Christopher Chan (004282878)
Joseph Zhou (604171655)

Grading Severity and Visualizing Saliency in Diabetic Retinopathy with an Inception Network: Literature Review

Diabetic retinopathy (DR) is a major cause of vision impairment in diabetic patients. In recent years, 28.5% of the adult U.S. population with diabetes contracted the complication, which placed 4.2 million individuals at risk of becoming forever blind [1]. It is thus crucial to be able to detect DR during onset to prevent such potentially extreme consequences of diabetes. We propose a solution to this archetypal machine learning classification problem with an Inception-based neural network and a complementary network visualization with salience maps for transparency and insight into our model's learning.

The primary method of detecting DR is manual examination from a trained clinician. In an effort to save time and resources, large fundus image datasets have been provided to researchers and data science competitors to develop ways to automate DR detection. A cutting-edge and relatively novel approach to solving this problem is building deep-learning neural networks. Google recently developed such a network and named their architecture Inception. When employed against the model that the ILSVRC2012, an object detection and image classification competition, winners used, Inception was able to achieve significantly more accurate classification results with merely 1/12 of the parameters to train the network [2]. This directly translates to less memory and computational resources, showing that it is possible to train more powerful networks on smaller budgets. Google also applied a form of their network to auto-classify into 2 classes DR fundus images based on the severity of the developing complication as presented in the image [3]. In a similar effort, we employ this architecture against the same problem but with 5 classes to efficiently diagnose patients' DR development into more possible stages of severity.

With the growing use of neural networks, there has also been a growing trend to make sense of these models via visualizations. This sub-field of explainable artificial intelligence is quickly emerging in an attempt to provide insight on how these neural networks are learning. This is valuable because by understanding what the computer finds important, the network can be retrained to ignore false or highlight missed, in this case, DR regions of interest. Zeiler *et al.* [4] proposed that deconvolutional neural networks (DNN) can be used for such insightful visualizations. DNNs can be thought of as reverse CNNs in that input signals to layers are estimated and constructed from feature maps. This idea of feature map reconstruction, however, can be achieved and generalized outside of convolutional layers with a gradient-based approach as Simonyan *et al*. [5] shows. This technique can be used to visualize the output grades of the final layer of the networks into salience maps which depict what the network deems important in a given image when classifying it into any given class. These salience maps are, in essence, what the model finds important in an image by its own global metric. For our network, we generate such salience maps to visualize what our network deems important for the DR severity classification task. Simonyan *et al.* also proposes a way actually numerically-generate class models [5] after the training phase for a network is complete. This technique provides illustrations of what the network understands for each class. In the case of the ILSVRC2013 dataset, images of a dumbbell and a goose, among many others, were depicted by creating an artificial image to maximize the class output score [5]. Such visualizations could aid in developing DR classification networks by understanding what the networks understands as severity indicators and after training it to produce better results.

Introducing machine-learning techniques to diagnose DR in diabetic patients is the key in effectively treating them while minimizing clinical resources for examination. Being able to visualize models to understand and improve them after for maximum precision is essential for such a health-related endeavor. Once accomplished, though, along with the right treatment and prevention measures, diabetic patients should expect to see little to no DR in their lives.

References

1. Department of Health and Human Services. Center for Disesase Control and Prevention. (2014). National Diabetes Statistics Report, 2014.
2. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. arXiv:1409.4842 [cs.CV]. 2014.
3. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016;316(22):2402-2410. doi:10.1001/jama.2016.17216
4. Zeiler, M. D., Krishnan, D., Taylor, G. W., & Fergus, R. (2010). Deconvolutional networks. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2010 (pp. 2528-2535). [5539957] DOI: 10.1109/CVPR.2010.5539957
5. Simonyan K, Vedaldi A, Zisserman A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv:1312.6034v2 [cs.CV] 19 Apr 2014.