Image Classification of Treatable Eye Diseases Through Combination Transfer Learning with ResNet-50 and SqueezeNet.

Richard Nardone

Department of Computer Science

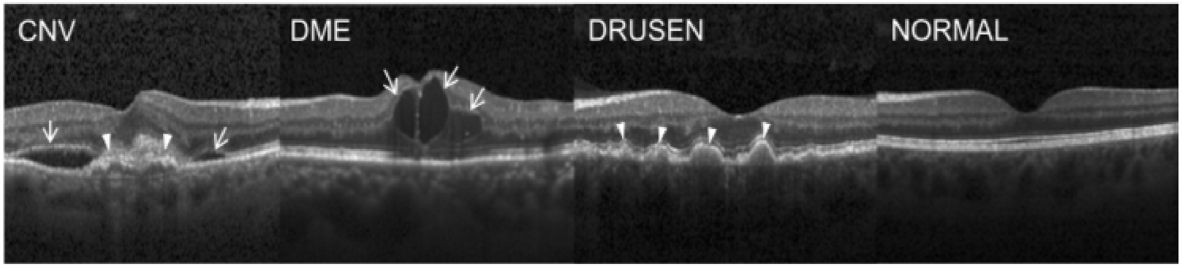
New Jersey Institute of Technology

Newark, NJ 07102

Contact: rmn9@njit.edu

Abstract

I present a new model developed through a form of transfer learning which relies on structural snippets of established and high performing models. The model I designed makes use of two structures extracted from ResNet-50 and SqueezeNet respectively. The model was trained on a dataset of 83,480 Optical coherence tomography (OCT) images of the retina with four unique classifiers for prediction, Normal, Drusen, Choroidal Neovascularization (CNV) and Diabetic Macular Edema (DME). The experimental results provide evidence that my model can accurately classify OCTs with a mean accuracy of 95.66% when tested on a dataset of 968 OCTs with even distribution among the four classifiers.



Sample OCTs with Annotations

I. INTRODUCTION

The medical industry is extremely reliant on highly trained professionals to hand read charts, medical imaging and physical specimens to correctly diagnose a patient. Advancements in data science and image processing provide a solution to support professionals. Trained image classification models can classify medical images to a high accuracy in place of highly trained professionals at a substantially faster rate.

Image classification techniques include neural networks, support vector machines, clustering, decision trees and more. For this application I chose to use a Convolutional Neural Network (CNN). "A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other." My initial focus was on implementing transfer learning, a process of using a pretrained network and adjusting it to work with the OCTs and classification. The issue with neural networks is they can be extremely large and the more layers and connections they contain the larger they are. For instance, AlexNet is a deep neural network that has 240MB of parameters which is fairly large by today's standards.

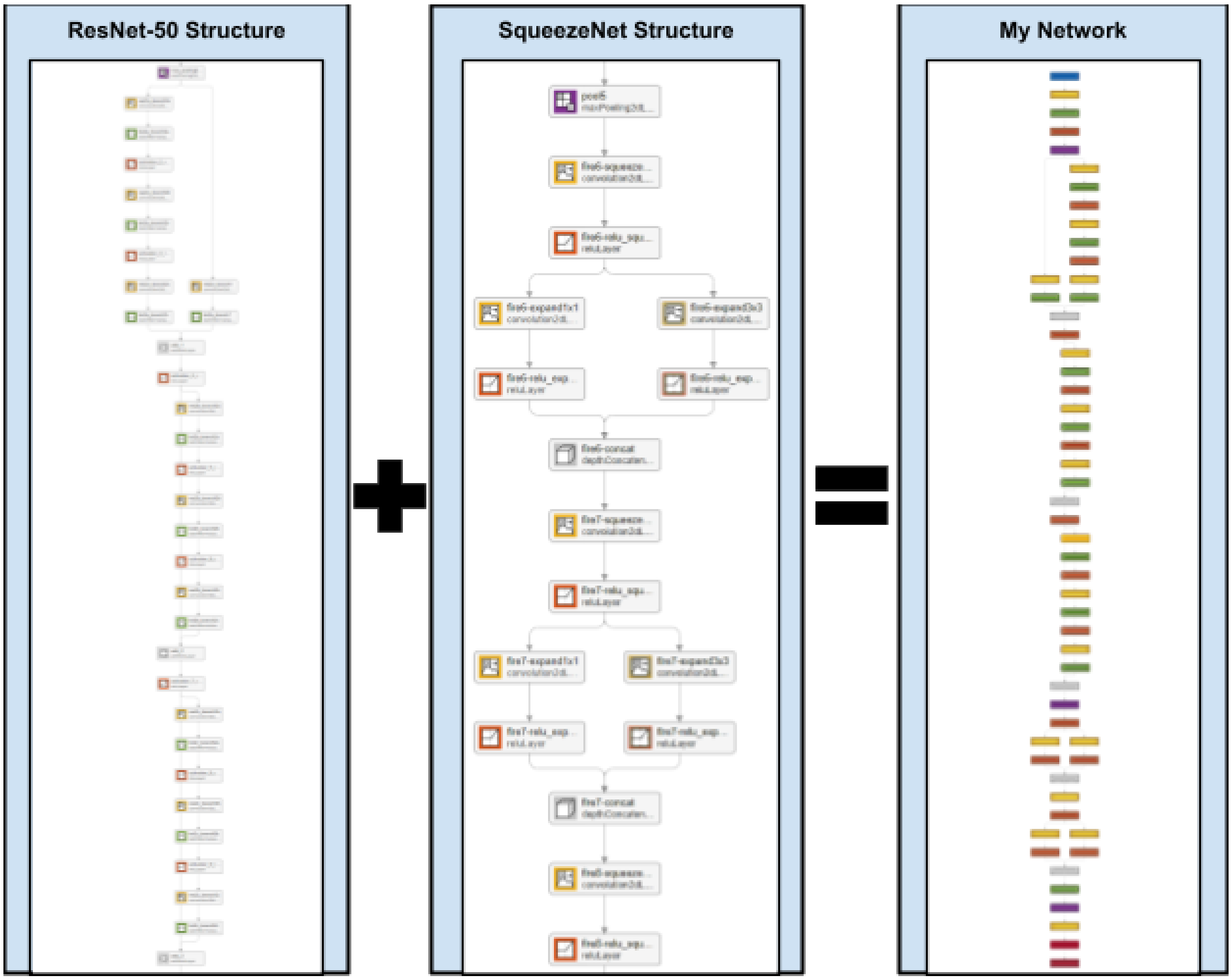
The structure of this paper will be as follows. In Section II, I will describe in a brief summary the existing methods of image classification. In Section III, I will give an in depth guide to the creation of the model and the thought process behind it. In Section IV, the experimental data, results will be presented and explored. In Section V, I will discuss and compare the results with those of the model I created, a well established model and the predictions made by clinical experts. In Section VI, I will draw conclusions from the presented information.

II. EXISTING NEURAL NETWORKS FOR IMAGE CLASSIFICATION

In the field of image classification neural networks are implemented as they yield the highest results with large image datasets. In computer vision specifically convolutional neural networks or CNNs are extremely effective as they focus on pixel detection and classification. Convolutional Neural Networks get their name from the convolution layer which is a key building block in any CNN. “The layer's parameters consist of a set of learnable filters (or kernels), which have a small receptive field, but extend through the full depth of the input volume. During the forward pass, each filter is convolved across the width and height of the input volume, computing the dot product between the filter entries and the input, producing a 2-dimensional activation map of that filter. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input.”[5]

For this experiment I chose to implement the SqueezeNet, a neural network descendant of AlexNet, a revolutional CNN used for image classification. SqueezeNet is an advantageous model because it reduces the number of parameters and outperforms the other neural networks. “The CNN basis to design the SqueezeNet: (1) replace 3×3 filters with 1×1 filters, (2) decrease the number of input channels to 3×3 filters, (3) downsample late in the network so that the convolution layers have large activation maps. The SqueezeNet is comprised mainly of Fire modules that are squeeze convolution layers with only 1 × 1 filters. These layers are then fed into an expand layer, which has a mix of 1 × 1 and 3 × 3 convolution filters” [4]

III. MODEL CREATION AND SUMMARY

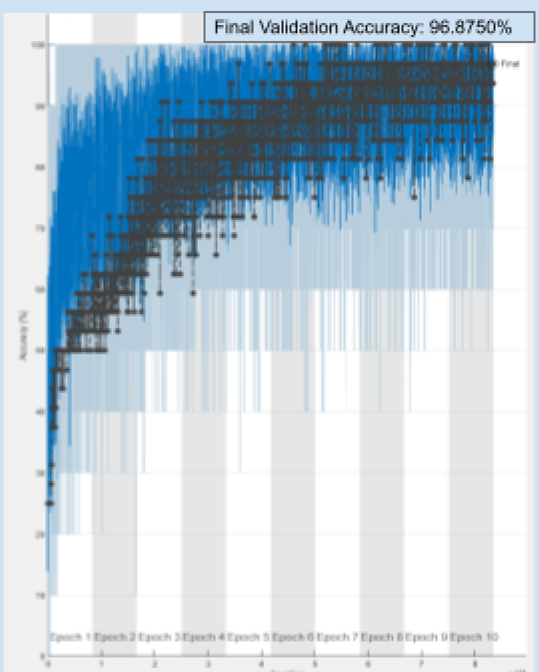


The neural network I have designed uses a form of transfer learning which takes unique structures from established networks and implements them into a smaller network to classify images. For this model I intended to keep it small due to the fact I would be training the model on my desktop workstation which only has one GPU. The two models I drew inspiration from were ResNet-50 and SqueezeNet. ResNet-50 is a deep neural network which contains 177 layers and 192 connections. SqueezeNet is a deep neural network which contains 68 layers and 75 connections. The two structures I have chosen to extract are outlined in the above figure. ResNet-50's structure is centered around 2-D concatenations. SqueezeNet's structure is centered around addition.

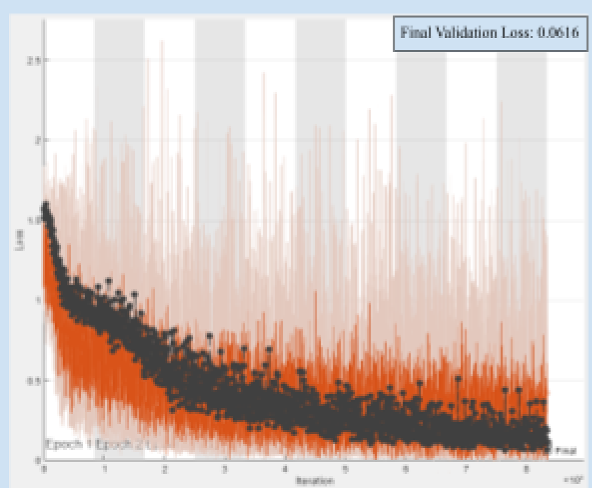
I implemented the structures with ResNet-50's structure being run through first and then SqueezeNet's structure. The resulting network contains 55 layers and 59 connections. There was a large importance on keeping the network layers and connections relatively small due to the lack of computing power I possess.

IV. EXPERIMENTAL RESULTS

Due to a lack of computational power the training process took a total of just under 270 minutes. The model ran through 10 epochs with a mini batch size of 10 and a validation frequency of 50 iterations. At the end of the training process the model received a final validation accuracy of 96.8750% and a final validation loss of 0.0616.

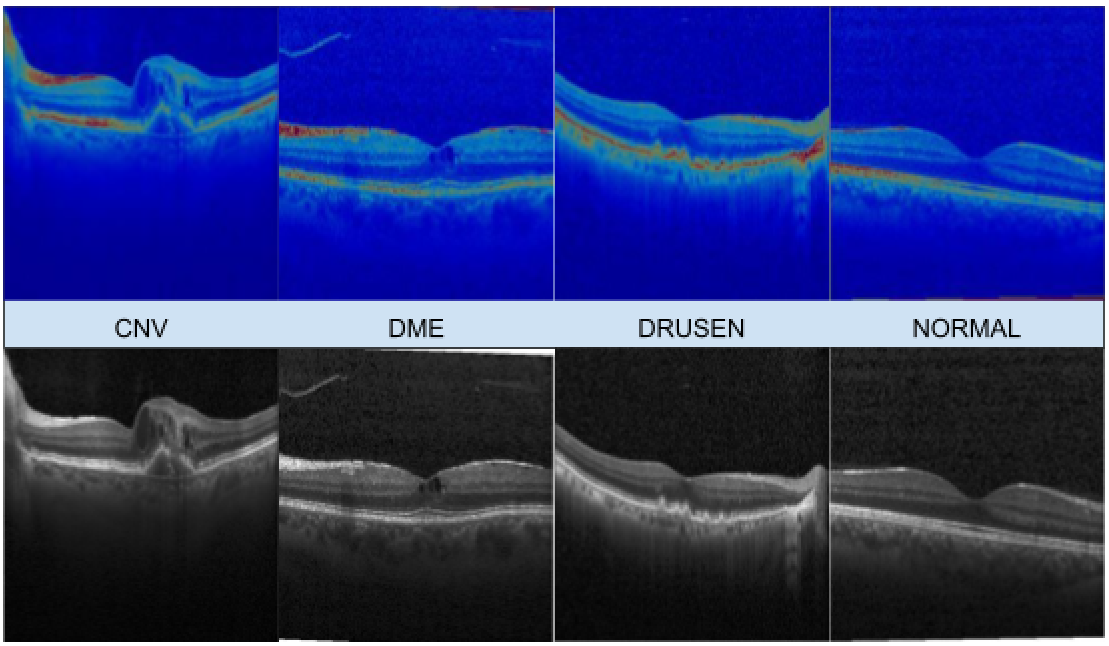


Validation Accuracy by Iteration graph of Neural Network



Validation Loss by Iteration graph of Neural Network

Post training the neural network's decision making for importance of features was visualized using imageLIME. The above image is a comparison of the images labeled and what the model selected as the important features. In comparison with the image presented in the abstract the model shows it is focusing on similar points to that of a clinical expert.

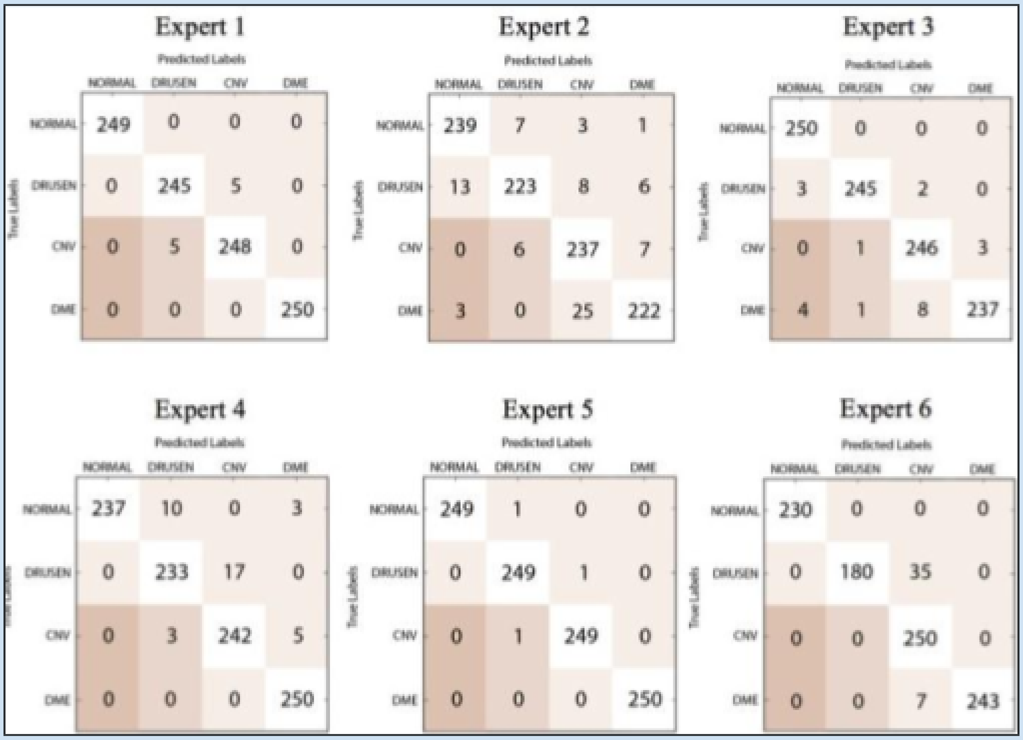


Comparison of imageLime (top) and original images (bottom)

The model was tested on a final dataset of 968 OCTs and received an accuracy of 95.66%. As the validation loss and validation accuracy were continuing to decrease and increase respectively at the end of the training I believe with more input data and some minor adjustments the model accuracy could be improved more.

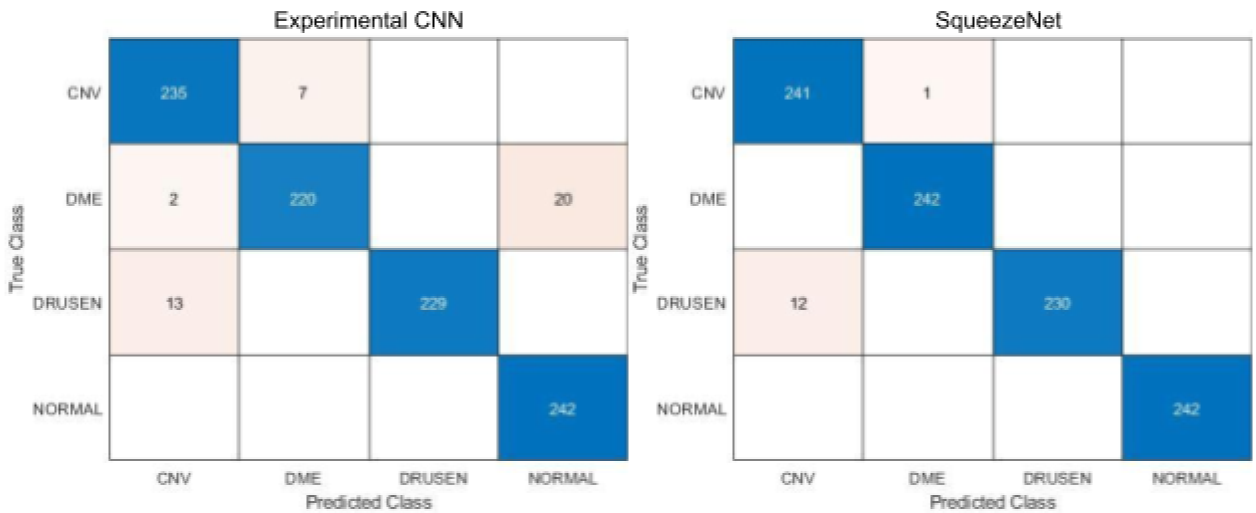
Expert confusion matrices sourced from Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning[1], we are provided valuable insight as to how our model is performing compared to the clinically trained experts. In comparison to the expert's analysis of the images, the model's performance was under achieving. The experts classification of the images performed at a mean accuracy of 96.77% and a range of 7.6% with a high of 99.70% and a low of 92.10%.

V. COMPARISONS WITH EXISTING METHODS

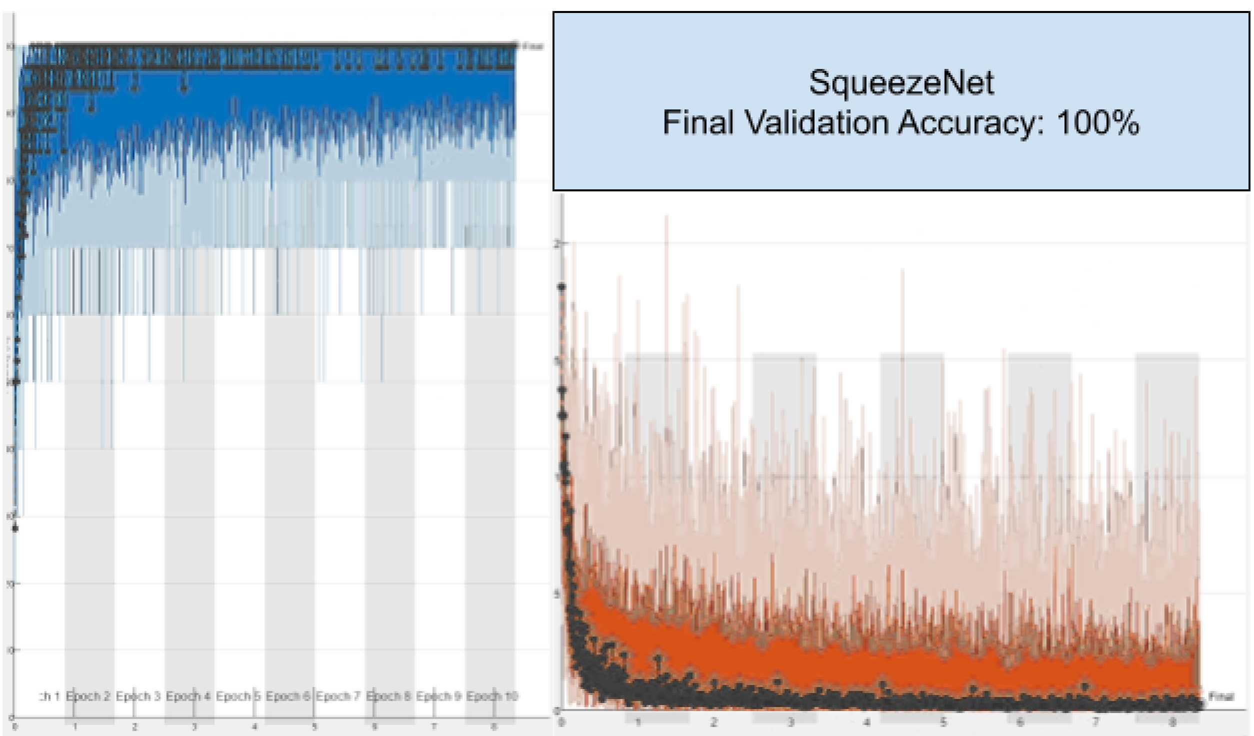


Expert Confusion Matrices

In comparison to the model presented in the paper “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning” my model performed 0.94% worse with a tenth of the epochs. In comparison with the expert diagnoses the model performed worse than 83.33% of the clinical experts. When compared with the average of the clinical expert diagnoses, 96.77% the model scored 1.11% worse.



SqueezeNet was then implemented and through the use of transfer learning I trained the neural network with the same data. The SqueezeNet was subject to the same training parameters as the experimental neural network. The SqueezeNet model outperformed the neural network I created by 3.00% and the average of the clinical diagnoses by 1.89%. In comparison with the expert diagnoses, the model performed worse than 33.33% of the clinical experts, an increase of 50% from the experimental neural network.



VI. CONCLUSIONS

In conclusion the SqueezeNet outperformed the experimental neural network and the majority of clinical experts in accuracy. While the experimental neural network did fall short of expectations it did not fall far behind the experts but in the real world experts need reliability and the SqueezeNet definitely provided more reliable results.

With more data, epochs and hypertuning parameters it may be possible to beat the experts in accuracy. An important concept which can be derived from this experiment is the use of non-traditional transfer learning techniques such as structure extracting to create smaller models which can perform adequately. With no prior experience with deep learning previous to this experiment I believe that it is feasible to say the models can be enhanced to outperform the experts and provide a noticeable change to the industry.

REFERENCES

[1] Kermany, Daniel S., et al. “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, Feb. 2018.

[2] https://www.kaggle.com/paultimothymooney/kermany2018

[3] https://www.cell.com/cell/fulltext/S0092-8674(18)30154-5

[4] H. Lee, I. Ullah, W. Wan, Y. Gao, and Z. Fang, “Real-Time Vehicle Make and Model Recognition with the Residual SqueezeNet Architecture,” *Sensors*, vol. 19, no. 5, p. 982, Feb. 2019

[5] Wikipedia contributors. "Convolutional neural network." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 28 Apr. 2022. Web. 1 May. 2022.