Grading Severity and Visualizing Saliency for Diabetic Retinopathy with an Inception Network

Christopher Chan (004282878) Joseph Zhou (604171655)

1 Introduction

We propose a solution to diagnose Diabetic Retinopathy (DR) severity with an Inception-based [1] convolutional neural network (CNN) and additional salience map visualizations for insight into our model's learning.

2 Method

Data

There are 5 different severity levels for DR classification according to ICO[2]. We obtained 88702 color fundus images from EyePac for training the CNN model. Each image has been graded by 3 certified readers and only unanimously-agreed upon results are taken.

Data Pre-processing

We pre-processed the images so that every image is truncated off of the redundant black area and resized to 800x800 pixels (Figure 1). During training, in order to improve training data size, reduce overfitting and therefore improve model performance, we augmented the images on the fly. We used image augmentation techniques, including shear, zoom, rotation, shift, and flip etc.

Model

We designed our CNN model based on Google's Inception[1] V3 architecture. We take all the layers from the input layer to 'mixed3' layer, including first 6 convolutional layers and the following 4 mixed layers('mixed0' to 'mixed3'). Every "mixed" layer is a concatenation of several different convolution layers with different kernel sizes. This design helps the model to learn features of different sizes quickly. Then we added 2 fully connected layers before the final output layer, which has 5 nodes corresponding to 5 DR classifications. The final output layer uses softmax as the activation function. All other layers uses Relu as the activation function. For every layer we used batch-normalization before activation.

Training

We initialized the weight with imagenet weight. Then we tried various optimizers. The upper layers first were trained first before training lower layers. Also since the data distribution of the classes is very disproportional, with the class 0 having more than 50k images while class 3 and 4 only have around 4k images, we oversampled the small class data to achieve the balance. Moreover, we used curriculum learning by designing an easier loss function first and before switching to the harder one.

Visualizing

To visualize the model, we used two approaches. First, we implemented a salience map construction as proposed by Simonyan *et al.* [3] in which the gradient of a specific class score with respect to each input pixel was determined. The maximum absolute value of each pixel across all channels (RGB) was selected as the salience value and overlayed on top of the original image. The

second method tested the occlusion sensitivity of the model as proposed by Zeiler *et al*[4]. To test whether or not the model is truly picking out signatures of certain DR severity classes, we occluded small patches from the original image. A score was obtained with this modified copy, and it was mapped onto the original image where the occlusion was. This process was repeated for different locations throughout the image and a class-specific heatmap was generated.

3 Evaluation

To evaluate the model, we used 3 metrics - accuracy, sensitivity, and specificity. Accuracy is defined as the amount of predictions exactly matching the human grading. Sensitivity and specificity are defined as the percentage of true negatives over all negative samples and the percentage of true positives over all positive samples respectively. Since there are five classes, we have multiple ways to define positive and negatives. We adopted 1 or above and 3 or above as the positives. We used 20% of the whole data as the test dataset, which amounts to 17740 images. To evaluate visualizations, we checked intuitively if the salient points of the blended images were in fact marked according to what should be clinically highlighted when diagnosing DR severity in patients.

4 Results

The results of our classifier based on the evaluation metrics mentioned are summarized in Figure 5. We were able to successfully create salience maps for the class and image-specific score gradient visualizations of our network; they can be seen in Figure 3. The result of the class and image-specific occlusion-based salience map can be seen in Figure 4.

5 Discussion

We found ‘Adam’ optimizer helps the learning converges faster than other optimizers like stochastic gradient descent. We found different initial learning rates are better to be big at the beginning training and small at later training stages. Besides, we found the network hard to converge if directly training the whole network, so we trained several top layers first before training the entire network. Regarding the sensitivity and specificity, since we have 5 classes, we need to group multiple classes as one class to proceed calculations. We focused on the difference between class 0 and above, and class 2 and above. We chose so because these two differences are more important that others. Firstly, class 0 is the normal class so it’s useful to find out whether a patient is normal or not. Secondly if the severity is beyond class 2 a patient needs very immediate treatment, but if not the patient don’t need any treatment [2].

The gradient-based visualization (Figure 3) suggested that the model needs to be further trained with images indicative of healthy retinas with blood vessel shapes or inherent retinal characteristics that can be confused with signs of DR development. As for the occlusion-based visualization (Figure 4), there were no noticeable salient patches visualized. While this could be attributed to code error or parameter misuse (such as occlusion patch size), what this may indicate is that the model is considering not specific areas that suggest a DR severity, but rather it is weighing the context of these areas more to classify the image, as mentioned by Zeiler *et al*[4]. Thus, the occlusion-based salience map may further suggest more explicit and distinct class data is needed for learning or even hyperparameter layer-level changes for the model to overcome this problem.

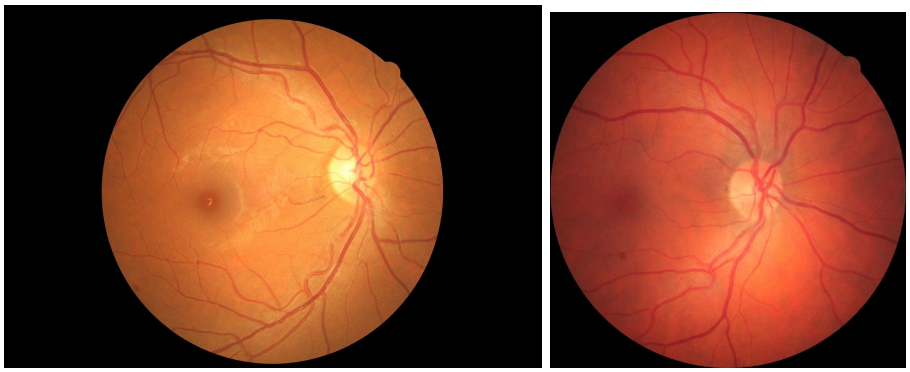

Figure 1. The left image is a non-preprocessed image. The right image is a pre-processed images.

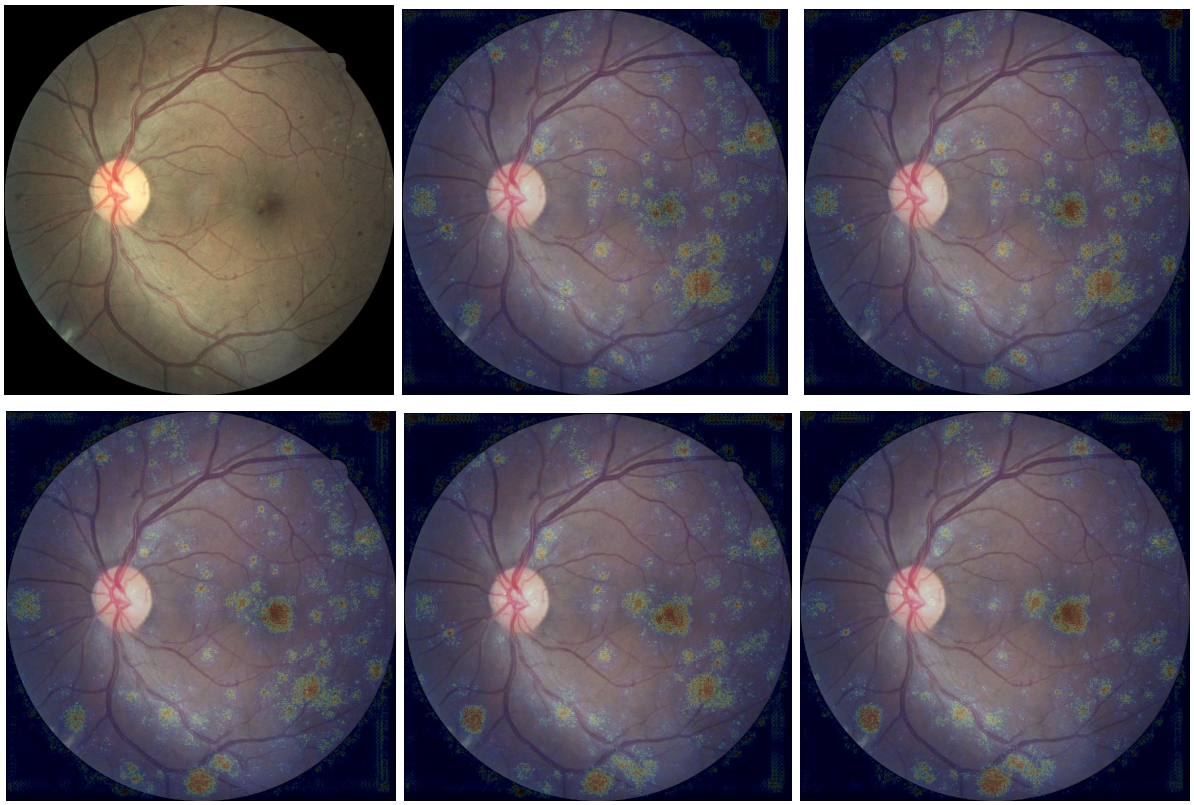

Figure 3. Salience maps from score gradient with respect to input image (class 3 severity).
From top left to bottom right, they are the original image and salient pixels for class 0,1,2,3,4 severity respectively.

Figure 4. Input Image (Left) and Salience Map from Class-Specific Occlusion-Sensitivity (Right). Only one is shown because they all look similar (small effective salience for each patch). These two look almost identical because the overlaid salience map on the right image is transparent and homogeneous (because no area is deemed salient by this occlusion approach), therefore it is hard to discern.

	Class 1 or Above	Class 2 or Above
Sensitivity	79.5%	84.7%
Specificity	90.2%	95.3%

Accuracy	79.5%

Figure 5. Model Performance

References

1. 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.
2. ICO Guidelines for Diabetic Eyecare. International Council of Opthamology. January 2017.
3. 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.
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