Deep Learning Based Sub-Retinal Fluid Segmentation in Central Serous Chorioretinopathy Optical Coherence Tomography Scans

Narendra Rao T J, Girish G N*, Abhishek R. Kothari and Jeny Rajan

***Abstract*— Development of an automated sub-retinal fluid segmentation technique from optical coherence tomography (OCT) scans is faced with challenges such as noise and motion artifacts present in OCT images, variation in size, shape and location of fluid pockets within the retina. The ability of a fully convolutional neural network to automatically learn significant low level features to differentiate subtle spatial variations makes it suitable for retinal fluid segmentation task. Hence, a fully convolutional neural network has been proposed in this work for the automatic segmentation of sub-retinal fluid in OCT scans of central serous chorioretinopathy (CSC) pathology. The proposed method has been evaluated on a dataset of 15 OCT volumes and an average Dice rate, Precision and Recall of 0.91, 0.93 and 0.89 respectively has been achieved over the test set.**

I. INTRODUCTION

Central serous chorioretinopathy (CSC) is a chorioretinal disorder of the eye characterized by serous detachment of the neurosensory retina, which is located at the posterior pole of the eye. CSC results in the accumulation of transparent fluid due to the defects at the level of the Retinal pigment epithelial (RPE), which allows serous fluid from the choriocapillaris to diffuse into the subretinal space between RPE and neurosensory retinal layers. Fig. 1 shows a pictorial representation showing sub-retinal fluid (SRF) pocket and a sample OCT B-Scan with SRF in CSC disorder. The fluid has the propensity to accumulate under the central macula, hence the name CSC [1], [2].

The advancements in diagnostic procedures in ophthalmology has revolutionized the diagnosis and management of patients with retinal diseases through various imaging techniques like colour fundus photography, fluorescein angiography and optical coherence tomography (OCT). Among these, in recent years, OCT has emerged as the primary diagnostic tool for the diagnosis of various retinal disorders including CSC. OCT provides an in-vivo, 3D cross-sectional view of the retina. The condition of CSC with SRF is clearly shown in the OCT scan with neurosensory retina being elevated with an optically empty space above the RPE [3] (see Fig. 1. (b)). The segmentation of this SRF is essential for the assessment of the severity and progression of the CSC disorder, preferably in an automated way.

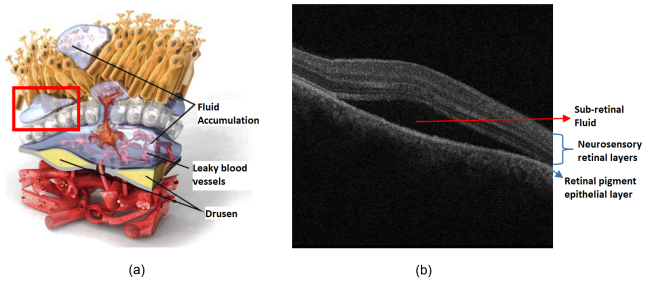

Fig. 1. (a) Pictorial representation of the accumulation of fluid within the retina[1] in case Age-related macular degeneration. The new blood vessel growth into the retina leads to blood and fluid leakage within the retina. The SRF is indicated by the bounding box. (b) Sample OCT scan showing the SRF accumulated due to CSC.

Few conventional methods have been proposed in the literature for automatic segmentation of SRF in CSC disorder using OCT scans. Wu *et al.* [4] presented an automated, 3D method for the segmentation of SRF regions in the case of neurosensory retinal detachment due to CSC in OCT scans. The method is based on the construction of probability map using the texture score, intensity score and distance between ILM and RPE layers using random forest classifier. Further, SRF regions are segmented by applying continuous max flow optimization on the probability map. An automatic, locally-adaptive method was proposed by Novosel et al. [5] for the segmentation of SRF in CSC disorder OCT scans. The approach utilizes the local attenuation differences in between the retinal layers near an interface and introduces new auxiliary interfaces so as to segment the fluid region.

Development of an automated retinal fluid segmentation technique with high enough accuracy is faced with its own set of challenges such as noise and motion artifacts present in OCT scans, varying sizes, shape and locations of fluid pockets within the retina, added with the variations that occur when dealing with OCT data captured using different OCT devices.

Recent trend in retinal fluid segmentation in general in OCT images is to employ deep-learning neural network architectures, specifically the convolutional neural networks (CNNs) as they can learn important low level features automatically and combine them to higher level features in next layers, which is inspired from the functioning of human

+This work was supported by the Science and Engineering Research Board (Department of Science and Technology, India) through project funding EMR/2016/002677.

* Corresponding author

Narendra Rao T J, Girish G N and Jeny Rajan are with the Dept. of Computer Science and Engineering, National Institute of Technology Karnataka, Surathkal, Karnataka, India. raonaren25@gmail.com girishanit@gmail.com jenyrajan@gmail.com

Abhishek R. Kothari is with Pink City Eye and Retina Center, Jaipur, India. dr.a.kothari@gmail.com

[1] from http://www.scienceofamd.org/wp-content/uploads/2012/04/Science_of_AMD%20Patient_Brochure.pdf

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

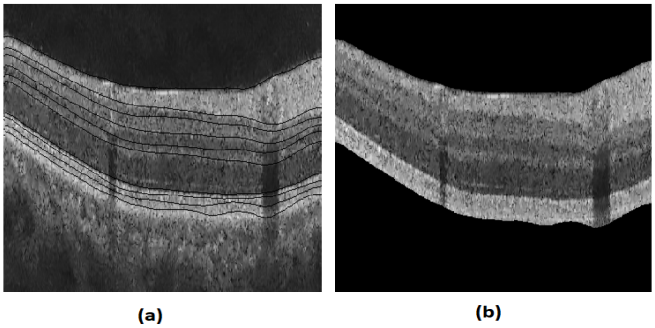

Fig. 2. (a) Sample OCT scan with the 11 layers marked for pre-processing. (b) Resultant image after neutralizing the background above the first layer (ILM) and last layer (RPE).

visual cortex in processing visual information and in object recognition. Hence, many deep-CNN based research works have been proposed lately towards retinal fluid segmentation such as [6], [7], [8], [9]. Inspired by their success, a fully convolutional neural network (FCNN) based method is proposed in this paper for the segmentation of SRF in CSC OCT scans.

II. PROPOSED METHOD

The proposed SRF segmentation method works in two stages – pre-processing stage followed by the fluid segmentation stage.

A. Pre-processing

OCT images suffer from granular natured speckle noise due to the phenomenon of coherence that occurs during OCT image formation. Speckle noise affects the signal-to-noise ratio (SNR) and contrast of OCT scans, thereby making the clinical diagnosis difficult [10]. This speckle ridden scans negatively affect the model learning process in turn resulting in poor segmentation. In order to deal with this issue, the region beyond the retina is neutralized/eliminated by making it a zero-intensity region. Since this region now does not contribute to the model learning process, the affect of noise on the segmentation is reduced by that extent. Opting for denoising to reduce noise is another alternate, but the former is chosen here since denoising smoothens the image resulting in loss of information. Cropping the unwanted background would also result in similar input, however, since the volumetric quantification of the retinal fluid would need the full scale scans, the original image dimensions are retained here.

In order to nullify the background, the first and the last retinal layers i.e., internal limiting membrane (ILM) and RPE layer are segmented. The IOWA reference algorithm [13] is utilized for this purpose which segments 11 retinal layers in order. The region beyond the ILM and RPE layers are explicitly converted into zero-intensity area as shown in Fig. 2. These background neutralized images form the input to the deep learning model for fluid segmentation.

B. Fluid Segmentation

A deep learning neural network architecture is employed for SRF segmentation. The proposed architecture is based on the FCN as described in the work of [14]. FCN assumes that the input is image and performs convolution operation on the input using the weights to extract abstract features from the image, while retaining the spatial arrangement information. It produces a prediction score matrix of the same size as input. This is used to build the binary segmentation output. The underlying network architecture is adopted from the popular U-Net architecture proposed by Ronneberger [15] and is based on the neural network used for intra-retinal fluid segmentation in [6]. This model takes into account the local as well as global features to generate an accurate segmentation map. The global features represent the location and approximate dimension of the fluid region, while the local features indicate the accurate region boundary.

The neural network used here consists of two paths – a contracting path and an expanding path similar to that in [15], [6], as shown in Fig. 3. The first path performs the convolution in 3 stages. Every level includes two convolutional layers followed by a downsampling operation using max-pooling function to gather a bigger receptive field. Each filter in the convolution layer can be thought of as feature identifier which identifies specific features like a dot, a straight line, a curve etc. These low-level features eventually develop into more complex attributes in following layers. The expanding path performs the reversing of activations to get the original resolution. Deconvolution layers which are trainable in nature do this up-sampling process over 3 stages. Moreover, in order to obtain an accurate segmentation, the features extracted in down-sampling phase are transferred and concatenated with the extracted features in the corresponding output of up-sampling blocks in the expanding path using skip connections.

The network is designed to accept input images of size 1024×512. In each block of the contracting path, there are two convolution layers and then a max pooling layer. The convolution filters are of size 3×3 throughout the model to allow for the discriminative feature extraction from the neighbourhood, at the same time maintain a lower count of parameters. There are a total of 14 convolutional layers in the network. The number of filters are doubled in every depth staring from 32 till 128. The Rectified Linear unit (ReLu) activation function is applied to each convolution output. The $2\times$ pooling layers progressively reduce the input size by half so as to reduce the number of parameters and network computations in turn reducing model over-fitting possibility.

In the expanding path, the activations produced are up-sampled by employing fractionally-strided convolution (deconvolution) operation. The number of filters reduce by half in every up-sampling block, staring from 128 till 32. This path is similar to the down-sampling path except that the deconvolution layers are placed instead of pooling layers. The last layer in the network is a 1×1 convolutional layer with softmax activation function which produces 1024×512

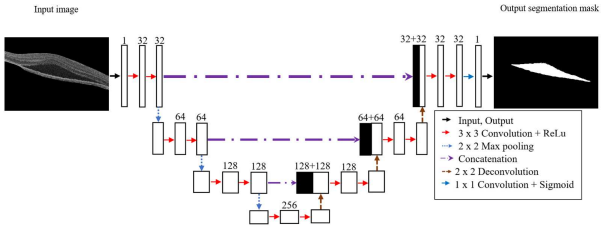

Fig. 3. The proposed network archtecture. It is to be noted that the input and out images are of size 1024×512.

sized output image, giving a pixel to pixel probability of belonging to fluid class or not. This binary output mask is compared with the groundtruth and the weights of the model are updated with the help of the binary cross-entropy loss function during back propagation, such that the loss is minimized in every epoch during the training process.

III. Experimental Details

The proposed FCN is implemented in Keras 1.0 [16] with Tensorflow back-end. The proposed method is tested on the privately developed dataset under the collaboration with Pink City Eye and Retina Center, Jaipur, India. The dataset consists of 15 OCT volumes acquired from CSC patients using a Zeiss manufactured Cirrus OCT device. Each volume contains 128 B-scans with a dimension of 1024×512 in vertical and horizontal directions respectively. The volumes are obtained over 6×6 mm of the macula.

The dataset is divided in to training, validation and testing sets for training, validation and performance evaluation of the FCN model respectively. The training set is made up of five OCT volumes, validation contains two volumes, whereas testing set consists of eight volumes. In all, the model is trained with 640 B-scans, validated with 256 B-scans and tested with 1024 B-scans. The original image dimensions are retained (1024×512) and the full scale scans are passed onto the model for training and prediction after normalizing them to zero mean and unit variance. As mentioned above, binary cross-entropy loss function is applied during the network training to obtain minimum loss and *Adam optimizer* [17] is employed to update the weights during model back propagation process. A suitable learning rate of 5×10^{-4} is empirically chosen using short random search for model learning. The network weights are initialized using *He_normal initializer* [18].

The proposed neural network is trained from scratch and no pre-trained weights are used. In order to generalize the model despite having less number of training samples, data augmentation concept is employed with horizontal flipping, height, width and zoom shifts, and random shear. This also helps in reducing the possibility of model overfitting. The augmentation occurs on the fly, thereby avoiding the need for additional memory. After training the model for 100 epochs the training and validation loss no longer reduce and the training is halted at this point.

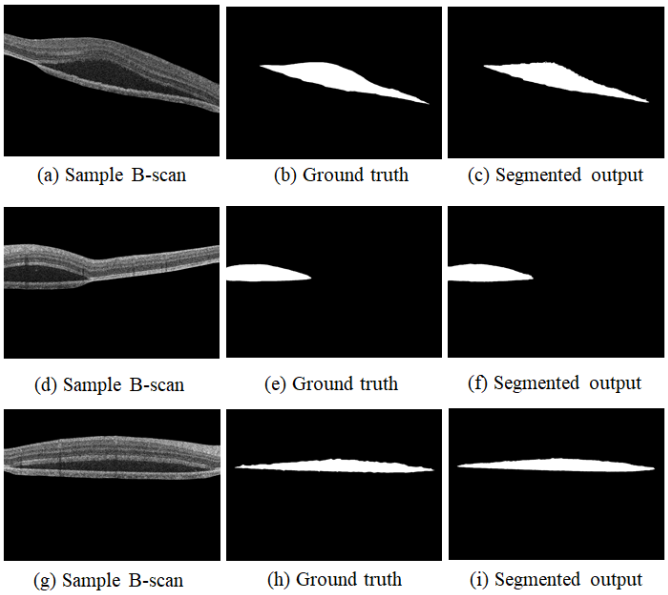

Fig. 4. Sample segmentation output of the proposed method on a B-scan from test set.

TABLE I
DC, PRECISION AND RECALL OBTAINED BY THE METHOD ON THE TEST SET.

Vol	DC	Precision	Recall
1	0.950	0.948	0.954
2	0.963	0.946	0.980
3	0.938	0.908	0.969
4	0.777	0.933	0.665
5	0.874	0.911	0.839
6	0.865	0.908	0.827
7	0.986	0.983	0.988
8	0.928	0.955	0.901
Average	**0.910**	**0.936**	**0.890**

IV. Results and Discussion

The performance of the method is analysed using 3 primary metrics generally used in evaluating the segmentation performance viz. Dice co-efficient (DC) [6], precision [6] and recall [6]. All these measures are obtained through the pixel-wise comparison between predicted output and reference ground truth marked by an expert ophthalmologist.

The results in terms of DC, precision and recall on the test set is shown in Table I. An average DC of 0.91, precision of 0.93 and recall of 0.89 are obtained. It is to be observed that except for V4, V5 and V6 all the volumes gave a very high DC, precision and recall of above 90%. In case of V4, V5 and V6 the fluid pockets were relatively small in nature and all these volumes were acquired from one subject at different times. Also, the training set had relatively less number of such scans with small volume of accumulated fluid. This issue may be solved by considering more scans with small volume regions in the training set so that the model learns such samples better.

In order to justify the neutralizing of the background

TABLE II

DC, PRECISION AND RECALL OBTAINED BY THE METHOD ON THE TEST SET. NO PRE-PROCESSING HAS BEEN PERFORMED IN THIS CASE ON THE TRAIN, VALIDATION AND TEST SETS.

Vol	DC	Precision	Recall
1	0.808	0.951	0.702
2	0.784	0.972	0.657
3	0.856	0.906	0.811
4	0.758	0.963	0.624
5	0.872	0.949	0.806
6	0.749	0.710	0.791
7	0.634	0.985	0.473
8	0.882	0.962	0.815
Average	**0.793**	**0.924**	**0.710**

during pre-processing of the input, the performance of the model with the background intact was analysed. The network was trained with a training set of raw input images and also validated and tested with raw scans. The results of the prediction on the eight test volumes are shown in Table II. It can be noted that a drop of 11.7% in DC, 1.2% in precision and 18% in recall is observed. This indicates that the pre-processing has reduced the impact of noise on model training and prediction.

V. CONCLUSION

In this paper, a method for automatic detection and segmentation of sub-retinal fluid in CSC affected OCT scans has been presented. To reduce the impact of noise in the background region on segmentation, the unwanted background around the retina is converted into a zero-intensity region as a pre-processing step. The pre-processed data forms the input to a fully convolutional neural network which learns the necessary features from the input during training. After successful training the model is evaluated against a test set of unseen data. The method yielded appreciable results with a DC of 0.91, precision of 0.93 and recall of 0.89 on the test set. As future work, in order test the robustness and generalizability of the model, it is intended to train and test on a larger dataset with scans from multiple vendors. Combining deep learning using neural network with conventional feature extraction techniques to develop a hybrid neural network for better segmentation result is another possibility that will be focused on in future.

REFERENCES

[1] D. T. Liu, A. C. Fok, W. Chan, T. Y. Lai and D. S. Lam, Central Serous Chorioretinopathy, in Retina, 5th ed. vol. 2, Elsevier BV, 2013, pp. 1291-1305, ch. 72.

[2] W. Maria, I. C. Munch, P. W. Hasler, C. Prünte and M. Larsen, Central serous chorioretinopathy, Acta ophthalmologica, vol. 86, no. 2, 2008, pp. 126-145.

[3] A. Mehreen, J. S. Duker, Optical coherence tomography–current and future applications, Current opinion in ophthalmology, vol 24, no. 3, 2013, pp. 213.

[4] M. Wu , W. Fan, Q. Chen, Z. Du, X. Li, S. Yuan and H. Park, Three-dimensional continuous max flow optimization-based serous retinal detachment segmentation in SD-OCT for central serous chorioretinopathy, Biomedical optics express, vol. 8, no. 9, 2017, pp. 4257-4274.

[5] J. Novosel, Z. Wang, H. de Jong, M. van Velthoven, K. A. Vermeer and L. J. van Vliet, Locally-adaptive loosely-coupled level sets for retinal layer and fluid segmentation in subjects with central serous retinopathy, in Proceedings of IEEE International Symposium on Biomedical Imaging, 2016, pp. 702–705.

[6] G. N. Girish , B. Thakur, S. R. Chowdhury, A. R. Kothari, and J. Rajan. Segmentation of Intra-Retinal Cysts from Optical Coherence Tomography Images using a Fully Convolutional Neural Network Model. IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 1, 2019, pp. 296–304.

[7] F. G. Venhuizen, B. van Ginneken, B. Liefers, F. van Asten, V. Schreur, S. Fauser, C. Hoyng, T. Theelen and C. I. Sánchez, Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography. Biomedical optics express, vol 9, no. 4, 2018, pp. 1545-1569.

[8] S. Yadav, K. Gopinath and Jayanthi Sivaswamy, A Generalized Motion Pattern and FCN based approach for retinal fluid detection and segmentation, 2017.

[9] D. Liu, X. Liu, T. Fu and Zhou Yang. Fluid region segmentation in OCT images based on convolution neural network, in Ninth International Conference on Digital Image Processing (ICDIP 2017), vol. 10420, 2017, pp. 104202A, International Society for Optics and Photonics.

[10] O. Aydogan, A. Bilenca, A. E. Desjardins, B. E. Bouma and G. J. Tearney, Speckle reduction in optical coherence tomography images using digital filtering, JOSA A vol 24, no. 7, 2007, pp. 1901-1910.

[11] K. Li, X. Wu, D. Z. Chen and M. Sonka, Optimal surface segmentation in volumetric images-a graph-theoretic approach, IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 1, Jan 2006, pp. 119–134.

[12] M. D. Abramoff, M. K. Garvin and M. Sonka, Retinal imaging and image analysis, IEEE Reviews in Biomed. Eng., vol. 3, 2010, pp. 169–208.

[13] M. K. Garvin, M. D. Abramoff, 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 Trans. Med. Imag., vol. 28, no. 9, 2009, pp. 1436–1447.

[14] J. Long, E. Shelhamer and T. Darrell, Fully convolutional networks for semantic segmentation, in Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 3431–3440.

[15] O. Ronneberger, P. Fischer and 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.

[16] F. Chollet, "Keras," 2015.

[17] K. Konstantinos, L. Chen, C. Ledig, D. Rueckert and Ben Glocker, Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI, Ischemic stroke lesion segmentation, vol. 13, 2015, p. 46.

[18] K. He, X. Zhang, S. Ren and Jian Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026-1034.