Convolutional Neural Network for Retinal OCT Classification

Aaron Chang, Alex Lo, Anusha Saha, Viswajith Rajagopalan, Vivan Shah
Department of Computer Science, The University of Texas at Dallas

Abstract

Optical Coherence Tomography (OCT) is a Retinal imaging technique used to detect retinal abnormalities. However, the analysis of such images to reach aretinal diagnosis serves to be a difficult and time-consuming task for medical professionals. This project aims to develop the most efficient convolutional neural network model to classify these OCT images. This CNN is then benchmarked against a fully-connected neural network model to demonstrate a CNN's performance in tasks that require spatial context.

Retinal OCT Imaging

Background

OCT is a retinal imaging technique which uses light waves and their reflections to build a cross-section of the eye. This cross-section image is used to look at the different layers of the retina and identify any signs of abnormalities to diagnose optical diseases. For the purposes of this project, we focused on 4 types of classifications – Choroidal Neovascularisation (CNV), Diabetic Macular Edema (DME), Drusen present in Age-related Macular Degeneration (DRUSEN), and a Normal Retina (NORMAL). Figure 1 depicts an OCT Image of each of these abnormalities, along with a normal retina for a comparison. It also highlights the identification signs for each through white arrows.

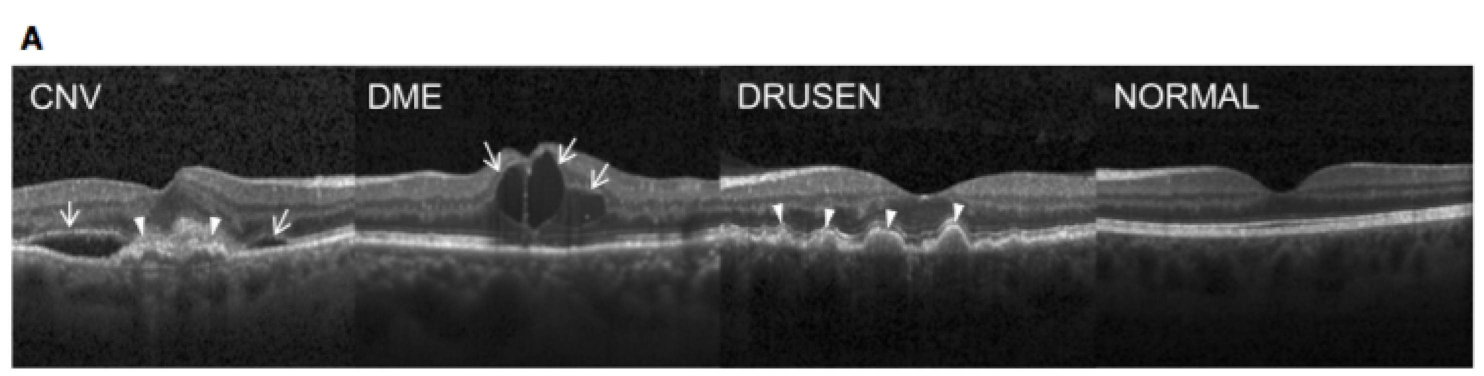


Figure 1: (Left to Right) Choroidal Neovascularisation, Diabetic Macular Edema, Drusen present in Age-related Macular Degeneration, Normal Retina

Dataset

The dataset being used is split into the 4 aforementioned classifications (CNV, DME, DRUSEN, NORMAL). The OCT images were gathered from a set of cohorts from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and Beijing Tongren Eye Center. Each and every image went through a thorough tiered grading system that contained several layers of trained graders. This was done for the correction and verification of image labels. Additionally, to account for any potential human errors in grading, a validation subset of 993 scans was graded separately by two ophthalmologists with a senior retinal specialist settling any disputes. This dataset was then published on Kaggle, where it was downloaded by our team.

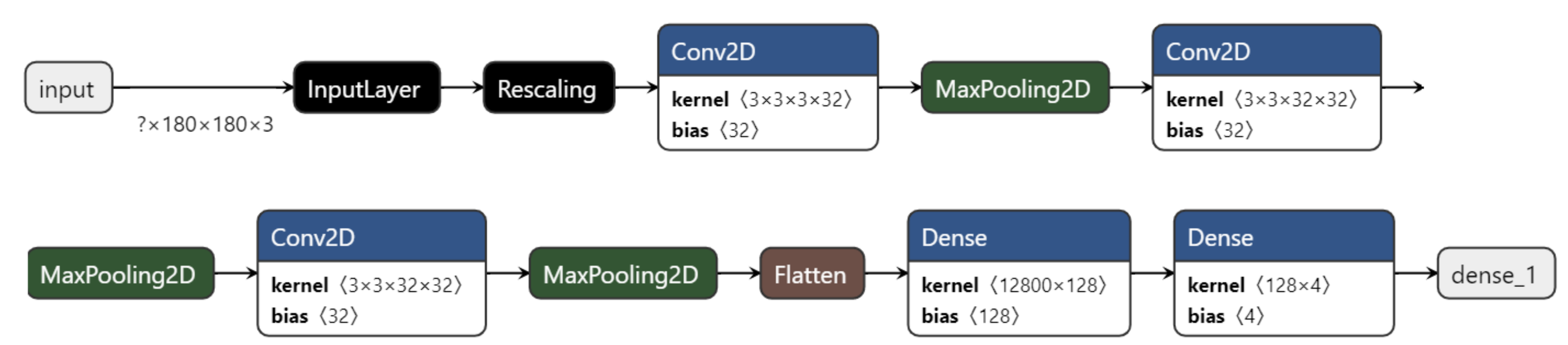
Models

Our models are constructed through Keras, a Python library used for deep learning. We experimented with different sets of layers and measured our model's effectiveness using accuracy, F_1-score, and sparse categorical cross-entropy loss metrics.The models are compiled with their respective training sets to produce the required metrics.

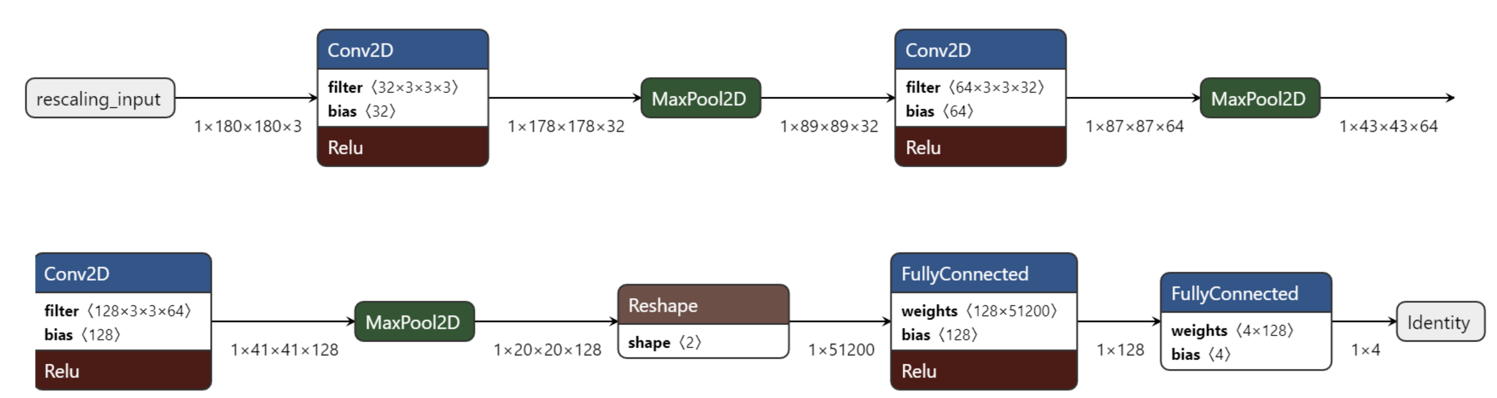
Summary of Models

Model Name	Accuracy	Loss	F1-Score	Total Parameters
1.6M CNN	.9375	.0957	.9452	1,658,436
6.6M CNN	.9504	.1642	.9502	6,647,492
15.4M CNN	.9483	.1808	.9475	15,495,876
500K Dense	.4256	1.3782	.1194	488,137
1.7M Dense	.2500	4.852	.1000	1,768,009

1.6M CNN Model



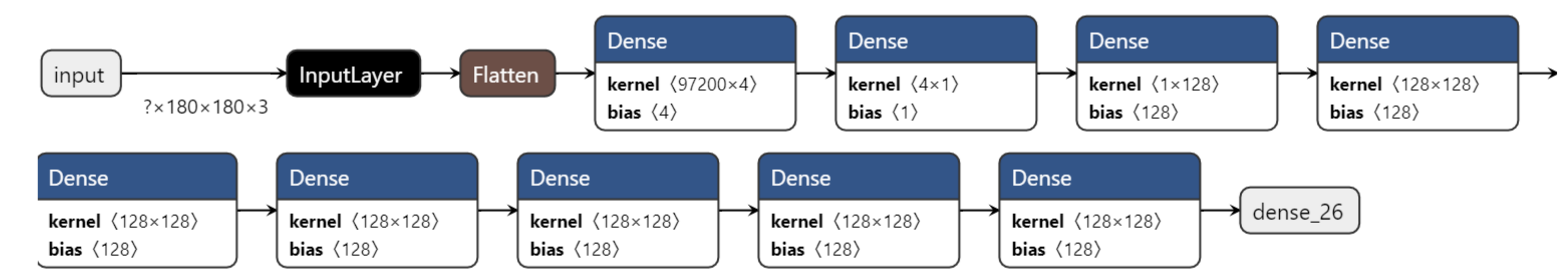
6.6M CNN Model



15.4M CNN Model

The 15.4M CNN Model consists of two CNN layers, along with MaxPooling layers, a flattening layer, and 6 dense layers. It has over 9 times as many parameters as the 1.6M CNN model. However, we noticed diminishing returns as the number of parameters went up, with classification accuracy increasing only 1.08% in comparison to the 1.6M CNN model.

500K Dense



1.7M Dense

The 1.7M Dense model is similar to the 500K Dense model, but with 3x the parameters. It performs significantly worse (68.75% lower accuracy) in comparison to the 1.6M CNN model. Surprisingly, increasing the parameters of the dense model caused it to perform even worse than the baseline 500K model (17.56% lower classification accuracy). This suggests that Dense models lack certain localized spatial information which is critical to this task.

Conclusion

Retinal disease and age-related macular degeneration affect up to 10 million Americans each year and more worldwide. Due to the widespread nature of this affliction, it takes time and effort to go through each retinal OCT image, effort that can be spent elsewhere. Our novel CNN solution reaches 93.75% accuracy without compromising our classification accuracy as evidenced by our F_1-score.

Future Work

Future work concerns deeper analysis of our models and the deployment of our model within the medical field. Implementation of countermeasures against false positives and false negatives need to be included, as well as more thorough testing of the model. Tweaking of the hyper-parameters would also be worked on in the future in order to increase accuracy and F_1-score.

References

[1] Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018). *Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification*. Mendeley Data, V2