

Document downloaded from:

<http://hdl.handle.net/10251/124182>

This paper must be cited as:

Fuentes-Hurtado, FJ.; Morales, S.; Mossi García, JM.; Naranjo Ornedo, V.; Fedulov, V.; Woldbye, D.; Klemp, K.... (2018). Deep-Learning-based Classification of Rat OCT images After Intravitreal Injection of ET-1 for Glaucoma Understanding. En Intelligent Data Engineering and Automated Learning – IDEAL 2018. Springer. 27-34.
https://doi.org/10.1007/978-3-030-03493-1_4



The final publication is available at

https://doi.org/10.1007/978-3-030-03493-1_4

Copyright Springer

Additional Information

Deep-Learning-based Classification of Rat OCT images After Intravitreal Injection of ET-1 for Glaucoma Understanding

Félix Fuentes-Hurtado¹, Sandra Morales¹, Jose M. Mossi¹, Valery Naranjo¹, Vadim Fedulov², David Woldbye², Kristian Klemp², Marie Torm³, and Michael Larsen³

¹Instituto de Investigación e Innovación en Bioingeniería, I3B, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain.

² Dept. of Neuroscience and Pharmacology, University of Copenhagen, Copenhagen, Denmark

³ Dept. of Ophthalmology, Rigshospitalet-Glostrup, Glostrup, Copenhagen, Denmark

⁴ Laboratory of Neural Plasticity, Department of Neuroscience, University of Copenhagen

Abstract. Optical coherence tomography (OCT) is a useful technique to monitor retinal damage. We present an automatic method to accurately classify rodent OCT images in healthy and pathological (before and after 14 days of intravitreal injection of Endothelin-1, respectively) making use of the DenseNet-201 architecture fine-tuned and a customized top-model. We validated the performance of the method on 1912 OCT images yielding promising results ($AUC = 0.99 \pm 0.01$ in a $P = 15$ leave-P-out cross-validation). Besides, we also compared the results of the fine-tuned network with those achieved training the network from scratch, obtaining some interesting insights. The presented method poses a step forward in understanding pathological rodent OCT retinal images, as at the moment there is no known discriminating characteristic which allows classifying this type of images accurately. The result of this work is a very accurate and robust automatic method to distinguish between healthy and a rodent model of glaucoma, which is the backbone of future works dealing with human OCT images.

Keywords: optical coherence tomography, deep-learning, glaucoma

1 Introduction

Optical coherence tomography (OCT) has become the most important retinal imaging instrument in ophthalmology for the early diagnosis of pathologies as it is high-resolution and non-invasive. The retina is organized in layers, and it has been demonstrated that changes in this structure may imply ophtalmic, neurodegenerative and vascular disorders [1]. Examples of these diseases are age-related macular degeneration (AMD), diabetic macular edema (DME) or glaucoma. Furthermore, as part of the central nervous system (CNS), the retina

is subject to a variety of specialized immune responses similar to those happening in the spinal cord and the brain. CNS disorders such as Alzheimer’s disease, Parkinson’s disease, multiple sclerosis or stroke have been associated to changes in the retinal structure [2].

Glaucoma is the second leading cause of blindness globally. The disease is caused by increased intra-ocular pressure resulting in irreversible damage to the optic nerve head (ONH). Systematic in depth ONH diagnostics after a positive glaucoma screening would save 4M cases of blindness yearly worldwide but a cost-effective test is not possible with existing technology. A leading candidate technology is optical coherence tomography (OCT). OCT is label free and non-invasive and can be implemented in compact and easy to handle systems with the potential to be used worldwide as diagnostic approach. Currently even high quality commercial systems typically achieve about 3-5 μm . However glaucoma screening requires a very high axial resolution for the posterior segment of the eye; perhaps as small as 1 μm . This paper is framed within the European project GALAHAD, which aims at building a new ultra-high resolution (UHR-) OCT system in a very cost-effective manner and developing new machine learning algorithms to analyse both existing data and new UHR- and polarisation sensitive (PS-) OCT images to identify the early stages of glaucoma.

Although the purpose of OCT is human retina imaging, rodents are sometimes utilized in research studies of ophthalmologic diseases and for treatment evaluation which would be unfeasible in humans. However, human and rodent retinas present significant differences mainly due to the difference in size. For example, rodent lenses are relatively larger, they have no macula or fovea. In addition, the thickness of rodent retinal layers is close to the resolution of standard OCT devices, which makes image processing of rodent OCT images even more challenging. For example, while the ganglion cell layer (GCL) in human eyes has a thickness of 20-60 μm , it is not visibly distinguishable in rodent OCT images (about 2 μm of thickness). The contrary occurs with the inner plexiform layer (IPL), which is relatively thick in rodent retinas (about 60 μm).

Endothelin-1 (ET-1) is a potent vasoactive peptide that causes vasoconstriction of retinal vessels and subsequent ischemia, which contributes to the degeneration of the retinal layers. The ischemia effects are similar to those present in glaucoma patients, which makes rat model of intravitreal injection of ET-1 suitable for research studies of the glaucoma disease and for testing the effectiveness of new treatments to slow down the layer degeneration process.

Some works have already attempted to classify retinal OCT images from different databases in healthy and a variety of pathologies. For example, in [3] the authors accurately detect AMD and DME using multiscale histograms of oriented gradient (HOG) descriptors as feature vectors of a support vector machine (SVM) in a database of more than 3.000 OCT images. [4] employed the VGG16 architecture to classify more than 80.000 OCT images in healthy or AMD with an area under the curve (AUC) of 92.77%. [1] achieved 99%, 89% and 86% in classifying healthy, AMD and DME OCT images from 45 patients by fine-tuning the GoogLeNet architecture. As for glaucoma detection, [5] achieved 93.1% of

accuracy on 102 patients feeding a random forest with the features extracted by a convolutional neural network (CNN). Finally, [6] obtained an $AUC = 0.88$ in discriminating glaucoma from healthy in 899 patients by applying a logistic regression on different features, such as retinal nerve fiber layer thickness, inter-ocular difference or age. At the light of these results, it remains unknown which features are enough discriminating to accurately classify glaucoma in OCT images [7].

In this paper, an automatic method to distinguish between healthy and a rodent model of glaucoma where damage to the inner retina is induced by intravitreal injection of endothelin-1 is presented. Its aim is to develop a robust method for glaucoma modelling in rat OCT images to then adapt it to be used with humans. Furthermore, this is a first step to better understand which features are the most discriminating at characterizing glaucoma in OCT images.

2 Methods

2.1 Data

A set of 89 Sprague-Dawley rats were used in the study. Rat OCT images were taken with the Micron IV equipment (Phoenix Research Labs, Pleasanton (USA), <http://www.phoenixreslabs.com/products/oct2/>). On average, 16 images per rat were recorded giving place to a total of 1912 rat OCT images (Fig. 1). Rats were anesthetized previously to image acquisition. Several OCT images were acquired before and after intravitreal injection of ET-1. In particular, a follow-up 14 days after injection was performed, when the maximum expression of retinal layer modification is produced according to the rodent model of glaucoma induced by ET-1 injection [8].

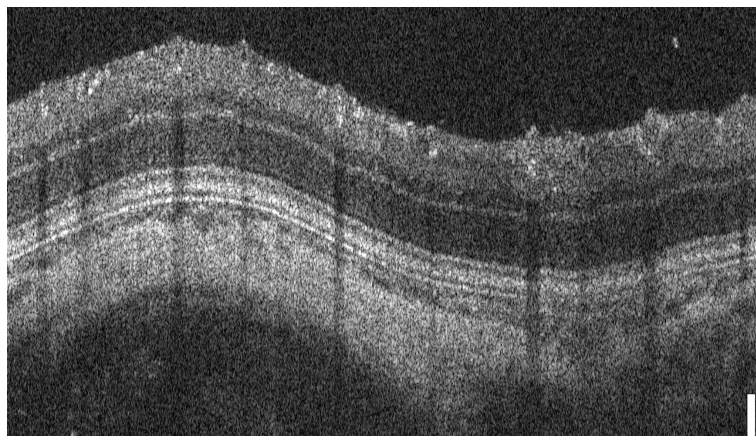


Fig. 1: Example of a healthy rat OCT acquired with the Micron IV equipment.

2.2 Deep Learning

Deep learning techniques have become the dominant machine learning approach for a wide variety of problems. Differently to traditional machine learning approaches, where features employed to solve the problem are hand-crafted by a human (feature engineering) and thus require a large amount of time and resources, deep learning methods learn these features by themselves through high amounts of data. Although it was first introduced over 20 years ago [9], it has been over the last few years when deep learning has led to an improvement of the state-of-the-art in several fields of artificial intelligence and machine-learning, such as image processing, voice processing and language processing applications. This was possible due to the increase of computational power achieved by GPUs and the availability of huge amounts of labeled data. Within the scope of image processing, CNNs are the most used approach [1].

In contrast to other deep neural networks which base their learning in updating the weight of their neurons, CNNs learn matrices (kernels or filters) to convolve across inputs to extract features from 3D tensors, normally images, and feed them to the following layer. Thus, first layers of CNNs learn to detect basic patterns, such as edges or color blobs of different hue, while layers towards the end of the model learn to detect more complex and problem-specific patterns. While convolutional layers are the heart of CNNs, other kind of layers use to be interspersed with them. Some of these layers are activation layers, such as rectified linear units layer (ReLU), which applies a fixed activation function to their input, pooling layers, which down-sample the spatial dimensions of data, or batch normalization layers, which normalize each input channel across a batch in order to speed up training and reduce the sensitivity to network initialization. A full explanation of the layers employed in deep learning is out of the scope of this paper.

DenseNet architecture. Many CNNs process data serially, that is, each layer operates only upon the output of the previous layer. In contrast, the DenseNet architecture allows each layer of the network to directly process the outputs of all previous layers. This schema permits features extracted in the early layers of the network to propagate across the network without being perturbed by subsequent layers, so the later layers can use this knowledge. Moreover, the availability of unaltered feature maps from initial layers in the final ones prevents the gradient “vanishing” problem and makes training via back-propagation more efficient [9].

The DenseNet architecture is composed by so called Dense Blocks and Transitions. Figure 2 shows its global architecture.

Transitions consists of a batch normalization layer and an 1x1 convolutional layer followed by a 2x2 average pooling layer [9]. On the other hand, the Dense Blocks are composed by sequences of Batch Normalization, ReLU and Convolutional (BN-ReLU-Conv) layers.

In this work, we use the DenseNet-201 architecture, composed by four Dense Blocks consisting of two BN-ReLU-Conv sequences of 1x1 and 3x3 kernel sizes. The number of filters of each Dense Block is 6, 12, 48 and 32, respectively.

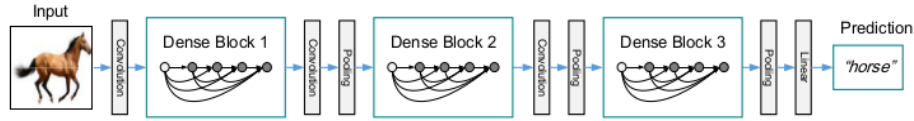


Fig. 2: A deep DenseNet with three Dense Blocks and two transitions (Convolution + Pooling layers).

Furthermore, we added a top model consisting of 5 layers: a dense layer of 512 units with ReLU activation followed by a 30% dropout layer, a dense layer of 128 units with ReLU activation followed by a 30% dropout layer and a final dense layer of 2 units with softmax activation, which is the one giving the predictions.

Fine tuning of existent architectures. While deep learning has revolutionized the field of computer vision, it is sometimes difficult to employ it due to the lack of large datasets, as it requires tens of thousands of examples work effectively. To overcome this limitation, transfer learning [1] has been introduced, in which a pre-trained model is employed for a different task than the one for which it was trained.

2.3 Classification approach

Our approach for classifying the OCT images in healthy (Day 0 or before injection) and pathological (Day 14 after injection) is based on the application of the DenseNet-201 with a customized top model (explained above). Concretely, we first train the network from scratch and then perform fine tuning to compare their behaviour. We implement no pre-processing step, just a resizing of OCT images to 224x224 pixels. The employed hardware is an nVidia Titan Xp GPU with 12GB of GDDR5X memory and 3840 CUDA cores. The network was configured to use the Adam optimizer [10] with a learning rate of 2×10^{-4} , a batch size of 16 images per batch and 40 epochs.

2.4 Evaluation methods and metrics

We use a leave-P-out cross-validation, where P refers to sets of rats mutually exclusive (that is, a single rat appears only in one set). The reason for this kind of cross-validation is to ensure that the training set and the test set contained images from mutually exclusive groups (i.e. there is no single rat which contributed both the training and test set). This decision was made to prevent the network to learn all type of images. Rats are distributed in cages and groups. Thus, a rat labeled as C1G1 identifies the rat 1 in cage 1, the one labeled C1G5 identifies rat 5 in cage 1, and so on. There are a total of 30 sets of rats which vary in the number of OCT images recorded. Then, to perform the validation, 15 folds of 2 sets of rats were made, trying to keep the folds balanced (i.e. with a similar number of healthy and pathological images). Table 1 shows the distribution of the 15 folds performed.

Table 1: Distribution of rat OCT images for the leave-P-out cross-validation. C stands for Cage, G for group, h for number of healthy images, p for number of pathological images and % for the percentage of healthy images.

	F1		F2		F3		F4		F5		F6		F7		F8		F9		F10		F11		F12		F13		F14		F15	
C	19	27	20	29	19	30	23	27	19	29	23	28	20	30	19	27	23	30	20	28	20	29	24	28	24	30	24	23	20	19
G	5	4	2	1	2	1	1	2	1	2	3	3	3	2	4	1	4	4	1	1	5	4	2	2	1	3	3	2	4	3
h	8	32	8	25	11	25	14	24	11	21	19	21	9	20	10	22	12	18	7	27	8	16	18	18	20	13	19	18	10	14
p	37	11	33	10	46	11	54	11	38	10	53	10	25	10	26	11	30	10	17	16	18	10	39	12	41	10	38	30	18	25
%	45.45	43.42	40.00	36.89	40.00	38.83	45.31	46.38	42.86	50.75	46.15	41.38	39.29	35.24	35.82															

3 Results

In this section, we show and compare the results obtained with the network trained from scratch and fine-tuned. All parameters were exactly the same for both experiments, the only difference was that first the experiment were performed randomly initializing the weights of the network, and then we trained the network pre-initializing it with the ImageNet weights (i.e. we fine-tuned it). Figure 3 shows the progress of the training of both networks in terms of validation accuracy. As can be observed, the behaviour of the fine-tuned network is better than the one trained from scratch. The fine-tuned network achieves a validation accuracy in the range of its final accuracy at epoch 30, whereas the network trained from scratch does so at epoch 55 and its variability is higher.

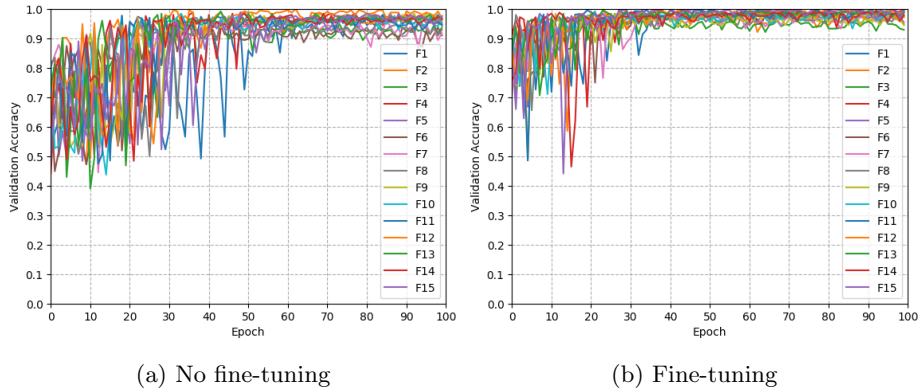


Fig. 3: Progress of the accuracy obtained in the validation set during training for each of the 15 folds.

As for the test set, Figure 4 shows the ROC curve and Table 2 presents the sensitivity, specificity, accuracy and area under the curve (AUC) of the 15 folds performed for the test set.

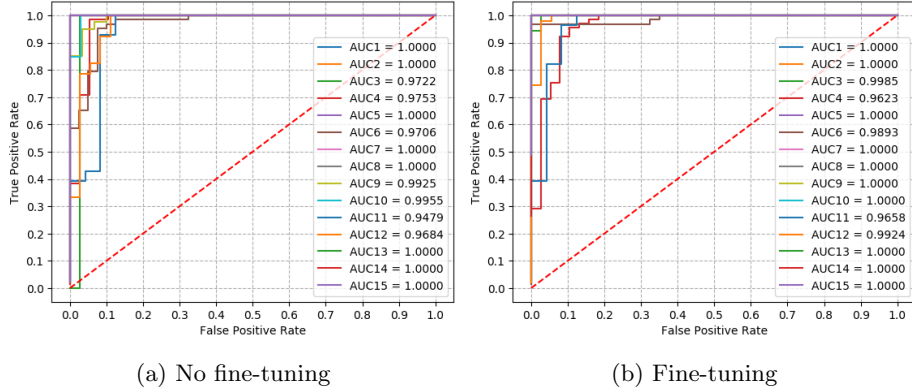


Fig. 4: ROC curve for each of the 15 folds.

Table 2: Metrics of the test set for the 15 folds performed, mean and standard deviation. The metrics presented include: area under the curve (AUC), accuracy (Acc.), sensitivity (Sn), specificity (Sp) and F1-score (F1).

	No fine-tuning					Fine-tuning				
	AUC	Acc.	Sn	Sp	F1	AUC	Acc.	Sn	Sp	F1
F1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
F2	1.0000	0.9342	1.0000	0.8958	0.9180	1.0000	0.9737	1.0000	0.9556	0.9688
F3	0.9722	0.9889	1.0000	0.9818	0.9859	0.9985	0.9778	0.9722	0.9815	0.9722
F4	0.9753	0.9709	0.9730	0.9697	0.9600	0.9623	0.7476	0.5968	0.9756	0.7400
F5	1.0000	0.9875	0.9697	1.0000	0.9846	1.0000	1.0000	1.0000	1.0000	1.0000
F6	0.9706	0.7670	0.6333	0.9535	0.7600	0.9893	0.9320	0.8511	1.0000	0.9195
F7	1.0000	0.9062	1.0000	0.8537	0.8846	1.0000	0.9062	1.0000	0.8537	0.8846
F8	1.0000	0.9710	0.9412	1.0000	0.9697	1.0000	0.9710	0.9412	1.000	0.9697
F9	0.9925	0.9571	0.9655	0.9515	0.9492	1.0000	0.9714	0.9375	1.0000	0.9677
F10	0.9955	0.9104	1.0000	0.8462	0.9032	1.0000	1.0000	1.0000	1.0000	1.0000
F11	0.9479	0.9423	1.0000	0.9032	0.9333	0.9658	0.9423	1.0000	0.9032	0.9333
F12	0.9684	0.8506	0.7447	0.9750	0.8434	0.9924	0.9080	0.8333	0.9778	0.8974
F13	1.0000	0.9524	0.8919	1.0000	0.9429	1.0000	1.0000	1.0000	1.0000	1.0000
F14	1.0000	0.9905	0.9737	1.0000	0.9867	1.0000	0.9810	0.9487	1.0000	0.9737
F15	1.0000	1.0000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
μ	0.9882	0.9419	0.9395	0.9553	0.9348	0.9939	0.9541	0.9387	0.9765	0.9485
σ	0.0162	0.0615	0.1044	0.0528	0.0638	0.0121	0.0635	0.1055	0.0415	0.0668

4 Conclusions

We have presented a method to accurately classify rodent OCT images in healthy or glaucomatous by using the DenseNet-201 architecture and a customized top-model. As can be observed in Figure 3, the fine-tuned approach is faster to converge and has less variability. We achieve a mean $AUC = 0.9939$, with 0.9387 of sensitivity and 0.9765 of specificity. Although this work has been performed on

rodent OCT images, this is a first step to a very promising human OCT screening system, since rodent OCT images are more challenging than human's due to their characteristics and resolution. In future works, we will study the activation of the layers in the network to better understand what characteristics allow it to be accurate, which will suppose important insights in glaucoma understanding. Furthermore, we will try to improve the system by applying some pre-processing steps to the images. Finally, the system will be modified to work with humans.

Acknowledgments. 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 Xp GPU used for this research. This work was supported by the Project GALAHAD [H2020-ICT-2016-2017, 732613].

References

1. Karri, S., Chakraborty, D., Chatterjee, J.: Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. *Biomedical optics express* **8**(2) (2017) 579–592
2. Pekala, M., Joshi, N., Freund, D.E., Bressler, N.M., et al.: Deep learning based retinal oct segmentation. *arXiv preprint arXiv:1801.09749* (2018)
3. Srinivasan, P.P., Kim, L.A., Mettu, P.S., et al.: Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomedical optics express* **5**(10) (2014) 3568–3577
4. Lee, C.S., Baughman, D.M., Lee, A.Y.: Deep learning is effective for the classification of oct images of normal versus age-related macular degeneration. *arXiv preprint arXiv:1612.04891* (2016)
5. Muhammad, H., Fuchs, T.J., De Cuir, N., De Moraes, C.G., et al.: Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *Journal of glaucoma* **26**(12) (2017) 1086–1094
6. Virgili, G., Michelessi, M., Cook, J., Boachie, C., et al.: Diagnostic accuracy of optical coherence tomography for diagnosing glaucoma: secondary analyses of the gate study. *British Journal of Ophthalmology* (2017) *bjophthalmol-2017*
7. Hood, D.C.: Improving our understanding, and detection, of glaucomatous damage: an approach based upon oct. *Progress in retinal and eye research* **57** (2017) 46–75
8. Nagata, A., Omachi, K., Higashide, T., et al.: Oct evaluation of neuroprotective effects of tafluprost on retinal injury after intravitreal injection of endothelin-1 in the rat eye. *Investigative ophthalmology & visual science* **55**(2) (2014) 1040–1047
9. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Volume 1. (2017) 3
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)