Retinal Fluid Segmentation and Classification in OCT Images Using Adversarial Loss Based CNN

Ruwan Tennakoon, Amirali Khodadadian Gostar, Reza Hoseinnezhad and Alireza Bab-Hadiashar

School of Engineering, RMIT University, Melbourne, Australia
{ruwan.tennakoon,amirali.khodadadian,reza.hoseinnezhad,abh}@rmit.edu.au

Abstract. This paper proposes a novel method in order to detect the presence and obtain voxel-level segmentation for three fluid lesion types (IRF/SRF/PED) in OCT images provided by the ReTOUCH challenge. The method is based on a deep neural network consisting of encoding and de-coding blocks connected with skip-connections which was trained using a combined cost function comprising of cross-entropy, dice and adversarial loss terms. The segmentation results on a held-out validation set shows that the network architecture and the loss functions used has resulted in improved retinal fluid segmentation.

Keywords: retinal fluid segmentation, Optical Coherence Tomography (OCT), OCT segmentation, macular edema, deep learning, GAN

1 Introduction

Macular Edema (ME) is the swelling of the macular region of the eye, caused by fluid buildups due to disruptions in blood-retinal barrier [1]. ME is prevalent in advanced stages of many retinal diseases including: diabetic retinopathy (leading cause of new-onset blindness in many industrialised countries [2]), age related macular degeneration (the leading cause of irreversible blindness in people aged 50 years or older in the developed world [3]) and retinal vain occlusion. Although methods such as anti-vascular endothelial growth factor (anti-VEGF) therapy are successfully used in treating patients with ME, availability of robust and sensitive imaging biomarkers would lead to better prediction of treatment requirements as well as enabling personalised treatment regimes that reduces the burden on patient and health care systems [7].

Since its advent in early 90's, optical coherence tomography (OCT) is becoming a valuable tool in the diagnosis and management of retinal disorders including ME [4]. Total retinal thickness measured from OCT images, is extensively used for diagnosing ME. However, it is shown that the retinal fluid volume provides a more accurate indication of vascular permeability in ME [7]. In order to have a pragmatic clinical setting for retinal fluids detection from OCT cross sections an automated image processing algorithm is necessary.

Automated fluid segmentation in OCT images is challenging because of number of reasons namely: high variability in the appearance of pathology on images, high speckle noise and motion artefacts inherent in OCT images (specially in images with severe pathology). Unlike computer tomography (CT) images, the OCT image pixel value does not have an absolute scale. Furthermore, image capturing procedures (i.e. scan patterns, resolution, filtering,) vary widely between different vendors. These factors introduces additional challenges in designing methods that work across images acquired using hardware produced by different vendors.

To stimulate research in this area, the Retinal OCT Fluid Challenge (ReTOUCH), has made available a relatively large dataset of spectral domain OCT scans, acquired using machines from three different vendors, with accompanying reference annotations. The objective of the challenge is to predict the presence and the segmentation of three classes of retinal fluid buildups: Intra-retinal fluid (IRF), Sub-retinal fluid (SRF) and Pigment Epithelial Detachment (PED).

In this paper we propose an end-to-end trained deep learning based retinal fluid segmentation method that works well across 3D-OCT images acquired using hardware from multiple vendors. The proposed methods does not require any additional information such as layer segmentation for training or pre-processing. Many image segmentation methods use a post-processing step based on conditional random fields (CRF) to smooth the segmentation map. However, due to computational complexity such methods can only encode first order information. The proposed method uses a adversarial network (that is learned simultaneously with the segmentation network) based loss function that can encode higher order relationships between image pixels. Hence eliminating the need for an additional post processing steps.

2 Background

2.1 Retinal Fluid Segmentation

Numarous semi or fully automatic methods for retinal fluid segmentation has been proposed in literature. Quellec et al. [13] proposed a method for identifying the 2-D footprint of IRF, SRF as well as PED in the macula, that relies on the characterisation of ten automatically segmented intra-retinal layers as well as 3-D textural features. This method was extended to 3D segmentation using a graph-theoretic approach in [14]. In [15] authors employed a kernel regression based classification method in conjunction with graph theory and dynamic programming (GTDP) framework in order to to identify fluid and eight retinal layer boundaries in clinical images containing severe DME.

These algorithms operate on 2D OCT slices, only. Recently, several new methods have emerged which uses volumetric images. Using volumetric image enables the integration of spatial context from multiple B-scans, thus is expected to perform better. Wang et al. [5] presents a volumetric OCT fluid segmentation method which applies fuzzy clustering and level-set technique.Whereas, the method proposed in [6] combines unsupervised feature representation and heterogeneous spatial context with a graph-theoretic surface segmentation to generate a 3D fluid segmentation and layer segmentation simultaneously.

2.2 Deep Learning for Medical Image Segmentation and Classification

Since winning the ImageNet competition in 2012 , deep-learning method has gained lot of attention in computer vision community with many applications in image segmentation. The main challenge in applying deep-learning to medical imaging is the lack of large training image sets with ground-truth annotations. U-Net architecture [12], comprising a stacked convolution based encoder followed by deconvolution based decoder, proposed for segmentation of neuronal structures in electron microscopic stacks, is known to perform well in medical imaging applications with limited data. A key contributor to the success of U-Net is the skip connection between encoder and decoder which enables the network to produce high resolution segmentation maps.

Recently, several deep-learning based volumetric retinal fluid segmentation methods have been proposed where their architecture was inspired in part by U-Net. Roy et.al [10] proposed an end-to-end learning framework – ReLayNet – for segmentation of multiple retinal layers and delineation of fluid pockets in eye OCT images. However, this method requires ground-truth segmentation of both retinal layer and fluid segmentation during training . Lee et al. [11] also used U-Net inspired deep network for fluid segmentation without layer segmentations. The drawback with these methods is that they could not differentiate between fluid types (i.e. inter-retinal, sub-retinal).

3 Proposed Method

Given a set of 3-dimensional OCT scans $\mathcal{X} = [x^{(i)}]_{i=1}^{N}; x^{(i)} \in \mathbb{R}^{h_i \times w_i \times z_i}$ and the corresponding ground truth segmentations $\mathcal{S} = [s^{(i)}]_{i=1}^{N}; s^{(i)} \in \mathbb{R}^{h_i \times w_i \times z_i}$, our objective here is to learn two parametrised functions: 1) $f_{jc}(x^{(i)}; \theta_s)$ that predicts the probability of each voxel, $j \in \mathcal{I}$, in an unseen input image, $x^{(N+i)}$, belonging to a particular class $c \in [0 \dots 3]$ (0: no pathology, 1: IRF, 2: SRF, 3: PED) 2) $g_c(x^{(i)}; \theta_g)$ that predicts the probability that an image $x^{(N+i)}$ contains class c pathology.

3.1 Network Architecture

Segmentation network: We use a deep neural network, inspired by U-Net [12] architecture, to model the function $f_{jc}(x^{(i)}; \theta_s)$. The proposed architecture differs form the traditional U-Net in following ways (apart from the changes in filter shape and number of layers):

- A batch normalisation layer is added after each block of convolutions/deconvolution which is known to improve training efficiency.
- Dropout is employed at each skip connection to prevent over-fitting.
- Final convolution layer of U-Net only use the output from the final decoder block whereas, the proposed network uses data from multiple scales which incorporate information from a large field-of-view.

- An adversarial network is used during training to encode higher order relationships between image regions.

An overview of the network architecture is given in Fig. 1. The individual network architectures are also described in the supplementary materials. The CNN is trained using *the combined cost function* given in the section 3.2. The results show that the modifications made in the architecture, significantly improved the overall accuracy of the network. The proposed network is fully-convolutional, therefore, training could be carried out using patches (of shape $[256, 128, 3]$) and during testing, the network can take the whole OCT image as input and produce the corresponding mask.

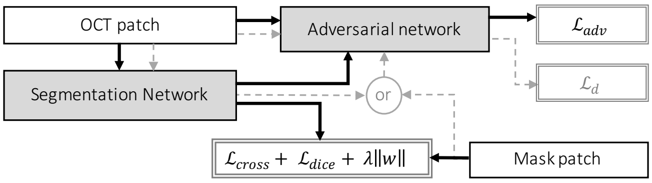

Fig. 1. Overall segmentation network architecture for the training phase. Dashed lines show the information flow of the adversarial network training step (Section 3.4).

Classification network: The classification network take the last layer features of the segmentation network as input and send them trough a NN with a convolutional and a global average pooling layer followed by 3-dense layers with soft-max activations (one for each fluid type).

3.2 Loss Function

The overall loss function used to train the segmentation network is as follows:

$$\mathcal{L} = \mathcal{L}_{cross} + \lambda_d \mathcal{L}_{dice} + \lambda_a \mathcal{L}_{adv} + \lambda_w \left\| w \right\|_2 . \tag{1}$$

Here $\|w\|_2$ is the l2-norm of the CNN weights and λ_j are hyper parameters. The first part of the loss function, $\mathcal{L}_{cross}$ is the class balanced categorical cross entropy loss, given by:

$$\mathcal{L}_{cross} = \mathbb{E}_i \left[-\sum_{c \in C} w_c^{(i)} \sum_{j \in \mathcal{I}} s_{jc}^{(i)} \ln f_{jc}(x^{(i)}) \right] . \tag{2}$$

Commonly, the fraction of an OCT image containing pathology is small. Therefor, using the unbalanced cross entropy loss for segmentation would result in a trivial solution (i.e. predict all pixels as belonging to the background class). To overcome this issue we used a ground truth dependent normalisation coefficient $w_c^{(i)} = {}^1/\sum_{j \in \mathcal{I}} p_{jc}^{(i)}$. However, using only this would lead to high false positives rate in the no-pathelogy class. Therefore we added a second term which is a smoothed version of dice index that account for false-positives of no-pathology class:

$$\mathcal{L}_{dice} = \mathbb{E}_i \left[1 - \left(\frac{2 \sum_j s_{j\bar{0}}^{(i)} f_{j\bar{0}}(x^{(i)})}{\sum_j s_{j\bar{0}}^{(i)} + \sum_j f_{j\bar{0}}(x^{(i)})} \right) \right] \tag{3}$$

where $\bar{0}$ refers to has-pathology.

Many image segmentation algorithms use a post-processing step that utilise a generative model such as a CRF to filter the resulting segmentation masks. However, typical CRFs used in literature only incorporate pairwise relationships due to computational complexity. In this work we utilise conditional adversarial loss that is capable of incorporating higher order information [9]. The adversarial loss [8] is obtained using a second parametrised model, $d\left(x^{(i)}, \hat{s}^{(i)}; \theta_d\right)$, that takes an image $(x^{(i)})$ and a segmentation mask $(\hat{s}^{(i)})$ as input and predict the probability that the segmentation mask is from the ground truth segmentation set $(\mathcal{S})$. The adversarial model $d(\cdot)$ is trained to differentiate segmentation maps generated by the model, $f(\cdot)$ from those in the ground-truth $\mathcal{S}$ using loss function: $\mathcal{L}_d = -\mathbb{E}_i \left[z \ln d\left(x^{(i)}, \hat{s}^{(i)}\right) + (1-z) \ln\left(1 - d\left(x^{(i)}, \hat{s}^{(i)}\right)\right)\right]$, where $z = 1$ if $\hat{s}^{(i)} \in \mathcal{S}$ and 0 otherwise. The segmentation model is then trained to fool the adversarial network by adding the following loss:

$$\mathcal{L}_{adv} = \mathbb{E}_i \left[-\ln d\left(x^{(i)}, f(x^{(i)})\right)\right] \tag{4}$$

3.3 Image Pre-processing and Patch Extraction

The OCT images in the training-validation set have been acquired using four OCT devices from three different vendors (Cirrus: Zeiss Meditec, Spectralis : Heidelberg Engineering, T-1000 and T-2000: Topcon). As a result, high level of variability in voxel intensities (i.e. range and distribution) exist. During preprocessing the voxel values were normalised to range $[0 - 255]$ followed by a histogram equalisation step using a randomly selected OCT scan as the template. Next, a median filter was applied to Topcon and Cirrus images along the z-direction (Spectralis images contain significantly lower speckle noise). The OCT scans from different vendors have different element spacing. Therefore, we resized the images in the second dimension using Spectralis images as the reference.

A typical OCT image contains large volume of uninteresting regions (outside the retinal layers) and extracting patches uniformly distributed across the whole volume would result in a large number of uninformative patches. Therefore, in our implementation we used image entropy to identify the regions on the image with useful information and patches were sampled such that they have high probability to include these regions. The patches were also augmented by rotations and flipping. It should be noted that our method does not use any additional information (i.e. layer segmentations, models pre-trained on OCT images in training or for pre-processing tasks such as retinal flattening). However if such information is available they can be used to improve the performance of the proposed method.

3.4 Training

Segmentation network: 3D-patches extracted from OCT images using the procedure explained in Sec. 3.3 were compiled into mini-batches. The parameters of the segmentation net was trained for 100 epoch using Adam optimiser to obtain the final CNN weights (θ_s, θ_d) as follows:

```
for each mini-batch:
    Train $f_{jc}(x^{(i)}; \theta_s)$ for one iteration using $\mathcal{L}$ ($\theta_d$ fixed)
    Train $d\left(x^{(i)}, \hat{s}^{(i)}; \theta_d\right)$ for one iteration using $\mathcal{L}_d$
```

Classification network: Once the segmentation network has converged, its weights were frozen and the the classification network parameters (θ_g) were trained using categorical cross entropy loss. After training is completed, each training OCT image was partitioned into overlapping B-scan segments (3 slices for each segment) and each partition was send through the combined network. The output after the average pooling layer for each segment is then aggregated using simple statistical functions (i.e. mean, max, min and median) to produce a combined image-wise feature vector. These image-wise feature vectors are used to learn three random forest classier that output whether a fluid-type is present in an image or not (10-fold cross validation was used in training the RF).

4 Results and Discussion

4.1 Data-set and Implementation

Seventy volumetric OCT scans, provided in the ReTOUCH challenge, was used to develop the proposed method. The data was randomly divided into two non overlapping segments of 75%-25% for training and validation respectively (validation set: 18 images, 6 from each vendor). The method was implemented using *Keras* library with *Tensor-flow* backend. The code is publicly available at: `https://github.com/RuwanT/retouch` . The segmentation results were evaluated on the validation set using the following 3D metric: dice-index (DI) and absolute volume difference (AVD).

4.2 Validation Results

Fluid Segmentation: The validation results of the proposed method trained with the combined cost function (RF-SNET-Full) is reported in Table. 1 together with those obtained using: typical U-Net (UNET-Lc), proposed architecture trained with only $\mathcal{L}_{cross}$ (RF-SNET-Lc) and with $\mathcal{L}_{cross}+\mathcal{L}_{dice}$ (RF-SNET-Lcd). The results show that the proposed modifications has resulted in significant improvement in segmentation results with the proposed method obtaining a overall DI of 0.75 and an AVD of 0.1. We see some variation in vendor specific segmentation results, with DI for "Spectralis" images reaching 0.83 whereas the other two only obtaining DI of 0.74 and 0.66. The individual fluid type segmentation results for RF-SNET-Full are also shown in the table. It should be noted that the DI and AVD are skewed when there are no positive voxels in the image. To overcome this, we used the predictions from the classification network to refine positive predicted voxels in the segmentation. Qualitative analysis (provided with the supplementary materials) showed that most errors occur near pathology boundaries.

Predicting presence of pathology: The results in Table. 2 show that the prediction model has achieved an overall accuracy of over 90% in predicting the presence of pathology in 3D-OCT images.

Table 1. Validation results for fluid segmentation. First four rows of the table compare different models using DI and AVD calculated on pathology vs no-pathology.

	Dice Index				**Absolute Volume Difference**			
	Spectralis	**Cirrus**	**Topcon**	**Mean**	**Spectralis**	**Cirrus**	**Topcon**	**Mean**
UNET-Lc	0.30	0.18	0.06	**0.18**	0.64	0.76	0.67	**0.69**
RF-SNET-Lc	0.80	0.65	0.57	**0.67**	0.07	0.22	0.33	**0.20**
RF-SNET-Lcd	0.83	0.69	0.67	**0.73**	0.03	0.17	0.15	**0.12**
RF-SNET-Full	0.83	0.74	0.66	**0.75**	0.03	0.12	0.16	**0.11**
				RF-SNET-Full				
IRF	0.64	0.74	0.68	**0.69**	0.23	0.16	0.09	**0.16**
SRF	0.63	0.77	0.62	**0.67**	0.33	0.19	0.22	**0.25**
PED	0.85	0.88	0.82	**0.85**	0.09	0.06	0.08	**0.08**

Table 2. Validation results on detecting the presence of individual fluid typesin 3D-OCT.

	IRF	**SRF**	**PED**	**Combined**
Accuracy	94%	83%	100%	93%
Sensitivity	100%	80%	100%	94%
Specificity	75%	87%	100%	90%

5 Conclusion

We propose a deep learning based method to solve the retinal fluid segmentation and classification problem in OCT images. The proposed method is based on a modified version of U-Net architecture trained using a combined loss function consisting of a adversarial loss term. Validation results show that the proposed method has been successful in predicting the presence and voxel-level segmentation of retinal fluids.

References

1. Marmor, M.F.: Mechanisms of fluid accumulation in retinal edema. In: 2nd International Symposium on Macular Edema, pp 35–45. Kluwer Academic Publishers (2000)
2. World Health Organization: Prevention of blindness from diabetes mellitus: report of a WHO consultation in Geneva, Switzerland. World Health Organization (2006)
3. Pascolini, D.: Global update of available data on visual impairment: a compilation of population based prevalence studies. Ophthalmic epidemiology, 67–115 (2004)
4. Trichonas, G. and Kaiser, P.K.: Optical coherence tomography imaging of macular oedema. British Journal of Ophthalmology, 98(Suppl 2), ii24–ii29 (2014)
5. Wang, Jie et al.: Automated Volumetric Segmentation of Retinal Fluid on Optical Coherence Tomography. Biomedical Optics Express 7.4, 1577–1589. PMC. (2017)
6. A. Montuoro, S. Waldstein, B. Gerendas, U. Schmidt-Erfurth, and H. Bogunovi?: Joint retinal layer and fluid segmentation in OCT scans of eyes with severe macular edema using unsupervised representation and auto-context, Biomed. Opt. Express 8, 1874–1888 (2017)
7. S. M. Waldstein, A.-M. Philip, R. Leitner, C. Simader, G. Langs, B. S. Gerendas, and U. Schmidt-Erfurth,: Correlation of 3-dimensionally quantified intraretinal and subretinal fluid with visual acuity in neovascular age-related macular degeneration, JAMA Ophthalmol. 134, 182–190 (2016)
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.,: Generative adversarial nets. In: Advances in neural information processing systems, 2672–2680 (2014).
9. Luc, P., Couprie, C., Chintala, S. and Verbeek, J., : Semantic segmentation using adversarial networks. arXiv preprint arXiv:1611.08408. (2016)
10. Roy, A.G., Conjeti, S., Karri, S.P.K., Sheet, D., Katouzian, A., Wachinger, C. and Navab, N.,: ReLayNet: Retinal Layer and Fluid Segmentation of Macular Optical Coherence Tomography using Fully Convolutional Network. arXiv preprint arXiv:1704.02161. (2017)
11. Cecilia S. Lee, Ariel J. Tyring, Nicolaas P. Deruyter, Yue Wu, Ariel Rokem, and Aaron Y. Lee, : Deep-learning based, automated segmentation of macular edema in optical coherence tomography, Biomed. Opt. Express 8, 3440-3448 (2017)
12. Ronneberger, O., Fischer, P. and Brox, T.,: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, 234–241. Springer, Cham. (2015)
13. Quellec, G., Lee, K., Dolejsi, M., Garvin, M.K., Abramoff, M.D. and Sonka, M.,: Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in SD-OCT of the macula. IEEE transactions on medical imaging, 29(6), 1321–1330. (2010)
14. Chen, X., Niemeijer, M., Zhang, L., Lee, K., Abrmoff, M.D. and Sonka, M.,: Three-dimensional segmentation of fluid-associated abnormalities in retinal OCT: probability constrained graph-search-graph-cut. IEEE transactions on medical imaging, 31(8), 1521–1531. (2012)
15. Chiu, S.J., Allingham, M.J., Mettu, P.S., Cousins, S.W., Izatt, J.A. and Farsiu, S.,: Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. Biomedical optics express, 6(4), 1172–1194. (2015)