

Towards Automatic Glaucoma Assessment: An Encoder-decoder CNN for Retinal Layer Segmentation in Rodent OCT images

Rocío del Amor^{*}, Sandra Morales^{*}, Adrián Colomer^{*}, José M. Mossi[†], David Woldbye[‡],
Kristian Klemp[§], Michael Larsen[§], and Valery Naranjo^{*}

^{*}*Instituto de Investigación e Innovación en Bioingeniería, I3B, Universitat Politècnica de València, Valencia, Spain*

Email: {madeam2, sanmomar, adcogra, vnaranjo}@i3b.upv.es

[†] ITEAM Research Institute, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain.

[‡] Laboratory of Neural Plasticity, Department of Neuroscience, University of Copenhagen, Denmark.

[§] Dept. of Ophthalmology, Rigshospitalet-Glostrup, Glostrup, Copenhagen, Denmark.

[§] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

Abstract—Optical coherence tomography (OCT) is an important imaging modality that is used frequently to monitor the state of retinal layers both in humans and animals. Automated OCT analysis in rodents is an important method to study the possible toxic effect of treatments before the test in humans. In this paper, an automatic method to detect the most significant retinal layers in rat OCT images is presented. This algorithm is based on an encoder-decoder fully convolutional network (FCN) architecture combined with a robust method of post-processing. After the validation, it was demonstrated that the proposed method outperforms the commercial Insight image segmentation software. We obtained results (averaged absolute distance error) in the test set for the training database of $2.52 \pm 0.80 \mu\text{m}$. In the predictions done by the method, in a different database (only used for testing), we also achieve the promising results of $4.45 \pm 3.02 \mu\text{m}$.

Index Terms—Optical coherence tomography, rodent OCT, layer segmentation, convolutional neural network, glaucoma assessment.

I. INTRODUCTION

Optical coherence tomography (OCT) is an important imaging modality used to capture various aspects of biological tissues, such as structural information, blood flow, elastic parameters, change of polarization states and molecular content [1]. OCT uses the principle of low coherence interferometry to generate two or three-dimensional images of biological samples by obtaining high-resolution cross-sectional backscattering profiles [2]. OCT is widely used as a retinal imaging modality since it is a non-invasive technique capable of capturing the structure within the retina at the micron scale. The retina is organized into layers and a change in this structure have been associated with ophthalmic, neurodegenerative and vascular disorders [3]. Some of these changes are related to the layer thickness, so the segmentation in this type of images is essential. Manual OCT segmentation is extremely time-consuming, tedious, and is subjected to inter and intra observer

variability. Automated segmentation holds the potential to reduce the time and effort required to delineate the retinal layers and also to provide repeatable and quantitative results. For all of the above, the development of automatic segmentation algorithms is very useful to obtain a rapid and effective diagnosis in some ocular diseases as glaucoma, where the thickness of some layers can be affected. In recent years, deep learning has had a substantial impact on medical imaging. Specifically for OCT segmentation, some recent studies have proposed the use of convolutional neural networks (ConvNets) to face the task of layer segmentation [4]. Leyuan Fang et al. proposed a new method named CNN-GS for automatic segmentation of nine layer boundaries on human retinal OCT images. This method included a CNN model to extract useful features of specific retinal layer boundaries and trained a corresponding classifier to delineate eight layers. After that, class probabilities for each pixel are used to refine the boundaries. To assess the effectiveness of their technique, they validated their method on 60 OCT volumes (2915 B-scans) from 20 human eyes. As quantitative performance metrics, they calculated the mean thickness difference (in pixels) between the automated and manual segmentation for all layers and calculated the absolute value of these differences [5]. Olaf Ronneberger et al. proposed a network called U-net for segmentation of neuronal structures in electron microscopic stacks. The architecture was composed of a contracting path to capture context and a symmetric expanding path that enabled precise localization. A set of 30 images (512x512 pixels) from the larva ventral nerve cord was used as training. Also, data augmentation were introduced to reduce the number of annotated images [6]. Lee et al. developed a CNN that detected intraretinal fluid (IRF) on OCT. The CNN used was a modified version of the U-net autoencoder architecture with a total of 18 convolutional layers and a sigmoid function as final activation function that generated a prediction of a binary segmentation map. Manual segmentation of 1,289 OCT macular images were used for training and cross-validation. This model segmented

This work was supported by the Project GALAHAD [H2020-ICT-2016-2017, 732613].

images with a 0.911 cross-validated Dice coefficient [7]. It should be noted, no recent studies have been focused on the segmentation of the rodents retina layers using deep learning techniques. So far, this kind of methodology has been only proposed to segment human retinal layers and the architectures proposed for humans by themselves do not obtain satisfactory results in rodents because the human and rodent retina present significant differences. This paper presents a method based on CNNs to detect the most significant retinal layer boundaries in rat OCT images which are represented in Fig.1.

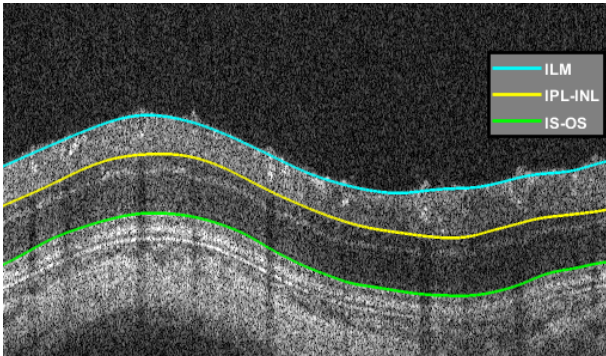


Fig. 1. Retinal layer boundaries segmented (ILM: Internal Limiting Membrane, IPL: Inner Plexiform Layer, INL: Inner Nuclear Layer, IS: Inner Segment, OS: Outer Segment).

Specifically, it consists in an encoder-decoder fully convolutional network (FCN) architecture, built upon ReLayNet [8]. Originally Relaynet was developed to segment human retinal layers and, although human and rodent retina presents significant differences, this study demonstrates that an algorithm used to detect several layers of the human retina could be adapted to capture the most significant features in rats OCT followed by the robust post-processing algorithm proposed in this work.

II. MATERIALS AND METHODS

A. Rodent OCT images

For this study, we used two private databases of images which were acquired at different conditions and belong to different rodent. In both cases, the permission for animal experiments was granted by the Danish Council for Animal Experimentation and rat OCT images were taken with the Micron IV equipment (Phoenix Research Labs, Pleasanton, USA). To carry out both studies, the images of rats were acquired before and after intravitreal injection of endothelin-1 (ET-1). ET-1 is a vasoconstrictor of retinal vessels that causes ischemia. So, this injection contributes to the degeneration of the retinal layers, simulating the conditions of a retinal disease. This fact makes the rat ET-1 intravitreal injection model suitable for research studies about retinal layer monitoring. In particular, the first database (RatsOne) is composed of 129 OCT images, belonging to ten rats of 1024x1024 pixels with 0.9775 $\mu\text{m}/\text{pixel}$. The second database (RatsTwo) contains 130 OCT images of 1024x1024 pixels with 0.9775 $\mu\text{m}/\text{pixel}$ of 12 rats. The most significant retinal layer boundaries that

were visibly distinguishable on these images were manually segmented by an expert (Fig.1). The RatsOne database is used for training and validation purposes. The RatsTwo database is used to do a second test and guarantee that our model is useful to different sets of images. So, the goal of the method proposed in this paper will be to segment those three-layer boundaries (ILM, IPL-INL and IS-OS) and generalise their results.

B. Fully convolutional neural networks (FCN)

The segmentation of the retinal layers proposed in this paper can be interpreted as a classification problem. Given a pixel of an OCT image $p = f(x, y)$, the goal is to assign it into its corresponding label k in the label space $L = 1, \dots, N$ being N the number of possible classes. In this work, $N = 4$, i.e. the number of boundary layers to be segmented plus one. To achieve this task, an encoder-decoder fully convolutional network built upon ReLaynet [8] is proposed in this paper (see Fig. 2).

1) *Encoder-decoder blocks*: Each encoder block is composed by four layers, in sequence: convolution layer, batch normalization layer, ReLU activation layer and max-pooling layer. To capture the transitions between the different retinal layers, 64 rectangular kernels of 7x3 are defined in the convolutional layers of all encoder blocks. Note that zero padding is used to preserve the spatial dimensions of the activation maps after each convolutional layer. Then, batch normalization technique is applied to avoid overfitting during the training procedure [9] and ReLU activation function is employed to introduce non-linearities and resulting in much faster training for large networks. Max-pooling operation is introduced in the last layer of each encoder block to condense the feature information reducing the spatial dimensions. This reduction depends on the kernel size and the stride used, in this case, the dimensions are reduced by a half since max-pooling kernel with stride $s = 1$ is used. In addition, to maintain spatial consistency, the pooling indexes of this operation are achieved and transferred to the corresponding unpooling layer in the decoder block.

Each decoder block consists of five layers, in sequence: unpooling layer, concatenation layer, convolution layer, batch normalization and ReLU activation function. The unpooling layer upsamples the feature maps from the previous decoder block to a double resolution by using the achieved pooling indexes corresponding to the matched encoder block. After this step, a concatenation of the upsampled feature maps with the corresponding output feature maps of the matched encoder block is performed to enrich the information and avoiding vanishing gradient problems. Finally, convolutional layer, batch normalization and ReLU are applied to the concatenated feature map.

The final decoder block consists of a convolutional layer with 1x1 kernel and the softmax activation function. This part is responsible to associate each pixel to one of the four possible classes (i.e. upper bound, RNFL+GCL+IPL, INL+OPL+ONL+IS-OS, bottom bound). Note that the target layers are shown between the boundaries exposed in Fig 1.

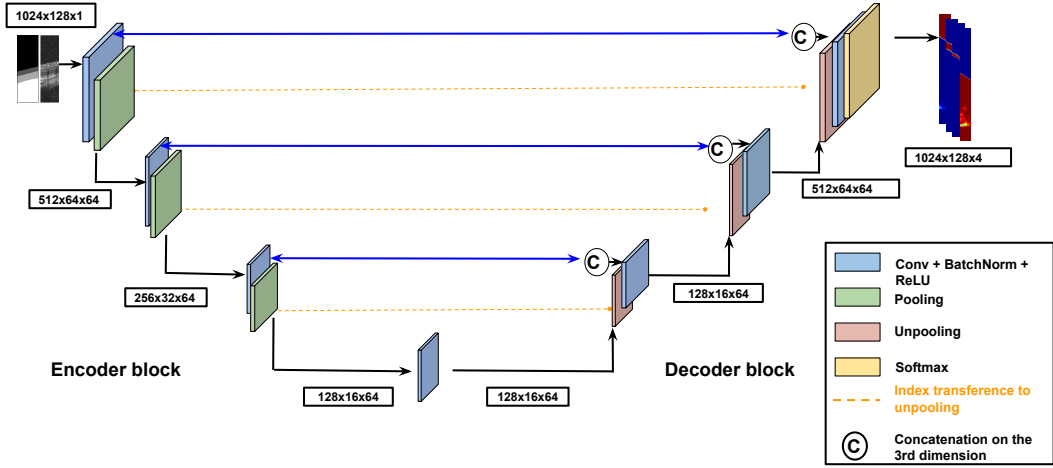


Fig. 2. The encoder-decoder architecture proposed to address the segmentation task.

2) Training process:

Data conditioning. Because the images used for training have large dimensions, a patch-wise learning methodology is applied in order to avoid memory problems. In this case, it is employed 1024×128 patches. Furthermore, we augment the sliced data by introducing random geometric transformations such as croppings, horizontal flips, rotations and translations that permits reducing the number of images annotated. In the testing stage memory requirements are more permissive allowing to predict a test sample in two slices of 1024×512 with a Titan V GPU.

Data partitioning. Attending to Section II-A, the dataset used to train, RatsOne, is composed by 129 images coming from ten different rodents. With the aim of avoiding biased results due to a specific partition of the database into training and test subsets and obtaining the segmentation of the 129 images, an external K -fold cross-validation technique was carried out. Specifically, $K = 8$ partitions were created. The images belonging to rats 1-7 composed seven individual partitions while the images from the last three rats were grouped forming the eighth fold to avoid imbalance problems (see Table I). Consequently, in the training process, $K - 1$ different folds in each external iteration were used while the remaining partition was utilized to test the model performance. In addition, an internal leave-one-out cross-validation was carried out, using the images from one different training fold in each internal iteration as validation set. As a result of this process, $(K - 1) \times K$ models were learned. This technique guarantees reliable results and robust models.

Loss function. The proposed network is trained by optimizing a weighted multi-class logistic function. This logistic loss provides a probabilistic measure of similarity between the prediction and the ground truth. Let $p_c(x)$ be the probability for the pixel x of belonging to the class c and $g_c(x)$ the ground truth probability, the basic logistic loss function can be defined as $L(x) = -\sum_x \omega(x)g_c(x)\log(p_c(x))$. The weights $\omega(x)$ are introduced in the loss function with two

TABLE I
DISTRIBUTION OF THE RAT IMAGES OF RATS ONE THROUGHOUT THE EIGHT FOLDS.

| Rat ID | #images | Fold |
|-------------|---------|------|
| C23 GP1A A1 | 16 | 1 |
| C23 GP1A A2 | 17 | 2 |
| C23 GP1A A3 | 16 | 3 |
| C23 GP1A A4 | 16 | 4 |
| C24 GP2B A1 | 16 | 5 |
| C24 GP2B A2 | 16 | 6 |
| C24 GP2B A3 | 16 | 7 |
| C24 GP2B A4 | 4 | 8 |
| C25 GP1A A1 | 8 | 8 |
| C25 GP2B A3 | 4 | 8 |

objectives: (i) compensating the effects produced by the imbalance character of the classes (ii) boosting the accurate segmentation of the retinal layer boundaries. The pixels near to the tissue-transition regions are the most challenging cases to be correctly segmented due to diffused boundaries, speckle noise and the limited OCT resolution. Given the cumulative frequency f_c of class c in the training data (i.e. the prior class probability), the logic operator I , a ground truth image M and the 2D gradient operator ∇ , the $\omega(x)$ term is formulated as $\omega(x) = \sum_c I(M(x) = c) \cdot \frac{\mathbf{f}}{f_c} + \omega_0 \cdot I(|\nabla M(x)| > 0)$, where $\mathbf{f} = [f_1, f_2, \dots, f_c]$ is a vector containing all frequencies. The first term models median frequency balancing and compensates for the classes imbalance problem by enhancing classes with low probability. The second term establishes higher weight on anatomical boundary regions to emphasize on the correct segmentation of contours. ω_0 balances both terms.

Hyper-parameter configuration. The proposed network is learned through the stochastic gradient descent (SGD) optimizer using mini batches of eight samples. A momentum value of 0.9 and the learning rate is initially established to 0.001 and reduced by one order after every 20 epochs. The training stage is composed by sixty epochs and the model that minimizes the validation loss is chosen as the best one. The Titan V GPU was used to carry out this process.

3) *Post-processing*: After the training and the testing processes, a segmentation map per image is obtained. The second column of Fig. 3 shows some segmentation maps corresponding to representative images of the RatsOne and RatsTwo databases. From these segmentation maps, the objective is to obtain the target boundaries (third column) that correspond to those which delimitate the different classes. To accomplish this goal, a post-processing step is carried out for all the segmentation maps. Firstly for each layer, the component (defined by 8-connectivity) with the greater area is chosen. Secondly, a closing operation defined as $\phi_B(x) = \epsilon_B(\delta_b(X))$ with a disk (B) as structural element is applied. Finally, due to the speckle noise affecting the OCT image, each layer is softened using an 8-degree polynomial.

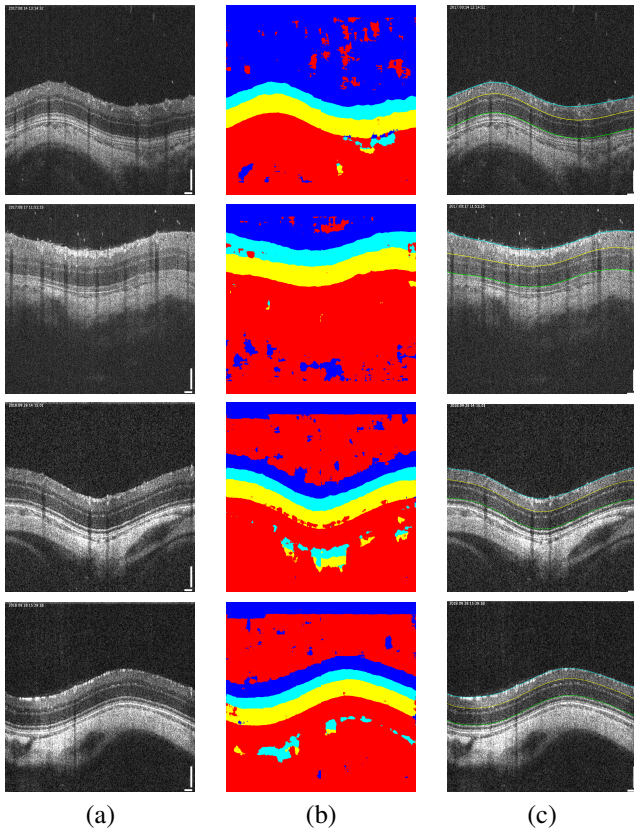


Fig. 3. Representative images from RatsOne and RatsTwo databases: (a) OCT images, (b) corresponding segmentation maps and (c) boundary segmentation results. The two top rows show RatsOne images and the two bottom rows show images from RatsTwo database.

III. RESULTS

To evaluate the performance of the proposed method, RatsOne and RatsTwo databases were manually segmented. The results obtained on RatsOne database were compared to the commercial Insight software from Phoenix Research Labs, which offers a tool for automatic segmentation. To obtain a quantitative analysis of the results, different measures were computed: absolute distance error between layer boundaries, layer thickness error and Dice similarity coefficient (see Table II-III).

TABLE II

ABSOLUTE DISTANCE ERRORS BETWEEN BOUNDARIES (MEAN AND STANDARD DEVIATION IN μm) CALCULATED COMPARING THE RESULTS OF THE PROPOSED METHOD AND INSIGHT WITH THE GROUND TRUTH ON RATSONE IMAGES.

| Boundary | Proposed | InSight |
|----------|-----------------|------------------|
| ILM | 1.93 ± 0.55 | 6.97 ± 4.59 |
| IPL-INL | 3.21 ± 0.72 | 10.49 ± 4.06 |
| IS-OS | 2.41 ± 1.11 | 8.67 ± 5.90 |
| Overall | 2.52 ± 0.8 | 8.71 ± 4.85 |

Note that InSight software does not detect some layers under analysis on some database images. So, only the automatically segmented layers were taken into account for absolute distance error computation (ILM is only detected in 113, IPL-INL in 26 and IS-OS in 70). Because of this, layer thickness error and Dice coefficient were not computed making use of InSight results.

Due to the inefficiency of the InSight software over the RatsOne database, RatsTwo images were not automatically segmented with this software, only absolute distance error, layer thickness error and Dice similarity coefficient between ground truth and our method was computed, see Table III-IV. These experiments were executed in MATLAB 2018b and MatConvNet framework [10].

IV. DISCUSSION

A wide validation of the proposed method was performed in the previous section. Regarding the RatsOne dataset, it must be emphasised that, unlike the proposed method, the commercial software Insight does not segment a fixed number of layers and its segmentation is not sufficiently accurate. Regarding the RatsTwo database, this was only used for testing and not for training purposes. It should be remarkable that, although the reported results were not as good as those reached on the RatsOne database, they are really promising taking into account the high variability of the RatsTwo database. Thereby, it is demonstrated that the proposed model could be generalised and applied to different datasets. This would avoid to do a manual segmentation or to train a new model if the time, hardware or software necessary to do that is not available. As it was mentioned in Section II, the proposed method is built upon ReLayNet architecture [8]. The results obtained in this paper demonstrate that an adequate adaptation of algorithms and good post-processing can help to generalize other deep learning algorithms to properly work on different images.

V. CONCLUSION

The use of OCT in rats presents interesting advantages for new drug testing. Therefore, the development of algorithms that allow the automatic analysis of this type of images is of importance.

Nowadays, OCT is very used as a retinal imaging modality. This fact allows the study of retinal diseases such as glaucoma which is characterized by the loss of RNFL thickness. In this paper, an algorithm to segment the rodent retinal layers is

TABLE III

THICKNESS ERRORS AND DICE COEFFICIENT (MEAN AND STANDARD DEVIATION) CALCULATED COMPARING THE RESULTS OF THE PROPOSED METHOD WITH THE GROUND TRUTH (GT) FOR IMAGES FROM RATSOne AND RATSTwo DATABASES.

| Layer | Average thickness (μm) | | | | Thickness errors (μm) | | Dice | |
|---------------------|-------------------------------------|----------|---------|----------|------------------------------------|----------------|------------------|-----------------|
| | RatsOne | | RatsTwo | | RatsOne | RatsTwo | RatsOne | RatsTwo |
| | GT | proposed | GT | proposed | proposed | proposed | proposed | proposed |
| RNFL+GCL+IPL | 66.81 | 67.37 | 72.34 | 67.85 | 1.70 ± 1.28 | 5.5 ± 3.8 | 0.96 ± 0.01 | 0.94 ± 0.02 |
| INL+OPL+ONL | 97.59 | 97.44 | 95.26 | 96.89 | 2.14 ± 1.76 | 4.42 ± 4.4 | 0.97 ± 0.01 | 0.95 ± 0.03 |
| Overall | 164.41 | 164.8 | 167.6 | 164.74 | 1.92 ± 1.52 | 4.96 ± 4.1 | 0.97 ± 0.002 | 0.95 ± 0.03 |

TABLE IV

ABSOLUTE DISTANCE ERRORS BETWEEN LAYER BOUNDARIES (MEAN AND STANDARD DEVIATION IN μm) CALCULATED COMPARING THE RESULTS OF THE PROPOSED METHOD WITH THE GROUND TRUTH OF THE RATSTWO DATABASE

| Boundary | proposed |
|----------------|-----------------|
| ILM | 3.09 ± 0.94 |
| IPL-INL | 5.25 ± 3.31 |
| IS-OS | 5.04 ± 4.81 |
| Overall | 4.45 ± 3.02 |

proposed. This permits to segment the RNFL+ GCL+IPL layer that can be used as an indicator for glaucoma diagnosis. Moreover, this type of algorithms reduces the subjectivity of manual methods making faster and less tedious the segmentation task.

After a wide validation using two different databases, we demonstrated that CNN architectures proposed to segment human retinal layers can be adapted to successfully perform this task on rodent OCT images.

ACKNOWLEDGMENT

Animal experiment permission was granted by the Danish Animal Experimentation Council (license number: 2017-15-0201-01213). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research.

REFERENCES

[1] R. Kafieh, H. Rabbani, M. Abramoff, and M. Sonka, "Intra-retinal layer segmentation of optical coherence tomography using diffusion map," in

2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013, pp. 1080–1084.

- [2] D. C. Debuc, "A review of algorithms for segmentation of retinal image data using optical coherence tomography," in *Image Segmentation*. IntechOpen, 2011.
- [3] M. Pekala, N. Joshi, D. E. Freund, N. M. Bressler, D. C. DeBuc, and P. M. Burlina, "Deep learning based retinal OCT segmentation," *CoRR*, vol. abs/1801.09749, 2018.
- [4] A. Yazdanpanah, G. Hamarneh, B. R. Smith, and M. V. Sarunic, "Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 484–496, Feb 2011.
- [5] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search," *Biomedical optics express*, vol. 8, no. 5, pp. 2732–2744, 2017.
- [6] 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.
- [7] C. S. Lee, A. J. Tyring, N. P. Deruyter, Y. Wu, A. Rokem, and A. Y. Lee, "Deep-learning based, automated segmentation of macular edema in optical coherence tomography," *Biomedical optics express*, vol. 8, no. 7, pp. 3440–3448, 2017.
- [8] 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 networks," *Biomed. Opt. Express*, vol. 8, no. 8, pp. 3627–3642, 2017.
- [9] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, 2015, pp. 448–456.
- [10] A. Vedaldi and K. Lenc, "Matconvnet – convolutional neural networks for matlab," in *Proceeding of the ACM Int. Conf. on Multimedia*, 2015.