A Supervised Joint Multi-layer Segmentation Framework for Retinal Optical Coherence Tomography Images using Conditional Random Field

Arunava Chakravarty[1], Jayanthi Sivaswamy

Centre for Visual Information Technology, International Institute of Information Technology Hyderabad, India-500032

Abstract

Background and Objective: Accurate segmentation of the intra-retinal tissue layers in Optical Coherence Tomography (OCT) images plays an important role in the diagnosis and treatment of ocular diseases such as Age-Related Macular Degeneration (AMD) and Diabetic Macular Edema (DME). The existing energy minimization based methods employ multiple, manually handcrafted cost terms and often fail in the presence of pathologies. In this work, we eliminate the need to handcraft the energy by learning it from training images in an end-to-end manner. Our method can be easily adapted to pathologies by re-training it on an appropriate dataset.

Methods: We propose a Conditional Random Field (CRF) framework for the joint multi-layer segmentation of OCT B-scans. The appearance of each retinal layer and boundary is modeled by two convolutional filter banks and the shape priors are modeled using Gaussian distributions. The total CRF energy is linearly parameterized to allow a joint, end-to-end training by employing the Structured Support Vector Machine.

Results: The proposed method outperformed three benchmark algorithms on four public datasets. The *NORMAL-1* and *NORMAL-2* datasets contain healthy OCT B-scans while the *AMD-1* and *DME-1* dataset contain B-scans of AMD and DME cases respectively. The proposed method achieved an average unsigned boundary localization error (U-BLE) of 1.52 pixels on

Email addresses: arunava.chakravarty@research.iiit.ac.in (Arunava Chakravarty), jsivaswamy@iiit.ac.in (Jayanthi Sivaswamy)

[1]corresponding author

NORMAL-1, 1.11 pixels on *NORMAL-2* and 2.04 pixels on the combined *NORMAL-1* and *DME-1* dataset across the eight layer boundaries, outperforming the three benchmark methods in each case. The Dice coefficient was 0.87 on NORMAL-1, 0.89 on *NORMAL-2* and 0.84 on the combined *NORMAL-1* and *DME-1* dataset across the seven retinal layers. On the combined *NORMAL-1* and *AMD-1* dataset, we achieved an average U-BLE of 1.86 pixels on the ILM, inner and outer RPE boundaries and a Dice of 0.98 for the $ILM\text{-}RPE_{in}$ region and 0.81 for the RPE layer.
Conclusion: We have proposed a supervised CRF based method to jointly segment multiple tissue layers in OCT images. It can aid the ophthalmologists in the quantitative analysis of structural changes in the retinal tissue layers for clinical practice and large-scale clinical studies.

Keywords: Optical Coherence Tomography, Conditional Random Field, Structured Support Vector Machines, Diabetic Macular Edema, Age-Related Macular Degeneration.

---

1. Introduction

Optical Coherence Tomography (OCT) is a non-invasive imaging modality that provides a 3D, cross-sectional view of the tissue lining the retina using infra-red light reflectivity [1]. It plays an important role in the diagnosis of ocular diseases such as glaucoma [2], Diabetic Macular Edema (DME) [3] and Age-Related Macular Degeneration (AMD) [4]. The OCT volumes are composed of multiple cross-sectional images called the *B-scans*. Each B-scan comprises a series of image columns called the *A-scans* that lies in the direction parallel to the propagation of light into the tissue. The intra-retinal tissue is a multi-layered structure which transforms light into neural signals for further use by the brain. It is commonly divided into 7 adjacent layers [5], separated by 8 boundaries as depicted in Fig. 1. The boundaries ordered from the top to bottom are the : i) Inner Limiting Membrane (ILM) separating the vitreous and Nerve Fiber Layer(NFL), ii) NFL/GCL boundary separating NFL from the Ganglion Cell and Inner Plexicon layer (GCL-IPL), iii) IPL/INL separating GCL-IPL from the Inner Nuclear Layer (INL), iv) INL/OPL separating INL from the Outer Plexiform Layer (OPL), v) OPL/ONL separating OPL from the Outer Nuclear and Inner Segment (ONL-IS) region, vi) IS/OS separating ONL-IS from the Outer Segment (OS) vii)RPE_{in} separating OS from the Retinal Pigment Epithelium (RPE)

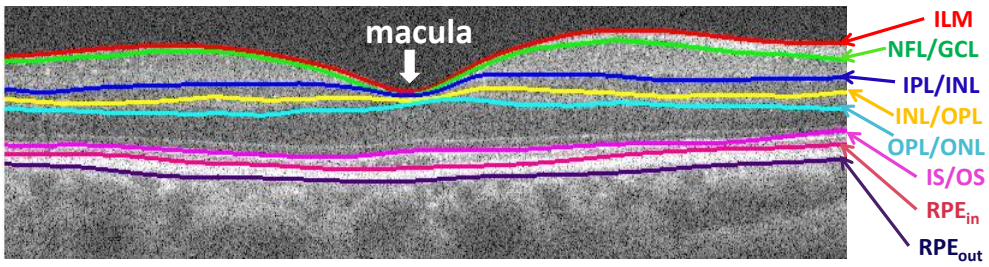

Figure 1: Retinal layer boundaries in a macular OCT B-scan.

layer and finally the viii) RPE_{out} boundary separating RPE from the choroid.

The accurate segmentation of these layers is necessary to quantify the morphological changes in the retinal tissue that characterize the presence and progression of various ocular diseases. These morphological changes have also been correlated to neuro-degenerative diseases such as Multiple Sclerosis [6]. While the development of spectral domain OCT has led to the fast acquisition of a large number of B-scans per OCT volume in a short duration, their manual segmentation is a tedious, time-consuming and subjective task. Most commercial OCT systems are equipped to segment only two or three layers and often fail in the presence of pathologies [7].

The main challenges in the automated layer segmentation are depicted in Fig. 2 and consist of the speckle noise, vessel shadows, indistinct layer boundaries, inter-scanner variations and the presence of pathologies. The vessel shadows are caused by the occurence of a retinal blood vessel at the beginning of an A-scan which absorbs the reflected light from locations beneath it [8]. Ocular diseases also lead to significant changes in the tissue morphology. In AMD, the drusen deposits in the RPE layer lead to irregularities and undulations in the RPE_{in} boundary [4] as depicted in Fig. 2 b. DME is characterized by the presence of fluid-filled regions [9],[10] in the OPL and INL layers around the macula leading to the swelling of the retinal tissue as shown in Fig. 2 c.

In this paper we extend our preliminary work in [11] by refining the method, adapting it to images with fluid-filled regions associated with DME and finally presenting a more comprehensive experimental evaluation with cross-testing across datasets acquired using different OCT imaging scanners. We explore a supervised Conditional Random Field(CRF) based framework for the joint multi-layer segmentation in OCT B-scans. The CRF energy con-

sists of multiple cost terms to capture the appearance and the shape priors for each layer. The appearance is captured by two convolutional filter banks, one to give high response at specific layer boundaries and the other to capture the appearance of each intermediate tissue region. The shape priors on the boundary smoothness and the thickness of each layer are modeled using Gaussian distributions. The CRF energy is *linearly parameterized* to allow a joint, end-to-end training of its constituent cost terms (both the filter banks and the relative weights of the shape priors) by employing a Structured Support Vector Machine(StructSVM) formulation. Being supervised in nature, our method can be easily *adapted* to different pathologies by training it on appropriate images without the need for handcrafting.

To summarize, the contributions of this paper are as follows. First, we propose a joint segmentation framework based on a novel CRF formulation that extracts all retinal boundaries in a single optimization step. Second, a supervised strategy is explored to learn the CRF energy in a joint, end-to-end manner, eliminating the need to handcraft the individual cost terms or fine-tune their relative weights. Third, the robustness of the method has been demonstrated on data acquired across multiple centres with different scanners and at different resolutions. Finally, We have also evaluated the adaptibility of the proposed method to the morphological changes in the presence of pathologies related to AMD and DME. The MATLAB implementation of the proposed method is also being made available for public research use and can be downloaded from `https://researchweb.iiit.ac.in/~arunava.chakravarty/CRF_OCT`

2. Background

The initial attempts to segment the layers in retinal OCT images employed simple image processing techniques and focussed on the segmentation of only few (2-4) prominent layers. Each A-scan in the OCT slice was segmented individually based on peak, valley and/or signed gradient analysis of the intensity profile. This was followed by regularization across adjacent columns based on rule based heuristics [12] or iterative refinement by incorporating 3D information [13]. Their performance suffered due to the lack of a strong intensity gradient at the boundaries and overlapping intensities between the adjacent layers.

To overcome these challenges, deformable models have been proposed to incorporate shape priors in addition to the appearance of the retinal layers.

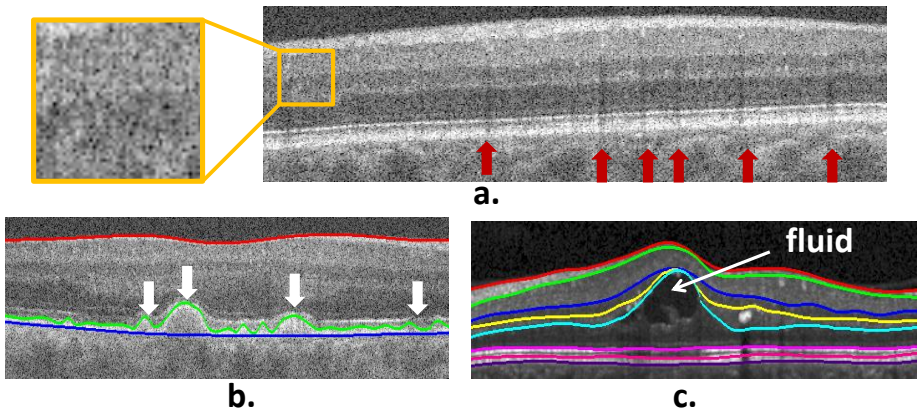

Figure 2: a. A retinal B-scan: a local patch is enlarged to depict the speckle noise and indistinct layer boundaries; the vessel shadows are indicated by red arrows. b. ILM(Red), RPE_{in} (Green) and RPE_{out} (Blue) boundaries in a B-scan with AMD; irregularities in the RPE_{in} boundary are indicated by white arrows. c. The 8 layer boundaries in a B-scan with DME; the fluid-filled region is indicated by a white arrow.

The edge based active contour models were explored in [14], [15]. While constraints on the layer thickness and boundary smoothness were imposed in [14] using a coupled level set framework, an approximate parallelism constraint between the adjacent layer boundaries was incorporated in [15]. The region based Chan-Vese model was adapted in [16] to simultaneously segment multiple layers using a circular arc based shape regularization. An active appearance based statistical model for the layer shape and texture has also been explored in [17]. However, these methods require a good initial estimate of the layer boundaries failing which the deformable model can be entrapped in a local Energy minima or require a high convergence time.

Graph based optimization methods have also been explored for layer segmentation. In [5] the layers in each OCT B-scan were segmented by finding the shortest path in a specially constructed undirected graph using the Dijkstra's algorithm. The nodes in the graph represented the pixels and the edge weights were defined using the gradient intensity and the eucledian distance between the pixels in the image. It employed a sequential approach, restricting the search space for each layer based on the previously segmented layers. This method has been further refined in [18] to reduce the computation time by leveraging the spatial dependency between the adjacent B-scans of an OCT volume. Another method explored in [19], [20],[21] employs an energy

minimization approach which is formulated as a Minimum Cost Closed Set (MCCS) problem on a specially constructed geometric graph. The Energy comprises multiple cost terms defined to capture the appearance and shape priors for each layer. However, each cost term is *handcrafted* manually and then combined using empirically determined relative weights.

Most of the deformable model or graph based energy minimization methods fail in the presence of pathologies. Few attempts have been made to adapt these methods by incorporating handcrafted disease specific modifications to handle a particular pathology. The Dijkstra's shortest path based method [5] has been adapted for DME in [22] by employing an explicit segmentation of the fluid-filled regions and attempts have also been made to adapt it for AMD cases in [23]. Similarly, the geometric graph based method has been adapted in [24] to handle Serous Pigment Epithelial Detachments and the coupled level set based deformable model [14] has been extended to handle DME cases in [25] by modelling the lesions as an additional space-variant layer delineated by auxiliary interfaces. However, designing a single method that works equally well on both healthy and abnormal cases without the prior knowledge of the presence and the type of abnormalities in the OCT B-scan still remains an open problem.

Recently, deep learning based methods have also been explored for this task. In [26], a Convolutional Neural Network(CNN) based on [27] was applied to 33×33 image patches extracted from the OCT B-scans to obtain the probabilities of the central pixel in the patch for each layer boundary. Alternatively, fully convolutional networks(FCN) have been employed in [28], [29] and [30] to obtain the probability maps for each tissue layer. Unlike the patch based method, FCNs estimate the per-pixel class probabilities for the entire image in a single forward pass.

The U-net architecture [31] was employed in [28] and [29] which consists of a contracting path of encoder blocks followed by an expansive path of decoder blocks. Additional skip connections are also employed in these architectures to directly provide the output feature maps from each encoder layer as input to the corresponding decoder layer. The *ReLayNet* architecture in [28] is comprised of an encoder with four convolutional layers of 64, 7×3 filters separated by max-pooling layers. The decoder consists of a sequence of four unpooling followed by convolutional layers (each with 64, 7×3 filters) to successively restore the resolution of the feature maps. Though the entire OCT B-scan was provided as input during testing, the *ReLayNet* was trained on sub-images obtained by slicing the B-scans width-wise into

a set of non-overlapping regions consisting of 64 A-scans. Alternatively, in [30], the Dense-net [32] architecture was explored where, for each layer, the feature-maps of all the preceding layers were used as direct inputs using skip connections.

The main advantage of CNNs is their ability to learn hierarchical features in a data-driven end-to-end manner. However, since CNNs pose segmentation as a pixel-labelling problem, they cannot explicitly incorporate any shape priors to capture the boundary smoothness or model the dependencies between the adjacent boundaries. As a result, complex post-processing methods have been employed to refine the probability maps obtained from the CNNs. For eg., in [26] and [29], the dijsktra's shortest path based energy minimization method was modified to refine the layer boundaries obtained from the CNN, while [30] employed a Gaussian process based regression for the same. Moreover, CNN architectures generally require a large amount of training data, often in the order of hundreds or thousands of training samples to prevent over-fitting.

In this work, we propose a method that combines the advantages of both the CNN and the energy minimization based methods. Similar to CNNs, the proposed supervised CRF formulation learns two sets of filter banks to capture the appearance of each tissue layer and their boundaries, thereby eliminating the need to handcraft the CRF energy. Additionally, it offers some key advantages over a CNN. In comparison to the CNN architectures in [26] and [28] which employ 85,578 and 900,170 network parameters respectively, the proposed CRF energy is linearly parameterized using only 5,438 learnable parameters. Fewer parameters reduce the risk of overfitting and allow the models to be trained using very few training samples. For example, only 87 and 55 training samples were employed to train the proposed method in the experiments reported in Sections 5.2 and 5.6 respectively.

Another key advantage of the proposed method over CNN is its ability to explicitly incorporate shape priors on the layer boundaries, thereby eliminating the need for any addititonal post-processing. The distance in height between the adjacent points in each layer boundary and the expected thickness of each tissue region are modelled as Gaussian distributions and large deviations of the segmentation result from these distributions are penalized within the proposed CRF energy. Hard constraints are also employed within the CRF framework to maintain the anatomically correct ordering of the tissue layers and prevent the intersection of the layer boundaries. The relative weights between the appearance and the shape prior based cost terms are

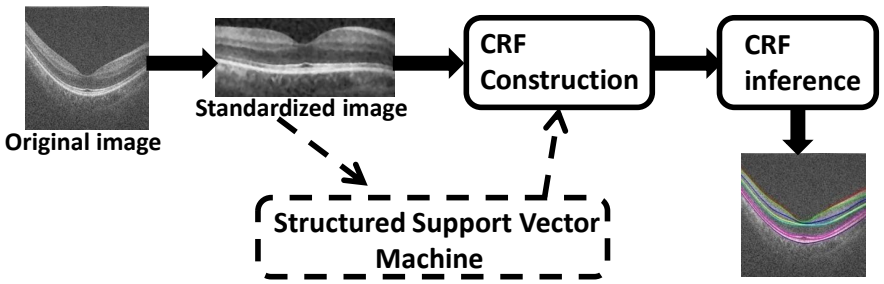

Figure 3: Overview of the proposed joint multi-layer segmentation pipeline. The supervised training is indicated by dashed lines.

automatically learned during the SSVM based end-to-end training. Unlike CNNs, the SSVM objective function used to train the CRF is convex and has a unique (global) optima which leads to a more efficient and robust training. As a result, our method has a low computational requirement and can be trained without a GPU.

The advantages of the proposed method over the existing energy minimization based methods are as follows. Since the learnt energy is optimally adapted to the segmentation problem, our method outperforms the existing energy minimization methods with similar but handcrafted energy cost terms such as [20]. Moreover, our CRF formulation efficiently incorporates both hard and soft constraints on the shape priors using a single undirected edge as opposed to the MCCS formulation that requires additional directed edges in the graph construction to incorporate soft constraints [21]. Finally, in this work all layers are jointly segmented in a single optimization step. In contrast, [20] takes a two-step approach by segmenting the outer layer boundaries (ILM, RPE_{in} and RPE_{out}) first, followed by the remaining inner layers, while [4], [5] segments each layer sequentially.

3. Methods

An overview of the proposed method is presented in Fig. 3. A preprocessing step described in Section 3.1 is applied to standardize both the training and test OCT B-scan images. The joint extraction of the multiple layer boundaries is formulated within a CRF based Energy Minimization framework in Section 3.2. During training, the CRF energy is linearly parameterized as detailed in Section 3.3 and its parameters are learnt in a supervised, end-to-end manner by posing it as a StructSVM optimization

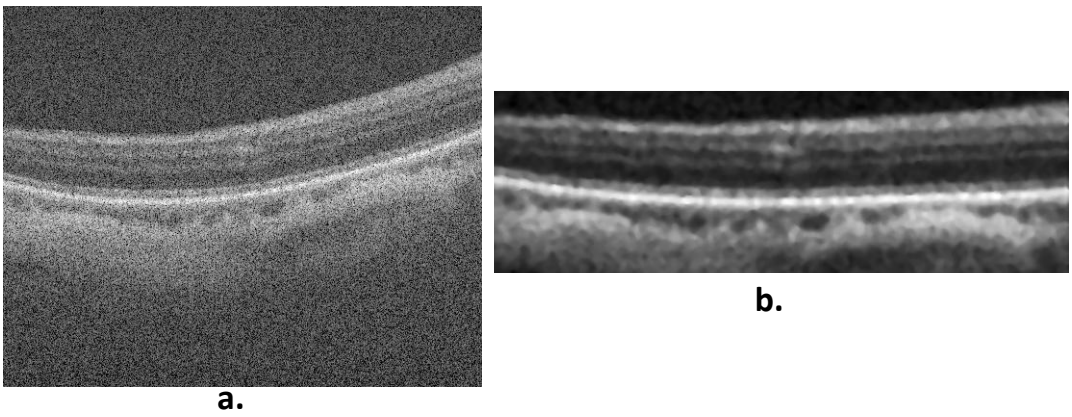

Figure 4: a. Raw OCT B-scan; b. Corresponding preprocessed Region of Interest.

problem as discussed in Section 3.4. During testing, the optimal labelling of the CRF is inferred to extract the multiple layer boundaries in a single optimization step.

3.1. Image Preprocessing

The preprocessing step involves the flatenning of the retinal curvature, extraction of the region of interest (ROI) containing the retinal tissue, reducing the speckle noise and intensity standardization of the OCT B-scans. Retinal curvature flatenning is a crucial preprocessing step that reduces the variations in the spatial location of the retinal tissue across the A-scans in an OCT image. It aids in obtaining a tighter ROI thereby reducing the time and memory requirements and also provides a more consistent shape of the layers for segmentation. Each B-scan is flatenned using the method employed in [5] as follows. At first, a rough estimate of the RPE_{out} boundary is obtained by fitting a quadratic polynomial to a set of candidate pixels. Since, RPE_{out} appears as the brightest boundary in the retinal OCT images, the candidate pixels are obtained by detecting the brightest pixel in each A-scan and removing the outliers. Then each column is shifted (cyclically) by an offset such that the detected boundary lies on a straight line.

The retinal tissue is surrounded by a dark background at the top and bottom in each B-scan (see Figure 4 a.). To estimate the ROI, the input image is smoothed by a large Gaussian filter with $\sigma = 9$ to reduce the effect of the speckle noise in the background and remove the dark regions within the retinal tissue. Then the ROI is estimated as the maximum extent of the

largest connected component obtained by thresholding the smoothed image at 0.3 after scaling the pixel intensities in the B-scan to [0,1].

Finally, the speckle noise in the OCT images is reduced by applying the speckle reducing anisotropic diffusion [33] for 30 iterations in time-steps of 0.1 and a histogram based intensity standardization scheme based on [34] is applied to handle the inter and intra-scanner intensity variations. The intensity of each B-scan is scaled to [0,1]. The ROI is resized to 190×600 to handle the variations in image resolution across the images.

3.2. Modelling Joint Multi-Layer Segmentation as a CRF

The joint multi-layer segmentation problem seeks to extract the L layer boundaries in an OCT B-scan image I of size $H \times Y$. Each boundary is labeled from $1 \leq l \leq L$ in the increasing order of its height along H as depicted in Fig. 5. The boundaries are uniformly sampled at N equidistant columns y_n along Y such that the l^{th} boundary intersects y_n at a height of $x_{l,n}$. Each $x_{l,n}$ can be interpreted as a discrete random variable that can take a value from the label set $\Omega = \{1 \leq i \leq H, i \in Z^+\}$, where Z^+ represents the set of positive integers. The set of all random variables is defined as the random field $X = \{x_{l,n} | 1 \leq l \leq L, 1 \leq n \leq N\}$. A feasible *labelling* denoted by $\mathbf{x} \in \Omega^{L \times N}$ can be obtained by assigning an arbitary label from Ω to each $x_{l,n}$ in X. Our objective is to define an energy $E(\mathbf{x}, I)$ over the Random Field X for each image I such that the optimal labelling that maximizes $E(\mathbf{x}, I)$ corresponds to the desired layer boundaries.

$E(\mathbf{x}, I)$ consists of multiple cost terms. A unary boundary cost $\varepsilon^{l}_{bnd}(x_{l,n})$ is defined for each $x_{l,n}$ to capture the likelihood that the l^{th} boundary passes through the point $(x_{l,n}, y_n)$ given the local appearance of the OCT B-scan around that location. The label of each $x_{l,n}$ is also dependent on its immediate neighbors $x_{l,n+1}$ on the same boundary and $x_{l+1,n}$ on the adjacent $(l+1)^{th}$ boundary resulting in a second order CRF with a maximum clique size of 2. An undirected graphical representation of the CRF for a local *4-neighborhood* is depicted in Fig. 5, where each node represents a random variable. By the Markovian property, the labelling of $x_{l,n}$ given the labelling of its immediate 4 neighbors is independent of all other nodes in the graph. The Intra-layer pairwise cost $\varepsilon^{l,n}_{intra}(x_{l,n}, x_{l,n+1})$ captures the smoothness and the similarity in appearance between the adjacent points on a boundary. Additionally, the pairwise Inter-layer Energy Cost $\varepsilon^{l,n}_{inter}(x_{l,n}, x_{l+1,n})$ is defined between the adjacent l and $(l+1)^{th}$ boundaries to enforce the correct layer ordering, the

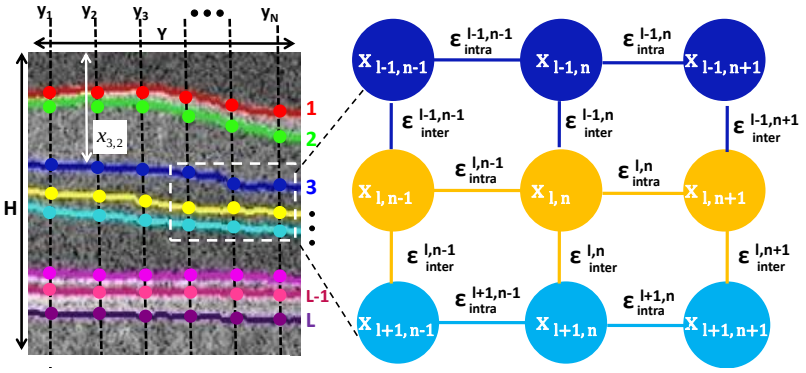

Figure 5: The Conditional Random Field Formulation for joint multi-layer OCT Segmentation.

expected layer thickness and capture the appearance of the tissue layer between them. All the three cost terms are dependent on the observed image I. To simplify the notation, the input arguments for the cost terms have been omitted in the rest of the paper and simply represented by ε^{l}_{bnd}, $\varepsilon^{l,n}_{intra}$ and $\varepsilon^{l,n}_{inter}$ respectively. Thus, the CRF inference problem for I is defined as

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

where $E_{bnd}(\mathbf{x}, I)$, $E_{intra}(\mathbf{x}, I)$ and $E_{inter}(\mathbf{x}, I)$ are the sum of all the unary, intra-layer and the inter-layer cost terms in the entire CRF respectively. During implementation, the CRF inference in eq. 1 is converted into a minimization problem by taking the negative of all the unary and pairwise cost terms, and solved using the Sequential-Tree Reweighted Message Passing algorithm (TRW-S) [35]. Instead of handcrafting the individual cost terms, we parameterize $E(\mathbf{x}, I)$ by a set of parameters θ which can then be learnt from a set of training images. Next, we develop an appropriate definition of $E_\theta(\mathbf{x}, I)$ for the multi-layer OCT segmentation problem.

3.3. Linear Parameterization of CRF Energy

Since E(x,I) is defined as a sum of the Unary and Pairwise Cost terms, its linear parameterization involves the decomposition of each of its cost terms

into a linear function. We define the individual cost terms in eq. 1 as,

$$\begin{aligned} E_{bnd}(\mathbf{x}, I) &= \mathbf{w}_{bnd}^{\top}.F_{bnd}(\mathbf{x}, I), \\ E_{intra}(\mathbf{x}, I) &= \mathbf{w}_{intra}^{\top}.F_{intra}(\mathbf{x}, I) \text{ and} \\ E_{inter}(\mathbf{x}, I) &= \mathbf{w}_{inter}^{\top}.F_{inter}(\mathbf{x}, I). \end{aligned} \tag{2}$$

The details of the linear decomposition of each cost term is detailed below in Sections 3.3.1, 3.3.2 and 3.3.3. Thus, the net CRF energy defined in eq. 1 can be rewritten in the linear form by substituting the definition of the individual cost terms from eq. 2 as

$$\begin{aligned} E_{\theta}(\mathbf{x}, I) &= \mathbf{w}_{bnd}^{\top}.F_{bnd}(\mathbf{x}, I) + \mathbf{w}_{intra}^{\top}.F_{intra}(\mathbf{x}, I) + \mathbf{w}_{inter}^{\top}.F_{inter}(\mathbf{x}, I) \\ &= \theta^{\top}.F(\mathbf{x}, I), \end{aligned} \tag{3}$$

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

3.3.1. Unary Boundary Cost

To capture the image appearance at each layer boundary, we aim to learn a convolutional filter bank $\{\mathbf{u_l}\}_{l=1}^{L}$. Each $\mathbf{u_l}$ is a $p \times p$ filter which should have a high response only at the pixels lying on the l^{th} boundary. Let $\mathbf{I_{l,n}}$ represent a $p \times p$ image patch centered at $(x_{l,n}, y_n)$. Both $\mathbf{u}_l$ and $\mathbf{I_{l,n}}$ are linearly indexed to $p^2 \times 1$ column vectors so that the response of the convolution filter at $(x_{l,n}, y_n)$ is obtained by their dot product. Thus, the boundary cost for each $x_{l,n}$ is defined as $\varepsilon_{bnd}^{l}(x_{l,n}) = \mathbf{u}_l^{\top}.\mathbf{I_{l,n}}$ and the total Boundary cost over the entire CRF is given by

$$E_{bnd}(\mathbf{x}, I) = \sum_{l=1}^{L}\sum_{n=1}^{N} \mathbf{u}_l^{\top}.\mathbf{I_{l,n}} = \sum_{l=1}^{L} \mathbf{u}_l^{\top} \left\{ \sum_{n=1}^{N} \mathbf{I_{l,n}} \right\} = \mathbf{w}_{bnd}^{\top}.F_{bnd}(\mathbf{x}, I), \tag{4}$$

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

3.3.2. Pairwise Intra-Layer Cost

The interaction between each pair of adjacent points $x_{l,n}$ and $x_{l,n+1}$ on the l^{th} boundary is modeled as a linear combination of a shape prior and an appearance term. The shape prior between $(x_{l,n}, x_{l,n+1})$ is a soft constraint that penalizes large deviations of the signed gradient of the height

values $(x_{l,n+1} - x_{l,n})$ to preserve the local smoothness of the l^{th} boundary. The deviation is modeled by a Gaussian function $d^{l,n}_{intra}(x_{l,n}, x_{l,n+1}) = \exp\left\{-\frac{1}{2}.(\frac{(x_{l,n+1}-x_{l,n})-\mu^{l,n}_{intra}}{\sigma^{l,n}_{intra}})^2\right\}$. The mean $\mu^{l,n}_{intra}$ and the standard deviation $\sigma^{l,n}_{intra}$ of the signed gradient are pre-computed for each layer l and column y_n using the ground truth layer markings of the training images.

The presence of abnormalities such as AMD and DME has an adverse effect on the boundary smoothness as depicted in Fig. 2 b,c. Hence, using the shape priors alone can lead to large segmentation errors in such cases. To overcome this, an additional term is introduced to favour labellings where the adjacent boundary points are also similar in appearance. The similarity term $S(x_{l,n}, x_{l,n+1})$ measures the dissimilarity between the two $p \times p$ image patches centered at the adjacent boundary points $(x_{l,n}, y_n)$ and $(x_{l,n+1}, y_{n+1})$. A histogram intersection [36] based dissimilarity measure is defined as $S(x_{l,n}, x_{l,n+1}) = 1 - min \sum_{k=1}^{255} min\{h_k(x_{l,n}, y_n), h_k(x_{l,n+1}, y_{n+1})\}$, where h_k represents the k^{th} bin of the normalized 255-bin histograms computed over the two image patches.

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

$$\begin{aligned} E_{intra}(\mathbf{x}, I) &= \sum_{l=1}^{L}\sum_{n=1}^{N-1}\left\{\alpha_l.d^{l,n}_{intra}(x_{l,n}, x_{l,n+1}) + \beta_l.S(x_{l,n}, x_{l,n+1})\right\} \\ &= \sum_{l=1}^{L}\alpha_l.\left\{\sum_{n=1}^{N-1} d^{l,n}_{intra}(x_{l,n}, x_{l,n+1})\right\} + \sum_{l=1}^{L}\beta_l.\left\{\sum_{n=1}^{N-1} S(x_{l,n}, x_{l,n+1})\right\} \\ &= \mathbf{w}^{\top}_{intra}.F_{intra}(\mathbf{x}, I). \end{aligned} \quad (5)$$

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

3.3.3. Pairwise Inter-Layer Cost

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

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

$$d_{inter}^{l,n} = \begin{cases} \exp\left\{-\frac{1}{2}.\left(\frac{(x_n^{l+1}-x_n^l)-\mu_{inter}^{l,n}}{\sigma_{inter}^{l,n}}\right)^2\right\}, & if\ T_{mn}^{l} \leq (x_{l+1,n} - x_{l,n}) \leq T_{mx}^{l} \\ -\infty,\ otherwise. \end{cases} \tag{6}$$

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

The second term in the $\varepsilon_{inter}^{l,n}$ seeks to capture the appearance of the intermediate tissue regions lying between the adjacent boundaries. Let R_l denote the tissue region between the l and $(l+1)^{th}$ boundary. The appearance of each of the $L-1$ regions is captured by a convolutional filter bank $\{\mathbf{v_l}\}_{l=1}^{L-1}$. Each $\mathbf{v}_l$ is a $p \times p$ filter that captures the appearance of R_l and has the maximum *average* filter response in each column y_n within R_l . The average filter bank response is obtained using the dot product $\frac{1}{|x_{l+1,n}-x_{l,n}|}(\sum_{j=x_{l,n}}^{x_{l+1,n}} \mathbf{v}_l^{\top}.\mathbf{I_{j,n}})$, where $\mathbf{I_{j,n}}$ represents a $p \times p$ image patch centered at $(x_{j,n}, y_n)$. Both $\mathbf{v}_l$ and $\mathbf{I_{j,n}}$ are linearly indexed to $p^2 \times 1$ column vectors.

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

$$
\begin{aligned}
E_{inter}(\mathbf{x}, I) &= \sum_{l=1}^{L-1}\sum_{n=1}^{N}\left\{\frac{1}{\mid x_{l+1,n} - x_{l,n} \mid}\left(\sum_{j=x_{l,n}}^{x_{l+1,n}} \mathbf{v}_l^{\top}.\mathbf{I_{j,n}}\right) + \gamma_l.d_{inter}^{l,n}\right\} \\
&= \sum_{l=1}^{L-1}\mathbf{v}_l^{\top}\left\{\sum_{n=1}^{N}\frac{1}{\mid x_{l+1,n} - x_{l,n} \mid}\sum_{j=x_{l,n}}^{x_{l+1,n}}\mathbf{I_{j,n}}\right\} + \sum_{l=1}^{L-1}\gamma_l\left\{\sum_{n=1}^{N} d_{inter}^{l,n}\right\} \\
&= \mathbf{w}_{inter}^{\top}.F_{inter}(\mathbf{x}, I). \qquad (7)
\end{aligned}
$$

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

3.4. The Structured Support Vector Machine Formulation

In Section 3.3, the proposed CRF energy for the joint multi-layer segmentation was linearly parameterized into $E_\theta(\mathbf{x}, I) = \theta^{\top}.F(\mathbf{x}, I)$ (eq. 3), where $F(\mathbf{x}, I)$ is known as the joint feature function. The problem of learning the model parameters θ during training can be posed as a structSVM [37] formulation. Let $\{I^{(k)}, \mathbf{x}^{(k)}\}_{k=1}^{K}$ denote a set of K training OCT B-scans, where $I^{(k)}$ denotes the k^{th} training image with the corresponding ground truth (GT) labelling $\mathbf{x}^{(k)}$. Let $\mathcal{Y}_k = \left\{\mathbf{x}|\mathbf{x} \in \Omega^{L\times N}\right\} - \left\{\mathbf{x}^{(k)}\right\}$ denote the set of all feasible but incorrect labellings. To quantify the segmentation error of $\mathbf{x}$, we define a loss function $\Delta(\mathbf{x}^{(k)}, \mathbf{x}) = \sum_{l=1}^{L}\sum_{n=1}^{N} \mid x_{l,n}^{(k)} - x_{l,n} \mid$ as the sum of the unsigned distances between the corresponding labels in $\mathbf{x}$ and the GT $\mathbf{x}^{(k)}$.

$E_\theta(\mathbf{x}, I)$ maps each feasible labelling $\mathbf{x}$ of an image I to a scalar score value. Our objective is to learn a θ such that for each I^k, i) the GT labelling has the maximum score, ie., $\mathbf{x}^{(k)} = \underset{\mathbf{x}}{\text{argmax}}\ E_\theta(\mathbf{x}, I^{(k)})$ and ii) the higher the loss $\Delta(\mathbf{x}^{(k)}, \mathbf{x})$ of a feasible labelling $\mathbf{x} \in \mathcal{Y}_k$, the lower is its energy $E_\theta(\mathbf{x}, I^{(k)})$ with respect to that of the correct labelling. This can be posed as the following StructSVM [37] formulation,

$$
\begin{aligned}
&\underset{\theta,\xi\geq 0}{\text{argmin}} \qquad \frac{\lambda}{2}\mid\mid \theta \mid\mid^2 + \frac{1}{M}\sum_{k=1}^{M}\xi_k \\
&\text{s.t.} \qquad \theta^{\top}.\left\{F(\mathbf{x}^{(k)}, I^{(k)}) - F(\mathbf{x}, I^{(k)}\right\} \geq \Delta(\mathbf{x}^{(k)}, \mathbf{x}) - \xi_k \quad \forall k,\, \forall \mathbf{x} \in \mathcal{Y}_k, \qquad (8)
\end{aligned}
$$

where ξ_k are the slack variables. The $L2$ regularization on θ is employed to ensure good generalization on unseen test images and λ is the regularization

weight. The constraints in eq. 8 ensure that the difference in E_θ between the GT and each incorrect labelling is greater than a margin which is scaled by the loss function.

Though eq. 8 is a convex Quadratic Programming(QP) Problem, it cannot be solved directly due to the extremely large number of constraints. For each training sample, there are an exponentially large number of possible incorrect labellings in $\mathcal{Y}_k$ resulting in a total of $\sum_k |\mathcal{Y}_k|$) constraints.

Hence an iterative Block Co-ordinate Frank Wolfe Algorithm [38] is employed to make the optimization tractable. In each epoch, for each image I_k, the $|\mathcal{Y}_k|$ constraints are replaced by a single *most violating* constraint obtained by keeping the θ fixed. Then θ is updated using gradient descent while considering the most violating constraint alone. We refer to [38] for further details on the update equations for θ and next discuss the method for obtaining the most violated constraint in our case.

By substituting $E_\theta(\mathbf{x}, I) = \theta^\top.F(\mathbf{x}, I)$, the constraints in eq. 8 for each I_k can be rearranged as $\xi_k \geq \Delta(\mathbf{x}^{(k)}, \mathbf{x}) - \left\{E_\theta(\mathbf{x}^{(k)}, I^{(k)}) - E_\theta(\mathbf{x}, I^{(k)}\right\}, \quad \forall \mathbf{x} \in \mathcal{Y}_k$. For a given θ, these $|\mathcal{Y}_k|$ constraints can be replaced by a single most violating constraint of the form $\xi_k \geq H_i$, where H_i is obtained by solving the *max oracle* optimization problem $H_i = \underset{\mathbf{x}}{\text{argmax}} \quad \Delta(\mathbf{x}^{(k)}, \mathbf{x}) - \left\{E_\theta(\mathbf{x}^{(k)}, I^{(k)}) - E_\theta(\mathbf{x}, I^{(k)}\right\}$. Since, $E_\theta(\mathbf{x}^{(k)}, I^{(k)})$ is the Energy of the GT labelling and independent of $\mathbf{x}$, the max oracle optimization problem reduces to

$$H_i = \underset{\mathbf{x}}{\text{argmax}} \quad \Delta(\mathbf{x}^{(k)}, \mathbf{x}) + E_\theta(\mathbf{x}, I^{(k)}) \tag{9}$$

Since in our case, $\Delta(\mathbf{x}^{(k)}, \mathbf{x})$ is separable at each $x_{l,n}$, eq. 8 can be solved using the TRW-S algorithm similar to the CRF inference in eq. 1. with an additional term $\mid x_{l,n}^{(k)} - x_{l,n} \mid$ added to the unary Boundary cost for each $x_{l,n}$.

4. Materials

The proposed method has been extensively evaluated on 4 public datasets that contain B-scans of healthy subjects as well as patients suffering from AMD and DME. The datasets are summarized in Table 1 and cover a range of image resolution, scanners and image quality. The *NORMAL-1* [5] and *NORMAL-2* [18] datasets contain B-scans from healthy subjects. Out of the 10 volumes in the *NORMAL-1* dataset, five volumes were acquired at a resolution of 400× 400 and the other half at 400 × 800 pixels respectively.

The *AMD-1* dataset [4] was acquired from 4 clinics at varying resolutions and contain B-scans of subjects suffering from intermediate AMD which is characterized by the presence of drusen and geographic atrophy. The B-scans in the *DME-1* dataset [22] contain fluid filled regions associated with DME.

Table 1: Dataset Description. The **voxel resolution** is reported in *axial, lateral, azimuthal* directions. # **Images** reports the (total number of B-scans/ acquired from the number of OCT volumes). # **GT** reports the number of layer boundaries for which the Ground Truth markings are available.

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

For each dataset, the manual ground truth (GT) markings by a senior grader is available for all the 8 boundaries with the exception of the *AMD-1* dataset for which the markings of only 3 clinically relevant boundaries, the ILM, RPE_{in} and RPE_{out} (see Fig. 1b.) are available. Due to the tedium involved in obtaining manual GT, only a few non-adjacent, linearly spaced B-scans from each OCT volume are provided in each dataset that encompass both the foveal and the peripheral regions. To evaluate the inter-observer variance, the manual marking from a second expert is also available for all the datasets except *NORMAL-1*, for which only a subset of 28 B-scans were marked by a second expert [5].

5. Results

In this Section, we present various experiments to validate our joint multi-layer OCT segmentation framework. Both boundary and region based metrics have been defined in Section 5.1 to evaluate the segmentation performance. In Section 5.2, a five-fold cross validation is performed on the *NORMAL-1* dataset to evaluate the performance on healthy OCT B-scans.

This is followed by cross-testing on the *NORMAL-2* dataset in Section 5.3 after training the CRF model on B-scans from the *NORMAL-1* dataset alone.

The proposed method is also evaluated in the presence of pathologies associated with AMD and DME in Sections 5.4, 5.5 and 5.6. Ideally, a single CRF model should be able to segment both healthy and abnormal cases without any prior knowledge about the presence of pathologies in the given image. Hence, these experiments are performed by combining the *NORMAL-1* dataset to the *AMD-1* dataset in Section 5.4 and the *DME-1* dataset in Section 5.5. The performance of our method when trained and evaluated on the *DME-1* dataset alone has also been presented in Section 5.6.

The performance of the proposed method is compared to the results obtained from three publicly available OCT segmentation softwares, CASEREL [39], the Iowa Reference Algorithm (IRA) [40] and OCTSEG [41]. The qualitative and quantitative results of these methods were obtained by running their publicly available implementations on each dataset using their default parameters and weights. CASEREL is based on [5] and provides the segmentation of seven layer boundaries (except RPE_{in}). IRA is based on [20] and segments eleven layer boundaries out of which the boundaries 1, 2, 4, 5, 6, 8, 10 and 11 correspond to our GT markings. OCTSEG is based on [42] and segments six layer boundaries with the exception of RPE_{in} and INL/OPL.

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

5.1. Performance Metrics

Let $\mathbf{x}^{gt}$ denote the GT and $\mathbf{x}$ denote the corresponding estimated layer boundary markings for a given test image I with H rows and Y columns. Using a notation similar to Section 3.2, $\mathbf{x} = \{x_{l,y} | 1 \leq x_{l,y} \leq H, 1 \leq l \leq L,\ 1 \leq y \leq Y,\ l, y \in Z^{+}\}$, where Z^{+} represents the set of positive integers and $x_{l,y}$ represents the height at which the l^{th} boundary passes through column y. The L boundaries divide the retinal tissue into $L-1$ adjacent layers. Let R_l denote the layer that lies between the l^{th} and the $(l+1)^{th}$ boundary.

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

the image. Thus, the unsigned BLE for the l^{th} boundary is defined as

$$U\text{-}BLE_l(\mathbf{x}, \mathbf{x}^{gt}) = \frac{1}{Y} \sum_{y=1}^{Y} \mid x_{l,y} - x_{l,y}^{gt} \mid . \tag{10}$$

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

$$S\text{-}BLE_l(\mathbf{x}, \mathbf{x}^{gt}) = \frac{1}{Y} \sum_{y=1}^{Y} \left(x_{l,y} - x_{l,y}^{gt} \right) . \tag{11}$$

While U-BLE gives a true measure of the segmentation error, S-BLE measures the overall bias of the method to over-estimate or under-estimate the boundary. This is because in S-BLE, the positive and negative errors across the columns tend to cancel each other out. A positive value of the S-BLE indicates that the GT tends to lie above the estimated boundary and viceversa. The U-BLE metric is similar to the loss function Δ used during training (in Section 3.4) with the exception that while the loss function is computed only at the N equidistant points used to represent the boundary, U-BLE is computed across all the columns in the image. Ideally, U-BLE and the magnitude of the S-BLE should be close to 0.

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

$$Dice(R_l, R_l^{gt}) = \frac{2.|R_l \cap R_l^{gt}|}{|R_l| + |R_l^{gt}|}. \tag{12}$$

The thickness of various retinal tissue layers provide clinically relevant information that aids in the detection and tracking the progresssion of ocular diseases [43]. Hence, LTE is defined to measure the average absolute difference in the thickness (in pixels) between the extracted and GT tissue regions across each column in the image. Since, $(x_{l+1,y} - x_{l,y})$ represents the thickness of the tissue region between the l^{th} and the $(l+1)^{th}$ boundary at

column y, LTE for the region R_l is defined as

$$LTE_l(\mathbf{x}, \mathbf{x}^{gt}) = \frac{1}{|Y|} \sum_{y=1}^{|Y|} \mid (x^{gt}_{l+1,y} - x^{gt}_{l,y}) - (x_{l+1,y} - x_{l,y}) \mid . \tag{13}$$

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

5.2. Performance on the NORMAL-1 dataset

The proposed method has been evaluated on the *NORMAL-1* dataset which consists of 108 B-scans of healthy subjects from 10 OCT volumes. A five-fold cross-validation was performed by randomly dividing the dataset into five parts, each containing the B-scans from 2 OCT volumes. Three parts consist of 22 B-scans and the remaining two parts contain 21 B-scans respectively. In each fold, the proposed method was tested on one part after being trained on all the remaining 86 or 87 B-scans. The various performance metrics for each boundary are reported in Table 2 and 3 respectively. Sample qualitative results are depicted in Figure 6.

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

5.3. Cross-testing Performance on the NORMAL-2 dataset

A cross-testing based evaluation has been performed to test the generalizability of the proposed method on unseen test data. In this experiment, our method was trained on the 108 B-scans from the *NORMAL-1* dataset and tested on the 110 B-scans in the *NORMAL-2* dataset. The quantitative and sample qualitative results are depicted in Tables 4, 5 and Figure 7 respectively.

Table 2: Unsigned and Signed Boundary Localization Errors (mean ± standard deviation in pixels) on the NORMAL-1 dataset. The best result in each column is indicated in bold.

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

Table 3: Layer Thickness Error in pixels and Dice coefficient (mean ± standard deviation) for 7 tissue regions on the NORMAL-1 dataset. The best result in each column is indicated in bold.

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

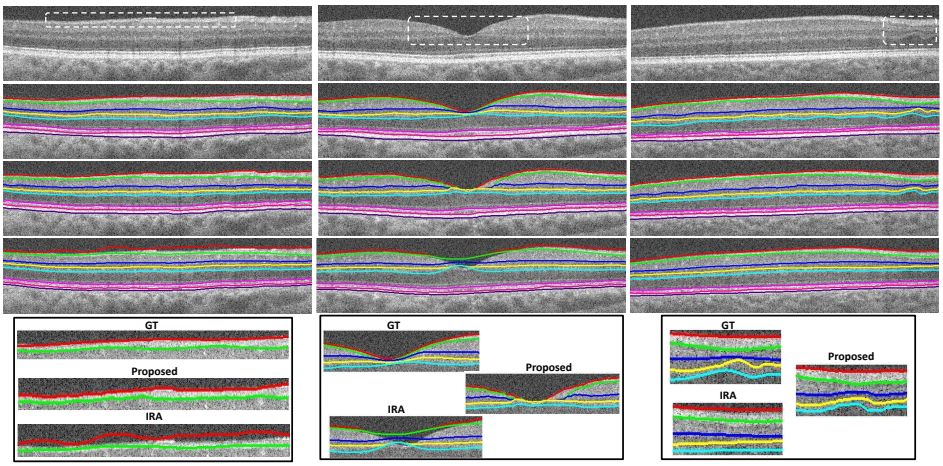

Figure 6: Qualitative results on 3 B-scans of healthy subjects from the *NORMAL-1* dataset is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark. Region within the white dashed rectangle in the first row is magnified at the bottom for comparison.

Table 4: Unsigned and Signed Boundary Localization Errors (mean ± standard deviation in pixels) on the NORMAL-2 dataset. The best result among the automated methods in each column is indicated in bold.

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

Table 5: Layer Thickness Error in pixels and Dice coefficient (mean ± standard deviation) for 7 tissue regions on the NORMAL-2 dataset. The best result among the automated methods in each column is indicated in bold.

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

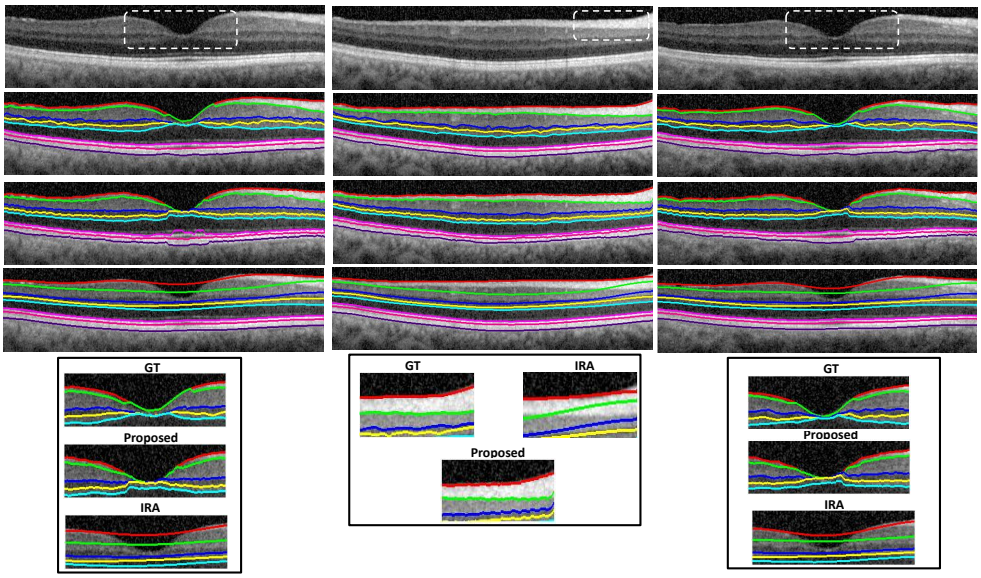

Figure 7: Qualitative results on 3 B-scans of healthy subjects from the *NORMAL-2* dataset is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark. Region within the white dashed rectangle in the first row is magnified at the bottom for comparison.

5.4. Performance in the presence of Age-Related Macular Degeneration

The proposed method has been evaluated in the presence of drusen and Geographic Atrophy associated with AMD by employing a five-fold cross validation on the combined *NORMAL-1* and *AMD-1* dataset. In each fold, the test set consists of the 65 B-scans(21 Normal + 44 AMD) obtained from two OCT volumes from the *NORMAL-1* and four volumes from the *AMD-1* dataset. The CRF is trained on 263 B-scans (87 Normal+176 AMD) in each fold, obtained from the remaining 8 volumes from *NORMAL-1* and 16 OCT volumes from the *AMD-1* dataset. Since the GT for only the ILM, RPE_{in} and RPE_{out} boundaries are available for the *AMD-1* dataset, our method has been evaluated on them. The quantitative results are reported in Tables 6, 7 and sample qualitative results on B-scans with AMD are presented in Figure 8. Among the benchmark methods, only IRA segments the RPE_{in} boundary, hence the regional metrics in Table 7 could only be compared against it. On the combined dataset, the U-BLE varies in the range of 1.18 to 2.38 pixels with a mean of 1.86 pixels (see Table 6) across all the three boundaries and the Dice is 0.98 for the ILM-RPE_{in} region and 0.81 for the RPE layer.

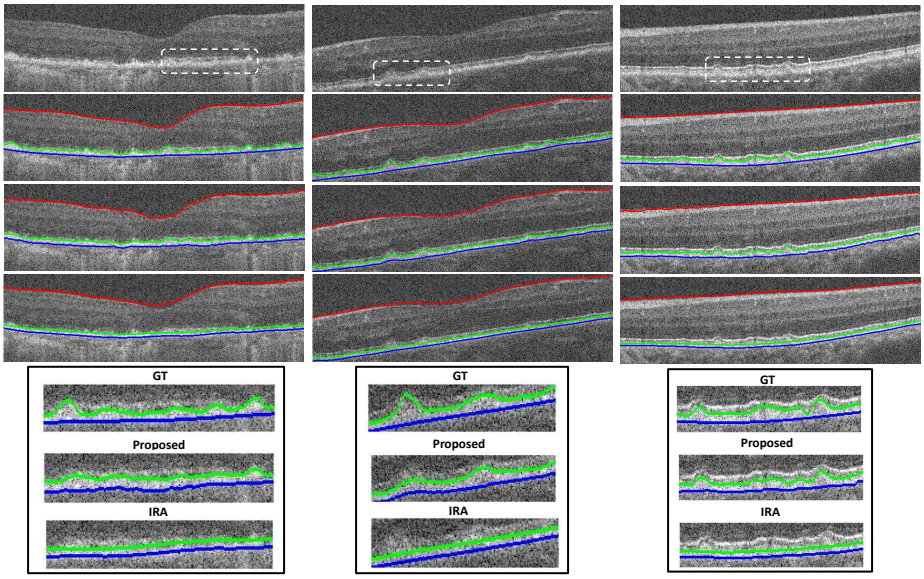

Figure 8: Qualitative results on 3 B-scans with AMD is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark.

Table 6: Unsigned and Signed Boundary Localization Errors (mean ± standard deviation in pixels) on the combined *AMD* and *NORMAL-1* dataset. The best result among the automated methods in each column is indicated in bold.

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

Table 7: Layer Thickness Error in pixels and Dice coefficient (mean ± standard deviation) for 3 layer boundaries on the combined AMD and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold.

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

5.5. Performance in the presence of Diabetic Macular Edema

To evaluate our method in the presence of fluid-filled regions associated with DME, a five-fold cross-validation has been performed on the combined *NORMAL-1* and *DME-1* dataset. The combined dataset was randomly divided into five parts. Each part consists of the B-scans from four OCT volumes; two volumes each from the *NORMAL-1* (21 OCT B-scans) and the *DME-1* dataset (22 B-scans). In each fold, the proposed method is tested on one part after learning a single CRF model to segment both the healthy and DME cases from the remaining 175 (87 Normal + 88 DME) B-scans. The performance of the proposed method on the combined dataset as well as the Normal and DME cases separately have been reported in Tables 8 to 11. Sample qualitative results on the B-scans with DME are presented in Figure 9. The benchmark algorithms being unsupervised cannot be re-trained and their performance on the *NORMAL-1* dataset remains the same as presented in the Tables 2 and 3. On the combined dataset, the U-BLE varies in the range of 1.15 to 3.23 pixels with a mean of 2.04 pixels (see Table 8) and the Dice varies in the range of 0.75 to 0.92 with a mean of 0.84 (see Table 11) across all the layers.

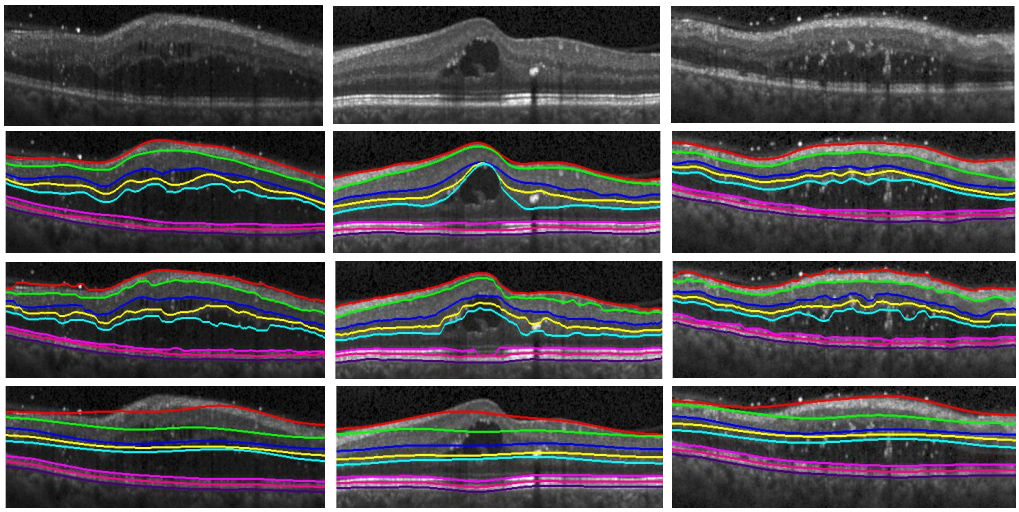

Figure 9: Qualitative results on 3 B-scans from the *DME-1* dataset is depicted in each column. 1^{st} **row :** Original OCT B-scan; 2^{nd} **row :** Ground truth markings; 3^{rd} **row :** Proposed Method; 4^{th} **row :** IRA benchmark.

Table 8: Unsigned Boundary Localization Errors (mean ± standard deviation) on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold.

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

Table 9: Signed Boundary Localization Errors (mean ± standard deviation) on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold.

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

Table 10: Mean Layer Thickness Error (mean ± standard deviation) in pixels for 7 tissue regions on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold.

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

Table 11: Dice coefficient (mean ± standard deviation) for 7 tissue regions on the combined DME and NORMAL-1 dataset. The best result among the automated methods in each column is indicated in bold.

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

5.6. Performance on DME cases alone

In this section, the proposed method has been evaluated on the *DME-1* dataset alone to compare its performance against some of the recent methods in [22], [28] that have been specifically designed to segment the OCT B-scans in the presence of DME through an explicit segmentation of the fluid-filled regions. A kernel regression (KR) based classification scheme was employed by the GTDP+KR method in [22] to explicitly segment the fluid-filled regions and the retinal layers which were further refined using a graph theory and dynamic programming (GTDP) framework.

A deep learning architecture called the *ReLayNet* was employed in [28] to segment both the fluid filled regions and layer boundaries. We followed the standard convention of splitting the *DME-1* dataset into the training and test sets as reported in [22], [28]. The B-scans from subjects 1-5 were used as the training set and the remaining B-scans from subjects 6-10 were used as the test set resulting in 55 B-scans in each set.

The mean Dice and the LTE metrics of the seven tissue regions are presented in Table 12. The results of the GTDP+KR has been reproduced from [22], while the performance of the *ReLayNet* and the second expert have been reproduced from [28]. The proposed method with a mean Dice coefficient of 0.85 across the seven tissue layers, outperforms the GTDP+KR method which has a mean Dice of 0.82. The performance of the proposed method is below the DL based *ReLayNet* (mean Dice= 0.90). The reason for this could be that the fluid-filled regions lead to large deviations from the shape priors on the expected layer thickness and smoothness of the inner layers. An explicit modelling of the fluid filled regions as a separate auxiliary boundary between the OPL/ONL and the IS/OS boundaries similar to [25] can be explored in the future within our CRF framework to address this issue.

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

6. Discussions

A Matlab implementation of our method takes around 9 seconds to process each B-scan on a i7 processor with 8 GB RAM. In this work, each B-scan

Table 12: Mean Dice coefficient and the layer thickness error (in pixels) for 7 tissue regions evaluated on the *DME-1* dataset alone.

	NFL	GCL-IPL	INL	OPL	ONL-IS	OS	RPE
Mean Dice coefficient							
GTDP+KR [22]	0.86	0.88	0.73	0.73	0.86	0.86	0.80
ReLayNet [28]	0.90	0.94	0.87	0.84	0.93	0.92	0.90
Expert 2	0.86	0.90	0.79	0.74	0.94	0.86	0.82
Proposed	0.86	0.88	0.77	0.76	0.94	0.88	0.87
Mean layer thickness error (pixels)							
GTDP+KR [22]	3.68	4.84	7.90	6.35	6.80	2.88	3.61
ReLayNet [28]	1.50	1.20	1.00	1.31	1.35	0.62	0.92
Expert 2	2.01	2.33	2.17	2.29	2.24	1.53	1.54
Proposed	2.56	2.54	2.39	2.13	2.25	1.42	1.42

is segmented independently similar to the existing methods in [4], [5] and [22]. Although the proposed CRF framework can be directly extended to 3D by adding an inter-slice pairwise term (similar to the intra-layer pairwise terms defined in eq. 5) between the corresponding boundary points on the adjacent B-scans, it could not be evaluated due to the unavailability of the groundtruth markings for the consecutive B-scans. Due to the tedium involved in obtaining manual GT, the four public datasets (*NORMAL-1*, *NORMAL-2*, *DME-1* and *AMD-1*) only provide the GT for a few non-adjacent, linearly spaced B-scans from each OCT volume. Hence, a 3D CRF model could not be trained. Moreover, in retinal OCT imaging, the pixel resolution across the B-scans is approximately 10 times coarser than the intra-slice lateral resolution in most of the image acquisition settings (see for eg., Table 1). Hence, regularization across adjacent B-scans only has a marginal effect on the segmentation performance while adversely impacting the computational and memory requirements.

Next, we discuss the impact of the tunable hyperparameters on the performance followed by a discussion of the results presented in Section 5.

Effect of Hyperparameters on Performance

The proposed method has three tunable hyperparameters, the column spacing between the adjacent points on the boundary, the size of the convolutional filters $\mathbf{u_l}$ and $\mathbf{v_l}$ which capture the appearance of the layer boundaries and the intermediate tissue regions respectively, and the regularization weight λ in eq. 8 used during training. In all the experiments, the column

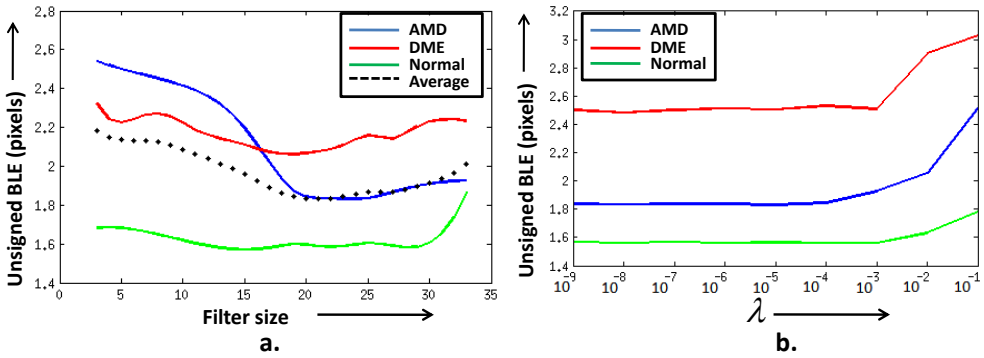

Figure 10: Variation in performance with respect to a) the filter size and b) the regularization weight λ. Lower values of U-BLE in pixels indicate better performance.

spacing was empirically fixed to 4 pixels which resulted in 150 control points to represent each boundary. The intermediate boundary points were obtained through b-spline interpolation. This choice provided a good trade-off between the computational efficiency and the interpolation error.

The exact size of the convolutional filters and the regularization weight were fixed experimentally. The size of all the convolutional filters were kept equal to minimize the number of tunable parameters. A small set of 15 B-scans were randomly selected from each of the *NORMAL-1*, *AMD-1* and the *DME-1* dataset. For each of the three sets, a separate CRF energy was trained using 8 B-scans and the remaining 7 B-scans were used as the validation set. At first λ was fixed to 10^{-4} and the filter size was varied in the range of 3 to 33 in steps of 2. Fig. 10 a. depicts the variation of the U-BLE in pixels against the varying filter size. The average performance across all the three datasets is depicted by the black dotted line. Overall, the performance of the proposed method was relatively stable with the U-BLE varying in the range of 1.6-2.6 pixels across the varying filter size for all the three sets. A filter size of 19×19 was found to be a robust choice across the normal and different pathological cases. Next, λ was varied in powers of 10 in the range of 10^{-9} to 10^{-1} keeping the filter size fixed at 19×19. Initially, the performance improved from 10^{-1} to 10^{-3} after which the U-BLE was relatively stable across all the three sets. Based on these experiments, the filter size and λ were fixed to 19×19 and 10^{-4} respectively, across all the experiments presented in Section 5.

Performance on *NORMAL-1* dataset: The results on the *NORMAL-1*

dataset in Tables 2,3 illustrates the good performance of our method on OCT B-scans of healthy subjects. The U-BLE across the eight boundaries (see Table 2) varies in the range of 1.09 to 1.95 pixels with a mean value of 1.52 pixels which improves on the IRA (with a mean U-BLE of 1.77) by 16% which is the second best performing method. The S-BLE is close to 0 pixels across all boundaries indicating the absence of any significant bias towards over or under-estimating a boundary. The Dice coefficient for the seven retinal tissue layers (in Table 3) varies in the range of 0.8 to 0.94 with a mean value of 0.87 and is consistently better than the other methods. In terms of LTE, the error of the proposed method is lower than the other methods on six out of the seven layers with a mean LTE of 2.00 pixels across all layers.

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

Performance on *NORMAL-2* dataset: On the *NORMAL-2* dataset, our cross-testing performance is within the inter-expert variance, performing better than or equivalent to the second expert on all layers in terms of the Dice coefficient (see Table 5) and all except the NFL layer in terms of the LTE (1.86 pixels in comparison to 1.42 pixels for the second expert). The proposed method doesnot show any significant bias towards over or under-estimation of any boundary as indicated by a S-BLE value close to 0 pixels (in Table 4) for all the boundaries.

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

In terms of the U-BLE metric reported in Table 4, OCTSEG is the best performing among the benchmark algorithms with a mean U-BLE of 2.23 pixels across the eight boundaries. The proposed method improves on the performance of OCTSEG by about 50% with a mean U-BLE of 1.11 pixels. The better performance of our method with respect to the IRA (U-BLE of 3.04 pixels) illustrates the advantage of learning the energy over the hand-crafted cost terms employed in IRA based on [20]. Moreover, while IRA

performs the segmentation in 2 steps, the outer (1,7,8) layer boundaries followed by the inner (2-6) ones, our method extracts all the 8 boundaries in a single step.

Performance in the presence of AMD: The average U-BLE (see Table 6) of the proposed method on the ILM and RPE_{out} boundaries considering the AMD cases alone is 2.03 pixels which is only a slight improvement over the performance of CASEREL with a U-BLE of 2.07 pixels. However, the RPE_{in} boundary which plays a crucial role in the detection of AMD is currently unavailable in the OCTSEG and CASEREL softwares. Considering all the three layers, our method outperforms IRA by 37% in terms of the average U-BLE across all the boundaries (average U-BLE of 2.08 pixels in comparison to 3.28 pixels for IRA) for the AMD cases. The S-BLE metric of our method is close to 0 for all the three boundaries and doesnot indicate any significant bias towards over or under-estimation. In terms of the Dice and LTE metric (see Table 7), the performance of the proposed method is similar to that of the second expert (LTE of 2.57 pixels compared to 2.60 for the second expert markings) on the tissue region between the ILM and RPE_{in} boundaries. However, the performance drops by 4% for the RPE layer in terms of Dice indicating a scope for further improvement. This is consistent with the fact that AMD leads to drusen deposits in RPE layer.

Performance on combined Normal and DME cases: When the proposed method is jointly trained on the Normal and DME cases, it outperforms the three benchmark methods on each of the 8 boundaries both in terms of the Dice coefficient (see Table 11) and the U-BLE (see Table 8) metric considering the DME cases alone.

The CASEREL is the best performing among the benchmark algorithms and the proposed method improves on its performance by 35% (2.50 pixels in comparison to 3.86 pixels for CASEREL) in terms of the average U-BLE across all boundaries and 16% in terms of the Dice coefficient (0.81 of the proposed method in comparison to 0.70 for CASEREL). However, in comparison to the second expert, the performance of our method is slightly lower indicating a scope for further improvement. In terms of Dice, our performance is comparable to that of the second expert on the last four layers but drops by approximately 5% for each of the first three layers. This is consistent with the observation that the fluid-filled regions tend to occur between the ILM and the INL/OPL boundaries. The absolute value of S-BLE is less than 1 pixel for all layers except IPL/INL and OPL/ONL for which there is a slight bias towards estimation of boundary to lie below the GT.

Effect of joint training on healthy Images: In Sections 5.4 and 5.5, the CRF models were learnt on combined datasets consisting of both healthy and abnormal cases. This allows the method to be employed to segment new test cases for which any prior information on the presence of pathologies is unavailable. However, the healthy and abnormal B-scans differ widely in their appearance and there is also a large deviation in the distribution of the layer thickness and boundary smoothness statistics. Hence, the model trained on the combined datasets can adversely affect the segmentation performance on the healthy images. However, the results indicate that there is no significant decrease in performance on the *NORMAL-1* dataset. The average U-BLE of 1.52 pixels across the eight boundaries when trained on healthy images alone (in Table 2) drops to 1.56 pixels when trained jointly with DME cases (in Table 8). Similarly, considering the ILM, RPE_{in} and RPE_{out} alone, the average U-BLE drops from 1.41 (in Table 2) to 1.42 pixels when the CRF is trained on the combined dataset with AMD cases (in Table 6).

Cross-testing across the AMD and DME cases: It may be interesting to note the performance of our method on unknown disease cases which have not been encountered during training. For this reason, the proposed method has been tested on the *AMD-1* cases after training on the combined *DME-1* and *NORMAL-1* datasets and viceversa. Since the *AMD-1* dataset only provides the GT for the ILM, RPE_{in} and RPE_{out} boundaries, the evaluation has been restricted to them.

The results of cross-testing on the *AMD-1* dataset indicate only a slight drop in performance in all the three boundaries. In comparison to the five fold cross-validation performance described in Section 5.4, the mean U-BLE increased from 1.24 to 1.54 pixels for the ILM, 2.20 to 3.13 pixels for the RPE_{in} and 2.82 to 3.53 pixels for the RPE_{out} boundaries respectively. Cross-testing on the *DME-1* dataset after training on the combined *AMD-1* and *NORMAL-1* datasets depicted a larger drop in performance in comparison to the 5 fold cross-validation results in Section 5.5. The U-BLE increased from 1.22 to 2.76, 1.34 to 4.20 and 1.09 to 3.26 pixels for the ILM, RPE_{in} and the RPE_{out} boundaries respectively. The DME and AMD have a significantly different effect on the shape priors. While AMD affects the smoothness of RPE_{in}, DME leads to large deviations in the expected layer thickness between the ILM and the RPE_{in} region. Thus, cross-testing across the disease classes is expected to result in a decrease in segmentation performance.

7. Conclusions

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

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

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

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

Acknowledgement

This research was supported in part by the doctoral fellowship provided by Tata Consultancy Services under their Research Scholarshp Program.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this article.

References

[1] R. Kafieh, H. Rabbani, S. Kermani, A review of algorithms for segmentation of optical coherence tomography from retina, Journal of Medical Signals and Sensors 3 (2013) 45.

[2] S. I. Niwas, W. Lin, X. Bai, C. K. Kwoh, C.-C. J. Kuo, C. C. Sng, M. C. Aquino, P. T. Chew, Automated anterior segment oct image analysis for angle closure glaucoma mechanisms classification, Computer Methods and Programs in Biomedicine 130 (2016) 65–75.

[3] D. Sidibé, S. Sankar, G. Lemaître, M. Rastgoo, J. Massich, C. Y. Cheung, G. S. Tan, D. Milea, E. Lamoureux, T. Y. Wong, et al., An anomaly detection approach for the identification of dme patients using spectral domain optical coherence tomography images, Computer Methods and Programs in Biomedicine 139 (2017) 109–117.

[4] S. J. Chiu, J. A. Izatt, R. V. O'Connell, K. P. Winter, C. A. Toth, S. Farsiu, Validated automatic segmentation of amd pathology including drusen and geographic atrophy in sd-oct images, Investigative Ophthalmology & Visual Science 53 (2012) 53–61.

[5] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, S. Farsiu, Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation, Optics Express 18 (2010) 19413–19428.

[6] R. Behbehani, A. A. Al-Hassan, A. Al-Salahat, D. Sriraman, J. Oakley, R. Alroughani, Optical coherence tomography segmentation analysis in relapsing remitting versus progressive multiple sclerosis, PloS One 12 (2017) e0172120.

[7] F. G. Schlanitz, C. Ahlers, S. Sacu, C. Schütze, M. Rodriguez, S. Schriefl, I. Golbaz, T. Spalek, G. Stock, U. Schmidt-Erfurth, Performance of drusen detection by spectral-domain optical coherence tomography, Investigative Ophthalmology & Visual Science 51 (2010) 6715–6721.

[8] S. M. Golzan, A. Avolio, S. L. Graham, Minimising retinal vessel artefacts in optical coherence tomography images, Computer Methods and Programs in Biomedicine 104 (2011) 206–211.

[9] G. Girish, V. Anima, A. R. Kothari, P. Sudeep, S. Roychowdhury, J. Rajan, A benchmark study of automated intra-retinal cyst segmentation algorithms using optical coherence tomography b-scans, Computer Methods and Programs in Biomedicine (2017).

[10] E. S. Varnousfaderani, J. Wu, W.-D. Vogl, A.-M. Philip, A. Montuoro, R. Leitner, C. Simader, S. M. Waldstein, B. S. Gerendas, U. Schmidt-Erfurth, A novel benchmark model for intelligent annotation of spectral-domain optical coherence tomography scans using the example of cyst annotation, Computer Methods and Programs in Biomedicine 130 (2016) 93–105.

[11] A. Chakravarty, J. Sivaswamy, End-to-end learning of a conditional random field for intra-retinal layer segmentation in optical coherence tomography, in: Medical Image Understanding and Analysis, 2017, pp. 3–14.

[12] H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, J. S. Schuman, Macular segmentation with optical coherence tomography, Investigative Ophthalmology & Visual Science 46 (2005) 2012–2017.

[13] T. Fabritius, S. Makita, M. Miura, R. Myllylä, Y. Yasuno, Automated segmentation of the macula by optical coherence tomography, Optics Express 17 (2009) 15659–15669.

[14] J. Novosel, G. Thepass, H. G. Lemij, J. F. de Boer, K. A. Vermeer, L. J. van Vliet, Loosely coupled level sets for simultaneous 3d retinal layer segmentation in optical coherence tomography, Medical Image Analysis 26 (2015) 146–158.

[15] F. Rossant, I. Bloch, I. Ghorbel, M. Paques, Parallel double snakes. application to the segmentation of retinal layers in 2d-oct for pathological subjects, Pattern Recognition 48 (2015) 3857–3870.

[16] A. Yazdanpanah, G. Hamarneh, B. R. Smith, M. V. Sarunic, Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach, IEEE Transactions on Medical Imaging 30 (2011) 484–496.

[17] V. Kajić, B. Považay, B. Hermann, B. Hofer, D. Marshall, P. L. Rosin, W. Drexler, Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis, Optics Express 18 (2010) 14730–14744.

[18] J. Tian, B. Varga, G. M. Somfai, W.-H. Lee, W. E. Smiddy, D. C. DeBuc, Real-time automatic segmentation of optical coherence tomography volume data of the macular region, PloS One 10 (2015) e0133908.

[19] M. Haeker, M. Abramoff, R. Kardon, M. Sonka, Segmentation of the surfaces of the retinal layer from oct images, Springer, 2006, pp. 800–807.

[20] M. K. Garvin, M. D. Abràmoff, X. Wu, S. R. Russell, T. L. Burns, M. Sonka, Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images, IEEE Transactions on Medical Imaging 28 (2009) 1436–1447.

[21] P. A. Dufour, L. Ceklic, H. Abdillahi, S. Schroder, S. De Dzanet, U. Wolf-Schnurrbusch, J. Kowal, Graph-based multi-surface segmentation of oct data using trained hard and soft constraints, IEEE Transactions on Medical Imaging 32 (2013) 531–543.

[22] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, S. Farsiu, Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema, Biomedical Optics Express 6 (2015) 1172–1194.

[23] M. A. Hussain, A. Bhuiyan, A. Turpin, C. D. Luu, R. T. Smith, R. H. Guymer, R. Kotagiri, Automatic identification of pathology-distorted retinal layer boundaries using sd-oct imaging, IEEE Transactions on Biomedical Engineering 64 (2017) 1638–1649.

[24] F. Shi, X. Chen, H. Zhao, W. Zhu, D. Xiang, E. Gao, M. Sonka, H. Chen, Automated 3-d retinal layer segmentation of macular optical coherence tomography images with serous pigment epithelial detachments, IEEE Transactions on Medical Imaging 34 (2015) 441–452.

[25] J. Novosel, K. A. Vermeer, J. H. de Jong, Z. Wang, L. J. van Vliet, Joint segmentation of retinal layers and focal lesions in 3-d oct data of topologically disrupted retinas, IEEE Transactions on Medical Imaging 36 (2017) 1276–1286.

[26] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, S. Farsiu, Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search, Biomedical optics express 8 (2017) 2732–2744.

[27] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[28] A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, N. Navab, Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks, Biomedical optics express 8 (2017) 3627–3642.

[29] A. Ben-Cohen, D. Mark, I. Kovler, D. Zur, A. Barak, M. Iglicki, R. Soferman, Retinal layers segmentation using fully convolutional network in oct images, 2017. URL: `https://www.rsipvision.com/wp-content/uploads/2017/06/Retinal-Layers-Segmentation.pdf`.

[30] M. Pekala, N. Joshi, D. E. Freund, N. M. Bressler, D. C. DeBuc, P. M. Burlina, Deep learning based retinal oct segmentation, arXiv preprint arXiv:1801.09749 (2018).

[31] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[32] G. Huang, Z. Liu, K. Q. Weinberger, L. van der Maaten, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, volume 1, 2017, p. 3.

[33] Y. Yu, S. T. Acton, Speckle reducing anisotropic diffusion, IEEE Transactions on Image Processing 11 (2002) 1260–1270.

[34] L. G. Nyúl, J. K. Udupa, X. Zhang, New variants of a method of mri scale standardization, IEEE Transactions on Medical Imaging 19 (2000) 143–150.

[35] V. Kolmogorov, Convergent tree-reweighted message passing for energy minimization, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (2006) 1568–1583.

[36] S. Lee, J. Xin, S. Westland, Evaluation of image similarity by histogram intersection, Color Research & Application 30 (2005) 265–274.

[37] T. Finley, T. Joachims, Training structural svms when exact inference is intractable, in: Proceedings of the 25th International Conference on Machine Learning, ACM, 2008, pp. 304–311.

[38] S. Lacoste-Julien, M. Jaggi, M. Schmidt, P. Pletscher, Block-coordinate frank-wolfe optimization for structural svms, in: 30th International Conference on Machine Learning - Volume 28, JMLR.org, 2013, pp. I–53–I–61.

[39] P. Teng, Caserel - an open source software for computer-aided segmentation of retinal layers in optical coherence tomography images, 2013. URL: `http://pangyuteng.github.io/caserel/`.

[40] K. Lee, M. D. Abramoff, M. Garvin, M. Sonka, et al., The iowa reference algorithms, version 3.8.0 (retinal image analysis lab, iowa institute for biomedical imaging, iowa city, ia), 2014. URL: `http://www.iibi.uiowa.edu/content/iowa-reference-algorithms-human-and-murine-oct-retinal-layer-analysis-and-display`.

[41] M. Mayer, et al., Octseg, version 4.0 (pattern recognition lab, friedrich-alexander-universitt erlangen-nrnberg), 2016. URL: `https://www5.cs.fau.de/research/software/octseg/`.

[42] M. A. Mayer, J. Hornegger, C. Y. Mardin, R. P. Tornow, Retinal nerve fiber layer segmentation on fd-oct scans of normal subjects and glaucoma patients, Biomedical Optics Express 1 (2010) 1358–1383.

[43] A. M. Syed, T. Hassan, M. U. Akram, S. Naz, S. Khalid, Automated diagnosis of macular edema and central serous retinopathy through robust reconstruction of 3d retinal surfaces, Computer Methods and Programs in Biomedicine 137 (2016) 1–10.