DETECTION OF DIABETIC MACULAR EDEMA IN OPTICAL COHERENCE TOMOGRAPHY SCANS USING PATCH BASED DEEP LEARNING

Abhishek Vahadane   Ameya Joshi   Kiran Madan   Tathagato Rai Dastidar

SigTuple Technologies Pvt. Ltd., Bangalore, India,

ABSTRACT

We propose a two step framework to automatically classify an OCT scan as indicative of Diabetic Macular Edema (DME) by detecting abnormal pathologies in OCT frames. The first step involves detection of candidate patches for fluid filled regions and hard exudates using image processing techniques. The second step is to predict a label for these candidate patches using deep convolutional neural network. In the final collation step, we aggregate the confidences of the CNN models and use a rule based method to predict the presence of DME.

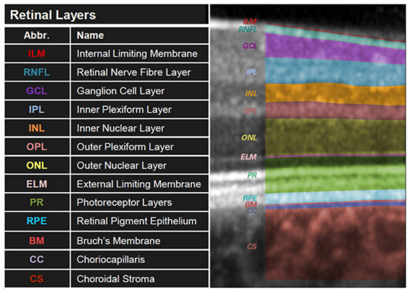
***Index Terms*—** Optical Coherence Tomography, Eye, Computer Aided Detection and Diagnosis

1. INTRODUCTION

Optical coherence tomography (OCT) is an imaging technique capable of capturing high resolution three dimensional images of biological tissue. One of its applications is in the acquisition of retinal OCT volumes. These OCT volumes are useful in diagnosis of retinal diseases and treatment planning [1]. There are different abnormal objects of interest *viz* hard exudates, cysts, drusen, vitreo-macular traction (VMT) etc. that can be present in retinal OCT [2]. These local abnormal structures are linked to eye diseases like macular edema and age-related macular degeneration.

In an OCT scan, fluid filled regions or FFRs are visible as dark spaces with well defined walls, found in between retinal layers. Hard exudates show up as hyper-reflective objects located between the ganglion cell layer (GCL) and external limiting membrane (ELM) layer as show in Fig. 1. Together, these two gross pathologies are indications of diabetic macular edema (DME). A normal OCT frame has smooth layers with no breaks or loss of layer continuity. Abnormal frames have non-smooth distorted layer structures. Dense speckle noise in some frames adds more challenges in detecting abnormal objects of interest.

In this paper, we propose a hierarchical approach to detect presence of DME in OCT scans. We detect potential candidates in the OCT frame using image processing followed by their classification into FFRs, exudates and unimportant classes using a set of deep convolutional neural networks (CNN). We further use a rule based approach to classify an OCT scan as being indicative of DME or not based on our classification of the input patches.



Retinal Layers

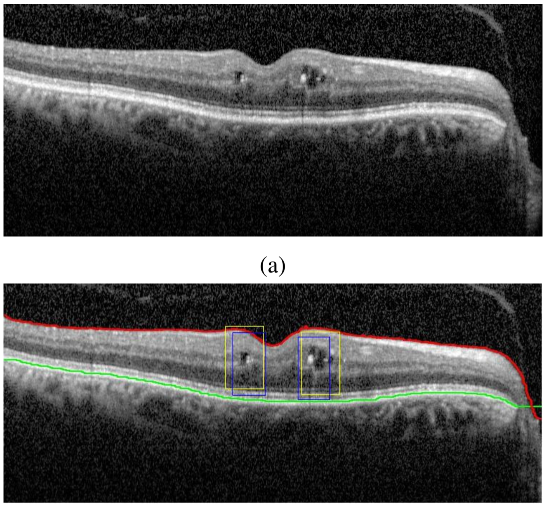
Abbr.	Name
ILM	Internal Limiting Membrane
RNFL	Retinal Nerve Fibre Layer
GCL	Ganglion Cell Layer
IPL	Inner Plexiform Layer
INL	Inner Nuclear Layer
OPL	Outer Plexiform Layer
ONL	Outer Nuclear Layer
ELM	External Limiting Membrane
PR	Photoreceptor Layers
RPE	Retinal Pigment Epithelium
BM	Bruch's Membrane
CC	Choriocapillaris
CS	Choroidal Stroma

Fig. 1: Typical retinal layers in OCT frame [3]

The paper is organized as follows: Related work is discussed in Section 2. Section 3 describes our methodology. Experimental results are presented in Section 4. Finally, Section 5 concludes the paper.

2. RELATED WORK

In this section, we talk about related work in literature for automated analysis of OCT scans. In [4], authors present a semi-automatic method to detect hard-exudates and ILM and RPE layers by limiting the search region and finding the shortest path within the search region [5]. Hard exudates are detected by a self adoption threshold and region growing. However, the method was not robust [4] and does not work in presence of variation in hard exudate appearance. Roy *et al.* in [6] use a jointly weighed loss function with a special architecture for simultaneously segmenting and classifying layers and fluid filled masses. Lee *et al.* [7] propose a U-Net based segmentation network for fluid filled regions. In [8], the authors model cystic changes using a motion based model along with a CNN to segment cystic changes in an OCT scan. The method fails to detect small cysts and is therefore less sensitive to earlier changes which is important for screening. In [9], ElTanboly *et al.* detect features corresponding to each layer in the OCT using Markov-Gibbs random fields and follow it up with deep classification network with auto-encoders to classify a scan as edematous. However, they only segment layers and do not detect any specific abnormalities.



(a)

(b)

Fig. 2: (a) Original frame, (b) visual results. Color coding – ILM: green; RPE: red; FFR: yellow; hard exudates: blue.

Deep learning techniques have been shown to outperform image processing and machine learning methods for natural image object as well as biomedical object classification [10]. Hence, we employ deep convolution neural networks (CNN). As OCT frames are quite large in size and resizing them leads to significant information loss, we use patches corresponding to the various abnormal regions as input to our deep neural models. This allows us to use lesser amount of data as each frame generates a large number of samples as well as create smaller models which can be deployed on an edge device.

3. GROSS PATHOLOGY EXTRACTION AND CLASSIFICATION

We present a two stage framework for detecting hard exudates and fluid filled regions (FFR). The first step is to extract candidate patches using image processing. The second step is to classify the candidate patches using deep CNN.

3.1. Unsupervised Candidate Detection

ILM and RPE detection: We first pre-process the OCT frame to crop the infrared fundus image in the OCT frame using trivial gradient and edge detection based techniques. In the segmented OCT frame, we detect nerve fibre layer (ILM) and retinal pigment epithelium (RPE) using Dijkstra's shortest path algorithm, representing pixels as nodes in a graph. In an OCT scan, layers are horizontally aligned and extend from left to right or right to left. Hence, we define the start and stop node to be any pixel on a zero padded left and zero padded right column respectively. The RPE is the lowermost hyper-reflective layer and therefore, has high intensity values. We define the cost matrix for RPE in Eq. 2. Let $I(x,y)$ be the pixel intensity in the OCT frame at position (x,y). Then, the cost matrix is given as:

$$C_{\mathrm{rpe}}(x,y) = 1-(\frac{I(x,y)}{255}+L(x,y)) \tag{1}$$

$$L(x,y) = T(x,y)\cdot\widetilde{y} \tag{2}$$

Here, $T(x,y)$ is the thresholded binary image with the threshold as twice the threshold given by Otsu's algorithm [11]. The factor of two is selected to retain only the relatively bright retinal layers. $\widetilde{y}$ is the normalized relative distance in the y-axis. We then use Dijkstra's algorithm to solve for the minimum cost path using $C(x,y)$ as the cost function, allowing us to localize the RPE layer. Similarly, we detect the ILM using the cost matrix in Eq. 4 using the assumptions that the ILM is the topmost layer as well as shows a significant gradient change with respect to the intra-ocular space above it.

$$C_{\mathrm{ilm}}(x,y) = 1-(E(x,y)\cdot(1-\widetilde{y})) \tag{3}$$

$$E(x,y) = G(x,y)\cdot T(x,y) \tag{4}$$

$G(x,y)$ is the Canny [11] edge map whereas $T(x,y)$ is the thresholded binary image. The product of the two, $E(x,y)$ represents the term corresponding to the large gradient change between the ILM and the hyloid region. We then set the pixel intensities above the detected ILM and below the detected RPE to zero so as to exclude them from the search region for further candidate detection. We show the results of the ILM and RPE detection in Fig. 2.

Fluid Filled Region Candidate Extraction: Fluid Filled regions are dark featureless regions in the intracellular space. We use this property to isolate candidates for FFR. Our method is based on the Otsu thresholding technique [11]. We enhance this method by taking the negative of our OCT frame and thresholding the image with factor $k\cdot T_{otsu}$ where T_{otsu} is the threshold calculated using Otsu's method. The factor k ensures that we are very sensitive in our FFR detection. In this paper, we use a empirically calculated value of k as 0.9.

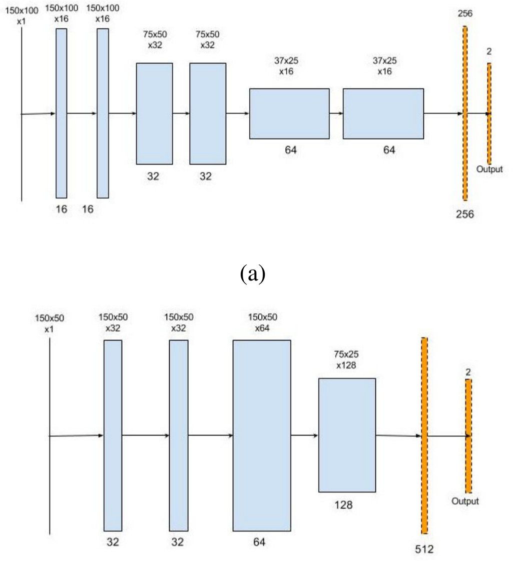
Hard Exudate Candidate Extraction: Since hard exudates are hyper-reflective regions in the intra-retinal area, we detect hard exudate candidates between ILM and RPE using local peak detection. The mathematical expression is depicted in Eq. 5.

$$(x,y) = \arg\max(I_{\mathrm{ilm}\to\mathrm{rpe}}(x,y), 2d+1) \tag{5}$$

Here (x,y) are the set of detected local peaks; computed as the maximum over a local $(2\times d+1)$ spatial window. We chose $d=10$ which approximately corresponds to radius of an average hard-exudate. $I_{\mathrm{ilm}\to\mathrm{rpe}}$ is the region between the ILM and RPE. We show examples of our detection algorithm in Fig.2.

3.2. Classification

The above module provides us with candidates of FFR and hard exudates. We then classify these candidates as being positive samples of the suggested class or not by using a set of patch based CNN classifiers. The output of each CNN is a



(a)

(b)

Fig. 3: **(a)FFR Classification Network** (b)**Exudate Classification Network**. The convolutional layers are represented by solid boxes) and MLP layers are represented by dashed boxes). Figure best viewed in color

softmax vector representative of the confidence of the model for the positive label. We then collate the confidence vectors with respect to a frame and use a learned rule to detect whether the frame shows signs of DME or not.

Fluid Filled Regions: We primarily adapt the VGG[12] architecture to classify 150×100 patches centered around the candidates detected in section 3.1. A rectangular patch is chosen because the OCT scan has different scales in the x and y axes. The CNN architecture used for classification of FFRs is shown in Fig.3. We augment our training by adding random gaussian noise ($\mu = 0.35$, $\sigma = 0.01$) in the convolutional layers for emulating speckle noise in images as well as regularizing the network. Finally, we use a dropout factor of 0.7 for our fully connected layers.

Hard Exudates: Similar to our approach for FFRs, we train a CNN for classifying hard exudate candidates into hard exudates or not. For hard exudates, we choose a patch of 150×50 corresponding to their average size. The size of each patch being small, we down-sample it using 2×2 maxpooling only once.

We use negative log likelihood function as the cost function and stochastic gradient descent [13] for optimization for training both the models. In order to detect if a frame shows signs of DME or not, we use a simple association rule derived from our training set as defined in Eqn.3.2.

$$\hat{Y} = \begin{cases} \text{DME}, & N_{\text{FFR},(p>0.8)} > 1 \bigvee N_{\text{HE},(p>0.75)} > 3 \\ \text{NoDME}, & \textit{otherwise} \end{cases} \tag{6}$$

4. EXPERIMENTS AND RESULTS

Our training data consists a set of 328 cases; taken using a Heidelberg Spectralis OCT Scanner [3], which we split into 317 training cases and 204 validation cases in a stratified manner based on the reports. Each case is a video consisting of multiple frames. We then extract 1827 frames, mostly corresponding to the central foveal region. These frames are annotated by two experts by marking out the region of interest. Conflicting annotations are resolved by arbitration. We show the results of the arbitration process in Table 2 by comparing each consultant with the post-arbitration ground truth. Further, we take the overlap of our candidate detection algorithms with the marked regions and label each patch as being an example of a label only if more than 50% of the patch is contained by a marked region. Additionally, we select 210 random frames from the validation cases for frame-wise and patch-wise analysis of our system.

FFR Training Methodology: For FFR, we extract 1369 patches of the positive class and 11063 patches of the negative class for training. We use a weighted negative log likelihood function with weights of $5:1$ in favor of the positive class in order to maintain class balance. We then train this model for 300 epochs with a learning rate of 0.1 and an exponential decay of 0.95.

Hard Exudate Training Methodology: The training data for hard exudates consists of 3651 positive samples and 12367 negative samples from 537 frames. We follow the same training methodology as that of FFRs apart from training the model for 500 epochs.

Label	Patchwise metrics(%)			Framewise metrics(%)		
	P	R	f_1	P	R	f_1
FFR	96.36	74.96	0.84	97.34	86.39	0.91
Hard exudates	95.08	89.73	0.91	96.04	88.20	0.92

Table 1: Quantitative Results for pathologies on the test set. P is the precision whereas R is the recall.

Label	FFR(%)			Exudates(%)		
	P	R	f_1	P	R	f_1
Consultant 1	84.1	79.3	0.81	66.2	76.7	0.71
Consultant 2	81.0	93.0	0.87	85.8	62.3	0.72

Table 2: Pre-arbitration expert Inter-observer Variability. P is the precision and R represents recall. Statistics are calculated with respect to the ground truth after arbitration.

Quantitative results: We validate the performance of our deep CNN based classification using patch and frame level

Method	Precision(%)	Recall(%)	$f1-$score
Our Method	96.43	89.45	**0.9281**
Athar *et al.* [14]	81.60	95.33	0.8793

Table 3: Quantitative Results for pathologies on the test set

recall and precision metrics on a test set of 421 frames randomly selected from the 204 validation cases. The results of our system are shown in Table 1. The precision, recall and $f1-$scores can be compared with the variability of each consultant from the generated ground truth as shown in Table 2. As evident, our system shows a more consistent performance on the test set. The $f1-$score for hard exudates correlate better with the ground truth in comparison with both consultants, while for FFRs it lies in between. We also show the results of our confidence based rule for predicting whether a frame showcases signs of DME or not. In Table 3, we present a comparison of our approach with a simple weak labeling based approach [14] on the same set where we use a CNN for simultaneous classification and localization. The $f1-$scores clearly show that our patch based algorithm performs better than using the original image in its entirety.

Visual results: We show the visual results on a challenging frame selected from our database. Fig. 2 shows the detected candidates as boxes and our classification of the same. The yellow boxes correspond to detected FFRs and the blue boxes correspond to the detected hard exudates. The red line depicts the ILM where as the green line depicts the RPE. As our method predicts gross abnormalities in different local scenario and irrespective of variations, we imply that CNNs have learned the feature representation invariant to size, shape and translation. The primary advantage of using patches over the weak labeling method is that the networks can be less complex due to the focused nature of the method. In addition, our method is highly extensible as any new pathology can be added to the system simply by training a binary classification model for the same.

5. CONCLUSION

In this paper, we consider a patch based approach to classify OCT frames as being indicative of DME or not. We detect and classify patches corresponding to hard exudates and fluid filled regions using image processing and deep learning. We also prove that this approach works better than using a frame level deep neural network classifier as well as other related methods as shown in Section 2. In the future, we want to show generalization capability of our approach on different OCT scanners. Also, we would like to extend the current framework detect other abnormal structures *viz* vitreo-macular traction, epiretinal membranes, drusen etc., representing various other disorders such as age related macular degeneration.

6. REFERENCES

[1] James G Fujimoto et al., “Optical coherence tomography (oct) in ophthalmology: introduction.,” *Optics express*, vol. 17, no. 5, pp. 3978–3979, 2009.

[2] Jay S Duker, Nadia K Waheed, and Darin Goldman, *Handbook of Retinal OCT: Optical Coherence Tomography E-Book*, Elsevier Health Sciences, 2013.

[3] “Heidelberg engineering,” `https://media.heidelbergengineering.com`.

[4] Sijie Niu et al., “Multimodality analysis of hyper-reflective foci and hard exudates in patients with diabetic retinopathy,” *Scientific Reports*, vol. 7, 2017.

[5] Stephanie J Chiu et al., “Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation,” *Optics express*, vol. 18, no. 18, pp. 19413–19428, 2010.

[6] Abhijit Guha Roy et al., “Relaynet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional network,” *arXiv preprint arXiv:1704.02161*, 2017.

[7] Cecilia S Lee et al., “Deep-learning based, automated segmentation of macular edema in optical coherence tomography,” *bioRxiv*, p. 135640, 2017.

[8] Karthik Gopinath et al., “Segmentation of retinal cysts from optical coherence tomography volumes via selective enhancement,” *arXiv preprint arXiv:1708.06197*, 2017.

[9] Ahmed ElTanboly et al., “A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images,” *Medical physics*, vol. 44, no. 3, pp. 914–923, 2017.

[10] Mark JJP van Grinsven et al., “Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images,” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1273–1284, 2016.

[11] Rafael C Gonzalez, Richard E Woods, et al., “Digital image processing,” 1992.

[12] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[13] Pedro Pinheiro and Ronan Collobert, “Recurrent convolutional neural networks for scene labeling,” in *International Conference on Machine Learning*, 2014, pp. 82–90.

[14] ShahRukh Athar et al., “Weakly supervised fluid filled region localization in retinal oct scans,” in *International Symposium on Biomedical Imaging*, 2018, p. in press.