2023 CMBEC45
Vancouver, B.C.
May 16-18, 2023
CMBES/SCGB

Segmentation of retinal layers on OCT scans using deep learning

I. Giraud[1] and L. Duong[1]

[1] Laboratoire d'imagerie interventionnelle, École de Technologie Supérieure, Montreal, Canada

Abstract— **This past decade, the use of artificial intelligence, and more precisely deep learning, has been really efficient in image processing, especially in object detection and image segmentation. In the medical field, retinal imaging represents an important area of research with a great clinical interest. Indeed, the observation of the retinal layers is helpful in the diagnosis, treatment and monitoring of plenty of retinal pathologies. In this context, this project was focused on using deep learning for OCT retinal layers segmentation. To do so, a UNET-VGG16 model has been employed and the method was evaluated on a Duke OCT database of 4780 B-scans. It succeeded in segmenting three retinal layers with an IoU of 0.529 and a Dice coefficient of 0.685. To go further, the use of data augmentation, preprocessing and post processing functions could solve some issues and improve the method.**

Keywords— **image segmentation, OCT, retinal layers, U-NET, VGG16**

I. Introduction

Retinal imaging has revolutionized the way to identify pathologies and monitor response to therapy. It is an eminent tool in the prediction, diagnosis and monitoring of the majority of retinal diseases, but also other pathologies such as cardiovascular and neurodegenerative diseases. Studies have besides shown that the changes in the retinal structure occur before the patient experiences vision problems. Consequently, a relevant scanning of the retina is crucial in numerous aspects of the healthcare domain.

Facing the lack of techniques for a depth diagnosis of the retinal structure, OCT has appeared to be one of the most used methods in ophthalmology for retinal observation. Indeed, by being completely non-invasive, it provides images without damaging the tissue. The OCT resolution is also higher than the other medical imaging techniques. Accordingly, it has become a widely-used and valuable method.

However, retinal OCT images are difficult to read and often suffer from speckle noise. The images can contain varying intensity and the movement of the patient during the process is also responsible for image distortion. All of these features result in a time-consuming diagnosis by the ophthalmology expert. Thus, there is a need for an automatic retinal image analysis in order to improve the diagnostic process and to deal with an ever-increasing number of patients suffering from retinal pathologies. Nevertheless, many traditional techniques of retinal segmentation have failed to extract accurately the desired information in the retina. The retina is composed of 10 distinct layers, and is only a few hundred micrometers thick. Thus, if the algorithm is not sufficiently precise, it can lead to an incorrect visual prediction of the disease. In this context, despite a great variety of algorithms for retinal OCT segmentation, a powerful use in clinical practice remains a significant challenge. For this purpose, deep learning approaches have been largely used in the last decade and tend to be a promising solution.

In 2017, Roy et al. proposed ReLayNet, a variation of the U-Net network [1]. The model has a fully convolutional deep architecture, for an end-to-end semantic segmentation of retinal OCT B-scan into 7 retinal layers and fluid masses. Although it achieves well performance in retinal segmentation, it is easily affected by image degradation and poor spatial resolution. Pekala et al. (2019) [2] proposed a set of fully convolutional networks together with a Gaussian process based regression for improving the quality of the estimates. Their model achieves comparable results with human annotation of five OCT retinal layers but has slightly overestimated the support of the retinal layers. In 2020, Li Q et al. proposed DeepRetina [3]. Used as a backbone, an improved version of Xception65 extracts and learns the characteristics of retinal layers to create feature maps next inputted to an ASPP module. The multiscale feature information obtained is then recovered to capture clearer retinal layer boundaries using an encoder-decoder module. The model successfully segments ten retinal layers but still has difficulties to achieve an accurate segmentation because of the low resolution of the retinal OCT image and noise-induced interference. In 2021, He et al. [4] combine the two steps of pixel classification and topological smoothing into a unique end-to-end deep learning process by directly modeling the distribution of the surface positions. The model succeded to delineated nine retinal surfaces but it is easily affected by bad quality images and artifacts. Moreover, their topology module updates the surfaces iteratively from top to bottom, which may affect the lower surfaces with bad estimates of the upper surface.

The different authors used different metrics and databases than our study. Therefore, comparison is difficult and further

experiments with a common dataset should be conducted to conclude the respective effectiveness.

II. Materials and Methods

A. Environnement

For this project, the python environment has been chosen, with the use of the Keras and the Tensorflow-GPU libraries. The requirements are the following : python = 3.10.8, keras=2.10.0, tensorflow=2.10.1.

B. Database

The database used in this study was the Duke University SD-OCT Dataset [5]. Only the B-scans from the subjects without AMD have been used. It represents 4780 images, randomly sorted in three sets following the 90%/10% distribution: 3824 scans for the training set, 478 for the testing set and 478 for the validation set. As the database was originally in .mat format for a MATLAB utilization, the first step was to transform the images into a .tif format, more suitable with a python environment. Each image has its corresponding mask, resulting from the segmentation of 3 of the retinal layers, which are the inner limiting membrane (ILM), the Bruch's membrane (BM) and the retinal pigment epithelium (RPE). The figure 1 presents one scan of the database with its mask superimposed to situate the boundaries.

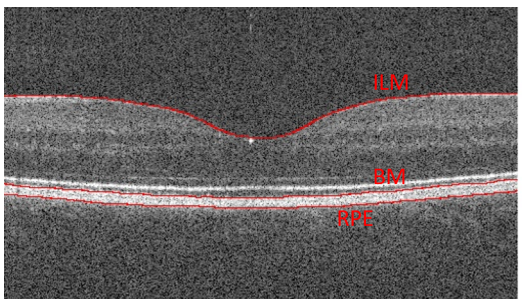

Fig. 1: OCT scan with the inner limiting membrane (ILM), the Bruch's membrane (BM) and the retinal pigment epithelium (RPE) labeled

C. Model Overview

The aim of the study was to build an end-to-end algorithm which takes in input an OCT B-scan and gives in output a mask presenting the three boundaries corresponding to the inner limiting membrane, the Bruch's membrane and the retinal pigment epithelium. The model is largely inspired by the structure of the U-NET network [6], with a down-sampling and an up-sampling sections. The encoder, or down-sampling, phase corresponds to the 17 first layers of the VGG16 network, with the ImageNet weights [7].

As this project is focused on image segmentation, the last 3 fully connected layers have been removed and replaced by a decoder phase, proper to the U-Net architecture, using skip connections. The final structure of the model is presented on the figure 2.

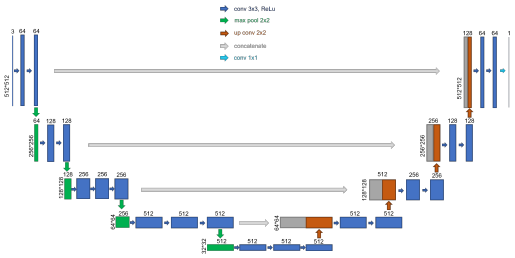

Fig. 2: Architecture of the proposed model

D. Metrics

For determining objectively the performance and quality of segmentation methods, metrics have to be used. During the training phase of the algorithm, the metrics used were the Precision, the Recall, the Dice coefficient and the Intersection over union parameter. During the validation process, the Accuracy has been also computed in addition of these parameters. The Dice similarity coefficient and the Intersection over Union are both widely used metrics for biomedical image segmentation due to their robustness to class imbalance.

Intersection-Over-Union (IoU): also called the Jaccard Index, it corresponds to the ratio of the area of overlap between the predicted segmentation and the ground truth (image labeled by an expert) among the area of union between the predicted segmentation and the ground truth.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where A represents the ground truth and B the segmentation predicted by the algorithm.

Dice coefficient: or F1 score, it is two times the area of overlap divided by the total number of pixels in both the predicted image and the ground truth.

$$Dice\ coef = \frac{2 * |A \cap B|}{|A| + |B|} \tag{2}$$

Lastly, the Dice loss coefficient, used during the training phase as the loss function, is determined by :

$$Dice\ loss = 1 - Dice\ coef \tag{3}$$

E. Training of the model

First, the database is split into the training, testing and evaluating datasets, according to the 90%/10% ratio. Both the OCT images and their labeled images are then read and resized into 512x512 images. After that, the proposed model is loaded and compiled using the Keras library. The function compile() takes in arguments the metrics presented in the precedent subsection, the dice loss as the loss function, and the Adam optimizer with a learning rate of 10^{-5}.

The model is then trained using the model.fit() method with an epoch's number of 25. ModelChekpoint callback is also used in conjunction with model.fit() to save the best version of the model computed along the training. Therefore, the model can be loaded later to continue the training from the state saved. The ReduceLROnPlateau function is also used to reduce the learning rate by 10 when a metric has stopped improving during 5 consecutive epochs.

F. Model evaluation:

The evaluation phase takes in input the 478 images of the evaluation set. Using the model.predict() function of the Keras library, a prediction is made for each of the input images for the three desired retinal layers. Then, to evaluate the performance of the model, the predicted masks are compared with the corresponding ground truth for all the B-scans. The evaluation is made using the metrics presented in the subsection D.

III. Results

After training the model on 3824 images and along 25 epochs, the evaluation has been made on 478 images. On the figure 3, three B-scans from the database are presented with their corresponding ground truth and predicted image. On the plot a), the predicted mask is very similar to the ground truth image, with is a promising result. The model has a good capacity to fit precisely with the retinal layers' boundaries. However, it can also be noticed that the model is not perfect. Indeed, with a bad contrast, as in the case b), all the boundaries have not been plotted entirely. In addition, it can be observed that when the retina presents abnormalities in one of its boundaries, such as the upper right of the inner limiting membrane (ILM) in the plot c), the model hardly perceives these differences.

In order to have a global idea of the efficiency of the segmentation, the mean of the evaluation parameters has been computed on all the validation set (478 images). The five metrics are: precision, recall, accuracy, IoU, and Dice coefficient. The results are presented in the table 1 below.

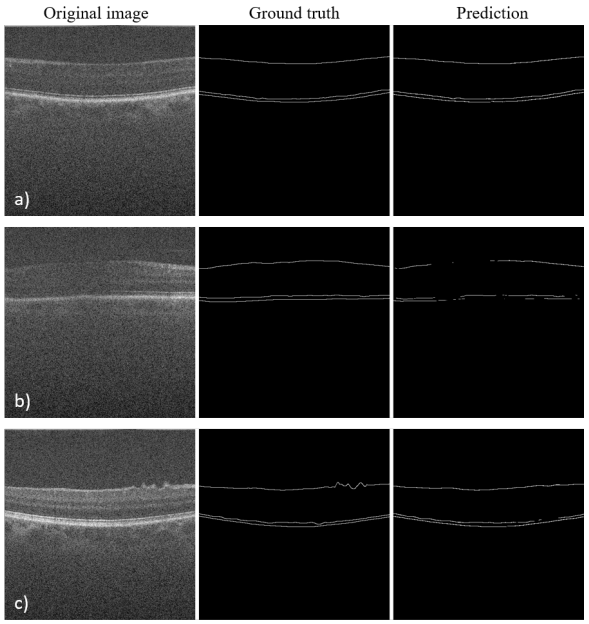

Fig. 3: In the left, 3 B-scans of the dataset with their corresponding ground truth and prediction mask, respectively in the center and the right of the figure

Table 1: Mean of the five metrics used for the validation of the model

Metrics	Mean value
Precision	0.685
Recall	0.686
Accuracy	0.996
IoU	0.529
Dice coef	0.685

Regarding the statistics, the accuracy is high, but this is also due to the class imbalance of the retinal masks. For the other metrics, the results are satisfactory but not good enough to affirm that the model is clinically effective. Among the metrics, the intersection over union parameter, IoU, is 0.529 and therefore relatively low. Improvement has to be done to enhance these results.

Moreover, the evolution of the metrics has also been measured all along the training of the model. It shows for each of the 5 metrics used (recall, precision, IoU, dice coefficient, and dice loss) their evolution along the epochs during the training and testing phases. On the training data, the metrics are still improving at the end of the 25 epochs. However, the same metrics on the testing data have stopped improving and begun to decrease in the last epochs. More than that, the recall has only been decreasing on the testing set, from 0.95 to lower than 0.70. This feature is a sign of over-fitting. When

the number of epochs used to train a neural network model is more than necessary, the model learns patterns that are specific to the training data. This makes the model unable to perform well on a new data set. Thus, this model gives high accuracy on the training set but fails to achieve good accuracy on the test set.

IV. Discussion

At the end of the project, the model was capable of predicting the 3 retinal layers expected. It has obtained first convincing results, but a lot of work can still be done to improve the performance. Indeed, some limitations can be noticed : the model struggles to perform an accurate segmentation in presence of images with low resolution or retinal layers abnormalities. Therefore, the model can not be generalized to relevant retinal diseases because of its weakness to adapt to changes. In order to ameliorate the model, the number of input images can be increased by the use of data augmentation with the ImageDataGenerator function from the Keras library. The model can also extend to the entire Duke dataset with the retinal images of age-related macular degeneration (AMD). By doing so, the model will generalize itself and be able to segment at the same time different profiles of retinal images. However, to take this a step further, the lack of unified public retinal OCT datasets with retinal profiles from various diseases makes it more difficult to create an algorithm that can produce accurate segmentation for each type of retinal OCT scan.

In parallel, it has been shown by looking at the results that the model suffers from over-fitting, as the metrics have continued to improve along the epochs on the training set when they were simultaneously stagnant or decreasing on the testing set. Therefore, enriching the database is also a good solution to limit the over-fitting. Optimizing the number of epochs to set is as well to take into consideration for avoiding overfitting in the continuation of this project.

Moreover, no pre-processing operation has be done on the dataset. Nonetheless, it has been noticed that some of the retinal images of the dataset suffer with low contrast or illuminations which then tend to make the segmentation more difficult by the algorithm. Indeed, it has been noticed that the model do not have the same performance for each scan. Therefore, preparing images should be a step further in this project. In the same way, post processing operations can be useful to enhance the segmentation and give in output better prediction masks.

V. Conclusion

The aim of the project was to segment OCT retinal images using deep learning. With this purpose, a Duke dataset has been utilized containing 4780 B-scans of normal retinas. These scans were labeled by an expert for three retinal layers : the inner limiting membrane (ILM), the Bruch's membrane (BM) and the retinal pigment epithelium (RPE). The algorithm proposed was created using a python environment and is largely inspired of the U-Net structure. It is also composed, for the encoder phase, of the 17 first layers of the VGG16 network, with the ImageNet weights.

After 25 epochs and without data augmentation, the model presented in this paper has been capable of predicted the three retinal layers, with a global Dice coefficient of 0.685 and IoU of 0.529. Recall, precision and accuracy have also been computed and are equal to respectively 0.686, 0.685 and 0.996. The model has still a lot of difficulties when dealing with low contrast and illuminations. However, a lot of aspects to work on has been found to improve the performance of the method, with among others, the use of data augmentation, as well as pre and post processing operations.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "Relaynet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional network," *CoRR*, vol. abs/1704.02161, 2017.
2. M. Pekala, N. Joshi, T. A. Liu, N. Bressler, D. C. DeBuc, and P. Burlina, "Deep learning based retinal oct segmentation," *Computers in Biology and Medicine*, vol. 114, p. 103445, 2019.
3. Q. Li, S. Li, Z. He, H. Guan, R. Chen, Y. Xu, T. Wang, S. Qi, J. Mei, and W. Wang, "DeepRetina: Layer Segmentation of Retina in OCT Images Using Deep Learning," *Translational Vision Science Technology*, vol. 9, no. 2, pp. 61–61, 12 2020.
4. Y. He, A. Carass, Y. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Structured layer surface segmentation for retina oct using fully convolutional regression networks," *Medical Image Analysis*, vol. 68, p. 101856, 2021.
5. S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, and C. A. Toth, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," *Ophthalmology*, vol. 121, no. 1, pp. 162–172, 2014.
6. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, 2015.
7. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.