



# Joint Segmentation and Uncertainty Visualization of Retinal Layers in Optical Coherence Tomography Images Using Bayesian Deep Learning

Suman Sedai<sup>(✉)</sup>, Bhavna Antony, Dwarikanath Mahapatra, and Rahil Garnavi

IBM Research - Australia, Melbourne, VIC, Australia  
ssedai@au1.ibm.com

**Abstract.** Optical coherence tomography (OCT) is commonly used to analyze retinal layers for assessment of ocular diseases. In this paper, we propose a method for retinal layer segmentation and quantification of uncertainty based on Bayesian deep learning. Our method not only performs end-to-end segmentation of retinal layers, but also gives the pixel wise uncertainty measure of the segmentation output. The generated uncertainty map can be used to identify erroneously segmented image regions which is useful in downstream analysis. We have validated our method on a dataset of 1487 images obtained from 15 subjects (OCT volumes) and compared it against the state-of-the-art segmentation algorithms that does not take uncertainty into account. The proposed uncertainty based segmentation method results in comparable or improved performance, and most importantly is more robust against noise.

**Keywords:** Retinal imaging · Bayesian segmentation · OCT Fully convolution neural networks · Uncertainty maps

## 1 Introduction

Optical Coherence Tomography (OCT) is a popular non-invasive imaging modality for retinal imaging. OCT provides volumetric scans of retinal layers for the diagnosis and evaluation of different diseases such as Glaucoma and Age related macular degeneration (AMD). For example, [1] have shown the correlation between outer retinal layer thickness and visual acuity in early AMD patients. It has also been shown that retinal layer features can be used to predict vision loss and progression [6].

The segmentation of retinal layers in OCT has been tackled in a number of ways, such as dynamic programming [13], graph-based shortest path algorithms [4], graph-based minimum  $s$ - $t$  cut formulations [8] and level sets [3, 14]. Machine-learning based approaches have also been proposed, where the retinal layer and boundary probability maps are detected using a trained classifier. The final segmentation is then obtained by imposing a model such as active contours [20] or minimum  $s$ - $t$  cut framework [12] on the soft labels.

In the past few years, Convolutional Neural Networks (CNNs) based methods such as Unet [15,17] and fully convolutional Densenet (FC-DN) [10] have achieved remarkable performance gain in medical image and natural image segmentation. The networks are trained end-to-end, pixels-to-pixels on semantic segmentation exceeded the most state-of-the-art methods without further machinery. [2,16] used Unet like network to perform pixelwise semantic segmentation of retinal layers. In another approach, [5], used CNN and graph search method for layer boundary classification. Once trained, these methods acts as a black box where one has to assume that the segmentation output is accurate which is not always the case. For example, the model will produce incorrect segmentation when the test image is different from the distribution of images used to train the model. This may happen when the model is trained using limited number of training images. In other scenario, the model will produce inaccurate segmentation when trained using normal images, yet pathologies are observed in the test image or the test image is noisy. Quantification of uncertainties associated with the segmentation output is therefore important to determine the region of incorrect segmentation, e.g., region associated with higher uncertainty can either be excluded from subsequent analysis or highlighted for manual attention. In another scenario, when the retinal layer segmentation map is used to diagnose the diseases such as AMD and Glaucoma, the uncertainty map can be used to determine the confidence of final automatic or clinical diagnosis.

Previous works have explored the uncertainty quantification in biomedical segmentation [9], however, these approaches do not utilize the representative power of deep learning. Recent research has shown that Bayesian probability theory offers a mathematically grounded technique to reason about uncertainty in deep learning models [7,11]. In this paper, we explore Bayesian fully convolutional neural network for segmentation and uncertainty quantification of retinal layers in OCT images. We experimentally demonstrate that in addition to the uncertainty based confidence measure, our method provides improved layer segmentation accuracy and robustness towards noise in the test images.

## 2 Methodology

We model two types of uncertainties for retinal layer segmentation; epistemic uncertainty and aleatoric uncertainty. The epistemic uncertainty captures the uncertainty related to the model parameters, e.g., when the model does not take into account certain aspect of the training data. Therefore, the epistemic uncertainty can be reduced by training the model using more images. Aleoretic uncertainty, on the other hand, captures the noise inherent in the images, therefore, it cannot be reduced with additional training images. We model the aleatoric uncertainty as an additional output variance for both deep learning networks.

We enhance fully convolutional Densenet (FC-DN) [10] for segmentation and uncertainty quantification of retinal layers. FC-DN is a fully convolutional neural network with several dense-blocks connected in encoder-decoder architecture

with skip connections across them which effectively combines coarse semantic features with fine image details for pixel-wise semantic segmentation. Each layer in the *dense block* is connected to all the preceding layers by iterative concatenation of previous *feature maps*. This allows all layers to access *feature maps* from their preceding layers which encourages heavy feature reuse. As a result, FC-DN uses less parameter and is less prone to over-fitting. The networks is then trained using the proposed class weighted Bayesian loss function by taking into account the output variance which is described in Sect. 2.1. Once the networks are trained, in the test phase, we use *dropout variational inference* technique [7] to compute the epistemic uncertainty which we describe in Sect. 2.2.

Let  $F_{\mathbf{W}}(X)$  be a FC-DN model parameterized by  $\mathbf{W}$  which takes input image  $X$  and produces the *logit* vector  $\mathbf{z}$  for each pixel as  $\mathbf{z} = F_{\mathbf{W}}(X)$ . The logit vector  $\mathbf{z}$  consists of *logits* for each class as  $\mathbf{z} = (z_1, \dots, z_C)$  where  $C$  is the number of classes i.e., number of retinal layers for segmentation. The final probability vector for a pixel  $\mathbf{y} = (y_1, \dots, y_C)$  can be computed by applying the *softmax* function over the *logits* as  $\mathbf{y} = \text{Softmax}(\mathbf{z})$ . The *softmax* function gives the relative probabilities between classes, but fails to measure the model’s uncertainty.

## 2.1 Bayesian Fully Convolution Network

Here we present a method to convert FC-DN to output the pixelwise uncertainty map in addition to the pixel-wise segmentation map. We name the proposed method Bayesian FC-DN (BFC-DN). In BFC-DN, we apply 1x1 convolution to the feature maps of last layers followed by *softplus* activation to output the variance  $\mathbf{v}$  for each pixel in addition to the *logit* vector  $\mathbf{z}$  i.e.,  $(\mathbf{z}, \mathbf{v}) = F_{\mathbf{W}}(X)$ . This variance gives aleatoric uncertainty of the model which the network learns to predict during the training. In addition, we include the *dropout layer* before every convolution layer which allows us to compute *epistemic uncertainty* which will be described in Sect. 2.2.

The output of the model is the Gaussian distribution  $\mathcal{N}(\mathbf{z}, \mathbf{v})$ . Computing the categorical cross entropy loss over this distribution is not feasible. Therefore, we approximated it using the monte-carlo integration. Given a set of training images and corresponding ground truth segmentation mask,  $D = \{X_n, Y_n\}_{n=1}^N$ , output *logit* for each sample in the mini-batch is perturbed  $T$  times with a Gaussian noise  $\epsilon_t \sim \mathcal{N}(0, \mathbf{v})$  as  $\hat{\mathbf{z}}_t = \mathbf{z} + \epsilon_t$  and the final pixel-wise bayesian loss is computed as:

$$L(W) = -\frac{1}{T} \sum_{t=1}^T \sum_{c=1}^C \beta_c \sum_{\forall Y_c} \log y_c^t \quad (1)$$

where  $y_c^t$  is obtained by applying *softmax* to the *logit* vector  $\hat{\mathbf{z}}_t$ ;  $Y_c$  denotes the pixels region of the  $c^{th}$  class in the ground truth  $Y$  and the scale factor  $\beta_c = 1/|Y_c|$  weights the contribution of each class to mitigate the class imbalances of different OCT layers and the background by increasing the weight of under represented classes while decreasing the effect of over represented classes. The

proposed Bayesian loss function encourage the network to minimize the larger losses by increasing the variance, therefore is more robust towards noise.

We train the proposed BFC-DN using the *bayesian loss* given by Eq. 1 for 40000 iterations. We have used mini-batch gradient descent and the Adam optimizer with momentum and a batch size of 2. The learning rate is set to  $10^{-5}$  which is decreased by one tenth after 10000 iterations of the training. Data augmentation is an important step in training deep networks. We augment the training images and corresponding label map masks through a mirror-image reflection and random rotation within the range of  $[-15, 15]$  degrees.

## 2.2 Segmentation and Uncertainty Quantification

Epistemic uncertainty is generally computed by assuming distribution over the network weights which allows the computation of distribution of class probabilities rather than point estimate [18]. Such methods require optimization over weights distribution and therefore is computationally expensive [7]. We adopt more practical approach introduced by [7] which is based on the *dropout variational inference*. We train the BFC-DN with a *dropout* layer before every convolution layer and use the *dropout* in test phase as well. Specifically, segmentation samples from the output predictive distribution are obtained by performing  $T$  stochastic forward passes through the network, i.e.,  $(\mathbf{z}^t, \mathbf{v}^t) = F_{\hat{\mathbf{W}}_t}(X)$ ,  $t = 1, \dots, T$  where  $\hat{\mathbf{W}}_t$  is an effective network weight after the *dropout*. In each forward pass, the fraction of network weights (denoted by *dropout rate*) are disabled and the segmentation score is computed using only the remaining weights. The segmentation score vector  $\bar{\mathbf{y}}$  and the aleatoric variance  $\bar{\mathbf{v}}$  is obtained by averaging the  $T$  samples, via *monte carlo integration*:

$$\bar{\mathbf{y}} = \frac{1}{T} \sum_{t=1}^T \text{Softmax}(\mathbf{z}^t) \quad (2)$$

$$\bar{\mathbf{v}} = \frac{1}{T} \sum_{t=1}^T \mathbf{v}^t \quad (3)$$

The average score vector contains the probability score for each retinal layers class, i.e.  $\bar{\mathbf{y}} = [\bar{y}_1, \dots, \bar{y}_C]$ . The overall segmentation uncertainty for each pixel can then be obtained as:

$$U(\bar{\mathbf{y}}) = - \sum_{c=1}^C \bar{y}_c \log \bar{y}_c + \bar{\mathbf{v}} \quad (4)$$

where the first term denotes *epistemic uncertainty* of the score computed as the entropy of the average score vector obtained by averaging  $T$  stochastic predictions (Eq. 2) and the second term is the uncertainty output produced by the network itself (Eq. 3). We set the *dropout* rate=0.4 and  $T = 50$  to allow sufficient sampling of network weights for final prediction.

For uncertain predictions, network assigns higher probabilities to different classes for different forward passes, resulting in higher epistemic uncertainty given by Eq. 4. For the certain predictions, network assigns higher probability to the true class for different forward passes, resulting in lower epistemic uncertainty. Since epistemic uncertainty is related to the model parameters weights, it can be reduced by observing more data. This is because, the network becomes robust towards weight dropout in test phase as it observes more data.

### 3 Experiments

The dataset [19] consists of 1487 images from 15 spectral-domain optical coherence tomography (OCT) volumes from unique normal subjects acquired on a Spectralis scanner. The size of each volume is  $512 \times 496 \times N_{slices}$  where  $N_{slices}$  is different for each volume and ranges from 49–100. All scans have axial resolution of  $3.87 \mu\text{m}$ . The ground truth has been obtained by manual annotation of the nine boundaries from eight retinal layers [12]. To facilitate the pixel-wise semantic segmentation, we convert the layer boundaries to the probability map for the eight layers regions and the background region. Therefore, the number of classes is  $C = 9$ .

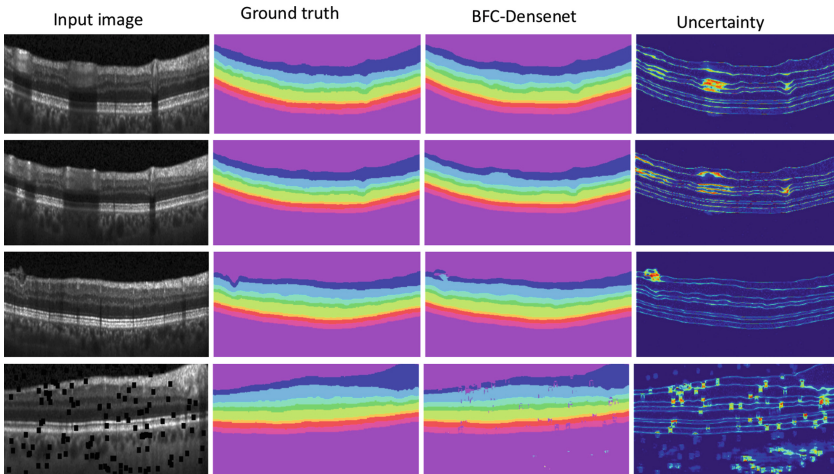
Out of 1487 images, we select 1116 images from 12 volumes to create a training set and remaining 291 images from 3 volumes for validation. We compare our method with the baseline FC-DN [10] which do not take into account uncertainty, i.e., the networks do not output aleatoric variance and segmentation is performed in a single forward pass by disabling the *dropout* in the test phase. To train these networks, we use non-bayesian class weighted cross entropy loss function which can be derived by setting  $T = 1$  and  $v = 0$  in Eq. 1.

**Table 1.** Performance of our proposed retinal layer segmentation method compared with the state-of-the-art Jégou et al. [10] and Lang et al. [12] segmentation methods.

Layer	Dice Coefficient		Boundary	Absolute error $\mu\text{m}$	
	FC-DN [10]	<b>BFC-DN</b>		Lang et al. [12]	<b>BFC-DN</b>
RNFL	$0.94 \pm 0.01$	$0.95 \pm 0.01$	ILM	$2.6 \pm 3.89$	$1.81 \pm 4.12$
GCL+IPL	$0.96 \pm 0.01$	$0.97 \pm 0.01$	RNFL-GCL	$4.0 \pm 6.11$	$3.6 \pm 6.3$
INL	$0.91 \pm 0.01$	$0.93 \pm 0.01$	IPL-INL	$3.78 \pm 4.41$	$2.6 \pm 2.73$
OPL	$0.90 \pm 0.01$	$0.91 \pm 0.01$	INL-OPL	$3.66 \pm 3.84$	$2.9 \pm 2.8$
ONL	$0.96 \pm 0.008$	$0.96 \pm 0.005$	OPL-ONL	$3.4 \pm 4.24$	$2.64 \pm 2.54$
IS	$90 \pm 0.01$	$0.91 \pm 0.01$	ELM	$2.79 \pm 2.68$	$2.44 \pm 2.4$
OS	$0.91 \pm 0.01$	$0.92 \pm 0.01$	IS-OS	$2.38 \pm 2.49$	$2.1 \pm 2.3$
RPE	$0.95 \pm 0.01$	$0.96 \pm 0.008$	OS-RPE	$4.16 \pm 4.13$	$3.8 \pm 3.3$
–	–	–	BrM	$3.87 \pm 3.69$	$3.14 \pm 2.81$

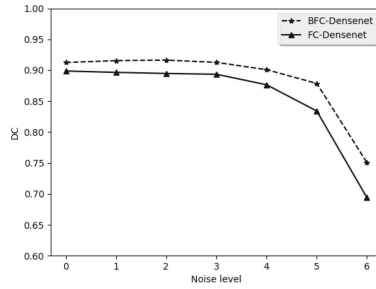
Table 1 compares the average Dice coefficient (DC) between the ground truth and predicted segmentation of the 8 layers using the proposed Bayesian method (BFC-DN) and non-Bayesian method (FC-DN [10]). The proposed method BFC-DN resulted in highest DC of 0.97 for GCL+IPL layer and lowest DC of 0.91 for OPL and IS layer. Moreover, BFC-DN resulted in improved segmentation for most of the layers in comparison to FC-DN. Table 1 also compares the average absolute error for 9 boundaries of our method with [12]. We observe that BFC-DN resulted in lower error than [12] which indicates proposed uncertainty based method is effective in segmenting retinal layers.

Figure 1 shows the examples of segmentation and uncertainty maps produced by our proposed method on few images from the validation set. It can be seen that our method produces pixel-wise uncertainty associated with the segmentation output where high uncertainty correlates with the inaccurate segmentation in the corresponding region. In order to validate the robustness of our proposed method against noise, we evaluate the performance by adding random block noise to the test images as shown in the last row image of Fig. 1. We observe that BFC-DN performs much better than FC-DN in presence of large noise levels as shown in Fig. 2. This demonstrates that BFC-DN is more robust towards the noisy images than FC-DN.



**Fig. 1.** Examples of retinal layer segmentation and uncertainty quantification using proposed BFCN-Densenet. (a) test images, (b) ground truth, (c) predicted segmentation map from FBC-Densenet (d) uncertainty map (warmer color denotes regions with higher uncertainty). The last row shows an example of layer segmentation in test images with added random block noise.

The average execution time for the retinal layer segmentation for BFC-DN is 2.5 s per image Tesla-K40 GPU which is somewhat slower than that of FC-DN



**Fig. 2.** Comparison of average segmentation performance of proposed BFC-DN with FC-DN [10] for different noise levels. The number of random block noise components at a given noise level is double than that at previous level.

which took 300 ms. This is because our model requires  $T$  forward passes in the test phase in contrast to FC-DN which requires one forward pass.

## 4 Conclusion

In this paper, we proposed a Bayesian deep learning based method for retinal layer segmentation in OCT images. Our method produces layer segmentation and corresponding uncertainty maps depicting the pixel-wise confidence measure of the segmentation output. Experimental results demonstrate that our method compares favorably with non-bayesian DL methods, particularly in the presence of noise and outperforms state of the art boundary based segmentation method. We have shown qualitatively that the resulting uncertainty maps correlates with the inaccuracies in segmentation output. The proposed method is applicable in determining the confidence of image analysis modules that utilizes the segmentation output for downstream analysis. Such uncertainty visualization can also be useful in computer-assisted diagnostic systems where clinician have additional insight about various measurements generated by the system to make necessary adjustments and make more informed decisions. Also, the resulting uncertainty map can be integrated within active learning systems to correct the segmentation output.

## References

1. Acton, J.H., Smith, R.T., Hood, D.C., Greenstein, V.C.: Relationship between retinal layer thickness and the visual field in early age-related macular degeneration. *Investig. Ophthalmol. Vis. Sci.* **53**(12), 7618–7624 (2012)
2. Apostolopoulos, S., Zanet, S.D., Ciller, C., Wolf, S., Sznitman, R.: Pathological OCT retinal layer segmentation using branch residual u-shape networks. *CoRR* abs/1707.04931 (2017)
3. Carass, A., Lang, A., Hauser, M., Calabresi, P.A., Ying, H.S., Prince, J.L.: Multiple-object geometric deformable model for segmentation of macular OCT. *Biomed. Opt. Express* **5**(4), 1062 (2014)

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. Fang, L., Cunefare, D., Wang, C., Guymer, R.H., Li, S., Farsiu, S.: Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomed. Opt. Express* **8**(5), 2732 (2017)
6. Farsiu, S., et al.: Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography. *Ophthalmology* **121**(1), 162–172 (2014)
7. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: *International Conference on International Conference on Machine Learning*, pp. 1050–1059 (2016)
8. Garvin, M.K., Abramoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M.: Automated 3D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Trans. Med. Imaging* **28**(9), 1436–1447 (2009)
9. Iglesias, J.E., Sabuncu, M.R., Leemput, K.V.: Improved inference in bayesian segmentation using monte carlo sampling: application to hippocampal subfield volumetry. *Med. Image Anal.* **17**(7), 766–778 (2013)
10. Jégou, S., Drozdal, M., Vázquez, D., Romero, A., Bengio, Y.: The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Honolulu, HI, USA, 21–26 July 2017*, pp. 1175–1183 (2017)
11. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: *Neural Information Processing Systems (NIPS)* (2017)
12. Lang, A., et al.: Retinal layer segmentation of macular OCT images using boundary classification. *Biomed. Opt. Express* **4**(7), 1133–1152 (2013)
13. Mishra, A., Wong, A., Bizheva, K., Clausi, D.A.: Intra-retinal layer segmentation in optical coherence tomography images. *Opt. Express* **17**(26), 23719–28 (2009)
14. Novosel, J., Thepass, G., Lemij, H.G., de Boer, J.F., Vermeer, K.A., van Vliet, L.J.: Loosely coupled level sets for simultaneous 3D retinal layer segmentation in optical coherence tomography. *Med. Image Anal.* **26**(1), 146–158 (2015)
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
16. Roy, A.G., et al.: Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional network. *CoRR* abs/1704.02161 (2017)
17. Sedai, S., Tennakoon, R., Roy, P., Cao, K., Garnavi, R.: Multi-stage segmentation of the fovea in retinal fundus images using fully convolutional neural networks. In: *ISBI*, pp. 1083–1086, April 2017
18. Springenberg, J.T., Klein, A., Falkner, S., Hutter, F.: Bayesian optimization with robust Bayesian neural networks. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (eds.) *Advances in Neural Information Processing Systems 29*, pp. 4134–4142. Curran Associates, Inc. (2016)



19. Srinivasan, P.P., et al.: Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed. Opt. Express* **5**(10), 3568–3577 (2014)
20. Vermeer, K.A., van der Schoot, J., Lemij, H.G., de Boer, J.F.: Automated segmentation by pixel classification of retinal layers in ophthalmic OCT images. *Biomed. Opt. Express* **2**(6), 1743–1756 (2011)