Visualizing the Indicators of Diabetic Retinopathy Learnt by Convolutional Neural Networks

Shikhar Srivastava[b], Siddarth Pratapneni[a], Sidharth R[a]*, Ashwin Abraham[a], Dr. Srikanth Prabhu[b], Dr. Sulatha V Bhandary[a]

[a] Kasturba Medical College, Manipal, Udupi, Karnataka 576104

[b] Manipal Institute of Technology, Manipal, Udupi, Karnataka 576104

Corresponding author Email: **tornadoalert@gmail.com*

Select Department: Technical Science and Health Science

This study proposes a novel application of visualizing features learnt by convolutional neural networks with the aim to further the understanding of Diabetic Retinopathy. A convolutional neural network is first trained to recognize and classify fundus images of diabetic and non-diabetic patients. The network is then visualized, using a technique of pixel optimization, to discover the features that the trained network looks for to classify the image. Through this novel application of network visualization, we show that critical features for diabetic retinopathy can be re-discovered, leaving great scope for its application in scarcely explored diseases under minimal resource constraints.

Keywords: Convolutional Neural Networks, Diabetic Retinopathy, Inceptionism, Image Classification, Artificial Intelligence, Deep Learning, Computer Vision.

1. Introduction

Diabetes is a disease that has a major prevalence in our population. One of the complications resulting from the disease is diabetic retinopathy (DR).

Elevated blood glucose levels (hyperglycemia), leads to weakening of the retinal blood capillaries. This can lead to outpouching of the capillaries at these points (microaneurysms). Such outpouchings can leak fluids and blood into the retinal tissue causing it to swell, leading to blurred, obscured vision. If untreated, this condition can escalate to blindness.

A commonly used method of diagnosing DR is by analyzing fundus images of the eye and classifying the condition, if present, into one of the following categories[1]: Mild, Moderate or Severe Non-Proliferative Diabetic Retinopathy and Proliferative Diabetic Retinopathy. [2]

Usually, an ophthalmologist analyzes these images and makes a diagnosis based on knowledge gained from previous studies. With the increasing capability of artificial intelligence, more specifically, Convolutional Neural Networks (CNNs) [3], this process of identifying and deciding if fundus scans are indicative of Diabetic Retinopathy, can be tasked to such a network. There have been many studies that have aimed to determine whether or not fundus images showed signs of Diabetic Retinopathy using computer vision. Some take the extra step to classify DR indicative images into one of the previously mentioned categories.[4]

CNNs have been long thought of as a 'black box', meaning it wasn't understood how the CNN works inside. More specifically, what features the artificial neurons look for to result in the classification of the image. Recently, with pioneering work that makes it possible to peer into the workings of a CNN by creating visualizations to maximize the activity of a certain class we can understand what the networks learn[5]. However, these visualizations were crude and did not show the features properly. In 2015, Google developed another method referred to as "Inceptionism". [6] This method used a pre-existing image and highlighted the features identified by the CNN for that image. The program did this by first identifying features within the image using a trained CNN, after which these features are exaggerated slightly. This new image is then fed back for a number of iterations leading to exaggerations of the important, classifying features as considered important by the trained CNN within the given image.

The aim of this study is to use this technology along with a library of classified DR images to first, train a CNN to recognize diabetic and non-diabetic images and secondly, visualize the trained CNN, to search for any new indicators of DR that may not have been used by physicians before to diagnose the disease. These factors may help in increasing the accuracy of diagnoses and allowing detection at an earlier stage.

2. Experimental Details

Collecting the Data and Preprocessing

Data for this experiment was drawn from a dataset provided via Kaggle and maintained by EyePacs. The data included several thousands of fundus images of the eye representing each stage of DR. From this reserve, we selected a couple thousand at random, which were free of artifacts, and equally represented all the stages of DR under study.

The dataset was prepared for three classes namely Normal, Moderate DR and Proliferative DR. The images were subsequently preprocessed by cropping black borders, local contrast normalization, and downscaling to make computation easier. With the aim to achieve a balanced dataset with sufficient variability, the dataset was augmented by randomly rotating images with 10% probability.

Training a CNN

The Google Inception v3 model, trained on the ILSVRC dataset [7], with over ten million annotated images, was fine-tuned on the Retinal dataset using Transfer learning [8]. To achieve this, the final layer of the network was removed, and replaced with a three node fully-connected softmax classifier, corresponding to the three-class dataset of Retinal images.

The weights and structure of the initial convolutional layers were fixed while those of the higher layers, including the layers preceding the classifier, were allowed to be fine-tuned. This is motivated by the observation that lower convolutional layers learn to identify generic features like edges and color blobs, that can be effectively reused [8]. Allowing the higher layers to be

fine-tuned with the softmax classifier, ensured that the network learned the higher abstractions specific to the current Retinal images, with sufficient inherent bias to prevent overfitting.

The model was trained for 4000 iterations in batches of 100 images, with a learning rate of 0.01, aimed to minimize the cross-entropy error of the softmax classifier.

10% of the dataset was used as the validation set during training, while the final model was evaluated on a test set of 200 randomly selected images. [Figure 1]

Visualization

The optimal model with its learned weights was saved after training, to perform the visualizations. The visualization technique used in this paper was introduced by Google as "Inceptionism" [9]. By iteratively optimizing the pixels of the input image for maximum activation of a specific layer of the network, we obtain a visualization of the network's interpretation of the image.

For a neuron i in the network, the activation Gi(x) is maximized for the input image x, by iteratively modifying the input image x as, $x \leftarrow x + \alpha * [\partial Gi(x)/\partial(x)]$.

The input pixels were each modified as a linear function of the gradient to the image until the subsequent input pixels produced maximum activation of the neurons in the specified layer of the network.

Along with the prior that the modified image must be a natural image, with high correlation among neighboring pixels, the pixels of the input image were optimized.

The obtained high-frequency image was iteratively normalized by the laplacian pyramid gradient normalization method [9], where a smoothness prior was added to the optimization function over several frequency octaves of the image.

This resultant normalized image was a visualization of the network's interpretation of the input image fed to it. Several retinal images were similarly fed as input to the network as above, and the subsequent visualizations were documented and interpreted.

3. Results and Discussion

The trained network was able to classify images from a randomly selected test set of 200 images, with an accuracy of 80%. The visualizations gave us an insight into what it had learned. Optimizing random pixels for highest activation of the Diabetic Retinopathy node gave us an image with artifacts that looked like blood vessels and small pinhead sized spots that resemble hard exudates. Optimizing the pixels of a fundus image to amplify the features of Diabetic Retinopathy provided better insight into what the network had learned.

Based on the "DeepDream" visualization of Diabetic Retinopathy, it was found that the key features that determined the activation of the Diabetic Retinopathy node were very similar, if not the same as what doctors currently use today to identify the disease.

A few images have been listed below, demonstrating how the network was able to enhance pre-existing indicators of DR in a fundus image. It was also able to add features to apparently normal fundus images, such that the resulting image activated the DR neuron in the softmax layer of the network. The features it added are pointed out [Figure 3-5] and closely represent indicators of DR currently used for diagnosis, i.e. microaneurysms, hemorrhages, neovascularization, hard exudates, venous changes, etc.[10]

Since some convolutional networks are better at object recognition than humans [11], it was expected that the network would learn new indicators of diabetic retinopathy. The trained network was successful at creating visualizations of many indicators of DR, that are currently employed clinically. It was able to learn several textbook features of DR, including hard exudates, microaneurysms, dot and blot hemorrhages, etc. Although the network didn't point towards any new indicators, the current model can significantly increase the speed and accuracy of diagnosis of Diabetic Retinopathy, especially at early stages when it is treatable.

The fact that the network has learned the current clinically relevant features of diagnosis from scratch is extremely significant, as the implications stretch across the field of medicine. This serves as a proof of concept that a Deep Convolutional Neural Network with sufficient labeled data and training time could learn the features of any particular disease, providing visualizable insights in a matter of days, which medical professionals could only accomplish after years of research. The visualizations enhanced certain blood vessels and elucidated many other fine changes in the fundus images. It might have significant practical implications if these 'enhanced' images can serve as a replacement for the fluorescein angiography, an invasive procedure done to visualize leaks and more subtle changes in vasculature which cannot be detected by the naked eye.

4. Conclusion

The neural network was able to recognize images with Diabetic Retinopathy with an accuracy of nearly 80%. The program was able to differentiate normal retinal images from diseased ones by looking for the presence of certain indicators which are similar to what doctors use today. The same was confirmed by an unbiased opinion from senior ophthalmologists, from a tertiary care center. The trained neural network was successfully able to differentiate between the various stages of DR. The visualizations provided valuable insight into the inner workings of the network, highlighting the key differentiating factors identified by the model. Studying these differentiating factors and correlating them to known indicators provides a deeper understanding of the pathology of the disease. Similarly, this technique can be applied to far more complex medical data, where existing knowledge and understanding is scarce. This approach can greatly

accelerate the advancements in the medical field and allow us to take larger steps towards better diagnoses.

Acknowledgments

We would like to thank Dr. Krishna Rao, KMC for his expert advice in the field of ophthalmology. Special thanks to Dr. Ramesh and Dr. Uma Maheswari, Aarthy Eye Hospital for their valuable input.

References and Notes

1. C. P. Wilkinson et al., Ophthalmology, vol. 110, no. 9, pp. 1677–1682, **(2003)**
2. Early Treatment Diabetic Retinopathy Study Research Group., ETDRS report number 10, Ophthalmology, vol. 98, no. 5 Suppl, pp. 786–806, May **(1991)**
3. Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, 13th Int. Conf. Control. Autom. Robot. Vis., vol. 2014, no. December, pp. 844–848, **(2014)**
4. California-Healthcare-Foundation, **(2015)**
5. J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, Int. Conf. Mach. Learn. - Deep Learn. Work. 2015, p. 12, **(2015)**
6. A. Mahendran and A. Vedaldi, **(2015)**
7. O. Russakovsky et al., Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, **(2015)**
8. Pan, Sinno Jialin, and Qiang Yang. IEEE Transactions on knowledge and data engineering 22.10, 1345-1359.**(2010)**
9. Mordvintsev, Alexander, Christopher Olah, and Mike Tyka. Google Research Blog. Retrieved June 20 **(2015)**
10. Wilkinson, C. P., et al. Ophthalmology 110.9, 1677-1682 **(2003)**
11. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; The IEEE International Conference on Computer Vision (ICCV), pp. 1026-1034 **(2015)**

Figure Captions

Figure 1. Accuracy of Network with iterative training. (Blue: Validation set; Orange: Training set)

Figure 2. Optimization of random pixels, synthesized by the trained CNN

Figure 3. Neovascularization (a), Venous Constriction (b), and Hard Exudates (c) shown by the network

Figure 4. Hemorrhage (a) and Hard Exudate (b) as synthesized by the network on normal fundus image

Figure 5. Microaneurysms (a,b) and Hard Exudates (c) as synthesized by the network on normal fundus image

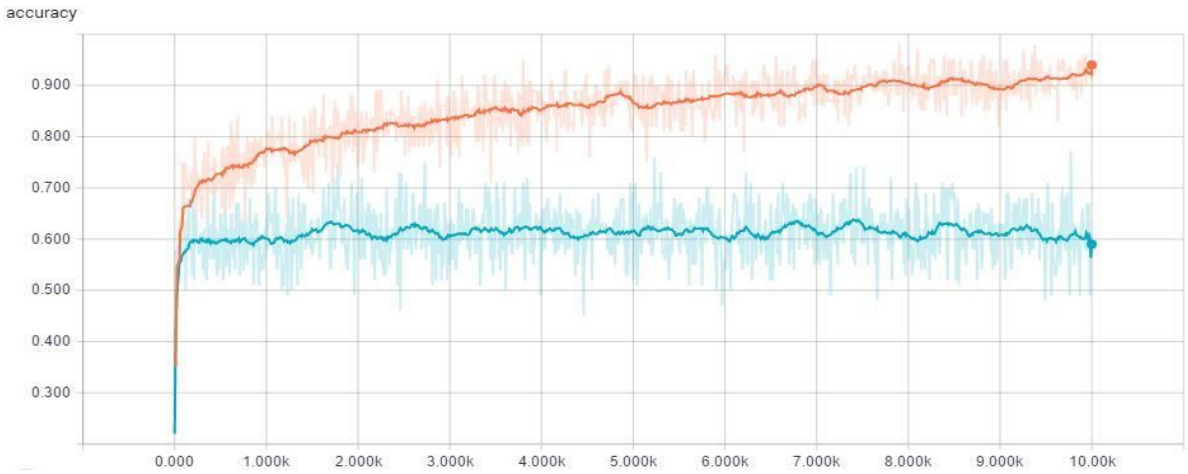

Figure 1

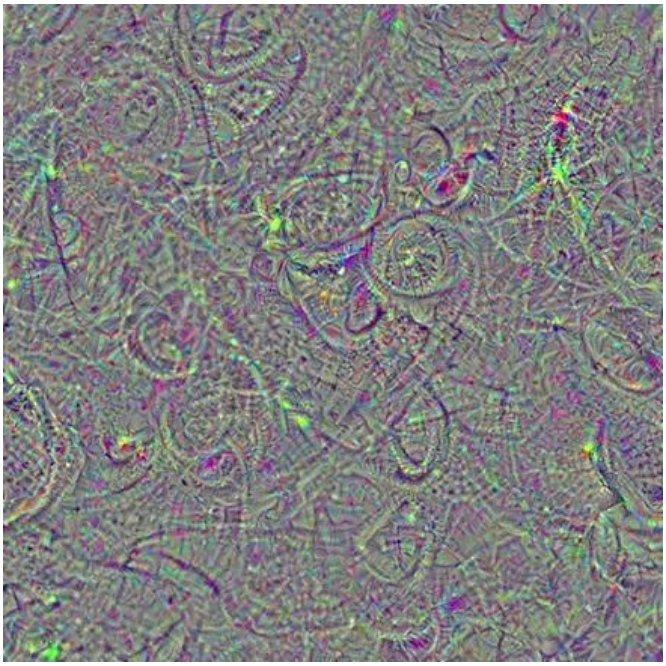

Figure 2

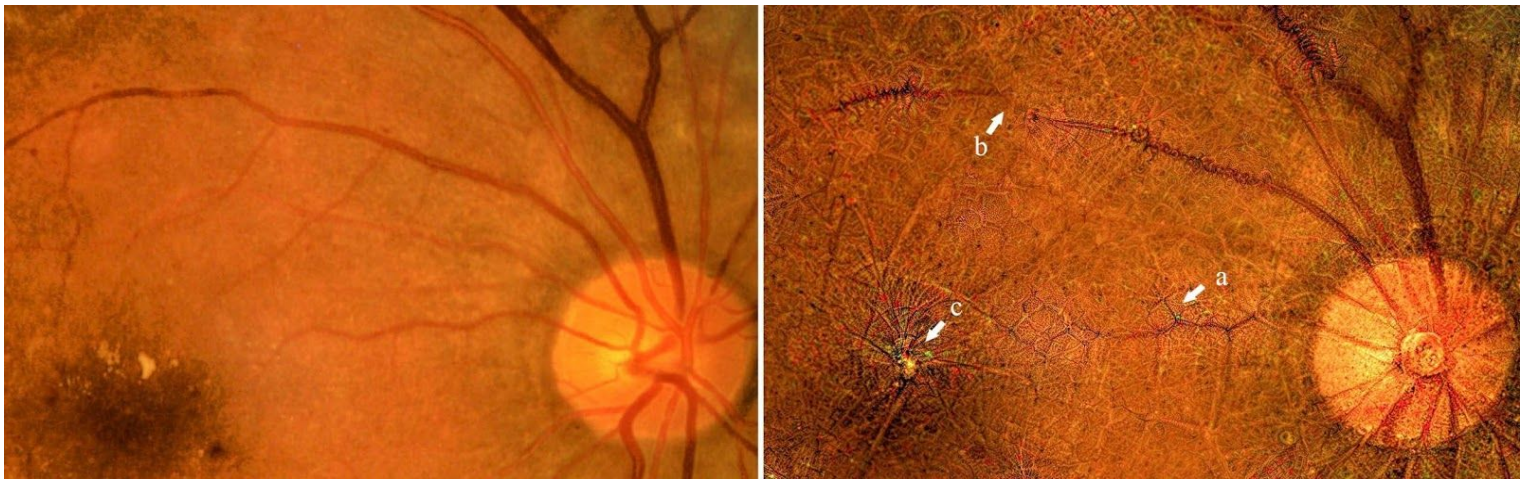

Figure 3

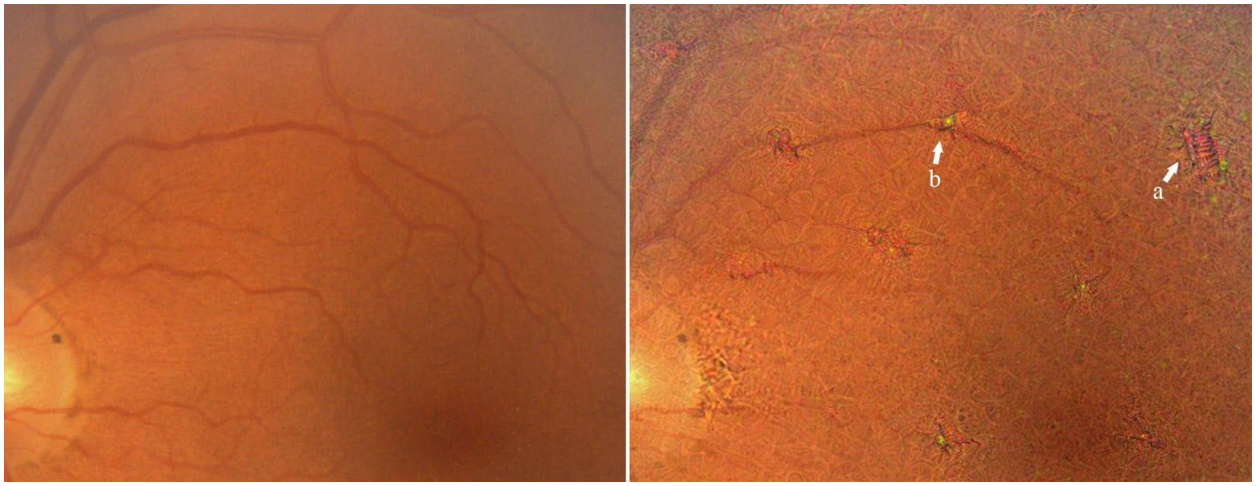

Figure 4

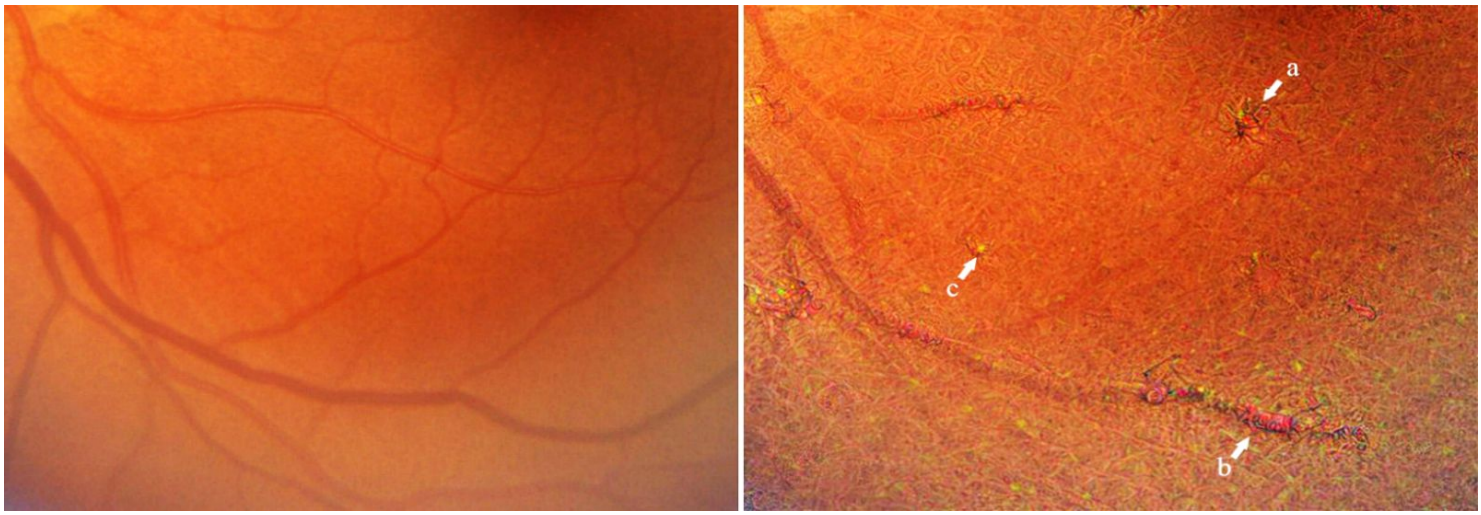

Figure 5