
Web Applicable Computer-aided Diagnosis of Glaucoma Using Deep Learning

Mijung Kim^{*†}
Jasper Zuallaert^{*†}

Olivier Janssens^{*}
Sofie Van Hoecke^{*}

Ho-min Park^{*†}
Wesley De Neve^{*†}

[†]Center for Biotech Data Science, Ghent University Global Campus

^{*}IDLab, Department of Electronics and Information Systems, Ghent University
mi.jung.kim@ugent.be

Abstract

Glaucoma is a major eye disease, leading to vision loss in the absence of proper medical treatment. Current diagnosis of glaucoma is performed by ophthalmologists who are often analyzing several types of medical images generated by different types of medical equipment. Capturing and analyzing these medical images is labor-intensive and expensive. In this paper, we present a novel computational approach towards glaucoma diagnosis and localization, only making use of eye fundus images that are analyzed by state-of-the-art deep learning techniques. Specifically, our approach leverages Convolutional Neural Networks (CNNs) and Gradient-weighted Class Activation Mapping (Grad-CAM) for glaucoma diagnosis and localization, respectively. Quantitative and qualitative results, as obtained for a small-sized dataset with no segmentation ground truth, demonstrate that the proposed approach is promising, for instance achieving an accuracy of 0.91 ± 0.02 and an ROC-AUC score of 0.94 for the diagnosis task. Furthermore, we present a publicly available prototype web application that integrates our predictive model, with the goal of making effective glaucoma diagnosis available to a wide audience.

1 Introduction

The revival of convolutional neural networks (CNNs) and the public availability of large-scale datasets like ImageNet [1] have led to significant performance improvements in computer vision. Unlike its predecessors, ImageNet contains more than one million natural images belonging to more than thousand classes, contributing substantially to the successful deployment of CNNs. Indeed, ImageNet for instance allows constructing CNN-based predictive models that are more accurate and that better generalize for the tasks of image segmentation and classification. Moreover, CNN-based predictive models have shown to be highly successful in medical image analysis [2–4]. For diabetic retinopathy [4], these models even managed to obtain a diagnosis accuracy that is as high as the diagnosis accuracy of human experts. However, most deep learning-based approaches towards medical image analysis face two major challenges. First, few medical image datasets are available with a volume sufficient to train a deep neural network from scratch, especially given that, by convention, the deeper the neural networks constructed, the better the predictions (i.e., the higher the classification accuracy). As a result, if not enough data are available, then the neural networks constructed may suffer from overfitting. Second, apart from labels for the purpose of classification, most medical image datasets do not have proper annotations available for the purpose of segmentation, given the time-consuming and error-prone nature of annotation efforts.

In our research, to tackle the two aforementioned challenges, we built a predictive model for computer-aided diagnosis (CAD) and localization of glaucoma, leveraging eye fundus images that are widely

used in present-day hospitals, given that these images can be acquired at a relatively low cost. Additionally, based on our CAD model, we developed a novel tool for glaucoma diagnosis that takes the form of a prototype web application¹. The main contribution of this paper stems from the novelty of our predictive model and its integration into a prototype web application. Indeed, although plenty of research efforts have been presented in the scientific literature on computer-aided diagnosis using fundus images, to the best of our knowledge, no previous research efforts have simultaneously targeted both diagnosis and localization using a dataset with no segmentation ground truth.

The rest of this paper is organized as follows. We explain our approach in Section 2. We subsequently discuss our experimental setup and results in Section 3. Finally, we conclude our research in Section 4. *Note that some of the research activities described in this paper have also been presented at The Second International Workshop on Deep Learning in Bioinformatics, Biomedicine, and Healthcare Informatics (DLB2H 2018).*

2 Architecture

Convolutional Neural Network. Introduced by Simonyan and Zisserman at Oxford’s Visual Geometry Group (VGG), VGG networks [5] are currently some of the most widely used CNN architectures. Among these different architectures, VGG16 and VGG19 are the most popular ones, thanks to their high classification accuracy. In our research, we decided to work with VGG16 over VGG19, given that the difference in effectiveness between the two is not significant, while VGG16 has less parameters to learn than VGG19. However, as mentioned in the introductory section, VGG16 may suffer from overfitting when directly applying this architecture to our eye fundus images. As such, to mitigate overfitting and to increase training effectiveness and efficiency, we applied a number of changes. To facilitate better model regularization, we introduced two techniques. First, we added dropout layers (ratio: 0.5) at the end of each convolutional block. Second, next to the use of default data augmentation techniques (that is, random cropping and random horizontal flipping), we also incorporated data augmentation techniques such as random brightness (using $\sigma \pm 0.8$). Furthermore, we adopted the adaptive moment estimation (ADAM) optimizer [6] instead of Stochastic Gradient Descent (SGD), allowing us to handle multiple drawbacks of SGD more efficiently, like a slow convergence and a high variance in parameter updates. Lastly, we used a pre-trained VGG16 model, so to be able to avoid a time-consuming weight initialization process and to considerably reduce the training time [7]. From an architectural point-of-view, we also applied a few more changes to the vanilla VGG16 network. First, we used fully convolutional layers instead of fully connected layers. Second, in order to meet our classification needs, the number of channels in the last fully convolutional layer was reduced from 1000 to 2.

Grad-CAM. Most sets of medical images do not have a ground truth (that is, golden standard annotations) for supervised segmentation or localization. As a result, localization needs to be treated as an unsupervised learning problem. A couple of deep learning-based approaches are frequently used to segment images in an unsupervised fashion. One such approach is Class Activation Mapping (CAM), as proposed by [8]. CAM allows visualizing what CNNs look at when they classify an input image, thus making it possible to visually understand why a CNN chooses one class over another. More recently, another algorithm was published in [9], called Gradient-weighted Class Activation Mapping (Grad-CAM), improving upon CAM. In particular, Grad-CAM uses the class-specific gradient information flowing into the final convolutional layer of a CNN to produce a coarse localization map of the important regions in a given input image. In our research, this localization process is activated only if the class is predicted as glaucomatous.

3 Experiments

Dataset: fundus images. The total number of eye fundus images in our training dataset is 1,080, among which 220 images were used for validation purposes. Another 110 eye fundus images unseen during training and validation were used for testing purposes. These images, with the horizontal resolution varying from 1,172 to 2,500 pixels and with the vertical resolution varying from 1,500 to 3,200 pixels, have been collected at Samsung Medical Center in 2016. All images were labeled by an experienced ophthalmologist. The proportion of normal and glaucomatous images for both

¹www.medinoid.org

training and testing was 1:1 (that is, the dataset used was perfectly balanced). This study followed all guidelines for experimental investigation in human subjects, was approved by the Samsung Medical Center Institutional Review Board, and adhered to the tenets of the Declaration of Helsinki.

Implementation. Our approach towards glaucoma diagnosis and localization was implemented in Python 2.7, using Tensorflow 1.7 and the , pre-trained VGG16 model over ImageNet from Tensorflow Slim². Furthermore, our approach was executed using two Intel(R) Xeon(R) E5-2620 2.4GHz CPUs and an NVIDIA GeForce GTX TITAN X GPU. Before feeding the input images to the network, they were center-cropped to remove excessive information not related to the classification task at hand (e.g., removal of a black background) and resized to $224 \times 224 \times 3$. In addition, for image normalization purposes, we changed the RGB means for our dataset as follows: 105.51, 54.52, and 16.19 (in RGB order). After changing the last layer of the pre-trained model into fully convolutional layer, the weights of all pre-trained layers were fine-tuned over our dataset. During fine-tuning, we used a batch size of 32, a learning rate of 0.0001, and data augmentation. We made use of 5-fold cross-validation for our experiments. The metrics used for evaluating the classification task are (1) accuracy, (2) recall, (3) precision, (4) F1 score, and (5) Receiver Operating Characteristic - Area Under the Curve (ROC-AUC).

Experimental results. Considering its effectiveness, early stopping is one of the most commonly used regularization methods in practice [10]. Thus, as shown in Figure 1, we early stopped the training at around epoch 80, coming with a validation loss that is the lowest and that is about to rise again, even though the training loss is still decreasing over time. Apart from an accuracy of 0.91 and an ROC-AUC score of 0.94, our model achieved a recall of 0.91, a precision of 0.92, and an F1 score of 0.91. As mentioned in Section 1, no method is currently available that simultaneously tackles classification and localization. Moreover, a direct comparison with already existing approaches is difficult, given a lack of publicly available datasets. In Table 1, some information can be found about the difference in effectiveness of our approach, as obtained for the dataset made available by Samsung Medical Center, and the effectiveness of the deep learning-based approach presented in [11], as obtained for the RIM-ONE dataset [12], for the classification task only.

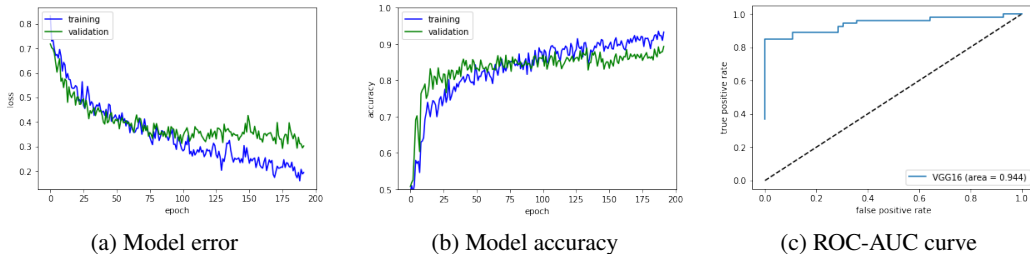


Figure 1: Model evaluation.

Table 1: Comparison with another deep learning model, using different datasets.

Model	Dataset	# positives	# negatives	Input size	Accuracy	AUC
[11]	RIM-ONE [12]	200	255	$227 \times 227 \times 3$	0.88	-
Proposed	SMC	540	540	$224 \times 224 \times 3$	0.91	0.94

Apart from the use of two different datasets, thus making a direct comparison impossible, we hypothesize that some of the differences in effectiveness can be attributed to the CNN architecture used and the way data augmentation has been used. In particular, [11] employs AlexNet, which was reported to be less effective for the ILSVRC classification task than VGG16 [5]. Furthermore, the approach of [11] did not apply any data augmentation technique. As a result, given the limited size of the dataset used, we believe chances are high that the model of [11] did not sufficiently generalize.

Regarding localization, since this paper presents the first attempt to localize glaucoma without a segmentation ground truth, there are no proper baselines to compare against. Therefore, we asked

²<https://github.com/tensorflow/models/tree/master/research/slim>

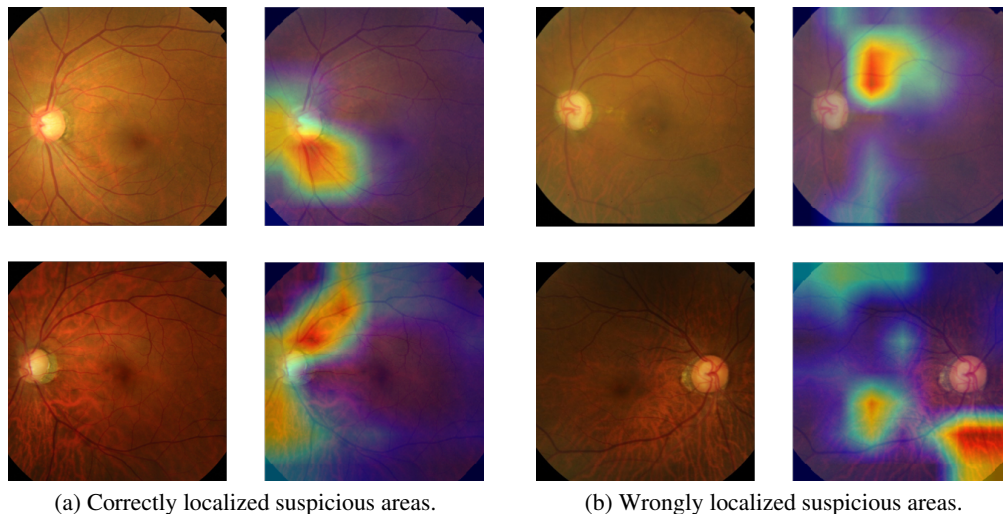


Figure 2: Localization results.

for the opinion of a clinician at Samsung Medical Center. That way, we found that most of the glaucomatous images were correctly localized, highlighting the corresponding suspicious area, as shown in Figure 2. However, we found that localization was also wrong for a few test examples (e.g., due to blur), thus pointing to the need for a more rigorous computational and clinical analysis. Lastly, as illustrated by Figure 3, we built a publicly available prototype web application that integrates our predictive model, with the goal of making glaucoma diagnosis available to a wide audience.

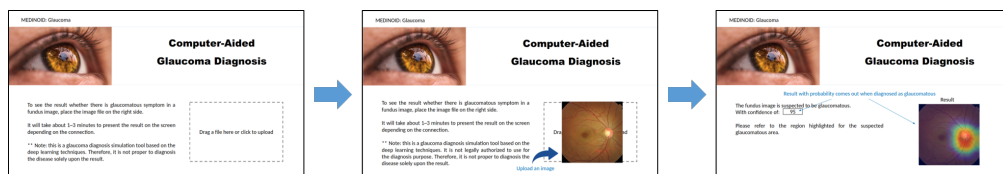


Figure 3: Prototype web application for glaucoma diagnosis. Users can provide this application directly with an input image using a drag and drop approach, as shown by the figure in the middle. Furthermore, the prototype web application shows the probability of having glaucoma, highlighting the suspicious region found.

4 Conclusions and Future Work

In this paper, we introduced a predictive model for computer-aided diagnosis of glaucoma, leveraging CNNs and Grad-CAM for diagnosis and localization, respectively. Our model has been developed by making use of small-sized dataset of fundus eye images, only coming with annotations for the diagnosis task, not having annotations for the localization task. Furthermore, we integrated our model into a publicly available prototype web application, with the ultimate goal of making effective glaucoma diagnosis available to a wide audience in a user-friendly way.

Acknowledgments

The research and development efforts described in this paper were funded by Ghent University, Ghent University Global Campus, Flanders Innovation & Entrepreneurship (VLAIO), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union. These efforts were also supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF), funded by the Ministry of Education (grant: 2017034834).

References

- [1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255, IEEE, 2009.
- [2] 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*, pp. 234–241, Springer, 2015.
- [3] L. Heinrich, J. A. Bogovic, and S. Saalfeld, "Deep learning for isotropic super-resolution from non-isotropic 3D electron microscopy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 135–143, Springer, 2017.
- [4] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [6] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [7] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, pp. 242–264, IGI Global, 2010.
- [8] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*, pp. 2921–2929, IEEE, 2016.
- [9] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?," *arXiv preprint arXiv:1611.07450*, 2016.
- [10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [11] B. Al-Bander, W. Al-Nuaimy, M. A. Al-Tae, and Y. Zheng, "Automated glaucoma diagnosis using deep learning approach," in *Systems, Signals & Devices (SSD), 2017 14th International Multi-Conference on*, pp. 207–210, IEEE, 2017.
- [12] F. Fumero, S. Alayón, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation," in *2011 24th international symposium on computer-based medical systems (CBMS)*, pp. 1–6, IEEE, 2011.