these centroids depicting blue for class ‘1’ and yellow for class ‘2’ (using the same colour for centroids affects the contrast). The contours reflect boundary conditions of different confidence levels 4.4. The black line(s) is (are) the hard decision boundary. The predictions of test dataset for the scenarios are illustrated in figure 4.4b and figure 4.4c. The shapes (circles and diamonds) represent ground truth classes, and colours represent (blue and red) predicted classes of the test set. So blue circles and red diamonds are the correct predictions, and any other combinations are false predictions. The figure 4.4b and figure 4.4c also illustrates the algorithm is capable of constructing linear and piecewise linear (nonlinear) boundaries simultaneously without the need of kernel knowledge.

4.4.3 Algorithm modelling characteristics and prediction for retinal pathologies

Following the training pipeline proposed by [76] each image in a trainset is translated to a feature as defined above and the subject level label is considered as image level label (weakly supervised). Concatenation of feature vectors and labels (scalar) for all images in trainset result in a data matrix and label vector. The classification training characteristics (three centroids) for first fold along each *iter* is illustrated in figure 4.5. The feature space being few thousands each samples is projected on to 2D space through t-SNE [138]. Following the notations as before the shapes are true classes and colors are predicted classes (to be noted that ambiguous images are removed during training). Diamonds, circles and plus are true first, second and third classes respectively. Red, blue and Green

4. Guided Fuzzy C Means: A supervised approach

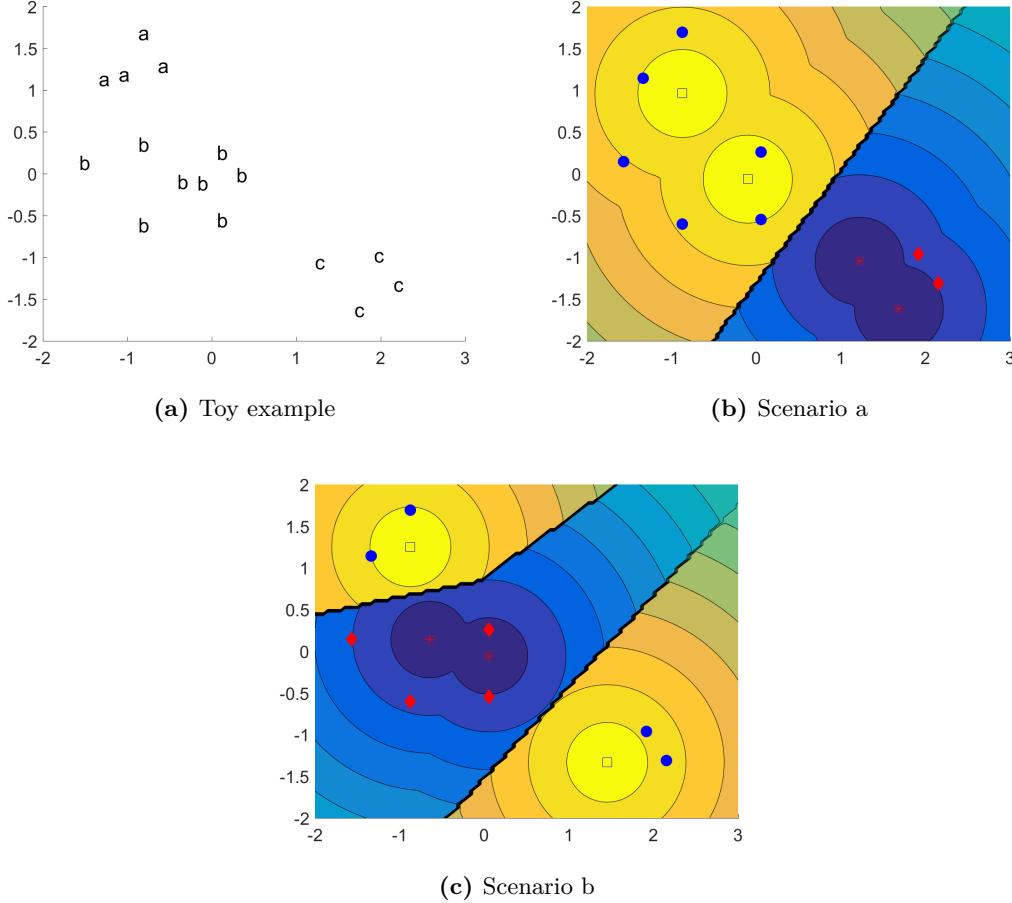


Figure 4.4: Illustrating toy example and performance of proposed approach on two scenarios

are predicted first, second and third classes respectively. Even though the first iteration has managed to classify majority of the samples there exists misclassification between samples from different classes. The same has been illustrated in magnified versions at each iteration. The right bottom image in figure 4.5 illustrates the convergence of the system.

The data standardising pipeline for trainset has been extended to test dataset and constructed test data matrix and test labels. During testing the class of closest mean vector (control point) is allotted as the prediction for the test sample. The image level

4.4. Results and Discussion

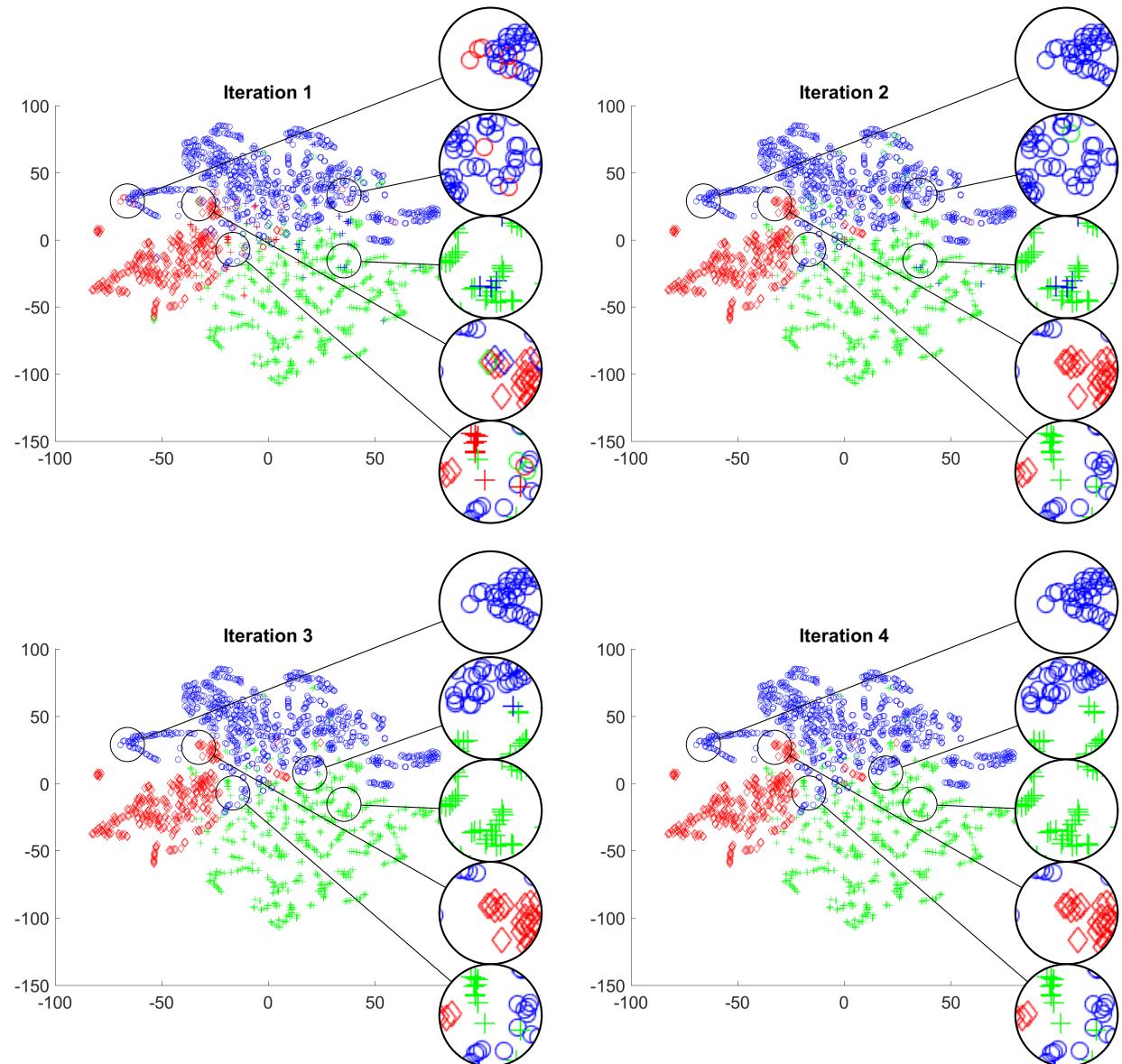


Figure 4.5: Illustrating the modeling capability of proposed approach for separating retinal pathologies along each iteration

4. Guided Fuzzy C Means: A supervised approach

Table 4.1: Decision pooling performance across proposed and baseline algorithms

		Normal	AMD	DME
Proposed	nc=2, m=.2	0.9905	0.8476	0.9333
	nc=3, m=1.2	1.0000	0.8667	0.9143
	nc=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

prediction accuracy of proposed approach, linear SVM, SVM with Gaussian kernel and SVM with RBF kernels are 0.868, 0.855, 0.452 and 0.452.

4.4.4 K-fold validation performance

The protocol proposed by [76] is employed for test accuracy, and decision pooling accuracy is commonly employed by the dataset for comparison. Every image in a pathological subject need not have a pathological state, so majority voting (decision pooling) of all image predictions in a subject is considered as the subject level prediction. The average accuracy of decision pooling for test sets across 15 fold validation is depicted in Table 5.1. To illustrate the impact of governing parameters over modelling the performance of proposed approach with different parameter combinations has been illustrated. Random forests (different number of trees) and SVM with different kernels are considered as baselines to illustrate the impact. It is to be noted structured SVM or SRF are not employed as they are meant for structured predictions.

4.5. Summary

4.5 Summary

Classification through proposed approach is capable of creating subspaces through piecewise wise linear (simultaneously) boundaries. The approach has utilised class subjective manifold representations and optimised the boundaries between classes through considering labels as auxiliary cost. Whereas conventional classifiers aim at a global separation of the manifolds, i.e., decision boundaries computation. The decision tree involves identification of decision rules based on entropy and SVM involves identification of hyperplane with maximum separability based on support vectors (boundary representation of all samples). One does not consider the representation of manifold and other considers the boundary sample representing the boundary but not the manifold. As illustrated by various empirical experimentations SVM is chosen for small training dataset, and the random forest is for large training dataset (as sampling statistics aligns with population statistics).

Proposed approach makes the system free from knowledge regarding manifolds structure or spaces and if required expert information can be induced through m . This has an impact on accuracies in current application as the selection of wrong kernel results in plunged accuracies. There is a basic difference between human intervention through kernels in SVM and through m in proposed approach. In the case of SVM, the expert should have knowledge of kernel to be selected. In case of the proposed approach if the expert does not know the m FCM can be replaced with K-means clustering. The control point (centroids) strategy employed has resilience to noise as proposed approach has the capability of identifying a separate control point for the outlier. Fractal Voronoi has the clear advantage through employing global to local modelling rather simultaneous modelling of global and local representations which is computationally expensive. The strategy also makes the approach free of decision boundary computations as bounds with respect to centroids on L2 space effectively act as decision boundaries. Collectively from a naive perspective the if feature space is considered to be a stable water surface then

4. Guided Fuzzy C Means: A supervised approach

centroids are an introduction of point force at different locations. The ripples represent the contours and colour representing various sources of forces. The intersection of ripples from various sources is treated as constructed decision boundary.

Chapter 5

Construction and Visualization of Data Driven Feature Quantifiers

5.1 Introduction

This chapter aims at the same application as previous chapter, i.e., classification of normal, AMD and DME subjects. However, achieves it through contributing to feature quantification block in image analysis pipeline. Classification is dependent on feature computed by feature quantifiers and any information loss during quantification affects the classifier performance. As an example, HOG feature descriptor has boosted the pedestrian detection solutions even with a standard decades old classifier [139]. The HOG features are widely employed for retinal classification as they are capable of representing geographical atrophy in terms of gradient orientations as illustrated in figure 5.1.

An effective feature quantifier can empower even a naive classifier to delineate the pathologies so in this respect a softmax classifier is considered.

5. Construction and Visualization of Data Driven Feature Quantifiers

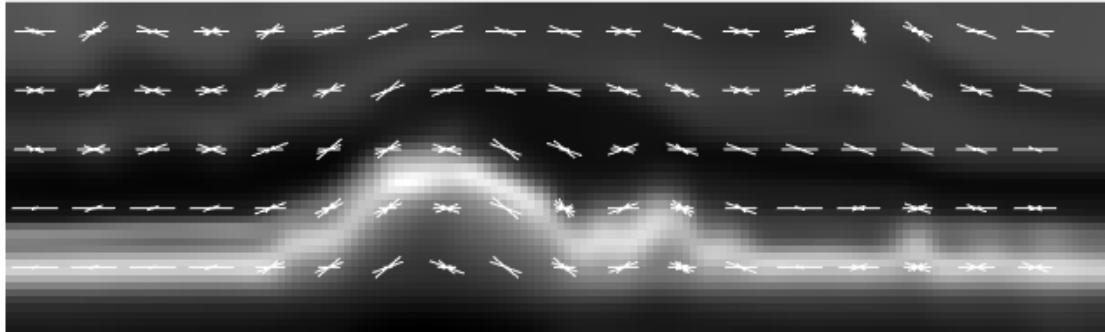


Figure 5.1: Illustrating the variations of gradient orientations along the geographical atrophy depicted by HOG descriptor

5.1.1 Need for learning representations as feature quantifiers

Conventional practice involves identification of representative features delineating target classes and formulate a descriptor which quantifies subjective features. It is always difficult to construct a mathematical expression representing anticipated feature, so multiple mathematical basis quantifying various partial aspects of the anticipated feature is considered. In simple senses, an orange can be quantified as round (eccentricity is one), medium size (radius is 3 cm), bright red colour, medium green colour, no blue colour, etc. Common hurdles with this approach are representative features for low-frequency patterns is difficult to attain, irrespective to pool of features considered some residual information is always held back upon consideration of new pathology the whole process is reiterated, etc. Such burden is reduced by employing existing salient feature descriptors for various applications and attained features are refined through feature subset selection (filter, wrapper or hybrid approaches) or feature extraction (principle component analysis, autoencoder, etc.) approaches to quantify anticipated features. Current algorithms in feature extraction process are capable of constructing anticipated features from raw information thus reducing the dependency on salient features.

5.1. Introduction

5.1.2 Automated representation learning

Feature extraction process has been an integral part of image analysis and transcended from choice to need in the analysis pipeline. Classical techniques involve construction of a transfer function to project computed salient features onto a feature space where samples from different classes are easily distinguishable by a classifier. There are both supervised and unsupervised sections of feature extraction so the proposed approach has taken the former approach due to its effectiveness. The approaches in past decade has illustrated the potentiality of considering raw information as input and constructing feature descriptor for computing the feature quantification and extraction in one transfer function. The existence of such transfer function is subjective to the anticipated features. Even if it exists, identification of such complex transfer function is computationally expensive. To overcome such hurdles, a hierarchical approach is employed where it is no longer a single transfer function but a sequence of transfer functions. So given the raw information, multi-stage approaches sequentially process the raw information through a set of transfer functions and such process is functionally equivalent to anticipated single transfer function. Convolutional neural networks (CNN), a common architecture employed in deep learning is one such approach for image analysis. From signal processing perspective convolving information with filters in a layer of CNN is equivalent to imposing a transfer function. CNN is like a Lego where multiple architectures can be generated with four types of building blocks, and not all combinations can construct anticipated transfer functions set, so the selection of architecture is crucial. GoogLeNet [25] architecture is considered as it established benchmark performance and it has less number of filter weights which makes it less prone to overfitting.

5.1.3 GoogleNet

The CNN architectures [21, 25, 140] for image classification applications involve combinations of the convolution block with learnable filters, activation block, cross-map

5. Construction and Visualization of Data Driven Feature Quantifiers

local response normalization (LRN) block, pooling block, fully connected (traditional neural network) block with learnable weights, dropout, and loss block as illustrated in figure 5.2. Given an image or the responses, the convolution block convolves with a set of filters (learnable), 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) of neighboring elements (image dilation or mean filtering), the fully connected layer is a traditional artificial neural network, and the loss block computes the error between the predicted class and actual class. The error computed in the loss layer is utilised to correct the filter weights of each convolution block with gradients computed through error back-propagation.

5.1.4 Visualizing learned representations

Interpreting learned filters has been a common trend in image analysis to validate if anticipated representations are modelled and understanding the characteristics of information processing pipeline. There are a wide variety of approaches depicting the feature representations from first layer filters such as hand picking the important filters, illustrating the saliency points in the input image, etc. There also interest exhibited for interpreting information from hierarchical layers as that has been a major drive for trusting deep learning. Such information always will not be as elegant as facenet (detecting human faces) but reflect distinguishable abstract patterns for different classes.

5.1.5 Identified approach

Learning representations from raw information through CNN needs ample amount of data to identify repetitive patterns or potential patterns. Transfer learning approach is employed as the crunch in image data is common in medical domain particularly retinal OCT images. Such approach needs a pretrained model, so GoogLeNet trained

5.1. Introduction

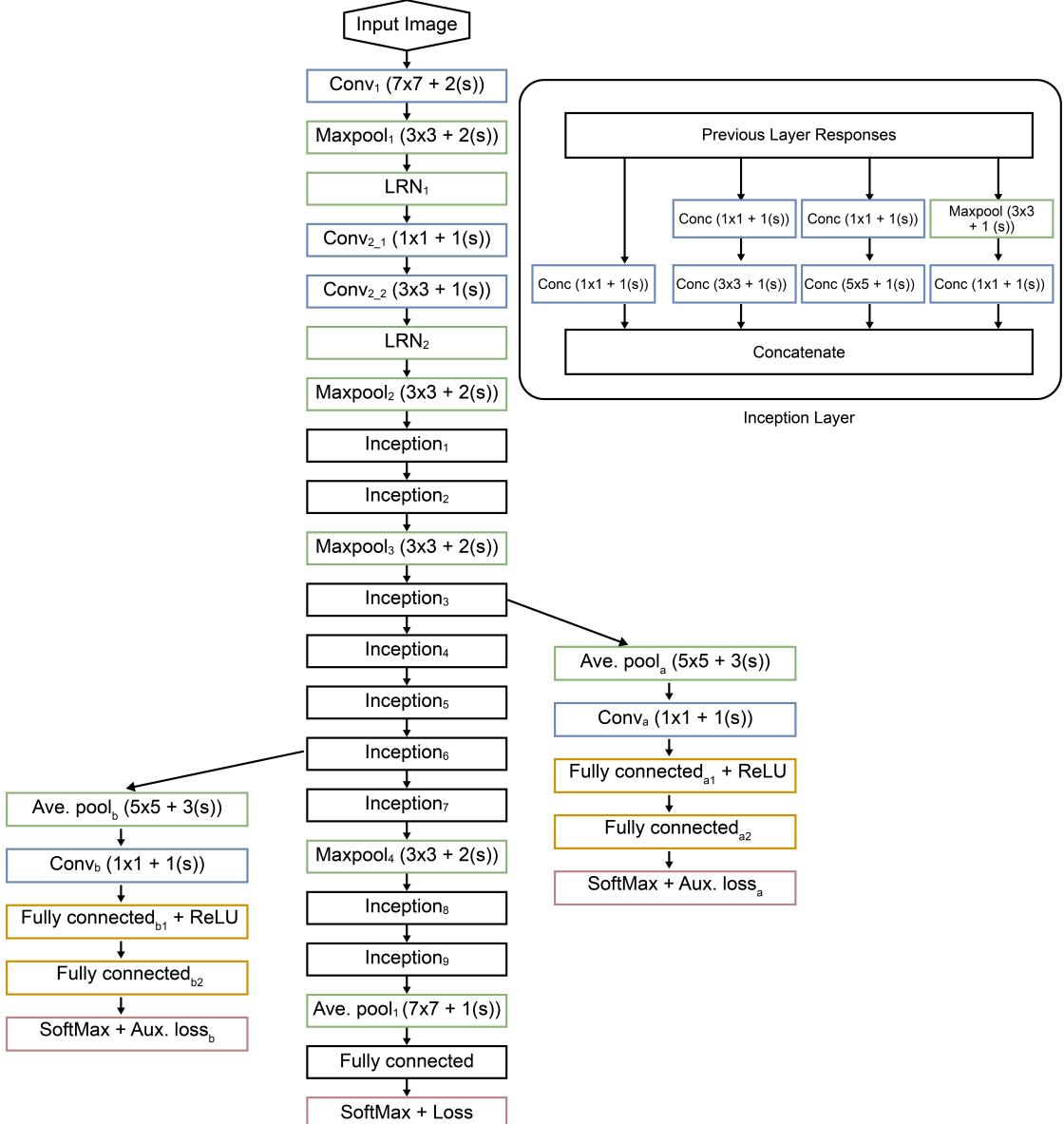


Figure 5.2: Classical GoogLeNet with illustration of employed inception block

on large image dataset for classification (ImageNet) has been employed. The reason for choosing GoogLeNet being less number of parameters or weights in comparison to other architectures. Overfitting is a common problem for transfer learning CNN architectures, particularly in the case of less training data. A common way around is choosing a model with less number of parameters. So GoogLeNet is employed as it has less number of

5. Construction and Visualization of Data Driven Feature Quantifiers

parameters in comparison to modern deep CNN architectures. It is established that transfer learning has improved performance to random initialization but need not be better to a classical approach, so HOG features are used for comparison.

Visualising representations is two-fold: Firstly, identification of potential filter in the first layer autonomously and interpreting its capability to describe salient feature and secondly gross interpretation of final layer representations to visually validate the difference in abstract patterns from different classes.

5.2 Preamble to Solution

Transfer learning for learning representations for a specific application is common practice in various domains including medical domain particularly in microscopic or CT images with underlying distribution as Gaussian. However, retinal OCT speckle distribution is double exponential, and image distribution is dependent on the adaptive compression techniques. The current approach explores the expandability of GoogLeNet to retinal OCT images and infused modifications to handle low probabilities at the final layer. It also deviates from conventional training strategies regarding data standardisation and optimisation for improved performance. The retinal flattening and preprocessing are similar to the previous chapter without the heuristically identified region of interest, i.e., cropping of the flattened retinal image with reference to the lower boundary of the retina. The multi-scale image is not considered as input as GoogLeNet inherently computes multi-scale in the inception block through convolving with filters of multiple receptive fields (filter size).

The first layer of GoogLenet has 64 filters, and identification of potential filter autonomously is crucial. The proposed approach is gradient based and is inspired from neural paint [141] as it corrects the filters which are contributing to the style. The initial filters depict feature descriptors similar to low-level information processing and act as a foundation for successive filters. The final responses also need to be depicted as they

5.3. Method

are the sample points in feature space and need to be illustrating interclass variance. The top layer feature descriptors aim for sparse representation, so gross representation is illustrated by the statistical combination (max operand) of all responses.

5.3 Method

5.3.1 Data preparation

In every image saturated regions (intensity value 255) are replaced with 10 and normalised. Retinal flattening is performed such that lower contour is shifted to 70% of the height. Each image is resized to 224×224 as employed GoogLeNet is trained on images with such dimensions. Images are filtered with BM3D filtering as OCT is constructed from speckle information. A three channel image is constructed by replicating the BM3D output as considered GoogLeNet is trained on colour images and each replicated output is treated as channel information. Following the strategy of [76] label for each image is allocated. Around 53% images are considered as trainset and remaining as a test set without the overlap of the subjects. Upon the division of dataset into train set and test set, ‘Mean image’ ($3 \times 224 \times 224$) of trainset is computed and subtracted from each trainset image and test set image.

5.3.2 Modifications to trained GoogLeNet on ImageNet

- Fully connected1 layer, fully connected layera2, fully connected layerb2 have three outputs
- All fully connected layers and convolutional layers in auxiliary branches weights are randomly initialized to learn feature space shifts between ImageNet data and OCT data
- All last layers have been changed from softmax to log softmax as the softmax probabilities are low and it avoids underflow of gradients.

5. Construction and Visualization of Data Driven Feature Quantifiers

- The new GoogLeNet is trained with Adagrad [142] optimizer rather to Adam or SGD (stochastic gradient descent) as Adam results in impulsive step lengths and SGD need expert introduced learning rates at different epochs.

5.3.3 Fine tuning

Fine tuning a pre-trained CNN with a mean subtracted train set and simultaneously predicting the accuracy on mean subtracted test set at each epoch is considered as an experiment.

Training phase involves breaking mean subtracted training set into batches of ≤ 64 images. On a forward pass of each batch, CNN predicts 64 labels from which loss is computed through negative log-likelihood. The loss is translated to gradients at each layer through chain rule in error back propagation which corrects the filter weights. The amount of correction to be incurred is determined by Adagrad optimizer. An iteration involves selecting a subsequent batch, forward pass, backward pass and weight correction. For current setup, training set is broken to 28 batches. A collection of 28 iterations is treated as an epoch and training is performed for 50 epochs, i.e., complete training data has been pushed though the CNN for 50 times.

In *Testing phase*, at each epoch mean subtracted test set is split into batches of ≤ 65 images. A forward pass of CNN predicts labels for each batch and concatenation of all such labels are considered as predicted labels for the test set. After completion of 50 epochs forward pass of each image though a CNN predicts corresponding label and mode (majority voting) of all image predictions from a subject is treated as subject level decision [76].

5.3.4 Identification of potential response

The first layer of finetuned CNN has numerous filters resulting in a large number of responses. In few cases for modifying the CNN architecture expert interpret the filters

5.4. Results and Discussion

and responses to identify repetitive filters, missing patterns, etc. Such response visualisation also helps a non-expert to understand the information processing at each layer. Conventionally, representative responses are hand picked. The current approach is automated and more inspired from the basic definition of pathology, i.e., deviation from being normal. So in test phase given an image and if fine tuned GoogLeNet predicts it as abnormal (AMD or DME) then considering a pseudo ground truth as normal (irrespective of being AMD or DME) to backpropagate the error. Now the filters which are more responsible for predicting the image label are expected to correct more through gradients. In other terms 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.

5.3.5 Visualization of abstract patterns

Top layer response patterns are abstract and sparse in nature, so all such information need to be accumulated for visualisation. 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 visualisation of abstract patterns. All responses after 20th layer are concatenated along the third axis and maximum operation along third axis results in a 2D map. That 2D map is considered as visualisation of abstract patterns.

5.4 Results and Discussion

5.4.1 Experimental setup

The Duke OCT classification dataset is considered. Torch API has been employed for training and testing of the CNN architectures along with the dependencies listed in chapter 3. A workstation with 12 GB GPU is required for experimentation as CNN architectures can be trained rapidly on GPU rather a CPU.

5. Construction and Visualization of Data Driven Feature Quantifiers

The train set is a collection of first eight subjects images from each class, i.e., AMD, DME, and normal (24 subjects in total) and corresponding labels are considered. Train set images are processed to generate mean subtracted images. The test set is remaining seven subjects images from each class, and corresponding labels are considered to compute accuracy. Trainset images are processed to generate mean subtracted images and divided into batches. A GoogLeNet trained on ImageNet dataset is fine-tuned on the trainset. Test accuracy over complete test set is computed after each epoch.

5.4.2 Repeatability

Fine-tuning GoogLeNet on a defined trainset for 50 epochs and computing prediction accuracy on a defined test set (training set and testing set have no overlap) at every epoch is treated as an experiment. The prediction (test) accuracy computation involves GoogLeNet predicting labels of all test set images and ratio between the number of correctly predicted labels and the total number of test images. For an experiment, CNN convergence and predictions are subjective to initialization. Random initialisation in fully connected layers results in the change of performance from experiment to experiment. Similarly, the dropout block also randomly nullifies (setting to zero) the percentage of responses. This also results in the changing of training model performance from experiment to experiment. Such randomness introduced during model fine-tuning produces variable results for each experimentation; thus, in practice, multiple (10) experiments are repeated, and the best model (96%) is stored. The test accuracy of each experiment for every epoch has been illustrated in figure 5.3.

5.4.3 Image level performance

Keeping the repeatability in mind, the model with 94% accuracy is considered for evaluations. The impact of the transfer learning is illustrated by comparison with GoogLeNet trained with random initialization at all filters. The plots in figure 5.4 present the predic-

5.4. Results and Discussion

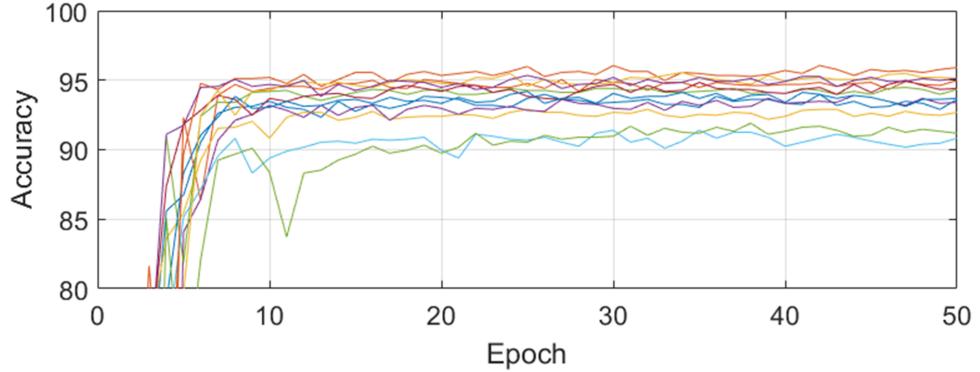


Figure 5.3: Illustrating the repeatability of the experiment

tion accuracy on the complete test set (remaining seven subjects) images and illustrates faster convergence with improved accuracy for transfer-learning approach.

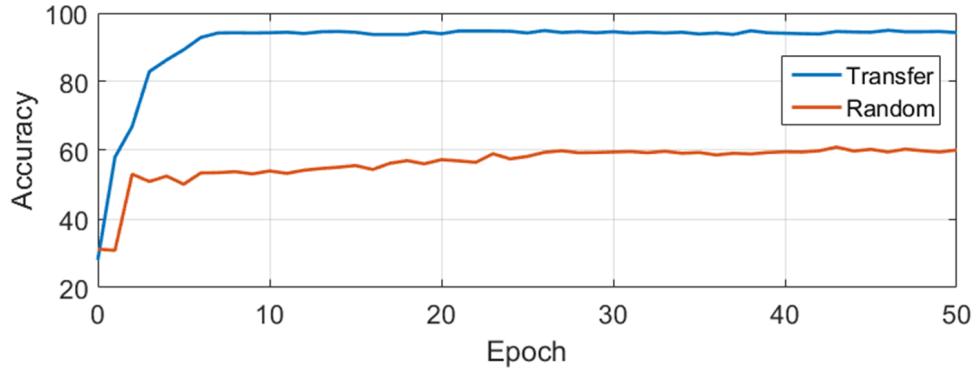


Figure 5.4: Illustrating the faster convergence and higher accuracy of transfer learned model

5.4.4 Visualization

Given a $3 \times 224 \times 224$ OCT image with AMD, the potential response at first layer is identified autonomously and are shown in figure 5.5. The Conv1 layer identifies the filter that quantizes curves, which incidentally happens to identify the deformations in the RPE layer along with other subtle information. Given a $3 \times 224 \times 224$ OCT image with DME, the potential response at first layer is identified autonomously and are shown in figure 5.5. The Conv1 potential filter identifies the deformations. The normal case

5. Construction and Visualization of Data Driven Feature Quantifiers

has squashed responses due to lack of deformations subjective to pathology. For better contrast each response is normalised to 0-1 and associated to psuedo colours where zero being blue and one being red.

The gross response of 20^{th} layer in GoogLeNet in the case of normal, AMD and DME are illustrated in figure 5.6. Unlike the first layer the final layers abstract patterns are clearly distinct in the case of various classes.

5.4.5 Baseline

The images in trainset are processed to generate BM3D filtered image. The process in [76] is employed for HOG feature computation without the cropping. A softmax classifier is employed for classification. This acts as a baseline to illustrate the predictive capability of constructed feature quantifiers from fine-tunes GoogLeNet. So grossly it is multi-scale HOG features vs. fine-tuned GoogLeNet filters.

5.4.6 Crossvalidation

Following the 15 fold validation as illustrated above a train set and test set for each fold are constructed. The average accuracy of decision pooling across all folds for proposed approach for normal, AMD and DME are 0.99, 0.89 and 0.86. The average accuracy of decision pooling across all folds for the baseline is 1, 0.9 and 0.78. In comparison to carefully crafted feature descriptors (HOG) the proposed approach is comparable in the case of normal and AMD but has a clear edge in case of DME.

The above mentioned performance is best performance of the CNN. This achieved by tweaking around the governing factors of the training process like class weights and loss weights. Class weights is helpful to compensate the unbalanced data, ie., imbalance in number of data points per class. As GoogLeNet has multiple losses (main and auxiliary loss), loss weights regulates the contribution of each loss. The variation in performance with various combination of class and loss weights are illustrated in table 5.1.

5.4. Results and Discussion

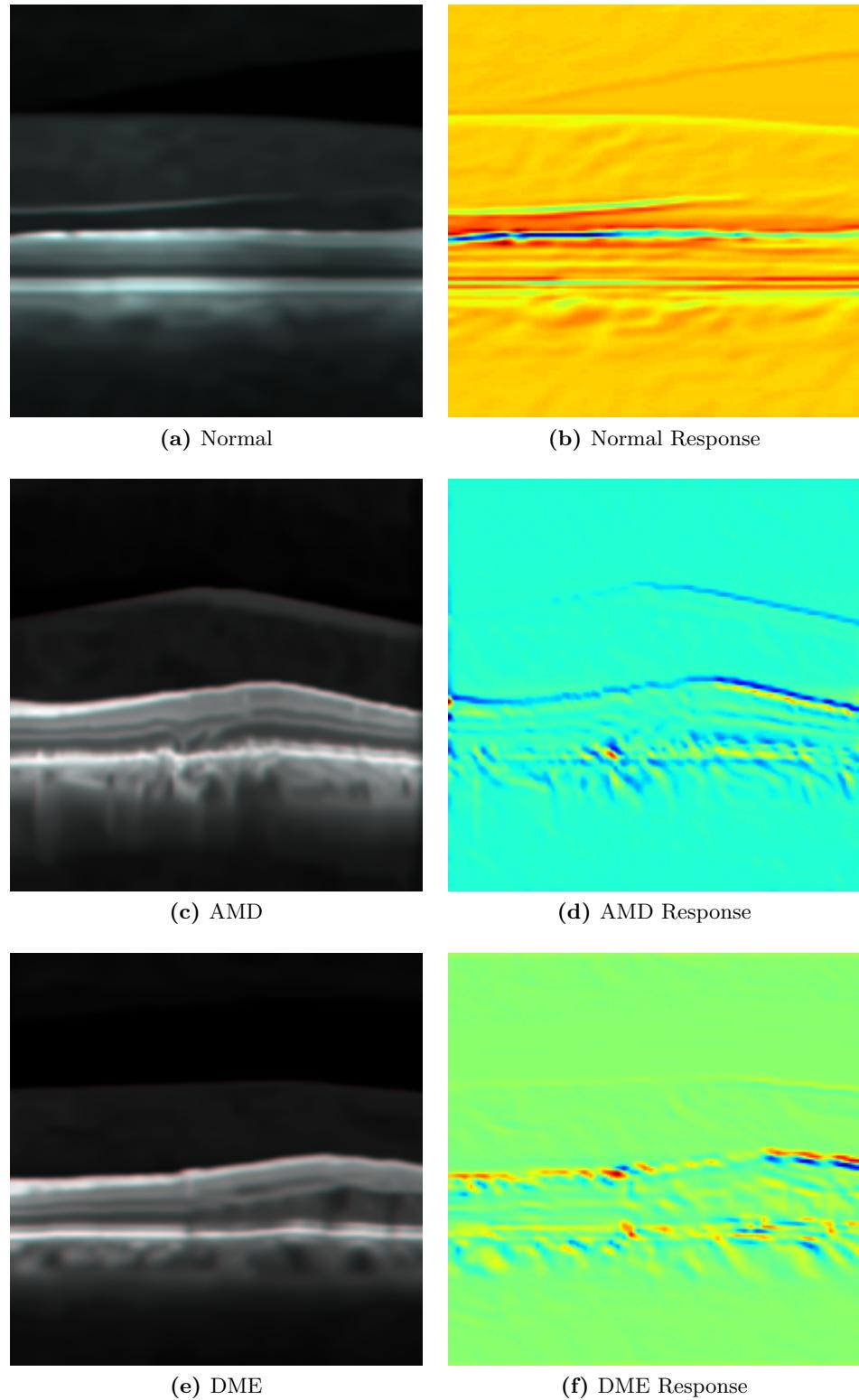
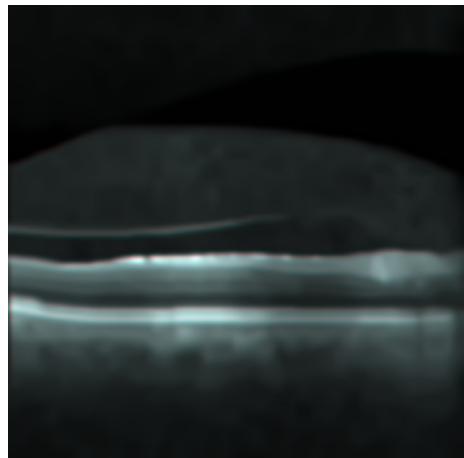
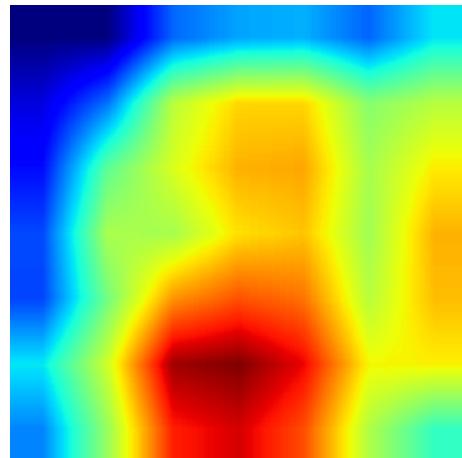


Figure 5.5: Visualizing the identified potential response for normal, AMD and DME case

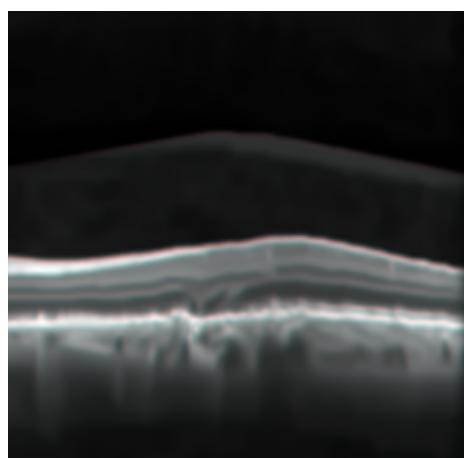
5. Construction and Visualization of Data Driven Feature Quantifiers



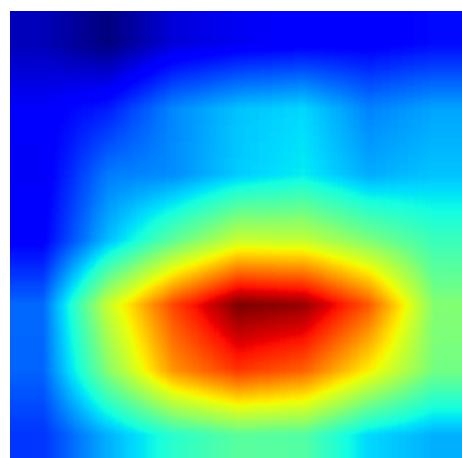
(a) Normal



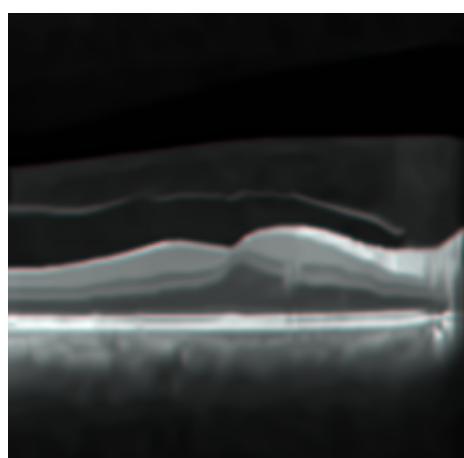
(b) Normal Response



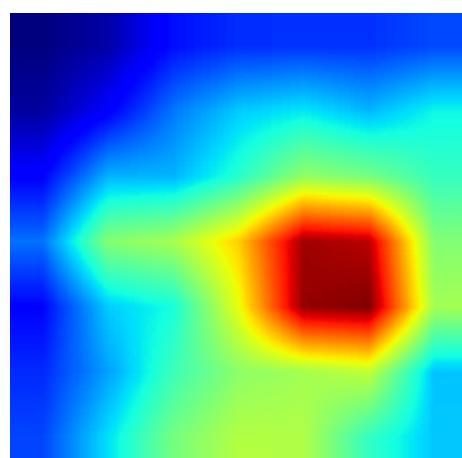
(c) AMD



(d) AMD Response



(e) DME



(f) DME Response

Figure 5.6: Vizualizing the maximum response for normal, AMD and DME case

5.5. Summary

Table 5.1: Crossvalidation accuracies 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

5.5 Summary

Deep learning (DL) has been successful in automatically crafting the feature quantifiers subjective to an image analysis application. DL algorithms require a large amount of data and a large number of epochs for convergence. To avoid such learning barriers models trained on similar datasets are fine tuned for various successful applications [143,144]. Transfer learning based classification or segmentation of medical images majorly involves microscopic and CT images. The underlying image distributions follow a Gaussian distribution similar to non-medical images. The current application involves fine tuning of models for OCT images whose imaging physics incorporate speckles for both information and noise.

In contrast to widely employed VGG architecture an effective architecture for image classification is identified and extended to the classification of retinal pathologies. Standard softmax classifier is employed as classifier but to overcome underflow in softmax ‘log’ is appended. The level of correcting weights based on loss is regulated by learning rate, and it should be finer with the progress of epoch. AdaGrad or ADAM optimizers are used to avoid human intervention in weight decay. However, ADAM is prone to blowing learning rates, so AdaGrad approach is considered. A common criticism on neural networks is non-deterministic models, i.e., upon the change of initialization the final model characteristics change. As the fully connected layer are randomly initialized to illustrate the model characteristics the same experiment is repeated. Conventionally

5. Construction and Visualization of Data Driven Feature Quantifiers

the final test accuracy is a surrogate for learning capabilities of the model. It builds up more confidence to the medical and biomedical communities if layer-wise signal processing characteristics are depicted. Conventionally the salient location in the input image is illustrated, but identification of potential filter or gross representation is not well explored. The initial filter response illustrates and visually validated by the expert if the potential filter (feature descriptor) has been constructed or not. The proposed automated selection is a ranking process so all filter responses can be arranged based on their potentiality. This is crucial because being a sequential (hierarchy) process an information lost in the initial layer cannot be employed by any top layer. The top layer filters have significance from the perspective 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.

Chapter 6

Low Rank CNN Filters for Simultaneous Image Reconstruction and Restoration

6.1 Introduction

The 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 micrometres. However, the same principles introduce stochastic noise which is also speckle in nature. This is a hurdle during interpreting the layers as core information is also speckle in nature. Classical approaches exploit the stochastic nature of the noise which can be removed through temporal averaging. So clinical OCTs are equipped for imaging a region of the retina for multiple instances and reconstruct a noise free image (HD image) through affine registration and averaging of multiple images (instances). An SD-OCT is capable of acquiring few hundred images in few seconds for averaging. Imaging human retina involves involuntary movements and averaged images reflects blurred noise free image. This restrains the number of B-scans (instances) and the number of A-scans

6. Low Rank CNN Filters for Simultaneous Image Reconstruction and Restoration

per B-scan to avoid motion blurs [145]. This has onset the need for single image based reconstruction and restoration algorithms.

6.1.1 Need for sparse representations

Computational imaging endeavours multiple approaches which include denoising [146, 147] the images initially and estimate the HD image (H) from sparsely sampled image (L) [148] or vice-versa. Alternative algorithms employ sparse dictionary learning or sparse coding. Such approaches are multi-image based and need neighbouring location images which need to intervene with imaging protocol of standard OCT, so current approach aims at single image based restoration. A comparison of all algorithms including image restoration, filtering and sparse coding (SC) approaches on a benchmark OCT dataset has depicted that sparse coding based approach have superior performance in case of OCT images [112]. Given a set of image patches a sparse coding iteratively identifies a dictionary through orthogonal matching pursuit with the capability of projecting image patches as sparse representations (vectors). Identification of sparse representations is the saliency of sparse coding and is induced as a regularizer in cost function during construction of dictionaries which is termed to be a reason for superior performance. Such dictionaries are capable of transforming image patches to a sparse vector and project back sparse representations to image patches. This is employed by computational imaging for super resolution.

6.1.2 Extending sparse representations for OCT denoising

For OCT application SC is wrapped as a supervised learning where image patches from sparsely sampled noisy images (l) and corresponding reconstructed noise free images (h) are required. Noise free patches are separated through clustering (K-means), corresponding noisy patches are assigned to the same cluster of noise free patches and centroids for each cluster are identified. In each cluster, dictionaries are learned independently

6.1. Introduction

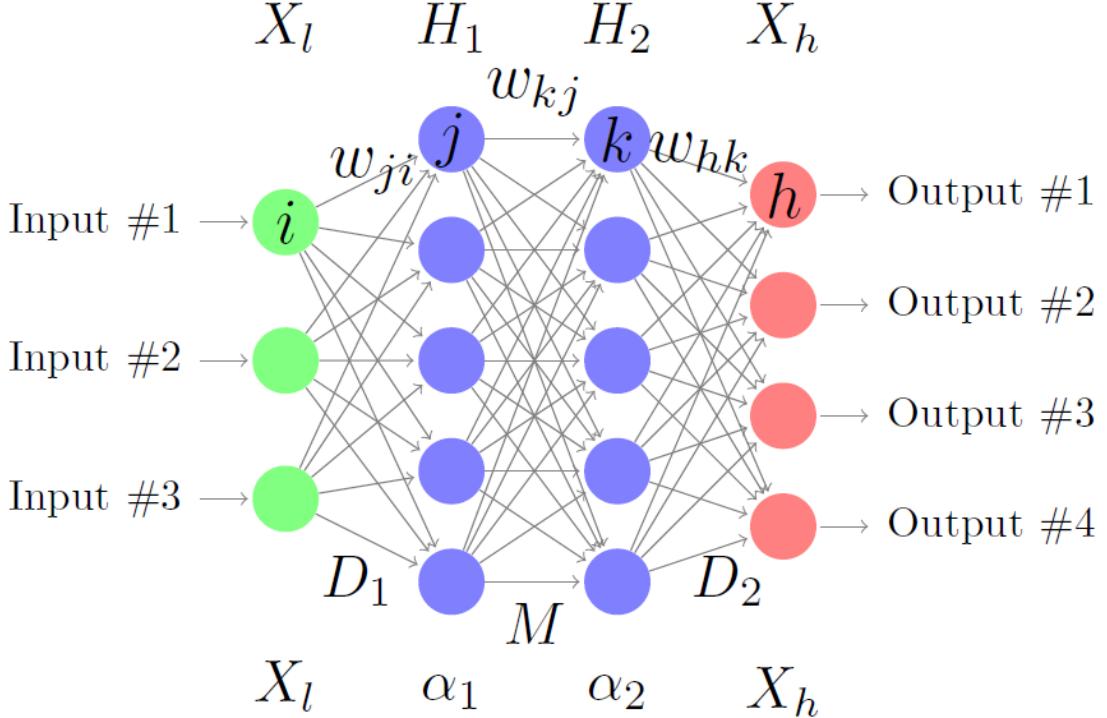


Figure 6.1: Illustrating the neural network functionally equivalent to sparse coding based super resolution

for noisy patches and noise free patches then each patch is transformed to sparse vector employing noisy patch based dictionary or noise free patch based dictionary, a mutual matrix is computed through SVD to transform sparse vector of the noisy patch to a sparse vector of noise free patch. So given a test noisy patch, the cluster it falls into is identified, corresponding noisy patch based dictionary transforms noisy patch to sparse vector, mutual matrix transforms sparse vector to sparse vector of noise free patch and the noise free patch based dictionary estimates noise free patch from the sparse vector of noise free patch. The above mentioned approach learns dictionaries and mutual matrix in mutually exclusive approach. This raises the need of multiple clusters each with dictionaries of 1000's of elements resulting in learning of few 100k parameters.

The figure 6.1 illustrates the neural representation of dictionaries for OCT application. Given sparsely sampled noisy image patches (X_l) and corresponding high dimen-

6. Low Rank CNN Filters for Simultaneous Image Reconstruction and Restoration

sional (X_h) patches, the the D_l , α_l , M, α_h and D_h are computed. From sparse coding perspective the D_l , α_l , M, α_h and D_h represents the dictionary for noisy patches, sparse vectors of noisy image patches, mutual matrix, sparse vectors of noise free image patches and dictionary for noise free patches respectively. From a neural network perspective its a multilayered perceptron architecture with two hidden layers where w_{jl} , w_{kj} and w_{hk} are expected to act as surrogates for D_l , M, and D_h respectively and are identified through gradient descent approach.

6.1.3 Identified approach

Sparse coding has been widely employed for reconstruction and denoising applications independently, so the current approach aims to combine the both. Any ANN architecture can be revamped as a CNN, and recent frameworks have extended above neural network architecture to a convolutional neural network 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 CNN intra-dependency within filters is a common aspect which can be exploited for further reduction of learnable parameters through signal processing.

In signal processing, separable filters is an approach of factorization to reduce the total number of filter elements (weights) through identification of low-rank filters. An illustration of separable filter employment is shown in figure 6.2 where the response of an image on the left-hand side and right-hand side are the same. This results in a reduction of total filter weights to be computed and the number of multiplications (convolution is a combination of multiplication and summation). For example, a 3×3 filter needs nine weights and multiplications on left side and only six weights and multiplications on the right side. The only draw back of separable filter is its sequential nature.

This approach i

6.2. Preamble to Solution

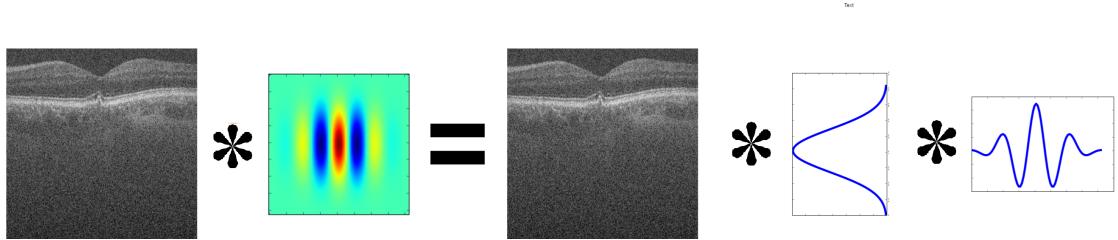


Figure 6.2: Illustrating separable filters

6.2 Preamble to Solution

An existing architecture for super-resolution can be extended to denoising application with modification on receptive fields of CNN. As conventional CNN employs convolution operation with 2D or 3D filters, the number of weights (parameters) can be further reduced through exploring filters with interdependency. Separable filter technique adaptation for the current application can further reduce parameters required. Learnable separable filters is an alternate approach with the hypothesis that common separable filters can exist and can be learned. In figure figure 6.3 illustrates common 1D filters for various 2D filters. For learning separable filters, geometrical constraint imposition is sufficient in case of deep learning [149], but such approaches considered lowest rank of the filter to be two (to resolve instability during training) and are sequential or parallel in nature. So the approach is aimed at single rank filters without trading off the accuracy. One common draw back of such constrain during training is the impulsive error rate which impacts the gradients for correction. This is resolved through selection of appropriate optimizer which keeps a track of gradients from previous iterations during error back propagation rather simple stochastic gradient descent.

6. Low Rank CNN Filters for Simultaneous Image Reconstruction and Restoration

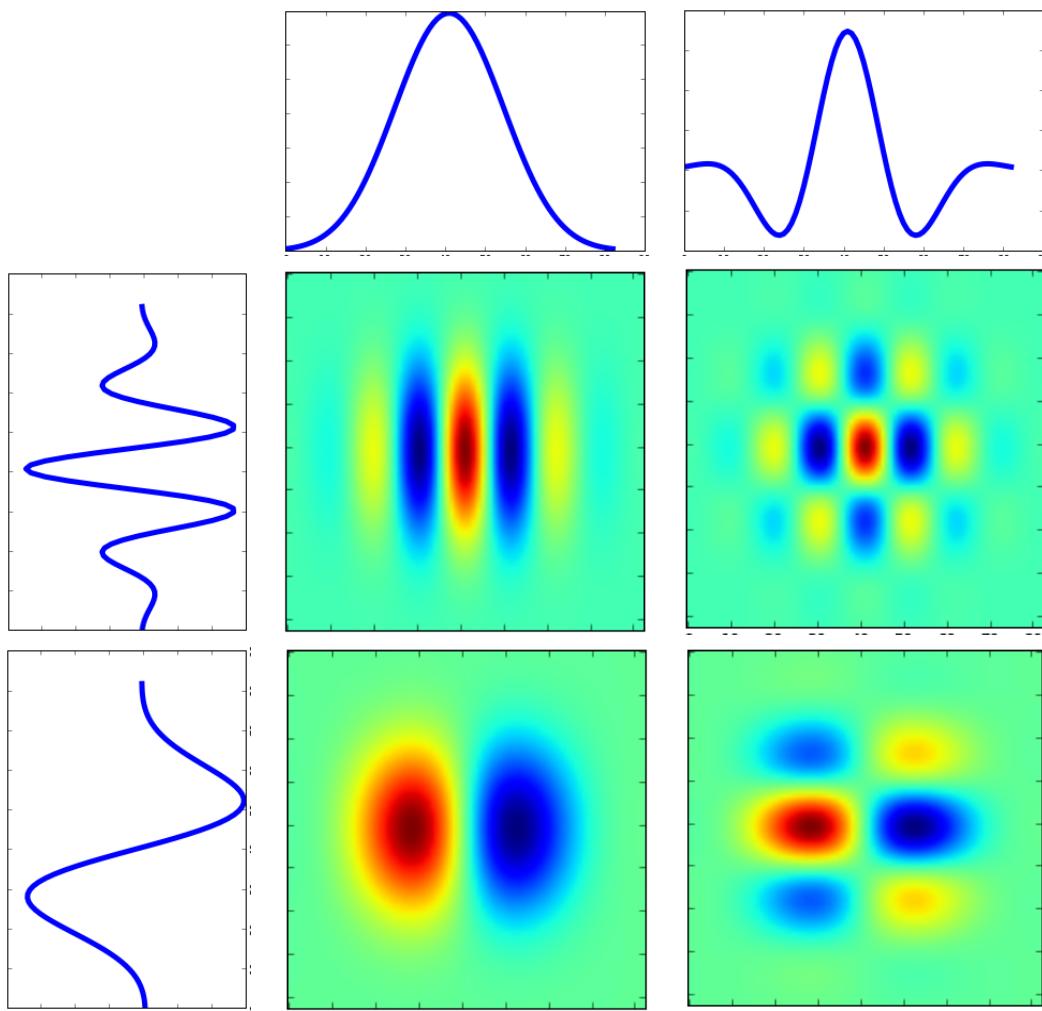


Figure 6.3: Illustrating the common separable filters for distinct 2D filters

6.3. Method

6.3 Method

6.3.1 Modifications considered during adapting SRCNN architecture

The sparse coding based CNN [150] architecture is a supervised approach so a set of interpolated low resolution image patches and corresponding true high resolution image patches are required. A CNN can be deployed over an image to generate a super resolved image as output so for current application entire image is considered as input so output is an entire image. This avoids image mosaicing from predicted patches to speed up the prediction process. For current approach a set of noisy sparsely sampled images I_L and corresponding true HD images I_H are required. Initially $I_L \in \mathbb{R}^{M \times N/2}$ is interpolated ($I_{LH} \in \mathbb{R}^{M \times N}$) to match I_H spatial resolution. A basic block of CNN involves convolution of input with i number of filters each with dimensions $h \times w \times d$ (height, width and depth) and non-linear transformation (Rectified linear unit) of each filter response. In deep learning taxonomy h and w are height and width of filters, d is number of input channels to filters and i is number of filters. The unification of both processes is represented in equation 6.1.

$$output^i = max(0, input * f_{j \times w \times d}^i) \quad (6.1)$$

It is to be noted that pooling in conventional CNN architectures is excluded for super resolution. A CNN is a cascade of convolution blocks where the output of each block is treated as input for subsequent block. The [150] didn't employ upsampling layer and compensated by interpolating the input patches to match the size of output patches so bicubic interpolation is employed as an initial estimate of missing A-scan from sparsely samples image. The filters size has been increased to increase the receptive field of the filters. The architecture input is interpolated noisy image and output image is restored noise free image. As considered input image is gray scale the input filter depth is set to '1'.

6. Low Rank CNN Filters for Simultaneous Image Reconstruction and Restoration

A SC equivalent CNN architecture (figure 6.4) involves learning cascade of three convolution blocks where U is equivalent of dictionary for noisy patch, V is equivalent of mutual matrix and W is equivalent of dictionary for noise free patches. So U transforms each pixel image in I_{LH} to LH sparse coefficients α_{LH} , V maps α_{LH} to sparse coefficients of noise free image α_H and W estimates each pixel of noise free image I_H . Extending the equation 6.1 to above blocks results in equations 6.2, 6.3 and 6.4 where h,w,d,i are user defined.

$$output_{block1}^i = \max(0, I_{LH} * U_{h \times w \times d}^i) \quad (6.2)$$

$$output_{block2}^i = \max(0, output_{block1} * V_{h \times w \times d}^i) \quad (6.3)$$

$$\hat{I}_H = \max(0, output_{block2} * W_{h \times w \times d}^i) \quad (6.4)$$

The number of filters and filter sizes are user defined but to match the mutual matrix functionality V filter size can be only 1×1 and number of W filters need to be one as predicted image has one channel. A forward pass of an interpolated noisy image through CNN is expected to predict a noise free image. The mean square error between estimated image and true noise free image is treated as error. In the pursuit of minimizing the error the U , V and W filters are corrected through error back propagation as shown in figure 6.4. The core distinction between [150] and proposed approach is in terms of application and inclusion of separability.

6.3.2 Factorization of filters

The separable filter approach utilizes the dependency in filter representations and aims to decompose a 2D filter into two unique vectors (1D). The response attained through

6.3. Method

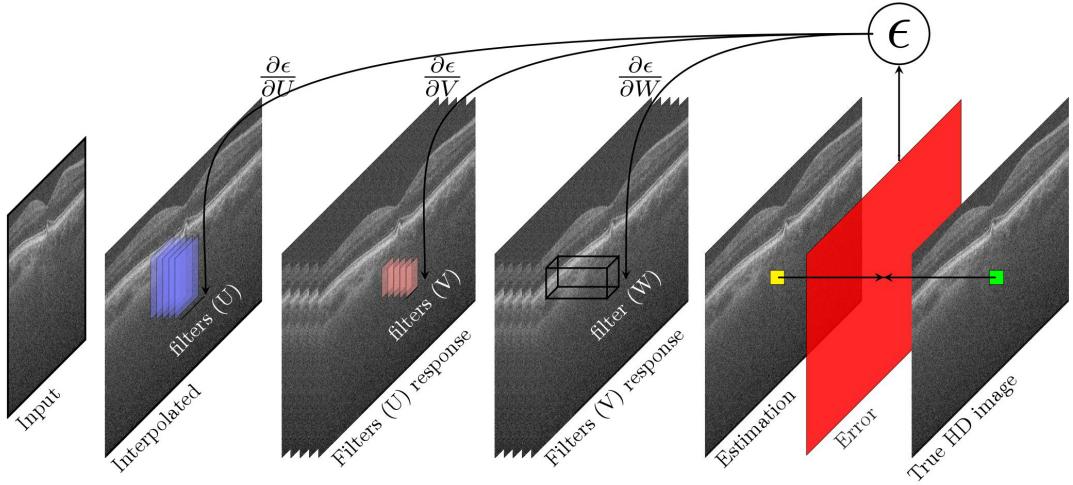


Figure 6.4: Adapted CNN for predicting noise free image

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

$$output_U^i = I_{LH} * U_{j \times 1 \times d}^i \quad (6.5)$$

$$output_U^i = \max(0, output_U * U_{1 \times w \times d}^i) \quad (6.6)$$

Instead of decomposing identified 2D filter a different branch learns separable filters with hypothesis that separable filters can be shared between multiple 2D core filters [149]. These architectures learn separable filters with geometrical constraint and depicted that core filters can be reconstructed with weighted combination of separable filters. But in the process of reducing parameters required the sequential steps have been increased. So alternate approaches have been explored to compute both 1D filters parallaly and combine at the end of the CNN before reconstruction in deeper architectures(similar to equation (6.13)). So the later approach reduces the sequential processing. Such approaches employ inference blocks for reconstruction of image information from gradients.

6. Low Rank CNN Filters for Simultaneous Image Reconstruction and Restoration

$$outputy_U = \max(0, I_{LH} * U_{h \times 1 \times d}^i) \quad (6.7)$$

$$outputy_V = \max(0, outputy_U * V_{h \times 1 \times d}^i) \quad (6.8)$$

$$outputy_W = \max(0, outputy_W * W_{h \times 1 \times d}^i) \quad (6.9)$$

$$outputx_U = \max(0, I_{LH} * U_{1 \times w \times d}^i) \quad (6.10)$$

$$outputx_V = \max(0, outputy_U * V_{1 \times w \times d}^i) \quad (6.11)$$

$$outputx_W = \max(0, outputy_W * W_{1 \times w \times d}^i) \quad (6.12)$$

$$\hat{I}_{LH} = inference(outputy_W, outputx_W) \quad (6.13)$$

The proposed approach fuses both approaches by only factorising U filter as illustrated in figure 6.5. The V filters of CNN can be accountable for the functionality of both weighted combination of responses and mutual matrix.

The U filters are replaced with separable filters in figure 6.4 through geometrical constraint. Employing such factorization through geometrical constrain makes the whole system impulsive and standard stochastic gradient descent approaches are prone to fail. Conjugate gradient apporach is employed to handle impulsive nature of the gradients while learning separable filters.

$$U_y \text{ responses} = \max(0, I_{LH} * U_{y \times 1 \times d}) \quad (6.14)$$

$$U_x \text{ responses} = \max(0, I_{LH} * U_{x \times w \times d}) \quad (6.15)$$

$$output_{block1}^i = (U_y \text{ responses}, U_x \text{ responses}) \quad (6.16)$$

where j, w, d and number of filters are user defined.

6.3. Method

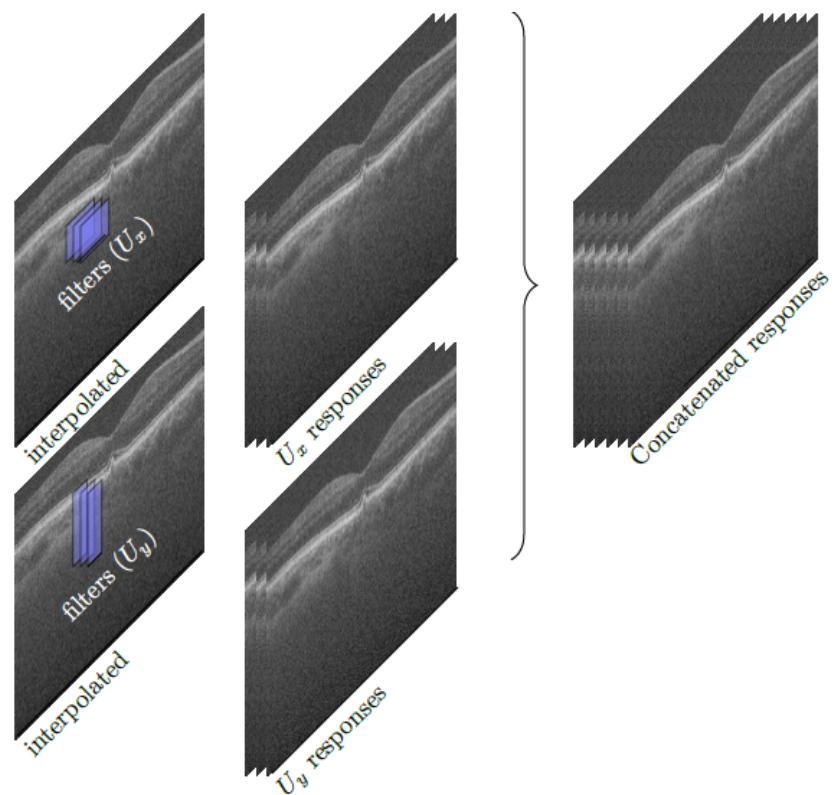


Figure 6.5: Proposed approach with similarity to separable filters

6.4 Results and Discussion

6.4.1 Experimental setup

A Duke SD-OCT restoration data set [112] comprising 10 training samples and 17 testing samples where each sample has a noisy image and noise free image. Experimentation setup proposed by considered by reduction of A-scans in noisy image by 50% and it aims to estimate corresponding noise free image. The computational complexity of experimentation is 8GB RAM and 2GB GPU with dependency over python and Torch.

6.4.2 Training

In current experimentation U is 120 filters with 15×15 size, V is U is 80 filters with 1×1 size and W is 1 filter with 15×15 size. Every noisy image is interpolated with bicubic kernel to size of noise free image and treated as input to CNN. Each image is considered as minibatch and gradients are corrected after an epoch. In contrast to conventional practice that a single image is discriminated to thousands of patches recent practices employ complete image due to convolution property of CNN. The training is performed for 1200 epochs and conjugate gradient is employed for correcting weights.

In case of modified architecture U_x is 60 filters with 15×1 size and U_y is 60 filters with 1×15 size. The responses are concatenated at the input of V , so it gets 120 responses at inputs which is same as SC based CNN architecture. The mean PSNR of estimated HD images of both train and test set at each epoch by Sparse coding CNN (SCCNN) and modified architecture are illustrated in figure 6.6.

Sample test images reconstructed through bicubic, proposed (modified SRCNN) and sparse coding are illustrated in figure 6.7.

6.4. Results and Discussion

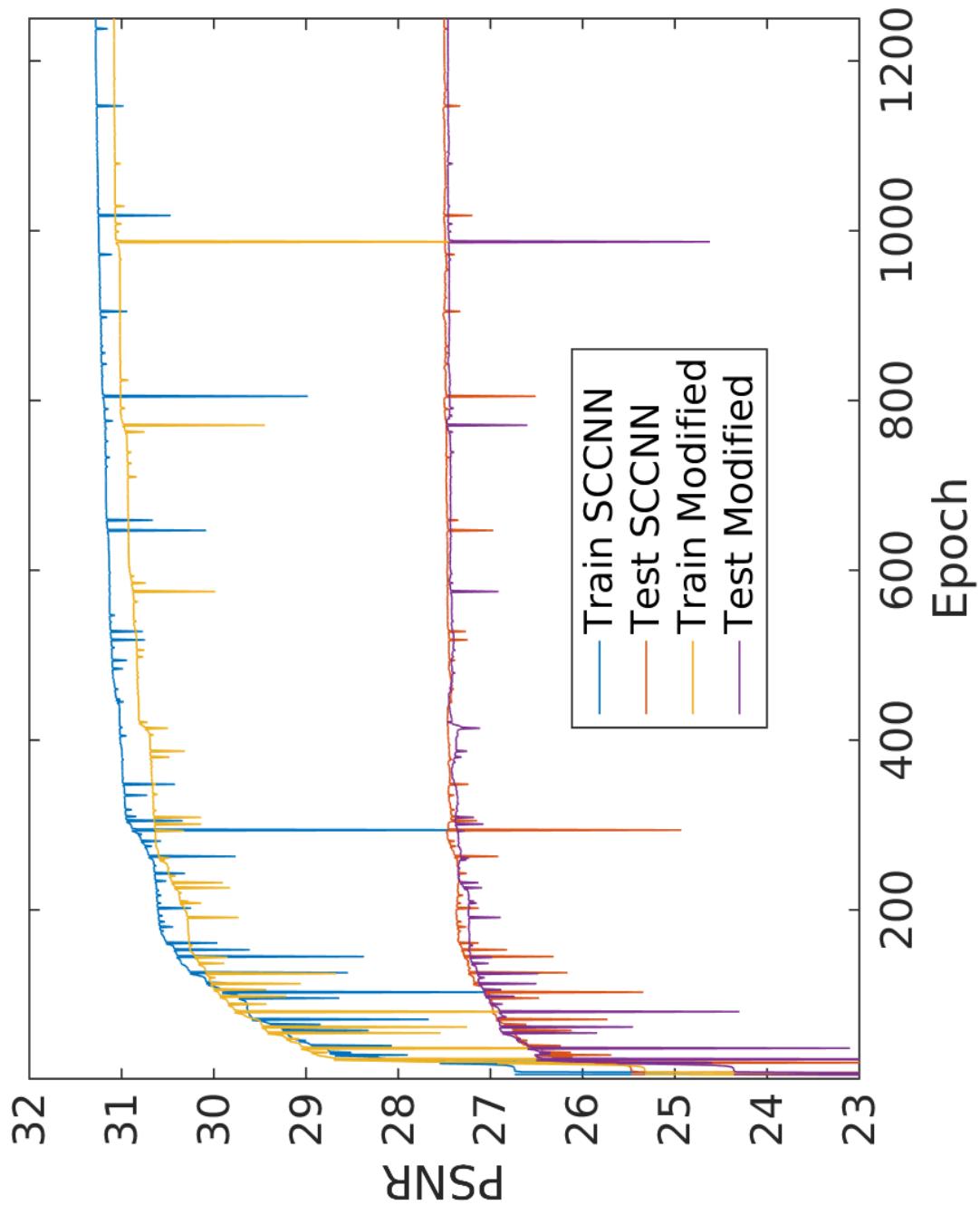
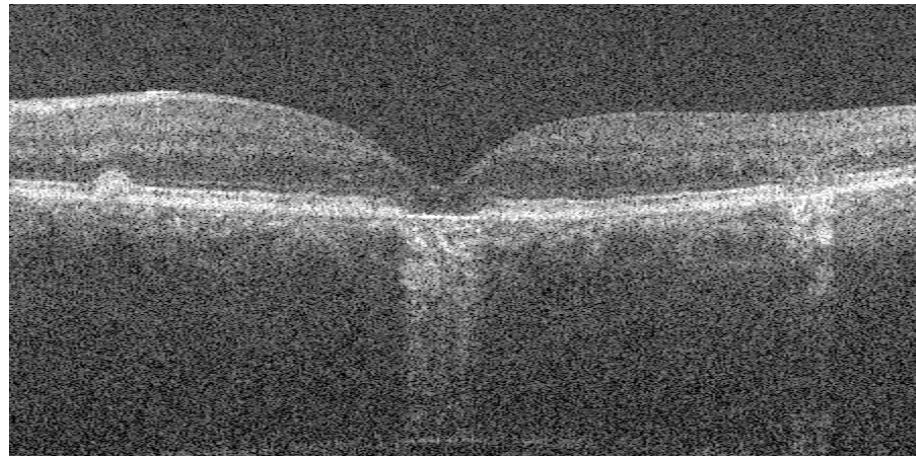
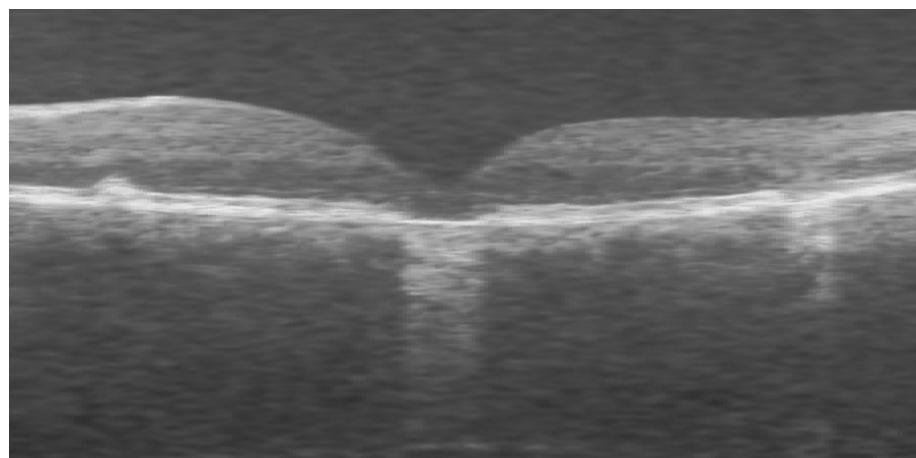


Figure 6.6: Learning characteristics of conventional and modified architecture on training and testing set at each epoch

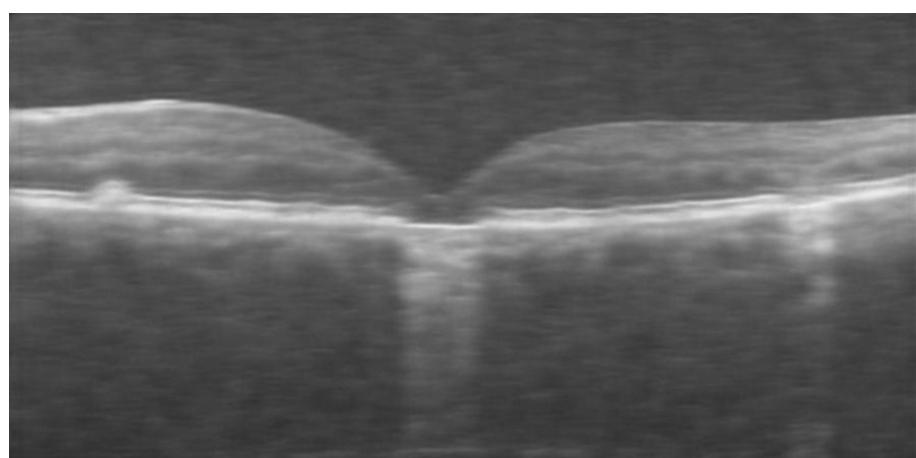
6. Low Rank CNN Filters for Simultaneous Image Reconstruction and Restoration



(a) Bicubic



(b) L Fang



(c) Proposed

Figure 6.7: Illustrates the reconstructed image through proposed and base line (L Fang)

6.5. Summary

Table 6.1: Comparison of adapted and modified CNN across various metrics

	PSNR	MSE	Max. Error	L2 norm	SSIM
SRR	28.24 (2.52)	115.58 (76.36)	92.52 (33.22)	0.99 (0.08)	0.69 (0.02)
SRR modified	28.15 (2.43)	116.68 (74.60)	90.03 (31.25)	0.99 (0.08)	0.68 (0.02)
SC	27.64 (2.24)	128.42 (76.72)	95.32 (27.65)	1.02 (0.08)	0.66 (0.02)
SRF(3)	26.17 (1.65)	168.71 (71.65)	198.70 (17.78)	0.98 (0.08)	0.61 (0.02)

6.4.3 Benchmark

The test set with 17 samples has been considered for validating adapted and modified architecture. SC and SRCNN without the separable filters have been considered as baselines. Every architecture is trained following the protocol for single image based restoration. Standard image quality metrics are considered for evaluation. SRCNN approach has the best qualitative performance in comparison to other approaches but proposed approach achieved similar performance with 45% less number of parameters.

6.5 Summary

Computational algorithms have been helpful to handle the constraints from instrumentation or imaging protocols and estimate the true information for effective screening. Sparse coding based approaches have been effective to restore sparsely sampled images so an existing CNN architecture whose functionality is identical to sparse coding. But it has been designed for superresolution applications. The employed CNN architecture and training procedure has been altered from size of receptive field to optimizer for simultaneous restoration and reconstruction application. A trained CNN is capable of estimating noise free images with reasonable image quality. The required parameters are multifold less than sparse coding pipeline and also has fast prediction time on employment of GPU. It also depicted that it has superior performance in comparison to other baselines in terms of various metrics ranging from signal strength to structure. Inspired from separable filters and edge aware filtering a geometrical constraint has been imposed over

6. Low Rank CNN Filters for Simultaneous Image Reconstruction and Restoration

first layer filters the filter architecture has been constrained geometrically. This has lead to construction of low-rank filter factorization which effectively reduced the number of parameters required. It has also been depicted that such imposition has negative effect on system stability and consideration of gradients from previous epoch equips optimizer to handle such impulse actions. It has been evident that redefined architecture has capability of predicting noise free and restored images without substantial trading off.

Chapter 7

Summary and Conclusion

The proposed work in the thesis has been motivated towards modelling of representations in various spaces for augmenting medical image analysis, and the individual block performances have been evaluated for retinal applications in comparison to various baselines. The block specific approach is considered based on the philosophy of Bellman's equation that optimal system can be achieved through starting at an optimal point and keep taking optimal decisions at each step. So, in future even if core algorithm in a block changes the proposed algorithms for other blocks stay as the optimal decision. The state of the art algorithms employed in each block of medical image analysis are based on machine learning but fails in case of pathological cases due to the incapability of learning multiple patterns.

7.1 Existing Challenges

Learning of representations is crucial for effective modelling which is commonly used by various blocks of image analysis for improved predictions. Representation space is subjective to the application or block (segmentation or classifier) as segmentation block is more concerned about representation in target space and classifier is more concerned about representative control points in the feature space. So multiple algorithms each

7. Summary and Conclusion

with the capability of modelling representations in different spaces need to be identified and adapted to retinal image analysis. As a proposed or adapted algorithm need not be effective for an application, each proposed algorithm needs to be validated against existing baselines. So proposed approaches are compared with various baselines for qualitative evaluation.

7.2 Contribution of Our Work

The proposed algorithms are capable segmenting layers even under pathological cases with large deformations, identification of pathological subjects with weakly supervised datasets, construction of data driven feature quantifiers to reduce the level of expert intervention and retrieving sparsely sampled images generated to avoid motion artefacts or to save storage memory. Such algorithms impact clinically in terms of accurate profiling of retinal layers width, automated screening of retinal pathologies so ophthalmologists can focus on pathological subjects, comparing the representations quantified by feature quantifier with standard clinical features and acquire images rapidly or store images in less space which impacts telemedicine. Along with augmenting the performance, the proposed approaches reflect reduced space complexity and time complexity.

A training framework has been developed to equip standard edge detection algorithms such as SRE to model layer specific edges in retinal OCT images. So given an image patch, each model is capable of predicting corresponding edge representations in target space. It initially validates the hypothesis that edges from different objects can be modelled differently. This subsequently augments segmentation performance for delineating layers which are pivotal for retinal layer profiles and thickness maps. Such segmentation algorithms are robust to layer deformation (due to pathology), noise (due to imaging principles) and artefacts (due to imaging protocol). The proposed framework can be employed for any other off the shelf edge prediction algorithms reflecting the flexibility of the proposed framework. The model for Bruch’s membrane aids for

7.2. Contribution of Our Work

retinal flattening which is a standard step in retinal OCT classification. Such flattened images are quantified using features for the construction of feature space.

A classifier inspired from Voronoi fractals has been developed to identify representative control points in feature space. This approach equipped the model to autonomously identify the feature space structures without the knowledge of kernels. This has been validated considering a toy example with various scenarios where the same algorithm constructed appropriate boundaries according to the feature space. This modelling capability of the classifier has been validated through modelling a two spiral example. Accurate identification of retinal pathologies has validated the real-world applicability of the proposed classifier.

Data driven feature quantifiers have augmented the performance in comparison to traditional feature descriptors. Existing data driven feature quantifiers has been adapted for retinal application through transfer learning. An appropriate CNN architecture with fewer parameters is considered as it less prone to overfitting. Being a neural network based algorithm, the consistency has been validated with multiple experiments. Such practices reduce the dependency of heuristics (by experts) involved during employment of traditional feature descriptors. Additionally, such architectures avail to visualise the step by step transformation of input space raw information to high-level abstract representations. Such abstract patterns augment a naive classifier to identify pathological subjects effectively. An automated approach has been proposed to identify potential response at each CNN block for easy visual interpretation of information processing and abstract patterns.

Additionally, a separable filters based restoration block is proposed to restore and filter sparsely sampled images simultaneously. Such blocks are employed for faster imaging, reduction of image storage and faster transmission of images for telemedicine. As sparse representation based algorithms have been successful for such applications a CNN architecture which is functionally equivalent to sparse coding but for super-resolution ap-

7. Summary and Conclusion

plication has been adapted for retinal OCT restoration application. The subtle changes like increasing the receptive field of CNN filters is needed to for current application and resulted in improved performance of restoring images. The core change of imposing a geometrical constraint on filters reduced the number of parameters with a permittable tradeoff of performance on testing set. During training, a complete image is considered instead of patches (conventional practice), and this saves up the time for image mosaicing based on predicted patches during testing.

The modelling back-end being ML the same algorithms can be employed for various datasets acquired from different OCT or different imaging protocol without much effort. Upon on arrival of a new dataset, the modelled representations can be further modified through transfer learning. Additionally, representations have the capability of modelling multiple patterns with limited training data. As proposed algorithms are supervised in nature, they cannot autonomously attribute a new pattern to a class like an unsupervised algorithm, but such dimension can be explored through semi-supervised approaches.

7.3 Future Scope of Work

A dataset common for all blocks needs to be acquired on a single machine for execution of all proposed algorithms in a pipeline. The data driven feature quantifiers need to have a classifier imbibed into the architecture, but at the current state, an end-to-end training between feature quantifier and the proposed classifier is infeasible due to architectural incompatibility. Extension of image level algorithms to a volumetric analysis by considering neighbouring images is trivial but needed for improved performance.

Individual chapter subjective future directions involve. The HOG features for SRE can be replaced with data drive feature quantifiers. Gradient boosted tree techniques can be extended to SRE so the models can adapt the criteria at each node upon arrival of new data. The representation learning based semantic segmentation can be an alternative for generation of delineated layers without the need for edge prediction and dynamic

7.3. Future Scope of Work

program framework.

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

Current practices widely employ CNN architectures for the construction of data driven feature quantifier. The interest has resulted in multiple CNN architectures which can be explored for adaptability to retinal pathology identification. As the architecture changes the optimal parameters (class weight, loss weight, etc.) combination needs to be explored. The visualisation also can be improved for effective interpretation of representations.

For restoration block, MSE loss is considered which aims for only mean signal match. As OCT have structures, combinations of SSIM and MSE can improve both mean signal and structure. The SC based CNN is only two layers deep, and it is empirically proven that deeper architectures improve performance so the number of layers in CNN architecture can be increased for improved performance.

Publications out of this work

Journal

- Karri, S.P.K., Chakraborty, D. 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, D. 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, D. 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., Garai, N., Nawn, D., Ghosh, S., Chakraborty, D. and Chatterjee, J., 2016, September. Retinal layer delineation through learning of tissue photon interaction in optical coherence tomography. In *Technology Symposium (TechSym)*, 2016 IEEE StudentsâŽ (pp. 46-51). IEEE.
- Karri, S.P.K., Garai, N., Nawn, D., Ghosh, S., Chakraborty, D. 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 StudentsâŽ (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.

References

- [1] M. Abramoff, M. Garvin, and M. Sonka, “Retinal Imaging and Image Analysis,” *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.
- [2] M. de Bruijne, “Machine learning approaches in medical image analysis: From detection to diagnosis,” *Medical Image Analysis*, vol. 33, pp. 94–97, 2016.
- [3] C. J. Burges, “A tutorial on support vector machines for pattern recognition,” *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [4] D. Parikh and C. L. Zitnick, “Finding the weakest link in person detectors,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1425–1432.
- [5] C. M. Bishop, *Pattern Recognition and Machine Learning*, 2006, vol. 4, no. 4. [Online]. Available: <http://www.library.wisc.edu/selectedtoc/bg0137.pdf>
- [6] F. Esposito and D. Malerba, “Machine learning in computer vision,” *Applied Artificial Intelligence*, vol. 15, pp. 693–705, 2001.
- [7] S. Wang and R. M. Summers, “Machine learning and radiology,” pp. 933–951, 2012.
- [8] T. M. Mitchell, *Machine Learning*, 1997, vol. 4. [Online]. Available: <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0070428077>
- [9] S. B. Kotsiantis, “Supervised machine learning: A review of classification techniques,” *Informatica*, vol. 31, pp. 249–268, 2007. [Online]. Available: http://books.google.com/books?hl=en&lr=&id=vLiTXDHR_sYC&oi=fnd&pg=PA3&dq=survey+machine+learning&ots=CVsyuwYHjo&sig=A6wYWvywU8XTc7Dzp8ZdKJaW7rc%5Cnpapers://5e3e5e59-48a2-47c1-b6b1-a778137d3ec1/Paper/p800%5Cnhttp://www.informatica.si/PDF/31-3/11{__}Kotsiantis

- [10] T. Joachims, T. Hofmann, Y. Yue, and C.-N. Yu, “Predicting structured objects with support vector machines,” *Communications of the ACM*, vol. 52, no. 11, p. 97, 2009.
- [11] J. Lafferty, A. McCallum, and F. C. N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” *ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning*, vol. 8, no. June, pp. 282–289, 2001. [Online]. Available: <http://repository.upenn.edu/cis/papers/159/> [Online]. Available: <http://dl.acm.org/citation.cfm?id=655813>
- [12] S. Nowozin, “Structured Learning and Prediction in Computer Vision,” *Foundations and Trends® in Computer Graphics and Vision*, vol. 6, no. 3-4, pp. 185–365, 2010.
- [13] P. Kotschieder, S. R. Bulo, H. Bischof, and M. Pelillo, “Structured class-labels in random forests for semantic image labelling,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2190–2197.
- [14] I. Tschantaridis, T. Hofmann, T. Joachims, and Y. Altun, “Support vector machine learning for interdependent and structured output spaces,” *International Conference on Machine Learning*, p. 104, 2004. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1015330.1015341>
- [15] P. Domingos, “Structured machine learning: Ten problems for the next ten years,” *Proceedings of 17th International*, pp. 1–4, 2007. [Online]. Available: <http://alchemy.cs.washington.edu/papers/domingos07d/domingos07d.pdf>
- [16] D. Marr, “Vision: A Computational Investigation of Visual Representation in Man,” *Phenomenology and the Cognitive Sciences*, vol. 8, no. 4, p. 397, 1982. [Online]. Available: <http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=12242&ref=nf>
- [17] M. A. Fischler and R. A. Elschlager, “The Representation and Matching of Pictorial Structures Representation,” *IEEE Transactions on Computers*, vol. C-22, no. 1, pp. 67–92, 1973.
- [18] M. Porat and Y. Y. Zeevi, “The Generalized Gabor Scheme of Image Representation in Biological Machine Vision,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 452–468, 1988.

References

- [19] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [20] G. E. Hinton, “Learning multiple layers of representation,” pp. 428–434, 2007.
- [21] A. Krizhevsky, I. Sutskever, and H. Geoffrey E., “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems 25 (NIPS2012)*, pp. 1–9, 2012. [Online]. Available: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- [22] P. Domingos, “A few useful things to know about machine learning,” *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.
- [23] X. Chen, D. P. Kingma, T. Salimans, Y. Duan, P. Dhariwal, J. Schulman, I. Sutskever, P. Abbeel, C. Science, and C. Science, “Variational Lossy Autoencoder,” in *International Conference on Learning Representation*, 2017, pp. 1–14.
- [24] P. Dollar and C. L. Zitnick, “Structured forests for fast edge detection,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1841–1848.
- [25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [26] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, “Convolutional neural networks for medical image analysis: full training or fine tuning?” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [27] A. Parker, “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*, 2003, p. i.
- [28] H. Kobayashi and S. Kohshima, “Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye,” *Journal of human evolution*, vol. 40, no. 5, pp. 419–435, 2001.
- [29] A. Roorda and D. R. Williams, “The arrangement of the three cone classes in the living human eye,” *Nature*, vol. 397, no. 6719, pp. 520–522, 1999.

References

- [30] R. R. A. Bourne, J. B. Jonas, S. R. Flaxman, J. Keeffe, J. Leasher, K. Naidoo, M. B. Parodi, K. Pesudovs, H. Price, R. A. White, T. Y. Wong, S. Resnikoff, and H. R. Taylor, “Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe: 1990-2010.” *The British journal of ophthalmology*, vol. 98, no. 5, pp. 629–38, 2014. [Online]. Available: <http://bjo.bmjjournals.org/content/98/5/629.long>
- [31] N. M. M. Bressler, “Age-Related Macular Degeneration Is the Leading Cause of Blindness,” *Jama*, vol. 291, no. 15, pp. 1900–1901, 2004.
- [32] D. Baylor and M. Fuortes, “Electrical responses of single cones in the retina of the turtle,” *The Journal of physiology*, vol. 207, no. 1, p. 77, 1970.
- [33] S. L. Polyak, “The retina.” 1941.
- [34] R. Klein, B. E. Klein, S. E. Moss, and K. L. Linton, “The beaver dam eye study: retinopathy in adults with newly discovered and previously diagnosed diabetes mellitus,” *Ophthalmology*, vol. 99, no. 1, pp. 58–62, 1992.
- [35] E. M. Frohman, J. G. Fujimoto, T. C. Frohman, P. A. Calabresi, G. Cutter, and L. J. Balcer, “Optical coherence tomography: a window into the mechanisms of multiple sclerosis,” *Nature Clinical Practice Neurology*, vol. 4, no. 12, pp. 664–675, 2008. [Online]. Available: <http://www.nature.com/doifinder/10.1038/ncpneuro0950>
- [36] A. London, I. Benhar, and M. Schwartz, “The retina as a window to the brain—from eye research to cns disorders,” *Nature Reviews Neurology*, vol. 9, no. 1, pp. 44–53, 2013.
- [37] T. A. Ciulla, A. G. Amador, and B. Zinman, “Diabetic Retinopathy and Diabetic Macular Edema: Pathophysiology, screening, and novel therapies,” *Diabetes Care*, vol. 26, no. 9, pp. 2653–2664, 2003. [Online]. Available: <http://care.diabetesjournals.org/content/26/9/2653.abstract>
- [38] L. A. Yannuzzi, M. D. Ober, J. S. Slakter, R. F. Spaide, Y. L. Fisher, R. W. Flower, and R. Rosen, “Ophthalmic fundus imaging: today and beyond,” *American journal of ophthalmology*, vol. 137, no. 3, pp. 511–524, 2004.
- [39] J. Cunha-Vaz, J. F. De Abreu, and A. Campos, “Early breakdown of the blood-retinal barrier in diabetes.” *British Journal of Ophthalmology*, vol. 59, no. 11, pp. 649–656, 1975.

References

- [40] A. Bird, N. Bressler, S. Bressler, I. Chisholm, G. Coscas, M. Davis, P. De Jong, C. Klaver, B. Klein, R. Klein *et al.*, “An international classification and grading system for age-related maculopathy and age-related macular degeneration,” *Survey of ophthalmology*, vol. 39, no. 5, pp. 367–374, 1995.
- [41] M. Adhi and J. S. Duker, “Optical coherence tomography—current and future applications.” *Current opinion in ophthalmology*, vol. 24, no. 3, pp. 213–21, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3758124/>
- [42] H. Cheng, G. Nair, T. A. Walker, M. K. Kim, M. T. Pardue, P. M. Thulé, D. E. Olson, and T. Q. Duong, “Structural and functional mri reveals multiple retinal layers,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 46, pp. 17 525–17 530, 2006.
- [43] A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser, “Optical coherence tomography - principles and applications,” *Reports on Progress in Physics*, vol. 66, no. 2, pp. 239–303, 2003. [Online]. Available: <http://stacks.iop.org/0034-4885/66/i=2/a=204?key=crossref>
- [44] J. Schmitt, “Optical Coherence Tomography (OCT): A Review,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 5, no. 4, pp. 1205–1215, 1999.
- [45] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, A. Et, and E. al., “Optical coherence tomography.” pp. 1178–81, 1991. [Online]. Available: [http://www.sciencemag.org/content/254/5035/1178/](http://www.sciencemag.org/content/254/5035/1178)
- [46] M. R. Hee, J. A. Izatt, E. A. Swanson, D. Huang, J. S. Schuman, C. P. Lin, C. A. Puliafito, and J. G. Fujimoto, “Optical coherence tomography of the human retina.” *Archives of ophthalmology*, vol. 113, no. 3, pp. 325–32, 1995. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC7887846/>
- [47] J. F. De Boer, B. Cense, B. H. Park, M. C. Pierce, G. J. Tearney, and B. E. Bouma, “Improved signal-to-noise ratio in spectral-domain compared with time-

- domain optical coherence tomography,” *Optics letters*, vol. 28, no. 21, pp. 2067–2069, 2003.
- [48] A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser, “Optical coherence tomography – development, principles, applications,” *Zeitschrift für Medizinische Physik*, vol. 20, no. 4, pp. 251–276, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0939388909001524>
- [49] P. H. Tomlins and R. K. Wang, “Theory, developments and applications of optical coherence tomography,” *Journal of Physics D: Applied Physics*, vol. 38, no. 15, p. 2519, 2005. [Online]. Available: [http://iopscience.iop.org/0022-3727/38/15/002{%}5Cnhttp://iopscience.iop.org/0022-3727/38/15/002/{%}5Cnhttp://iopscience.iop.org/0022-3727/38/15/002/pdf/0022-3727{ }38{ }15{ }002.pdf">http://iopscience.iop.org/0022-3727/38/15/002{%}5Cnhttp://iopscience.iop.org/0022-3727/38/15/002/{%}5Cnhttp://iopscience.iop.org/0022-3727/38/15/002/pdf/0022-3727{ }38{ }15{ }002.pdf](http://iopscience.iop.org/0022-3727/38/15/002)
- [50] R.-W. Lu, C. a. Curcio, Y. Zhang, Q.-X. Zhang, S. J. Pittler, D. Deretic, and X.-C. Yao, “Investigation of the hyper-reflective inner/outer segment band in optical coherence tomography of living frog retina,” *Journal of Biomedical Optics*, vol. 17, no. 6, p. 060504, 2012.
- [51] R. F. Spaide and C. A. Curcio, “Anatomical Correlates to The Bands Seen in The Outer Retina by Optical Coherence Tomography: Literature Review and Model,” *Retina*, vol. 31, no. 8, pp. 1609–1619, 2011.
- [52] M. Glocsmann, B. Hermann, C. Schubert, H. Sattmann, P. K. Ahnelt, and W. Drexler, “Histologic correlation of pig retina radial stratification with ultrahigh-resolution optical coherence tomography,” *Investigative Ophthalmology and Visual Science*, vol. 44, no. 4, pp. 1696–1703, 2003.
- [53] M. Mujat, R. C. Chan, B. Cense, B. Hyle Park, C. Joo, T. Akkin, T. C. Chen, and J. F. de Boer, “Retinal nerve fiber layer thickness map determined from optical coherence tomography images,” *Optics Express*, vol. 13, no. 23, pp. 9480–9491, 2005. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-13-23-9480>
- [54] M. Brar, R. Yuson, I. Kozak, F. Mojana, L. Cheng, D.-U. Bartsch, S. F. Oster, and W. R. Freeman, “Correlation between morphologic features on spectral-domain optical coherence tomography and angiographic leakage patterns in macular edema.” *Retina (Philadelphia, Pa.)*, vol. 30, no. 3, pp. 383–9, 2010.

References

- [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2870721/>{&}tool=pmcentrez{&}rendertype=abstract
- [55] S. Vujosevic, E. Midena, E. Pilotto, P. P. Radin, L. Chiesa, and F. Cavarzeran, “Diabetic macular edema: correlation between microperimetry and optical coherence tomography findings.” *Investigative ophthalmology & visual science*, vol. 47, no. 7, pp. 3044–3051, 2006.
- [56] R. L. M. Wong, J. W. Y. Lee, G. S. K. Yau, and I. Y. H. Wong, “Relationship between Outer Retinal Layers Thickness and Visual Acuity in Diabetic Macular Edema,” *BioMed Research International*, vol. 2015, pp. 1–5, 2015. [Online]. Available: <http://www.hindawi.com/journals/bmri/2015/981471/>
- [57] K. H. Kim, G. N. Maguluri, K. Cusato, R. B. Barlow, and J. F. D. Boer, “Monitoring mouse retinal degeneration with high-resolution spectral-domain optical coherence tomography,” *Journal of Vision*, vol. 8, no. 1, pp. 1–11, 2008.
- [58] D. Koozekanani, K. Boyer, and C. Roberts, “Retinal thickness measurements from optical coherence tomography using a Markov boundary model,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 9, pp. 900–916, 2001.
- [59] K. Lee, M. Abràmoff, M. Sonka, and M. K. Garvin, “Automated segmentation of intraretinal layers from spectral-domain macular OCT: reproducibility of layer thickness measurements,” in *Proceedings of SPIE Medical Imaging 2011 Biomedical Applications in Molecular Structural and Functional Imaging*, vol. 7965, 2011, pp. 796523–8. [Online]. Available: [&}{&}Agg=doi](http://link.aip.org/link/PSISDG/v7965/i1/p796523/s1)
- [60] Y. Lu, Z. Li, X. Zhang, B. Ming, J. Jia, R. Wang, and D. Ma, “Retinal nerve fiber layer structure abnormalities in early Alzheimer’s disease: Evidence in optical coherence tomography,” *Neuroscience Letters*, vol. 480, no. 1, pp. 69–72, 2010.
- [61] R. Kromer, N. Serbecic, L. Hausner, L. Froelich, F. Aboul-Enein, and S. C. Beutelspacher, “Detection of retinal nerve fiber layer defects in Alzheimer’s disease using SD-OCT,” *Frontiers in Psychiatry*, vol. 5, no. FEB, 2014.
- [62] A. Gaudric, V. Krivacic, and R. Tadayoni, “Outer retina capillary invasion and ellipsoid zone loss in macular telangiectasia type 2 imaged by optical coherence tomography angiography,” *Retina*, vol. 35, no. 11, pp. 2300–2306, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4644127/>{&}tool=pmcentrez{&}rendertype=abstract

- [63] L. M. Zangwill, C. Bowd, C. C. Berry, J. Williams, E. Z. Blumenthal, C. a. Sánchez-Galeana, C. Vasile, and R. N. Weinreb, “Discriminating between normal and glaucomatous eyes using the Heidelberg Retina Tomograph, GDx Nerve Fiber Analyzer, and Optical Coherence Tomograph,” *Arch Ophthalmol*, vol. 119, no. July, pp. 985–993, 2001.
- [64] W. K. Pratt, “Image Segmentation,” *Computer Vision*, pp. 579–622, 2000.
- [65] J. Delon, A. Desolneux, J. L. Lisani, and A. B. Petro, “A nonparametric approach for histogram segmentation,” *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 253–261, 2007.
- [66] O. J. Tobias and R. Seara, “Image segmentation by histogram thresholding using fuzzy sets,” *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1457–1465, 2002.
- [67] Y. Y. Boykov and M.-P. Jolly, “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, no. July, 2001, pp. 105—112.
- [68] L. Grady, “Random walks for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.
- [69] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [70] L. A. Vese and T. F. Chan, “A multiphase level set framework for image segmentation using the Mumford and Shah model,” *International Journal of Computer Vision*, vol. 50, no. 3, pp. 271–293, 2002.
- [71] D. C. Fernández, “Delineating fluid-filled region boundaries in optical coherence tomography images of the retina,” *IEEE Transactions on Medical Imaging*, vol. 24, no. 8, pp. 939–945, 2005.
- [72] H. Fujiyoshi, “Gradient-Based Feature Extraction -SIFT and HOG,” *Computer*, vol. 107, no. 206, pp. 211–224, 2007. [Online]. Available: <http://ci.nii.ac.jp/naid/110006423317/>

References

- [73] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abràmoff, and M. Sonka, “Three-dimensional analysis of retinal layer texture: Identification of fluid-filled regions in SD-OCT of the macula,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1321–1330, 2010.
- [74] Y.-w. Chen and C.-j. Lin, “Combining SVMs with Various Feature Selection Strategies,” in *Feature Extraction*, 2006, vol. 324, no. 1, pp. 315–324. [Online]. Available: http://link.springer.com/10.1007/978-3-540-35488-8_13
- [75] R. Kohavi and G. John, “Wrappers for feature subset selection,” *Artificial intelligence*, vol. 97, no. 1, pp. 273—324, 1997. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S000437029700043X>
- [76] P. P. Srinivasan, L. a. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. a. Izatt, and S. Farsiu, “Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images.” *Biomedical optics express*, vol. 5, no. 10, pp. 3568–77, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4206325/>
- [77] L. Yang and R. Jin, “Distance metric learning: A comprehensive survey,” *Michigan State Universiy*, vol. 2, no. 2, 2006.
- [78] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [79] S. T. Bow, *Pattern recognition and image preprocessing*. CRC Press, 2002.
- [80] G. Deng and L. Cahill, “An adaptive gaussian filter for noise reduction and edge detection,” in *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record*. IEEE, 1993, pp. 1615–1619.
- [81] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, “Bilateral Filtering: Theory and Applications,” *Foundations and Trends® in Computer Graphics and Vision*, vol. 4, no. 1, pp. 1–75, 2008. [Online]. Available: <http://www.nowpublishers.com/article/Details/CGV-020>
- [82] C. Tomasi and R. Manduchi, “Bilateral Filtering for Gray and Color Images,” *International Conference on Computer Vision*, pp. 839–846, 1998. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=710815&isnumber=15374>

- [83] C.-F. Westin, H. Knutsson, and R. Kikinis, “Adaptive Image Filtering,” *Handbook of Medical Image Processing and Analysis*, pp. 19–33, 2008. [Online]. Available: <http://dx.doi.org/10.1016/B978-012077790-7/50005-9>
- [84] K. He, J. Sun, and X. Tang, “Guided image filtering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.
- [85] K. Dabov, A. Foi, and V. Katkovnik, “Image denoising by sparse 3D transformation-domain collaborative filtering,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 1–16, 2007. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp={&}arnumber=4271520>
- [86] S. Ryan Fanello, C. Keskin, P. Kohli, S. Izadi, J. Shotton, A. Criminisi, U. Pattacini, and T. Paek, “Filter forests for learning data-dependent convolutional kernels,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1709–1716.
- [87] H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising: Can plain neural networks compete with bm3d?” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2392–2399.
- [88] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, “Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map,” *Medical Image Analysis*, vol. 17, no. 8, pp. 907–928, 2013.
- [89] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, “Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation.” *Optics express*, vol. 18, no. 18, pp. 19 413–28, 2010. [Online]. Available: <http://www.osapublishing.org/viewmedia.cfm?uri=oe-18-18-19413{&}seq=0{&}html=true>
- [90] M. A. Mayer, R. P. Tornow, R. Bock, J. Hornegger, F. E. Kruse, F. Mayer, MA., Tornow, RP., Bock, R., Hornegger, J. & Kruse, M. A. Mayer, R. P. Tornow, J. Hornegger, and F. E. Kruse, “Automatic Nerve Fiber Layer Segmentation and Geometry Correction on Spectral Domain OCT Images Using Fuzzy C-Means Clustering,” *Invest. Ophthalmol. Vis. Sci.*, vol. 49, no. 23, pp. E-abstract 1880, 2008. [Online]. Available: <http://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2008/Mayer08-FCC.pdf>

References

- [91] S. J. Chiu, J. A. Izatt, R. V. O'Connell, K. P. Winter, C. A. Toth, and S. Farsiu, "Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images," *Investigative Ophthalmology and Visual Science*, vol. 53, no. 1, pp. 53–61, 2012.
- [92] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomedical optics express*, vol. 6, no. 4, pp. 1172–1194, 2015.
- [93] P. P. Srinivasan, S. J. Heflin, J. A. Izatt, V. Y. Arshavsky, and S. Farsiu, "Automatic segmentation of up to ten layer boundaries in SD-OCT images of the mouse retina with and without missing layers due to pathology," *Biomedical Optics Express*, vol. 5, no. 2, pp. 348–365, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3920868/>
- [94] P. A. Dufour, L. Ceklic, H. Abdillahi, S. Schroder, S. De Dzanet, U. Wolf-Schnurrbusch, and J. Kowal, "Graph-based multi-surface segmentation of oct data using trained hard and soft constraints," *IEEE transactions on medical imaging*, vol. 32, no. 3, pp. 531–543, 2013.
- [95] F. Rathke, S. Schmidt, and C. Schnörr, "Probabilistic intra-retinal layer segmentation in 3-d oct images using global shape regularization," *Medical Image Analysis*, vol. 18, no. 5, pp. 781–794, 2014.
- [96] J. Comander, M. Gardiner, and J. Loewenstein, "High-resolution optical coherence tomography findings in solar maculopathy and the differential diagnosis of outer retinal holes," *American Journal of Ophthalmology*, vol. 152, no. 3, pp. 413–419, 2011.
- [97] Q. Dai and Y. Sun, "Automated layer segmentation of optical coherence tomography images," in *Proceedings - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011*, vol. 1, 2011, pp. 142–146.
- [98] S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, and C. A. Toth, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," in *Ophthalmology*, vol. 121, no. 1, 2014, pp. 162–172.

- [99] A. R. Fuller, R. J. Zawadzki, S. Choi, D. F. Wiley, J. S. Werner, and B. Hamann, “Segmentation of three-dimensional retinal image data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1719–1726, 2007.
- [100] M. Shahidi, Z. Wang, and R. Zelkha, “Quantitative thickness measurement of retinal layers imaged by optical coherence tomography,” *American Journal of Ophthalmology*, vol. 139, no. 6, 2005.
- [101] A. Ehnes, Y. Wenner, C. Friedburg, M. N. Preising, W. Bowl, W. Sekundo, E. M. Zu Bexten, K. Stieger, and B. Lorenz, “Optical Coherence Tomography (OCT) Device Independent Intraretinal Layer Segmentation.” *Translational vision science & technology*, vol. 3, no. 1, p. 1, 2014.
- [102] G. M. Somfai, E. Tátrai, L. Laurik, B. Varga, V. Ölvedy, H. Jiang, J. Wang, W. E. Smiddy, A. Somogyi, and D. C. DeBuc, “Automated classifiers for early detection and diagnosis of retinopathy in diabetic eyes.” *BMC bioinformatics*, vol. 15, no. 1, p. 106, 2014. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84899472145&partnerID=tZOTx3y1>
- [103] J. Massich, C. Y. Cheung, K. Alsaih, and G. Lema  , “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)*, no. August, pp. 1344–1347, 2016.
- [104] A. M. Syed, T. Hassan, M. U. Akram, S. Naz, and S. Khalid, “Automated diagnosis of macular edema and central serous retinopathy through robust reconstruction of 3D retinal surfaces,” *Computer Methods and Programs in Biomedicine*, vol. 137, pp. 1–10, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.cmpb.2016.09.004>
- [105] N. Anantrasirichai, A. Achim, J. E. Morgan, I. Erchova, and L. Nicholson, “Svm-Based Texture Classification in Optical Coherence Tomography,” *IEEE 10th International Symposium on Biomedical Imaging*, no. APRIL, pp. 1320–1323, 2013.
- [106] S. T. Devarakonda, K. K. Vupparaboina, A. Richhariya, J. Chhablani, and S. Jana, “Automated Detection of Retinal Disorders from OCT Images using Artificial Neural Network,” *2016 IEEE Annual India Conference (INDICON)*, 2016.
- [107] S. Sankar, D. Sidib  , Y. Cheung, T. Y. Wong, E. Lamoureux, D. Milea, and F. Meriaudeau, “Classification of SD-OCT volumes for DME detection: an

References

- anomaly detection approach,” *Medical Imaging - Proceedings of SPIE*, vol. 9785, p. 97852O, 2016. [Online]. Available: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2216215>
- [108] F. A. Medeiros, L. M. Zangwill, C. Bowd, R. M. Vessani, R. Susanna Jr., and R. N. Weinreb, “Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography,” *Am J Ophthalmol*, vol. 139, no. 1, pp. 44–55, 2005.
- [109] Q. Chen, T. Leng, L. Zheng, L. Kutzscher, J. Ma, L. De Sisterne, and D. L. Rubin, “Automated drusen segmentation and quantification in SD-OCT images,” *Medical Image Analysis*, vol. 17, no. 8, pp. 1058–1072, 2013.
- [110] A. Albarak, F. Coenen, and Y. Zheng, “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)*, 2013, pp. 59–64.
- [111] Y. Jian, R. J. Zawadzki, and M. V. Sarunic, “Adaptive optics optical coherence tomography for in vivo mouse retinal imaging.” *Journal of biomedical optics*, vol. 18, no. 5, p. 56007, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3710740/>
- [112] L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu, “Fast acquisition and reconstruction of optical coherence tomography images via sparse representation,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 11, pp. 2034–2049, 2013.
- [113] D. Alonso-Caneiro, S. A. Read, and M. J. Collins, “Speckle reduction in optical coherence tomography imaging by affine-motion image registration,” *Journal of biomedical optics*, vol. 16, no. 11, pp. 116027–1–1160275, 2011.
- [114] M. A. Mayer, A. Borsdorf, M. Wagner, J. Hornegger, C. Y. Mardin, and R. P. Tornow, “Wavelet denoising of multiframe optical coherence tomography data,” *Biomedical optics express*, vol. 3, no. 3, pp. 572–589, 2012.
- [115] H. S. Hou and H. C. Andrews, “Cubic Splines for Image Interpolation and Digital Filtering,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 6, pp. 508–517, 1978.

References

- [116] G. Gregori and R. Knighton, “A robust algorithm for retinal thickness measurements using optical coherence tomography (stratus oct),” *Investigative Ophthalmology & Visual Science*, vol. 45, no. 13, pp. 3007–3007, 2004.
- [117] M. K. Garvin, M. D. Abràmoff, R. Kardon, S. R. Russell, X. Wu, and M. Sonka, “Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search,” *IEEE Transactions on Medical Imaging*, vol. 27, no. 10, pp. 1495–1505, 2008.
- [118] D. C. Fernandez, N. Villate, C. A. Puliafito, and P. J. Rosenfeld, “Comparing total macular volume changes measured by optical coherence tomography with retinal lesion volume estimated by active contours,” *Iovs*, vol. 45, p. U61, 2004.
- [119] V. J. Srinivasan, B. K. Monson, M. Wojtkowski, R. A. Bilonick, I. Gorczynska, R. Chen, J. S. Duker, J. S. Schuman, and J. G. Fujimoto, “Characterization of outer retinal morphology with high-speed, ultrahigh-resolution optical coherence tomography,” *Investigative Ophthalmology and Visual Science*, vol. 49, no. 4, pp. 1571–1579, 2008.
- [120] K. Lee, M. D. Abràmoff, M. Niemeijer, M. K. Garvin, and M. Sonka, “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*, vol. 7626, pp. 76260V–76260V–8, 2010. [Online]. Available: <http://link.aip.org/link/PSISDG/v7626/i1/p76260V/s1{&}Agg=doi>
- [121] K. L. Boyer, A. Herzog, and C. Roberts, “Automatic recovery of the optic nerve-head geometry in optical coherence tomography,” *IEEE Transactions on Medical Imaging*, vol. 25, no. 5, pp. 553–570, 2006.
- [122] H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, and J. S. Schuman, “Macular segmentation with optical coherence tomography,” *Investigative Ophthalmology and Visual Science*, vol. 46, no. 6, pp. 2012–2017, 2005.
- [123] A. M. Bagci, M. Shahidi, R. Ansari, M. Blair, N. P. Blair, and R. Zelkha, “Thickness Profiles of Retinal Layers by Optical Coherence Tomography Image Segmentation,” *American Journal of Ophthalmology*, vol. 146, no. 5, 2008.
- [124] A. Mishra, A. Wong, K. Bizheva, and D. A. Clausi, “Intra-retinal layer segmentation in optical coherence tomography images.” *Optics Express*, vol. 17,

References

- no. 26, pp. 23 719–28, 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20052083>
- [125] D. C. Adler, T. H. Ko, and J. G. Fujimoto, “Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter,” *Optics letters*, vol. 29, no. 24, pp. 2878–2880, 2004.
- [126] V. Zlokolica, L. Jovanov, A. Pižurica, P. De Keyser, F. Dhaenens, and W. Philips, “Wavelet-based denoising for 3d oct images,” in *Optical Engineering+ Applications*. International Society for Optics and Photonics, 2007, pp. 66 960P–66 960P.
- [127] V. Gupta, C. C. Chan, C.-L. Poh, T. H. Chow, T. C. Meng, and N. B. Koon, “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, 2008, pp. 120–123.
- [128] M. Baroni, P. Fortunato, and A. La Torre, “Towards quantitative analysis of retinal features in optical coherence tomography.” *Medical engineering & physics*, vol. 29, no. 4, pp. 432–41, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1350453306001214>
- [129] R. Kafieh and H. Rabbani, “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, 2013, pp. 885 826–885 826.
- [130] L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu, “Sparsity based denoising of spectral domain optical coherence tomography images,” *Biomedical optics express*, vol. 3, no. 5, pp. 927–942, 2012.
- [131] H. Takeda, S. Farsiu, and P. Milanfar, “Kernel regression for image processing and reconstruction,” *IEEE Transactions on image processing*, vol. 16, no. 2, pp. 349–366, 2007.
- [132] K. Zhang, D. Tao, X. Gao, X. Li, and Z. Xiong, “Learning multiple linear mappings for efficient single image super-resolution,” *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 846–861, 2015.
- [133] A. Kanamori, M. Nakamura, M. F. Escano, R. Seya, H. Maeda, and A. Negi, “Evaluation of the glaucomatous damage on retinal nerve fiber layer thickness

- measured by optical coherence tomography,” *American journal of ophthalmology*, vol. 135, no. 4, pp. 513–520, 2003.
- [134] W. Drexler and J. G. Fujimoto, “State-of-the-art retinal optical coherence tomography,” pp. 45–88, 2008.
- [135] A. Savitzky and M. J. Golay, “Smoothing and differentiation of data by simplified least squares procedures.” *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [136] A. Bellet, A. Habrard, and M. Sebban, “A survey on metric learning for feature vectors and structured data,” *arXiv preprint arXiv:1306.6709*, 2013.
- [137] J. C. Bezdek, R. Ehrlich, and W. Full, “FCM: The fuzzy c-means clustering algorithm,” *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0098300484900207>
- [138] L.-v.-d. Maaten and G. Hinton, “Visualizing data using t-sne,” *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [139] N. Dalal and B. Triggs, “Histogram of Oriented Gradients for Human Detection,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, vol. 1, 2005, pp. 886–893. [Online]. Available: <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1467360&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs%2Fall.jsp%3Farnumber%3D1467360>
- [140] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [141] L. A. Gatys, A. S. Ecker, and M. Bethge, “A neural algorithm of artistic style,” *arXiv preprint arXiv:1508.06576*, 2015.
- [142] J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.
- [143] G. Mesnil, Y. Dauphin, X. Glorot, S. Rifai, Y. Bengio, I. J. Goodfellow, E. Lavoie, X. Muller, G. Desjardins, D. Warde-Farley *et al.*, “Unsupervised and transfer learn-

References

- ing challenge: a deep learning approach.” *ICML Unsupervised and Transfer Learning*, vol. 27, pp. 97–110, 2012.
- [144] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” in *Advances in neural information processing systems*, 2014, pp. 3320–3328.
- [145] S. Yun, G. Tearney, J. De Boer, and B. Bouma, “Motion artifacts in optical coherence tomography with frequency-domain ranging,” *Optics Express*, vol. 12, no. 13, pp. 2977–2998, 2004.
- [146] S. Chitchian, M. A. Mayer, A. R. Boretsky, F. J. van Kuijk, and M. Motamedi, “Retinal optical coherence tomography image enhancement via shrinkage denoising using double-density dual-tree complex wavelet transform,” *Journal of biomedical optics*, vol. 17, no. 11, pp. 116 009–116 009, 2012.
- [147] M. R. Avanaki, M. J. Marques, A. Bradu, A. Hojjatoleslami, and A. G. Podoleanu, “A new algorithm for speckle reduction of optical coherence tomography images,” in *SPIE BiOS*. International Society for Optics and Photonics, 2014, pp. 893 437–893 437.
- [148] L. Bian, J. Suo, F. Chen, and Q. Dai, “Multiframe denoising of high-speed optical coherence tomography data using interframe and intraframe priors,” *Journal of biomedical optics*, vol. 20, no. 3, pp. 036 006–036 006, 2015.
- [149] R. Rigamonti, A. Sironi, V. Lepetit, and P. Fua, “Learning separable filters,” in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 2754–2761.
- [150] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 184–199.

