RetiNet: Automatic AMD identification in OCT volumetric data

S. Apostolopoulos, C. Ciller, S. De Zanet, S. Wolf and R. Sznitman

Abstract

Optical Coherence Tomography (OCT) provides a unique ability to image the eye retina in 3D at micrometer resolution and gives ophthalmologist the ability to visualize retinal diseases such as Age-Related Macular Degeneration (AMD). While visual inspection of OCT volumes remains the main method for AMD identification, doing so is time consuming as each cross-section within the volume must be inspected individually by the clinician. In much the same way, acquiring ground truth information for each cross-section is expensive and time consuming. This fact heavily limits the ability to acquire large amounts of groundtruth, which subsequently impacts the performance of learning-based methods geared at automatic pathology identification. To avoid this burden, we propose a novel strategy for automatic analysis of OCT volumes where only volume labels are needed. That is, we train a classifier in a semi-supervised manner to conduct this task. Our approach uses a novel Convolutional Neural Network (CNN) architecture, that only needs volume-level labels to be trained to automatically assess whether an OCT volume is healthy or contains AMD. Our architecture involves first learning a cross-section pathology classifier using pseudo-labels that could be corrupted and then leverage these towards a more accurate volume-level classification. We then show that our approach provides excellent performances on a publicly available dataset and outperforms a number of existing automatic techniques.

keywords — Optical Coherence Tomography (OCT), Convolutional Neural Networks (CNN), Age-Related Macular Degeneration (AMD), pathology identification, ophthalmology, machine learning

1 Introduction

By and large, Optical Coherence Tomography (OCT) has reshaped the field of ophthalmology ever since its inception in the early 90s [1]. At its core, OCT uses infrared-light interferometry to image through tissue in order to characterize anatomical structures beyond their surface. Given its simplicity, affordability and safety, it is no surprise that its use has gained widespread popularity for both disease diagnosis and treatment. Similarly, its use has gained traction in other medical fields such as for histopathology and skin cancer analysis [2].

Indeed, with an ability to image the posterior part of the eye in 3D (e.g. the retina) at micrometer resolution, OCT imaging now allows for visualization of most retinal layers [3][4] and more importantly, numerous pathological markers, such as intraretinal fluid, drusens or cysts [5][6]. As illustrated in Fig. 1 such markers can be observed in OCT cross-sectional images, or B-scans and have
been linked to a number of eye conditions, including Age-Related Macular Degeneration (AMD) and Diabetic Retinopathy (DR) which currently affect over 8.7% of the world population and 159 million people worldwide, respectively [5, 7, 8]. Moreover, these pathologies are the major cause of blindness in developed countries [9]. Alarming, the number of people with either of these diseases is projected to skyrocket, with AMD affecting an estimated 196 million people by 2020 and 288 million people by 2040 [8]. Genetic factors, race, smoking habits and the ever growing world population are responsible for this pathology growth [10].

While OCT has gained significant importance in recent years for AMD and DR screening [11, 12], the process to do so remains time consuming however. In effect, 3D OCT volumes, also referred to as C-scans, are comprised of 50-100 cross-sectional B-scans. Traditionally, inspection of each B-scan is necessary in order to properly rule-out most retinal diseases. This process is particularly tedious not only due to its time-consuming nature, but also due to the multiple cross-sections that need to be inspected simultaneously to identify elusive and scarce traces of early-stage ocular diseases. In this context, automated algorithms for pathology identification in OCT volumes would be of great benefit for clinicians and ophthalmologists, as access to OCT devices becomes common and nation-wide screening programs commence [13].

Recently, research has given way to a variety of image processing methods for OCT imaging. Some of these have included: techniques for image denoising [14, 15, 16], strategies for improved image reconstruction [17, 18, 19, 20, 21, 22], dosimetry laser control systems [23, 24, 25] or instrument detection during surgical procedures [26, 27].

More specific to pathology identification, various groups have explored automatic detection of retinal pathologies using machine learning techniques, either focusing on segmentation of relevant pathological markers [28, 29, 30, 31, 32] or classification of 2D B-scans or 3D C-scans [33, 34, 35, 36]. While effective to some extent, most of these works have leveraged B-scan level groundtruth information in order to learn classification functions. These more detailed labels are unfortunately often not available and as such, limit the usability of these solutions.

To this end, we present a new strategy towards automatic pathology identification in OCT C-scans using only volume level annotations. To do this, we introduce a novel Convolution Neural Network (CNN) architecture, named RetiNet, that directly estimates the state of a C-scan solely using the image data and without needing additional information. At its core, our approach uses (1) a task-specific volume pre-processing strategy where we flatten and normalize the data in an OCT-specific manner, (2) we then train a 2D B-scan CNN using pseudo-labels that could be corrupted in order to pre-learn filters that respond to relevant image features and (3) reuse the learned features in a C-scan level CNN that takes a mosaic of B-scans as input and classifies the entire C-scan at once. Using a publicly available OCT dataset [5], we show that our approach is highly effective at separating AMD
from control subjects and outperforms existing state-of-the-art methods for image classification. In addition, we not only show that RetiNet outperforms excellent recent networks from the computer vision literature trained from scratch, but also surpasses the performance of state-of-the-art pre-trained networks with adapted filters. Last, we show how our approach provides high performances in terms of accuracy, learning pathology-specific filters capable to identifying pathological markers effectively.

The remainder of this article is organized as follows: The following section discusses the relevant related work. Sec. 3 then describes in detail our approach and the RetiNet architecture. Following this, we describe our experimental section and the evaluation of several baseline strategies in Sec. 4. We then conclude with final remarks in Sec. 5.

2 Related Work

We now briefly discuss a number of related works on the topic of OCT data classification.

In Venhuizen et al. [35] regions of interest are automatically extracted around the center of each C-scan via an intensity threshold. Principal Component Analysis (PCA) is then applied to each region for dimensionality reduction, followed by K-means clustering in order to build a Bag of Words (BoW) representation, which is then used in combination with a Random Forest classifier. The classifier is trained on a set of 284 AMD patients and healthy controls and evaluated on a balanced set of 50 AMD patients and 50 healthy controls.

The same dataset was previously used by Farsiu et al. [5], who developed a semi-automatic classification method for AMD patients. Given manually-corrected segmentations of Bruch’s Membrane (BM), the Retinal Pigment Epithelium (RPE) and Inner Limiting Membrane (ILM) layers, they calculated a number of metrics: total thickness of the retina; thickness between drusen apaxes and the RPE; abnormal thickness score; abnormal thinness score. From these, they trained linear regression models using different combinations of these metrics.

Srinivasan et al. [34] presented a method for classifying AMD, Diabetic Macular Edema (DME) and healthy C-scans using multiscale Histogram of Gradients (HoG) features and a Support Vector Machine (SVM) classifier. Each B-scan was first resized to a resolution of 246x256 pixels, denoised using the Block Matching and 3D filtering (BM3D) algorithm [15] and then flattened. While the dataset for this method was public, it is only available in a preprocessed form, which unfortunately limits our ability to compare to it.

More recently, Lemaitre et al. [36] followed in the direction of Liu et al. [33], by extracting 2D and 3D Local Binary Patterns (LBP) features from a set of 16 healthy and 16 patients suffering from DME. B-scans were denoised using the non-local means (NLM) algorithm [37] and flattened. The use of different linear and non-linear classifiers was then explored to identify which performed best.

Finally, Schlegl et al. [32] employed a 2D patch-based Convolutional Neural Network (CNN) to classify retinal tissue into Intra-retinal Cysts (IRC), Subretinal Fluid (SRF) and healthy categories, while providing information to the location of the pathology. They train their classifier using three different ground truths: weak-labeling, wherein a single label is applied to the whole C-scan; weak-labeling with semantic information, wherein coarse information about the location of the pathology is applied along with the weak label; full-labeling, wherein the classifier is trained on the per-voxel ground truth of the whole C-scan. The latter approach yields the best results, with 97.7%, 89.61% and 91.98% for the healthy, IRC and SRF classes, respectively. While the weak-labeling approach performs significantly worse, at 65.63%, 21.94% and 90.30%, respectively, this setting is the most closely related to the one in the present work.

More specifically, we present a novel method to automatically evaluate AMD or healthy volumes. Our strategy has two important advantages over existing methods: (1) it relies only on volume level labels to be trained and (2) it evaluates complete volumes in one shot, making it simpler to use. As we will show in Sec. 4, our approach allows for significant performance gains over these existing methods.
3 Our Approach

The overall goal in this work is to automatically evaluate whether an OCT volume contains AMD. The main challenges in tackling this problem lies in the fact that (1) relatively few volumes are typically available for training classification models even though volumes are large in size (e.g. $500 \times 1000 \times 100$ pixels) and (2) that labels denoting the presence of pathology are only available at the volume level and not at the cross-section level.

To perform effective volume classification, we will follow a Deep CNN approach and will describe in the following section our novel architecture to do so. In general, our approach relies on a three-stage process.

The first is an OCT-specific normalization and data-augmentation strategy for OCT volumes in order to improve overall generalization and classification performance. Here, we reduce image dimensionality and flatten OCT scans in order to regularize the data. Similarly, we make use of symmetries particular to the eyes in order to augment the data effectively. The second stage attempts to learn pathology-specific features at cross-section B-scan levels using volume-level labels. Here, we make use of a relatively simple network to learn filters that are relevant for 2D OCT image data. In the last stage, we remap the volume to a large image mosaic and train a new volume-level network by leveraging the previously learned filters that operate at the B-scan level.

We now begin by formalizing our problem and establish the necessary notation to precisely describe our strategy.

3.1 Notation and formulation

Without loss of generality, we assume that our training data $V = \{V_1, \ldots, V_N\}$ is comprised of $N$ OCT volumes. Each volume $V_n$ is of dimension $\{W \times H \times D\} = \mathcal{V}$, where a B-scan cross-section consists of a $W \times D$ image, with $D$ being the depth of the penetrating OCT light source. For a volume $V_n$, we denote $B_n^h, h = 1, \ldots, H$, as the $h^{th}$ B-scan in the volume.

Each volume in $V_n$ is associated with a class label $Y_n \in \{0, 1\} = \mathcal{Y}$ such that 0 corresponds to control volumes and 1 corresponds to pathological volumes (i.e. AMD). Our goal is to learn a classification function $f : \mathcal{V} \rightarrow \mathcal{Y}$ using the training set and labels available. Importantly, we assume that no information is available on the labels of $B_n^h$, as these are expensive to gather.

3.2 Data preprocessing

3.2.1 Normalization

As can be seen in Fig. 2 (top row), there is both high variability in positions of $B_n^h$ with respect to the anatomy and distortion of the retinal layers. In particular, the retinal layers may be tilted, shifted vertically, distorted due to the acquisition process and of varying intensity. In addition, significant portions of $B_n^h$ images contain little informative content, such as the area above the ILM and below
Figure 3: (top) RetiNet B network layout. Single B-scans are taken as input through the FEATURE layers and classified in the CLASSIFICATION layer. Weak labels derived from the the volume label are used to train this network. (bottom) RetiNet C network layout. B-scans from an entire OCT volume are concatenated vertically into a single image as input. The FEATURE layers are transferred from RetiNet B with their weights to this network. An additional adaptation block is added, composed of a repeated layers of convolutions, average pooling and batch normalization.

As such, in order to provide a more compact and consistent set of training volumes, we are interested in normalizing and reducing the size of the OCT volumes. Unfortunately, simply cropping each B-scan would be ill-suited, since the retina is curved and this would either remove most of the informative data or result in marginal rescaling. To this end, we propose an effective normalization or flattening strategy.

Our flattening approach consists in aligning the individual BM layers, rectifying for the eye curvature and normalizing for variations in volume intensities. To do so, we first detect the BM layer by applying an anisotropic filter on Bh using 200 diffusion iterations \[ c \left( \| \nabla I \| \right) = \frac{1}{1 + \left( \frac{\| \nabla I \|}{\kappa} \right)^2}, \]

where \( \kappa = 50 \) was empirically set for all experiments. We then compute the Difference of Gaussians (DoG) from the filtered responses and estimate the BM layer as maximal gradient pixels. Naturally, these responses are noisy and incorrect in some cases. For this reason, we fit a second-order polynomial model to the noisy responses using RANSAC outlier detection \[39\]. We then warp the estimated BM to a vertical line centered at 60% of the image height. In order to reduce the dimension of the image, we resize each Bh to be of smaller size \( w \times d, w < W, d < D \). As can be seen in Fig. 2, areas above the ILM and below the BM only contain noise. For this reason, we crop every Bh by discarding every voxel v in the C-scan so that \( \frac{d}{4} \leq v_d \leq \frac{3d}{4} \), with \( v_d \) being the depth of the voxel.

Finally, intensity variations in OCT are common when looking at acquisitions over different patients. In order to regularize across these variations, the voxel intensities are normalized to be zero mean and with a standard deviation of one.

3.2.2 Data augmentation

We use data augmentation to increase the number of samples in our training dataset and reduce overfitting \[40\ \[41\]. In particular, we take advantage of the bilateral symmetry of the eye to effectively double the number of samples. The resulting samples are biologically plausible, i.e. the optic disc, fovea and vessels remain at the correct spots relative to each other, and removes any latent sample bias due to different counts of left and right eyes in the dataset.
3.3 B-scan classification with weak labels: RetiNet B

Recall that our data is inherently volumetric and that the amount of available data is relatively small. Given this challenging learning context, we will first learn features that are efficient at detecting typical 2D OCT structures by learning to “classify” B-scans. While our labels are only at the volume level, we propose to learn a B-scan level classifier by using approximately correct, or “weak” labels to do so. In particular, we let \( \hat{Y}_h^n = Y_n \), where \( \hat{Y}_h^n \) is the weak label for \( B_h^n \). Note that for all control volumes, the \( \hat{Y}_h^n = 0 \) labels indicate the lack of an AMD diagnosis – individual B-scan cross-sections should be pathology free. Conversely, volumes from subjects diagnosed with AMD may contain a number of control or non-pathological B-scans. In particular, up to 50% of the labels could be incorrect for such volumes.

To learn this classification function, we proceed by constructing a feed-forward CNN whose architecture is illustrated in Fig. 3 (top). In this network, every single gray-scale \( B_h^n \) is fed as input and passed through a set of 7 convolutional layers with small kernels (3 × 3, 5 × 5) and max-pooling layers. We define this set of consecutive layers as the FEATURE layers.

Following these layers, classification is achieved by using two consecutive fully connected layers and using a soft-max activation with two outputs for our two classes (i.e. control and AMD). We denote these latter layers as the CLASSIFICATION layers. Throughout the entire network, convolutional and dense layers make use of leaky Rectified Linear Units (ReLUs) activations. From this point on, we refer to this network as RetiNet B.

To train this network, we first begin by initializing all layer parameters randomly using Glorot Uniform sampling [42]. We then make use of Extreme Learning [43], as it has been shown to increase regularization by forcing the convolutional layers to map to a broader features space. In practice, once the CLASSIFICATION layers are initialized, we do not allow them to change. That is, we freeze these layers and only allow the FEATURE layers to be modified during the learning phase. In Sec. 4.4, we show the effect of this learning strategy when compared to traditional regimes.

3.4 Volume classification: RetiNet C

As we will show later in our experiments, the performance of the above network is limited, as it must learn from weak labels and does not make use of volumetric information to make a final decision. More so, it is not possible to test if the classification is in fact correct as the true label per B-scan is not known.

For this reason, we proceed to a second stage that attempts to classify the complete C-scan in one shot. Our proposed network is depicted in Fig. 3. Instead of setting each \( B_h^n \) as an input channel, our network takes as input a vertical image of stacked B-scans, \( M_n \), where

\[
M_n = \begin{bmatrix} B_1^n & \cdots & B_h^n \\ \vdots \end{bmatrix},
\]

resulting in a \( \{w \times Dh\} \) sized image. Using the learned FEATURE layers from the previous section, we include these into our new network as they are invariant to the size of the input and because ideally, these have learned what anatomical and pathological structures are relevant. These are then followed by 5 consecutive blocks of convolutional layers, average-pooling and batch normalization. Finally, we add the CLASSIFICATION layer without transferring weights from RetiNet B. We define this network configuration as RetiNet C.

To train RetiNet C, we freeze the FEATURE layers, as these were learned on B-scans and should respond in the same way as above to preserve useful features extracted in the previous phase. The rest of the network is then trained using the true labels \( Y_n \) to learn the remainder of the network layers.

4 Evaluation

We now detail the performance of our strategy in the task of AMD classification in OCT volumes. We compare our approach to a number of existing state-of-the-art baselines coming from both the
OCT pathology identification literature and the more general computer vision literature. We also provide qualitative results of our method, illustrating the different activation maps produced by our network and show how the different stages of our approach benefit the overall performance.

4.1 Data set

Our method was trained and evaluated on the publicly available dataset from Duke University [5]. This dataset was made available to find methods to define quantitative indicators for the presence of intermediate AMD. In this set, 384 Spectral Domain OCT volumes are present, of which 269 volumes come from subjects with intermediate AMD while the remaining 115 subjects volumes were collected from healthy subjects. All scans are centered on the foveal pit. Each volume is acquired with 1000 A-scans per B-scan and 100 B-scans per volume. This results volume dimensions of $100 \times 1000 \times 500\text{px}^3$. In general, the volumes are not isotropic.

4.2 Baselines

To illustrate how each part of our strategy influences overall performance, as well as to compare how our approach performs in contrast to other existing techniques in the literature, we now outline a number of baselines to which we will compare to directly:

- **VGG19**: is the 19-layer variant of the deep CNN approach for image classification described in [44]. We pre-trained this network on the ImageNet dataset and fine-tuned the resulting filters using the OCT dataset. To do that, we modified the receptive field of the network to match our B-scan resolution of 384x298 and exchanged the classification layer of the network with a fully-connected layer of size 2.

- **ResNet**: similar to VGG19, we evaluated a pre-trained version of the 152-layer residual network described by He et al. [45]. Due to the highly tuned parameters of this network, we maintained the size of the receptive field at 224x224, opting instead to resize our input volume dimension to match. As before, we exchange the classification layer with a fully-connected layer of size 2.

- **DenseNet**: is a recent architecture network by Huang et al. [46], which extends the residual network concept using a complete graph of skip connections. We implemented DenseNet with 3 dense blocks and a growth rate of 12, and trained the entire network on the OCT dataset.

- **2DSeg**: is the patch-based classification scheme for pathological OCT identification described in Schlegl et al. [32], which we re-implemented and trained on the OCT dataset. Due to the lack of location information or per-voxel classifications in our ground truth data, we focused on the weak-labeling approach described in this same paper.

In addition to our complete RetiNet C approach, we also compare its performance to:

- **RetiNet B (extreme)**: This consists of the RetiNet B classifier described in Sec. 3.3 learned with Extreme Learning [43]. By comparing RetiNet to this baseline, we can see the performance gain provided by the RetiNet C network construction.

- **RetiNet B**: Similarly to RetiNet B (extreme), this classifier is identical in structure to the network of RetiNet B but trained without Extreme Learning.

In addition, we attempted to train both VGG19 and ResNet using only the OCT data, but given the large size of these networks and the small size of the dataset, this yielded in extremely poor classification methods. To avoid bias, we omit these methods from our experiments.

4.3 Experimental setup

We partition the dataset into five randomized, equi-sized subsets, using four for training and one for testing, for a total of five cross-validations per network. All methods were trained on the same partitions using the same folds. The random seed was preserved across all runs in order to remove any dataset-dependent bias.
Table 1: Summary of parameters used for training both RetiNet B and RetiNet C.

| Variable                              | Value       |
|---------------------------------------|-------------|
| Cross-validation folds                | 5           |
| Training epochs                       | 100         |
| Early stopping patience               | 15          |
| Adadelta decay rate $\rho$            | 0.95        |
| Mini-batch size (RetiNet B/RetiNet C) | 20/1        |

We trained each network for a maximum of 100 epochs per fold, using early stopping with a patience of 15 epochs to avoid over-fitting [47, 48]. We relied on the adadelta algorithm [49] to optimize the parameters of each network. All networks except that of 2DSeg were optimized by minimizing the categorical cross-entropy of their predictions versus the ground truth. 2DSeg was optimized by minimizing the mean squared error, as described in [32].

VGG19, DenseNet and ResNet networks were trained and evaluated at the B-scan level using a weak labeling scheme, where the label of the complete C-scan was applied to each B-scan of that subject. The final C-scan classification prediction was defined as the mean score of the B-scan level predictions. The maximum achievable B-scan level accuracy is limited to roughly 94% due to mislabelings of individual B-scans (i.e. the C-scan of an AMD patient may contain a number of healthy B-scans), as well as acquisition artifacts (i.e. blinks).

We provide a version of our RetiNet C implementation online[^1] A complete list of the parameters used can be found in Table 1. These were selected using experimental validation. In Fig. 4, we show the learning rate of our network with the above parameters on the training data and on the validation set.

4.4 RetiNet characterization

As an initial set of experiments, we are interested in characterizing the performance of our strategy. In Fig. 5, we directly compare performances between RetiNet C, RetiNet B and RetiNet B (extreme) in terms of classification performance. The results shown in this figure are attained with 5-fold cross-validation.

First, we show in Fig. 5 (top) the ROC curves of each of the three strategies. In addition to this traditional metric, we also show in Fig. 5 (bottom) the False Negative Rate versus the False Positive

[^1]: Visit [https://github.com/thefiddler/retinet](https://github.com/thefiddler/retinet) for an implementation of RetiNet C.
A number of interesting conclusions can in addition be drawn from these results. First, off-the-shelf computer vision networks that perform exceptionally well on natural images and ResNet can see that for a 1% false negative rate (i.e., misclassifying pathological as healthy), RetiNet C has a 0.05% false positive rate (i.e., misclassifying healthy as pathological). This is interesting in the context of screening because it indicates that a human would need to evaluate pathological scans in any case and that the false positive rate indicates the reduced proportion of healthy scans one would still need to examine should one use an automatic classification algorithm. That is, allowing for a 1% error in predicted subjects would only require a 5% inspection of the healthy population.

With this, we can observe that RetiNet B has difficulty in learning correctly the volume labels given that it is trained on weak labels. As illustrated in the training and validation loss plots in Fig. 4, we can see that RetiNet B effectively lacks generalization capabilities as it heavily overfits the data.

In contrast, RetiNet B (extreme), with its Extreme Learning framework, allows for a stronger regularization and mitigates a significant amount of overfitting that is present with RetiNet B. As such, the difference in classification performance between RetiNet B (extreme) and RetiNet C can be attributed to the weak labels and the 2D nature of the strategy.

In this sense, these results highlight that RetiNet C overcomes the lack of B-scan level labels and that weak labels can be exploited at the C-scan level. To illustrate what our network learns however, we can visualize the network activation maps in Fig. 7. Here we show four (two healthy and two AMD) examples of volumes and how our network responds to them. For each case, we show three B-scan from a volume ($h = 25, 50, 75$), the associated fundus view of the RPE retinal layer and the projection of the activation map of the last convolutional layer RetiNet C. In particular, we can see for the AMD cases that the activation maps are responding very strongly at different locations of the volume and differently to healthy volumes.

Last, the learning of this two-stage network appears to learn even after 70 epochs, indicating that the overfitting is most likely limited. At the same time, we notice that learning is in general consistently noisy with our framework and this is most likely due to the limited number of C-scans available during training.

### 4.5 Baseline comparison

Fig. 6 outlines the performance of RetiNet C and the baseline methods in terms of ROC, as well as FNR/FPR. Across both metrics, RetiNet C appears to outperform these baselines.

A number of interesting conclusions can in addition be drawn from these results. First, off-the-shelf computer vision networks that perform exceptionally well on natural images VGG19, DenseNet and ResNet, have a strong tendency to overfit this data. Second, DenseNet performs very similarly to both VGG19 and ResNet even though it is trained from scratch and converges very quickly. Both VGG19 and ResNet could not be successfully trained from scratch however due to the relatively
Figure 6: Comparison of RetiNet C mosaic to state-of-the-art image classification networks: VGG19, ResNet and DenseNet, as well as the weakly-labeled approach described in Schlegl et al. [32]. VGG19 and ResNet are pre-trained on ImageNet and fine-tuned on the OCT dataset. The results are after a 5-fold cross-validation. Comparisons are with respect to (left) False Negative Rate (FNR) versus False Positive Rate (FPR) curves and (right) ROC curves.

small dataset. Last, the 2DSeg approach, which was specifically developed for this application, appears to have difficulties generalizing when training with weak labels. This is consistent with authors conclusion as well [32].

From our experiments, we can see that RetiNet C achieves an Area Under the Curve (AUC) of of 99.7%. This compares favorably to the semi-automatic method of Farsiu et al. [5], which achieved an AUC of 99.17% in their reported best case and the automatic method of Venhuizen et al. [35] which achieved an AUC of 98.4%. While we do not compare to these methods directly, we report these scores from their published results on the same data.

In this sense, it appears as though RetiNet C provides a stable strategy capable of leveraging volumetric information and only uses the volume level labels for training. Fig. 8 shows a few examples where RetiNet C correctly and incorrectly predicts different volume labels.

5 Conclusions

In this article, we have proposed a novel strategy for automatic identification of AMD in OCT volumes. Our strategy is advantageous as it only requires volume-level labels as opposed to cross-sectional labels, making it far easier to train from a groundtruth acquisition point of view. Our approach involves a novel two-stage deep learning architecture that in the first phase, focuses on learning features that are domain specific and then focuses on the volume classification task in the latter phase.

We validated our approach using publicly available OCT data and compared the performance of our method against both techniques from the OCT domain and the computer vision literature. We showed that not only does our approach do well in terms of ROC performance, but that it also does well with respect to a more clinically relevant metric.

This being said, our method still has difficulties identifying mild AMD cases as shown in Fig. 8(e) where the difference between healthy and pathological is visibly challenging. In this sense, we will focus in the future on developing strategies for identifying early-stages of the disease and we will look at how diseases differ to one another.

References

[1] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, “Optical Coherence Tomography HHS Public Access,” Science. November, vol. 22, no. 2545035, pp. 1178–1181, 1991.
Figure 7: Activation maps of the last convolutional RetiNet C layer for four different volumes. The top row depicts two control subjects, while the bottom row depicts two AMD subjects. Each subject is described in three columns: (left) Three B-scans at slice 25, 50 and 75; (center) the top-down Fundus reconstruction of the RPE layer; (right) Activation map of the last convolutional layer of the RetiNet C configuration. The activation maps highlights pathological structures relevant for AMD.

[2] J. Welzel, “Optical coherence tomography in dermatology : a review,” Skin Res Technol., vol. 7, no. 1, pp. 1–9, 2001.
[3] M. D. Abramoff, M. K. Garvin, and M. Sonka, “Retinal Imaging and Image Analysis,” IEEE Transactions on Medical Imaging, vol. 3, no. 1, pp. 169–208, 2010.
[4] M. K. Garvin, M. D. Abràmoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, “Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images.” IEEE Transactions on Medical Imaging, vol. 28, no. 9, pp. 1436–1447, 2009.
[5] S. Farsiu, S. J. Chiu, R. V. O’Connell, F. A. Folgar, E. Yuan, J. A. Izatt, and C. A. Toth, “Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography,” Ophthalmology, vol. 121, no. 1, pp. 162–172, 2014.
[6] R. D. Jager, W. F. Mieler, and J. W. Miller, “Age-Related Macular Degeneration,” The New England Journal of Medicine, vol. 358, no. 24, pp. 2606–2617, 2008.
[7] J. W. Y. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S.-J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund, S. Haffner, R. F. Hamman, M. K. Ikram, T. Kayama, B. E. K. Klein, R. Klein, S. Krishnaiah, K. Mayurasakorn, J. P. O’Hare, T. J. Orchard, M. Porta, M. Rema, M. S. Roy, T. Sharma, J. Shaw, H. Taylor, J. M. Tielksch, R. Varma, J. J. Wang, N. Wang, S. West, L. Xu, M. Yasuda, X. Zhang, P. Mitchell, and T. Y. Wong, “Global prevalence and major risk factors of diabetic retinopathy,” Diabetes Care, vol. 35, no. 3, pp. 556–564, 2012.
[8] W. L. Wong, X. Su, X. Li, C. M. G. Cheung, R. Klein, C. Y. Cheng, and T. Y. Wong, “Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: A systematic review and meta-analysis,” The Lancet Global Health, vol. 2, no. 2, pp. e106–e116, 2014.
[9] N. Bressler, “Age-related macular degeneration is the leading cause of blindness,” JAMA, vol. 291, no. 15, pp. 1900–1901, 2004.
[10] Age-related Eye Disease Study Research Group, “Risk factors associated with age-related macular degeneration. A case-control study in the age-related eye disease study: Age-Related Eye Disease Study Report Number 3.” Ophthalmology, vol. 107, no. 12, pp. 2224–2232, 2000.
[11] N. M. Bressler, “Early detection and treatment of neovascular age-related macular degeneration,” The Journal of the American Board of Family Practice / American Board of Family Practice, vol. 15, no. 2, pp. 142–52, 2002.
[12] T. A. Ciulla, A. G. Amador, and B. Zinman, “Diabetic retinopathy and diabetic macular edema: Pathophysiology, screening, and novel therapies,” *Diabetes Care*, vol. 26, no. 9, pp. 2653–2664, 2003.

[13] U. Schmidt-Erfurth, V. Chong, A. Loewenstein, M. Larsen, E. Souied, R. Schlingemann, B. Eldem, J. Mones, G. Richard, F. Bandello, and S. European Society of Retina, “Guidelines for the management of neovascular age-related macular degeneration by the European Society of Retina Specialists (EURETINA),” *Br J Ophthalmol*, vol. 98, no. 9, pp. 1144–1167, 2014.

[14] D. C. Adler, T. H. Ko, and J. G. Fujimoto, “Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter,” *Opt. Lett.*, vol. 29, no. 24, pp. 2878–2880, Dec 2004.

[15] K. Dabov and A. Foi, “Image denoising with block-matching and 3D filtering,” *Electronic Imaging*, vol. 6064, pp. 1–12, 2006.

[16] M. Maggioni, V. Katkovnik, K. Egiazarian, S. Member, and A. Foi, “Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction,” *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 119–133, 2013.

[17] M. Wojtkowski, V. J. Srinivasan, T. H. Ko, J. G. Fujimoto, A. Kowalczyk, J. S. Duker, and D. J. Fujimoto JG, Kowalczyk A, “Ultrahigh-resolution, high-speed, Fourier domain optical coherence tomography and methods for dispersion compensation,” *Optics Express*, vol. 12, no. 11, p. 2404, 2004.

[18] R. J. Zawadzki, A. R. Fuller, S. S. Choi, D. F. Wiley, B. Hamann, and J. S. Werner, “Correction of motion artifacts and scanning beam distortions in 3D ophthalmic optical coherence tomography imaging,” *Ophthalmic Technologies XVII*, vol. 6426, no. x, p. 42607, 2007.

[19] M. F. Kraus, B. Potsaid, M. A. Mayer, R. Bock, B. Baumann, J. J. Liu, J. Hornegger, and J. G. Fujimoto, “Motion correction in optical coherence tomography volumes on a per A-scan basis using orthogonal scan patterns.” *Biomedical optics express*, vol. 3, no. 6, pp. 1182–99, 2012.

[20] M. Szkulmowski and M. Wojtkowski, “Averaging techniques for OCT imaging,” *Opt Express*, vol. 21, no. 8, pp. 9757–9773, 2013.

[21] A. Montuoro, J. Wu, S. Waldstein, B. Gerendas, G. Langs, C. Simader, and U. Schmidt-Erfurth, “Motion artefact correction in retinal optical coherence tomography using local symmetry,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014.

[22] P. Steiner, J. H. Kowal, B. Považay, C. Meier, and R. Sznitman, “Automatic estimation of noise parameters in Fourier-domain optical coherence tomography cross sectional images using statistical information,” *Applied Optics*, vol. 54, no. 12, pp. 3650–3657, 2015.

[23] H. H. Müller, L. Ptaszynski, K. Schlott, C. Debbele, M. Bever, S. Koizner, R. Birngruber, R. Brinkmann, and G. Hüttmann, “Imaging thermal expansion and retinal tissue changes during photocoagulation by high speed OCT,” *Biomedical Optics Express*, vol. 3, no. 5, p. 1025, 2012.

[24] P. Steiner, A. Ebner, L. E. Berger, M. Zinkernagel, C. Meier, J. H. Kowal, C. Framme, R. Brinkmann, S. Wolf, and R. Sznitman, “Time-Resolved Ultra-High Resolution Optical Coherence Tomography for Real-Time Monitoring of Selective Retina Therapy,” *Investigative Ophthalmology and Visual Science*, vol. 56, pp. 6654–6662, 2015.

[25] S. Zhinden, S. S. Kucur, P. Steiner, S. Wolf, and R. Sznitman, “Automatic assessment of time-resolved oc images for selective retina therapy,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 6, pp. 863–871, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11548-016-1383-6

[26] Y. K. Tao, S. K. Srivastava, J. P. Ehlers, and C. Clinic, “Microscope-integrated intraoperative OCT with electrically tunable focus and heads-up display for imaging of ophthalmic surgical maneuvers,” *Biomed Opt Express.*, vol. 5, no. 6, pp. 1342–1350, 2014.

[27] M. T. El-Haddad and Y. K. Tao, “Automated stereo vision instrument tracking for intraoperative OCT guided anterior segment ophthalmic surgical maneuvers.” *Biomedical optics express*, vol. 6, no. 8, pp. 3014–31, 2015. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4541528[&amp;tool=pmcentrez[&amp;rendertype=abstract]}
[28] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abr`amoff, and M. Sonka, “Three-dimensional analysis of retinal layer texture: Identification of fluid-filled regions in SD-OCT of the macula,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1321–1330, 2010.

[29] S. J. Chiu, J. A. Izatt, R. V. O’Connell, K. P. Winter, C. A. Toth, and S. Farsiu, “Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images,” *Investigative Ophthalmology and Visual Science*, vol. 53, no. 1, pp. 53–61, 2012.

[30] P. A. Dufour, L. Ceklic, H. Abdillahi, S. Schroder, S. De Zanet, U. Wolf-Schnurrbusch, and J. Kowal, “Graph-based multi-surface segmentation of OCT data using trained hard and soft constraints,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 3, pp. 531–543, 2013.

[31] F. G. Venhuizen, M. J. J. P. V. Grinsven, and C. B. Hoyng, “Vendor Independent Cyst Segmentation in Retinal SD-OCT Volumes using a Combination of Multiple Scale Convolutional Neural Networks,” in *Medical Image Computing and Computer Assisted Intervention - Challenge on Retinal Cyst Segmentation*, 2015.

[32] T. Schlegl, S. M. Waldstein, and U. M. Schmidt-erfurth, “Predicting Semantic Descriptions from Medical Images with Convolutional Neural Networks,” in *Information Processing in Medical Imaging*, 2015, pp. 437–448.

[33] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, and J. M. Rehg, “Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid with local binary patterns,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*. Springer, 2010, pp. 1–9.

[34] P. P. Srinivasan, L. a. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. a. Izatt, and S. Farsiu, “Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images.” *Biomedical optics express*, vol. 5, no. 10, pp. 3568–77, 2014.

[35] F. G. Venhuizen, B. van Ginneken, B. Bloemen, M. J. J. P. van Grinsven, R. Philipsen, C. Hoyng, T. Theelen, and C. I. Sanchez, “Automated age-related macular degeneration classification in oct using unsupervised feature learning,” in *Proc. SPIE*, vol. 9414, 2015, pp. 94141I–94141I–7.

[36] G. Lemaitre, M. Rastgoo, J. Massich, C. Y. Cheung, Y. Wong, E. Lamoureux, D. Milea, M. Fabric, G. Lemaitre, M. Rastgoo, J. Massich, C. Y. Cheung, T. Y. Wong, and G. Lema, “Classification of SD-OCT Volumes using Local Binary Patterns: Experimental Validation for DME Detection,” *Journal of Ophthalmology*, vol. 6, 2016.

[37] A. Buades, B. Coll, and J.-M. J.-M. Morel, “A non-local algorithm for image denoising,” *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, no. 0, pp. 60–65 vol. 2, 2005.

[38] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[39] M. A. Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381 – 395, 1981. [Online]. Available: [http://dx.doi.org/10.1145/358669.358692](http://dx.doi.org/10.1145/358669.358692)

[40] P. Y. Simard, D. Steinkraus, and J. C. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings*. vol. 1, no. Icdar, pp. 958–963, 2003. [Online]. Available: [http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?urlnumber=1227801](http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?urlnumber=1227801)

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.

[42] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 9, pp. 249–256, 2010.

[43] G.-B. Huang, Q.-y. Zhu, C.-k. Siew, G.-b. H. Á, Q.-y. Zhu, C.-k. Siew, G.-B. Huang, Q.-y. Zhu, and C.-k. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.
Figure 8: Example B-scans from correctly and incorrectly classified volumes. While (a-d) show correctly identified cases, (e-f) are incorrectly classified. Surprisingly, our approach correctly identifies (d) as non-AMD, even though it illustrates an epiretinal membrane and vitreoretinal traction, neither of which is AMD.

[44] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” ArXiv preprint arXiv:1409.1556, pp. 1–14, 2015.
[45] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Arxiv.Org, vol. 7, no. 3, pp. 171–180, 2015.
[46] G. Huang, Z. Liu, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” ArXiv preprint, pp. 1–12, 2016.
[47] L. Prechelt, “Automatic early stopping using cross validation: Quantifying the criteria,” Neural Networks, vol. 11, no. 4, pp. 761–767, 1998.
[48] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, “Exploring Strategies for Training Deep Neural Networks,” Journal of Machine Learning Research, vol. 1, pp. 1–40, 2009.
[49] M. D. Zeiler, “ADADELTA: An Adaptive Learning Rate Method,” arXiv, p. 6, 2012. [Online]. Available: [http://arxiv.org/abs/1212.5701]