Retinal Layer Segmentation using Dilated Convolutions

T. Guru Pradeep Reddy[1], Kandiraju Sai Ashritha[1], Prajwala T. M.[1], G. N. Girish[1], Abhishek R. Kothari[2], Shashidhar G. Koolagudi[1], and Jeny Rajan[1]

[1] Department of Computer Science and Engineering
National Institute of Technology Karnataka, Surathkal, India
[2] Pink City Eye and Retina Center, Jaipur, India

Abstract. Visualization and analysis of Spectral Domain Optical Coherence Tomography (SD- OCT) cross-sectional scans is a powerful tool in the detection and diagnosis of several retinal abnormalities. Quantitative analytic techniques like retinal thickness and volumetric analysis which are performed on cross-sectional images of retina are used in early diagnosis and prognosis of retinal diseases. However, segmentation of OCT images is a complicated task on account of certain factors like speckle noise, low image contrast and low signal to noise ratio amongst many others. Owing to the importance of retinal layer segmentation in diagnosing ophthalmic diseases, manual segmentation techniques have been proposed and adopted which suffer from erroneous boundary detection issues. This paper thus proposes a fully automated semantic segmentation technique that uses a encoder-decoder architecture to accurately segment the retinal layers into 8 classes.

Keywords: Retinal layer segmentation · Optical coherence tomography · Dilated convolutions · Deep learning · Retina

1 Introduction

Optical Coherence Tomography (OCT) has gained a lot of medical importance in retinal imaging and diagnosis of several ocular diseases. OCT is a non invasive technique which uses low coherence infrared light to generate the imagery of retinal structures [10]. The introduction of Spectral Domain OCT (SD-OCT) has led to greater imaging speed, greater resolving power, high scan density and speed which has proven to be beneficial in OCT data acquisition.

The SD-OCT cross-sectional scans have a typical resolution of about 6×6 mm and a sufficient depth penetration of about (0.5-2 mm)[1]. These scans have found their applications in skin imaging,[12] intra-vascular imaging[2], lumen detection[8], plaque detection[9] and retinal imaging[12]. A few clinical applications of the SD-OCT images are: detection and identification of sub-retinal fluid, neurosensory detachments, pigment-epithelial detachments and assessment of choroidal neovascular membranes[6]. Cross-sectional analysis and investigation is a powerful tool in the detection and diagnosis of several retinal abnormalities.

Retinal thickness analysis and volumetric analysis performed on cross-sectional imaging of retina are quantitative techniques in early diagnosis and prognosis of retinal diseases. In this way, OCT is an invaluable clinical assessment tool that can provide proper visualization of cross-sectional retinal scans and surgical outcomes, which can be used for early disease detection and treatment.
The interpretation of OCT images requires the knowledge of the the several retinal layers present in a normal retina (retina without the macular pathology). Fig.1. shows the enlarged cross-sectional retinal image with the various layers marked. Effective segmentation of these layers for retinal thickness calculations can be of clinical importance as that can be used in the diagnosis of diseases. Retinal layer segmentation has several use cases in opthalmology.

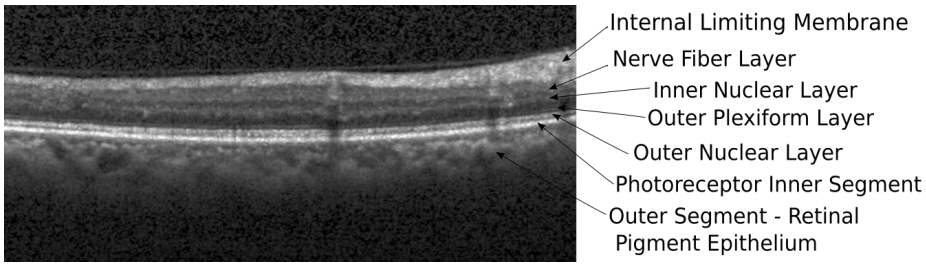


Fig. 1: Retinal layers

Segmentation of retinal layers is a challenging task due to constraints posed by presence of speckle noise, low image contrast, usage of morphological operators of irregular shape and low signal-to-noise ratio amongst others. Also inherent challenges during the acquisition of OCT images make it increasingly difficult to accurately segment the layers and detect the boundaries. But owing to the critical importance of retinal layer segmentation in diagnosing ophthalmic diseases, manual segmentation techniques have been adopted in most of the OCT related studies [3]. However, manual segmentation is exhaustive, tedious and prone to manual errors. Also, commercial algorithms available suffer with erroneous boundary and layer detection issues. A lot of effort and time has been invested by several researchers in order to come up with segmentation techniques that are automated requiring minimal human intervention.
Various approaches to segmenting retinal layer boundaries have been reported with different levels of success. Chiu et al.[4] presents an automatic approach for segmenting retinal layers in SD-OCT images using graph theory and dynamic programming(GTDP). It makes use of graph theory and dynamic programming for the segmentation of layered structures, which significantly reduces the processing time. The method proposed earlier in [4] for the quantitative analysis of retinal and corneal layer thicknesses is extended in [5] to segment retinal objects like cells and cysts. A transform that maps closed-contour features in the Cartesian domain into lines in the quasi-polar domain is used as a basis in this technique. The features extracted are used for segmentation using the approach described in [4].

The authors of [3] have proposed a method based on kernel regression (KR)-classification for the estimation of retinal layer and fluid positions. A fully automated algorithm based on KR and the previously described GTDP framework has been presented for the identification of fluid-filled regions and segmentation of seven retinal layers in SD-OCT images. KR-based classifiers have been created in the initial stages which are then used in the succeeding phases for guiding GTDP to provide more accurate results.

All of the above proposed algorithms are automatic segmentation approaches. However, they all use hand-designed features. It also needs to be considered that these manually extracted features are not discriminative enough for the differentiation of semantic object boundaries and unpredictable changes in low level image cues. Inspired by deep neural networks, a number of researchers have used approaches that replaced hand-designed features with deep features. Hence, introducing the deep learning techniques into the edge segmentation problem is reasonable and feasible. The authors of [7] present a deep-learning based approach to learn discriminative features for layer segmentation. With the learned deep features,they claim that layer segmentation can be implemented by transmitting them into any classifiers. They have employed the computationally efficient random structured forest to classify layers and background. Then, dynamic programming has been performed for the identification of the shortest path from the top pixel of the left column to the bottom pixel of the right column.

Inspired by the results obtained using the deep convolutional networks for the purpose of segmentation of retinal layers, the authors of [10] have proposed a new fully convolutional deep architecture, termed ReLayNet, for end-to-end segmentation of retinal layers and fluid masses in SD-OCT scans. The authors have proposed an approach that makes use of a contracting path of convolutional blocks (encoders) that learn a hierarchy of contextual features, followed by an expansive path of convolutional blocks (decoders) corresponding to the encoder blocks, for the purpose of semantic segmentation. ReLayNet is trained to optimize on a combined loss function comprising of a weighted logistic loss and Dice loss. This architecture is validated on a publicly available benchmark dataset and has been compared against five state-of-the-art segmentation methods which include two deep learning based approaches. As a result, this is considered as one of the best current state of the art approaches in comparison to all the previous approaches used for retinal layer segmentation.

Our paper proposes a fully convolutional network which makes use of dilated convolutions [13] for increasing the receptive field size without compromising on the dimensionality of the images. We have also introduced a weighting scheme that addresses the challenges posed by the segmentation of thin retinal layers and the class imbalance between the retinal layers and the background.

The paper is organized as follows: Section II describes the methodology proposed for segmentation. Section III describes how the model has been trained to learn the features using a joint loss and weighting scheme. Section IV details experimentation performed. Section V outlines qualitative and quantitative analysis of results. Section VI concludes the paper.

2 Methodology

The task to be carried out is to assign each pixel of the input OCT Retinal image to a class. Eight classes have been considered for the task of segmentation. This includes seven retinal layer classes and one background class. The architecture of the model is shown in Figure 2.

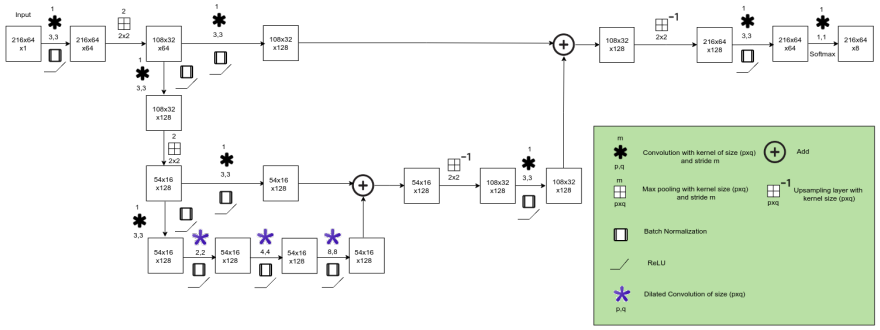


Fig. 2: Dilated Retinal Layer Segmentation Network

The model is symmetric in nature. It is fully convolutional and does not contain any fully connected layers. It is a variant of Encoder- Decoder architecture [10] with the introduction of dilated convolutions and skip connections. The skip connections carry out the task of transfer of features from an encoding block to its corresponding decoding block. The former part of the network is a contracting path consisting of a series of encoder blocks, followed by a series of dilated convolutions.
Dilated convolutions preserve the dimensions of the feature maps. The latter part of the network comprises of an expanding path consisting of a series of decoding blocks with skip connections from corresponding encoding blocks. The final classification block computes the probability of pixels belonging to different classes. The details and significance of each block can be explained as follows:

2.1 Contracting Path

The contracting path comprises of two encoder blocks. Each encoder block consists of 4 components : convolution layer, batch normalization layer, ReLU activation layer and a max pooling layer. The kernels in the convolution layer are of size 3×3. The number of convolution filters is set to 64 in the first encoder block which is doubled to 128 in the second owing to the reduction in dimensionality of the image as the feature maps are passed from one encoder block to the other. This also ensures a richer encoding of the representations as we move down the hierarchy (the complexity of the features increases as we move down the network). The feature maps are appropriately zero padded in order to preserve the

size of the image after the application of the convolution operation. The next layer is a batch normalization layer which reduces the ill effects of co-variate shift, makes learning faster and also provides a regularization effect, which, in turn, helps in prevention of over-fitting the train set. This layer is followed by a ReLU activation layer which introduces non linearity to the network required for the learning of complex mappings. The last layer is a max pooling layer of size 2×2, which reduces the dimensionality of the feature maps and increases the field of view.

2.2 Dilation Path

Although the max pooling layer incorporated in the contracting path increases the field of perception for the kernels, it compromises on the resolution of the feature maps due to the dimensionality reduction. This decrease of resolution is undesirable for the task of semantic segmentation. This is because it requires the usage of a max pooling layer which in turn, needs a corresponding upsampling layer that learns the complex mapping for an increase in size. Thus, dilated convolutions, also known as Atrous convolution [13] have been introduced to expand the size of the receptive field exponentially without compromising on the resolution. In the proposed model, 3 such dilated convolution layers have been used with dilation rates set to 2, 4 and 8 respectively. This ensures that an exponential increase in the receptive field size is obtained without increasing the number of parameters. The number of filters for all the dilated convolution layers is set to 128.

2.3 Expanding Path

The Expanding path consists of two decoder blocks, corresponding to the two encoder blocks. Each decoder block consists of 5 components: skip connection, upsampling layer, convolution layer, batch normalization and a ReLU activation layer. The skip connections consist of a convolution layer, batch normalization and ReLU activation layer each. These are applied on the output of the encoder and will be added pixel-wise with the output of corresponding decoder. There are two main advantages of using skip connections - they allow transfer of low level information from lower layers to higher layers in order to simplify the process of learning of the upsampling features and also aid an easier backward flow of gradients i.e they provide a direct gateway for the gradient to flow to lower layers which minimizes the problem of vanishing gradient. The next layer is an upsampling layer, which increases the dimensions by repeating the data based on the kernel size, which has been set to 2 in both the blocks. This is followed by a convolution layer with a kernel size of 3×3 and zero padding to preserve the dimension of feature maps. The number of convolution filters is set to 128 in the first decoder block and halved to 64 in the second.

2.4 Classification Path

The final part of the network is a classification block, to obtain the probability maps of pixels belonging to eight classes. A convolution layer of kernel size 1×1 with 8 filters is applied to the output of the final decoder block followed by a softmax activation layer. Softmax activation layer outputs the probability of a pixel belonging to either of the eight classes (seven different layers and background).

3 Loss function

The model is optimized on a combined loss function, inspired by [10]. It comprises of the following two loss functions:

3.1 Weighted multi-class logistic loss

In general, cross-entropy measures the similarity between the actual and predicted values in probabilistic terms. Cross entropy loss, also known as the logistic loss, penalizes the deviation of the predicted value from the actual values. At a pixel location x, if the predicted probability of the pixel belonging to class l is $p_l(x)$, $\omega(x)$ is the weight associated to pixel x and $g_l(x)$ is the actual label of the pixel x, then the weighted multi-class logistic loss $\tau_{logloss}$, is defined as (1):

$$\tau_{logloss} = -\Sigma_{x \epsilon \Omega} \omega(x) g_l(x) log(p_l(x)) \tag{1}$$

The weights associated with the pixels are to compensate for the unbalanced thicknesses of the different retinal layers.

Weighting scheme for the multi-class logistic loss Class imbalance is one of the major challenges associated with retinal layer segmentation. Figure 3 shows the histogram of the number of pixels belonging to each class. It can be observed that the background class dominates due to the large number of background pixels.

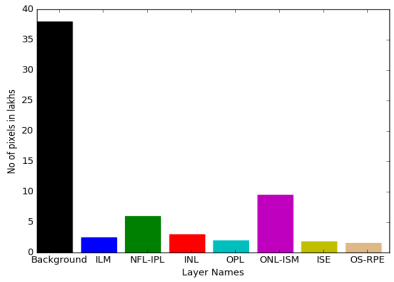


Fig. 3: Histogram portraying class imbalance

In the absence of any weighting scheme, there are two issues that arise : (i) The boundary pixels in the layer transition regions will be challenging to be segmented. (ii) The model will be biased towards the most dominant background pixels class.
The first issue is resolved by weighting all the pixels belonging to the retinal layer boundaries by a factor of w_1. This ensures that the pixel contributions from the retinal layer transition regions are weighted more relative to the interior retinal layer pixels. The experimental value of w_1 is 15.
The second issue is addressed with an introduction of weighting factor w_2. This weighs the contributions from the pixels belonging to under-represented classes (including the boundary pixels) and ensures that the convolution kernels are more sensitive to the gradient contributions from these pixels. As is evident from the histogram, the different classes are non-uniformly balanced relative to each other and hence the value of w_2 is varied for different layers.

Median Frequency Method The value of w_2 for each of the retinal layers is computed using the median frequency method. Here, the number of pixels belonging to each of the layers is computed as n_l. The median of these values is calculated as M. L represents the set of 7 retinal layer classes. The value of w_2 is calculated as (2):

$$w_2 = \begin{cases} M/n_l & \text{when } l(x) \in L \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Layer	w_2
ILM	11.459
NFL-IPL	5.63
INL	11.007
OPL	14.368
ONL-ISM	3.336
ISE	13.647
OS-RPE	16.978

Table 1: Computed value of w_2 for different layers

The final weighting scheme is thus derived as (3):

$$w = 1 + w_1 + w_2 \tag{3}$$

The weighting scheme ensures that the boundary pixels receive a weight higher than the maximum value of w_2 assigned to any of the layers. Retina comprises of hypo and hyper-reflective layers and it can be observed from Table 1 that the value of w_2 in hypo reflective layers is lower relative to hyper reflective layers. Hence, darker pixels are assigned lesser weights and the brighter ones are assigned higher weights ie: the weights assigned to the pixels are proportional to the intensity of the pixels. Figure 4 depicts the weighting scheme described.

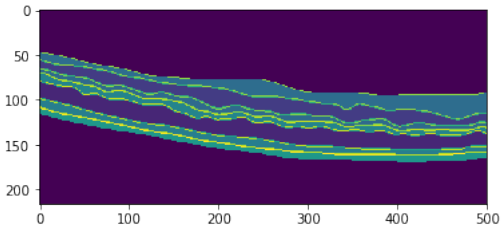


Fig. 4: Weighting scheme representation

3.2 Dice loss

Dice coefficient is used to measure the similarity between the predicted and actual labels, and is used popularly in segmentation to compare the performance of the predicted values with respect to the reference masks in medical applications. Dice loss $\tau_{diceloss}$, defined as 1 - *dice co-efficient* thus penalizes the deviation of the predicted values from the actual labels. It is defined as (4).

$$\tau_{diceloss} = 1 - 2\Sigma_{x\epsilon\Omega} p_l(x) g_l(x) / \Sigma_{x\epsilon\Omega} p_l(x)^2 + \Sigma_{x\epsilon\Omega} g_l(x)^2 \tag{4}$$

3.3 Combined loss

The overall loss function used for optimization of the proposed network is a weighted combination of the weighted multi-class logistic loss ($\tau_{logloss}$) and the dice loss ($\tau_{diceloss}$) along with an additional weight decay term [10]. Overall loss function is defined as (5):

$$\tau_{overall} = \lambda_1 \tau_{logloss} + \lambda_2 \tau_{diceloss} + \lambda_3 \|W^{(\cdot)}\|_F^2 \tag{5}$$

where λ_1,λ_2 and λ_3 are the weight terms and $\|W^{(\cdot)}\|_F^2$ represents the Frobenius norm on the weights W of the model.

4 Training

4.1 Dataset Description

The proposed architecture is experimented on the Duke DME SD-OCT dataset [3]. The scans of the dataset were obtained from 10 subjects with DME. The dataset consists of a total of 110 annotated images (11 B-scans per subject) with a resolution of 496×768. All these scans were annotated for the retinal layers by two expert ophthalmologists.

4.2 Data pre-processing

The annotations for different patients were of differently sized and hence pre-processing was required to obtain a uniformly sized dataset. Different bounds

were manually computed for the images of different patients and all the 110 images were sized to 496×500. It was observed that the majority of the pixels in these images comprised of background and cropping a major portion of the background would aid in faster and unbiased model training. On horizontal cropping of certain portions, the images were sized to 216×500. Thus, the dataset comprised of 110 images of size 216×500.
The dataset is relatively small in order to train a full scale fully convolutional network. Hence, in order to increase the size of the dataset, to reduce overfitting and to speed up the training process, the images were sliced vertically to obtain 770 images of size 216×64 each.
The OCT scans contain speckle noise, which can be approximated with a Gaussian distribution. Thus, the images were denoised using Unbiased Fast Non-local-Means algorithm [11]. The smoothing parameter has been set to 10 and the bias value is set to 15.

4.3 Experimental Settings

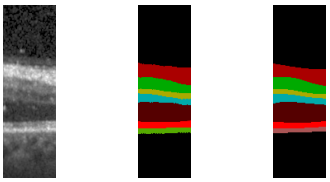
The dataset, thus consists of 770 images, out of which 540 were set aside for training and the remaining 230 were used for testing. In order to improve the model's ability to generalize, the train and test images were distributed uniformly across all the patients. The hyper-parameters for the loss function are set as $\lambda_1 = 1$, $\lambda_2 = 0.5$ $\lambda_3 = 0.01$. Every layer of convolution incorporated in the model has a L2 kernel regularizer with a weight decay = 0.01. The optimization was carried out using SGD optimizer with Nesterov Momentum and a batch size of 32 was used. The initial learning rate was set to 0.01, and it is reduced by a factor of 0.5 on plateau where the loss stops decreasing for more than 6 epochs and it is allowed to reach a minimum learning rate of 0.000005. Momentum of 0.9 was used while training. The model is trained for 200 epochs and takes an average of 25 seconds for training per epoch. All the experiments were performed on Intel Xenon CPU with 64GB RAM and Nvidia K4O GPU with 10GB dedicated memory.

5 Results and Analysis

5.1 Qualitative Analysis

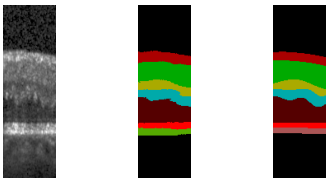
Figure 5 and Figure 6 show the sample OCT scans, model predictions and ground truths for qualitative analysis of the results obtained. In Figure 5, (a),(b),(c) show the OCT scan, model prediction and ground truth of a sample image with minimal retinal layer distortion and (a),(b),(c) in Figure 6 show the OCT scan, model prediction and ground truth of a sample with significant retinal layer distortion respectively.
It can be seen that the segmentations of the proposed model which make use of dilated convolutions for retinal layer segmentation, are of high quality and the retinal layer boundary demarcations are comparable to human expert clinicians.

The shape of the layer boundaries is highly preserved as per visual analysis. This is evident by comparing (b) to its corresponding ground truth in (c). The model has been able to successfully learn even the complicated edges and fine details as it can be seen in (b) and (c). Also, the predictions are free from speckle noise.



(a) Sample image (b) Model prediction (c) Ground truth

Fig. 5: Segmentation result with limited retinal layer distortion



(a) Sample image (b) Model prediction (c) Ground truth

Fig. 6: Segmentation result with significant retinal layer distortion

5.2 Quantitative Metrics

Dice co-efficient has been used as the quantitative metric for evaluation of the performance of the proposed model for retinal layer segmentation. The dice co-efficient score in (6) has been computed for every retinal layer for analysis of each of the layer segmentations.

$$DSC = \frac{2|Predicted \cap Actual|}{|Predicted| + |Actual|} \tag{6}$$

Comparative analysis Quantitative comparison of the dice score metrics has been carried out against the benchmark paper [10] and they have been tabulated in 2.

Layer	Proposed	ReLayNet
Background	**0.99**	**0.99**
ILM	0.89	**0.90**
NFL-IPL	**0.94**	**0.94**
INL	**0.88**	0.87
OPL	**0.87**	0.84
ONL-ISM	**0.96**	0.93
ISE	**0.94**	0.92
OS-RPE	**0.90**	**0.90**

Table 2: Comparative analysis with ReLayNet. The best performance is shown by **bold**.

It can be observed that the proposed models have comparable dice scores for most of the retinal layers. The OPL and ONL-ISM show a significant improvement over the corresponding ReLayNet dice scores. It can be noted that the proposed model has performed well for layers which are even particularly challenging to segment, namely INL and OPL layers. This increase in performance can be owed to the introduction of the weighting scheme that weighs the contributions from the thin retinal layers and boundaries.

6 Conclusion

In this paper, we have introduced Dilated retinal layer segmentation network, which is fully convolutional. The proposed model classifies every pixel in the given OCT image to either background or one of the seven retinal layer classes. Combined loss function which is a weighted sum of dice loss and multi-class logistic loss has been used. Pixel-wise weighting scheme has been adopted to solve class imbalance problem and to accurately segment the pixels in the layer transition regions. Dice score has been used for quantitative performance evaluation and the same metric has been reported for every layer. Our model provides a fast response and hence can be incorporated for real time applications. Aiding to its fully convolutional nature, it can be used independent of the test image dimensions. One major challenge faced is the noise present in the dataset induced during the OCT image acquisition. Thus, devising a method which can denoise the images in real time is an open-ended research problem. Another limitation under concern is the scarcity of data, which can be resolved by acquisition of a larger annotated dataset from the clinical experts.

Acknowledgement

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

Authors would like to thank Vision and Image Processing (VIP) Lab, Department of Biomedical Engineering, Duke University for providing DME dataset.

References

1. Anger, E.M., Unterhuber, A., Hermann, B., Sattmann, H., Schubert, C., Morgan, J.E., Cowey, A., Ahnelt, P.K., Drexler, W.: Ultrahigh resolution optical coherence tomography of the monkey fovea. identification of retinal sublayers by correlation with semithin histology sections. Experimental eye research **78**(6), 1117–1125 (2004)
2. Bouma, B., Tearney, G., Yabushita, H., Shishkov, M., Kauffman, C., Gauthier, D.D., MacNeill, B., Houser, S., Aretz, H., Halpern, E.F., et al.: Evaluation of intracoronary stenting by intravascular optical coherence tomography. Heart **89**(3), 317–320 (2003)
3. Chiu, S.J., Allingham, M.J., Mettu, P.S., Cousins, S.W., Izatt, J.A., Farsiu, S.: Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. Biomedical optics express **6**(4), 1172–1194 (2015)
4. Chiu, S.J., Li, X.T., Nicholas, P., Toth, C.A., Izatt, J.A., Farsiu, S.: Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation. Optics express **18**(18), 19413–19428 (2010)
5. Chiu, S.J., Toth, C.A., Rickman, C.B., Izatt, J.A., Farsiu, S.: Automatic segmentation of closed-contour features in ophthalmic images using graph theory and dynamic programming. Biomedical optics express **3**(5), 1127–1140 (2012)
6. Coscas, G.: Optical coherence tomography in age-related macular degeneration. Springer Science & Business Media (2009)
7. Fu, T., Liu, X., Liu, D., Yang, Z.: A deep convolutional feature based learning layer-specific edges method for segmenting oct image. In: Ninth International Conference on Digital Image Processing (ICDIP 2017). vol. 10420, p. 1042029. International Society for Optics and Photonics (2017)
8. Roy, A.G., Conjeti, S., Carlier, S.G., Dutta, P.K., Kastrati, A., Laine, A.F., Navab, N., Katouzian, A., Sheet, D.: Lumen segmentation in intravascular optical coherence tomography using backscattering tracked and initialized random walks. IEEE journal of biomedical and health informatics **20**(2), 606–614 (2016)
9. Roy, A.G., Conjeti, S., Carlier, S.G., Houissa, K., König, A., Dutta, P.K., Laine, A.F., Navab, N., Katouzian, A., Sheet, D.: Multiscale distribution preserving autoencoders for plaque detection in intravascular optical coherence tomography. In: Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. pp. 1359–1362. IEEE (2016)
10. Roy, A.G., Conjeti, S., Karri, S.P.K., Sheet, D., Katouzian, A., Wachinger, C., Navab, N.: Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. Biomedical optics express **8**(8), 3627–3642 (2017)
11. Sudeep, P., Palanisamy, P., Rajan, J., Baradaran, H., Saba, L., Gupta, A., Suri, J.S.: Speckle reduction in medical ultrasound images using an unbiased non-local means method. Biomedical Signal Processing and Control **28**, 1–8 (2016)
12. Yeh, A.T., Kao, B., Jung, W.G., Chen, Z., Nelson, J.S., Tromberg, B.J.: Imaging wound healing using optical coherence tomography and multiphoton microscopy in an in vitro skin-equivalent tissue model. Journal of biomedical optics **9**(2), 248–254 (2004)
13. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)