Retina Pathology Diagnosis

INTRODUCTION TO NEURAL NETWORKS

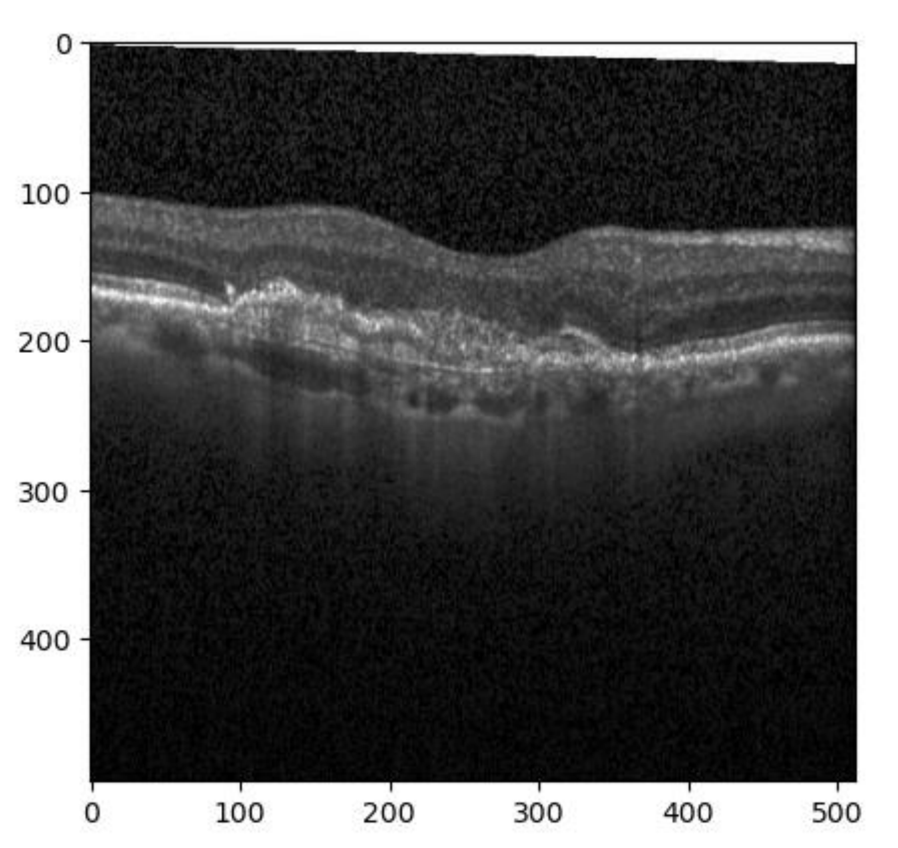
https://github.com/slilly4/OCT-Retina-Classification

Problem Statement

- This project attempts to classify optical coherence tomography (OCT) images of the retina. OCT is a non-invasive imaging technique that uses light to generate cross-sectional scans. When applied to the retina, a number of pathologies can be identified. Traditionally, diagnoses have been made by a qualified medical professional. However, with the power of neural networks, it should be possible to automatically detect the health of the retina.
- In addition to attempting to achieve a high classification accuracy on this task, I would also like to explore the use of transfer learning to speed up training on CNNs.

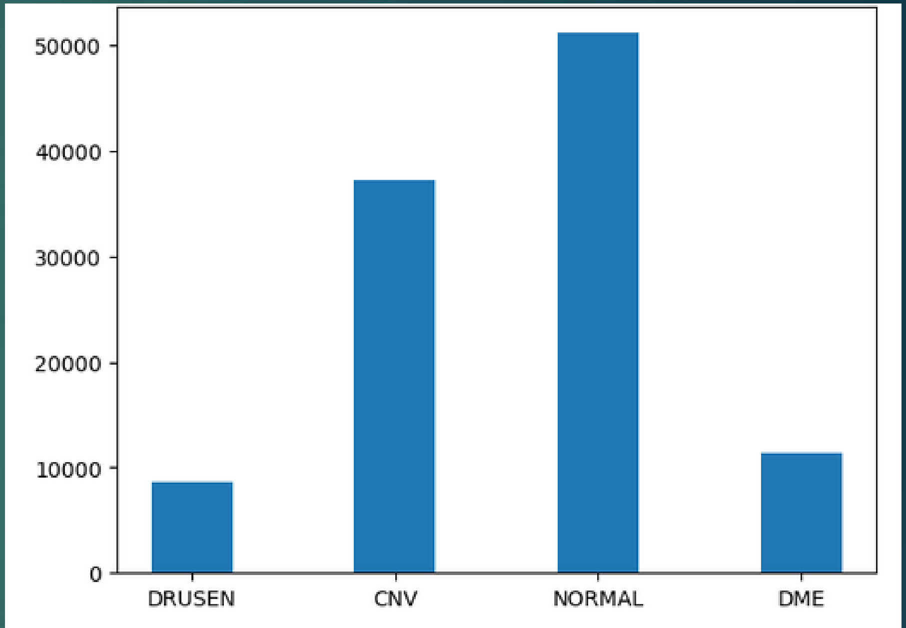
Data

- The OCT images used in this project were initially employed for research conducted by Daniel S. Kermany and other researchers at the University of California San Diego and the Guangzhou Women and Children's Medical Center. Their results were published in the 2018 article, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning."
- Citation:
  - *Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images", Mendeley Data, V3, doi: 10.17632/rscbjbr9sj.3*



Data Attributes

- 108,309 images in 4 classes
  - Choroidal Neovascularization (CNV)
  - Diabetic Macular Edema (DME)
  - Drusen
  - Normal
- Imbalanced
- 3 channels, but inconsistent dimensions
- Data will be normalized; flip, zoom, and shear to be added for diversity
- Dimensions standardized to 496x496 pixels



Convolutional Neural Network (CNN) Approach

- VGG-Style Model: 4 repeating layers of convolution and pooling with increasing filters.
  - 327,060 parameters
- ResNet-Style Model: 4 blocks of residual layers, each of which includes “skips” to allow to for deeper networks.
  - 153,076 parameters
- ResNet50V2: Pre-Trained on ImageNet dataset
  - 23,572,996 parameters

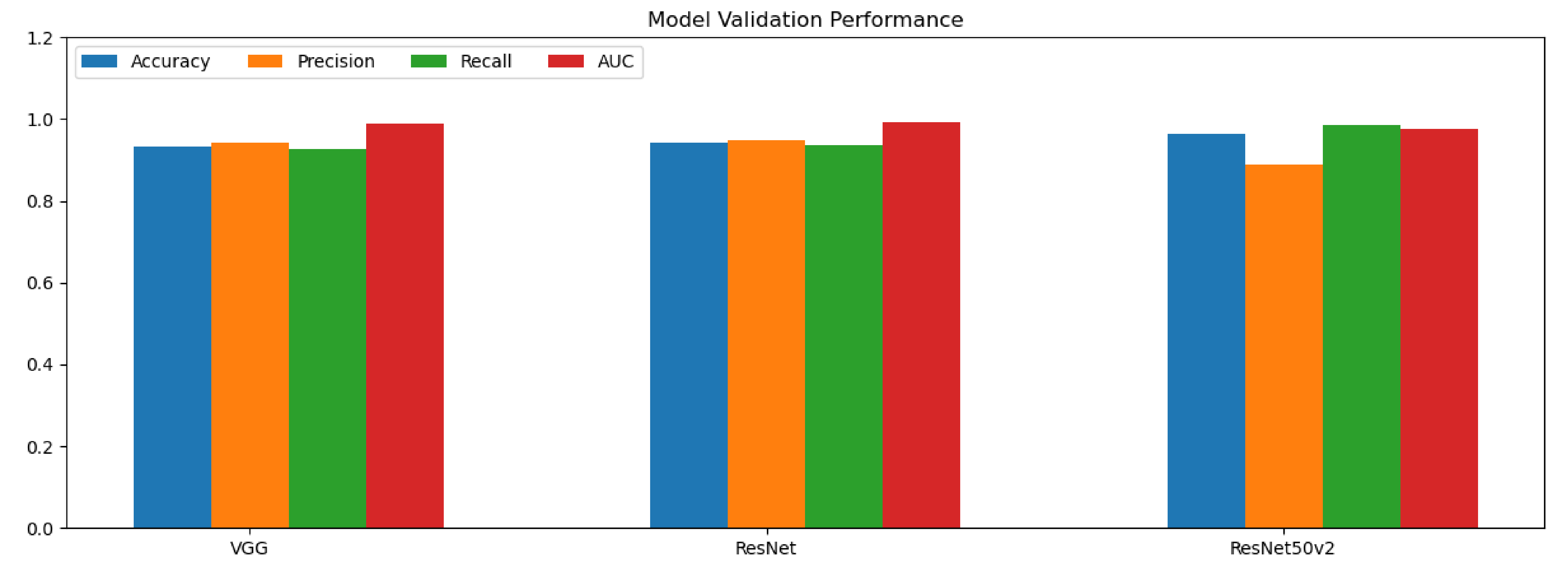
Validation Metrics

- While the pre-trained ResNet50v2 model has over 23.5 million parameters, it achieved only marginally better gains in categorical accuracy than our baseline models, and it performed worse in Precision and AUC than the other models.

Metric	VGG-Style	ResNet-Style	ResNet50v2
Loss	0.2237	0.1750	0.1081
Accuracy	0.9343	0.9420	0.9648
Precision	0.9417	0.9484	0.8878
Recall	0.9262	0.9369	0.9849
AUC	0.9886	0.9933	0.9779
# Parameters	327,060	153,076	23,572,996
Training Time	35,299s	45,972s	42,710s
Accelerator	GPU	TPU	TPU

Accuracy, Precision, Recall, AUC

- While accuracies where high, Recall lagged in ResNet and Precision in ResNet50v2. Unbalanced data can skew performance.



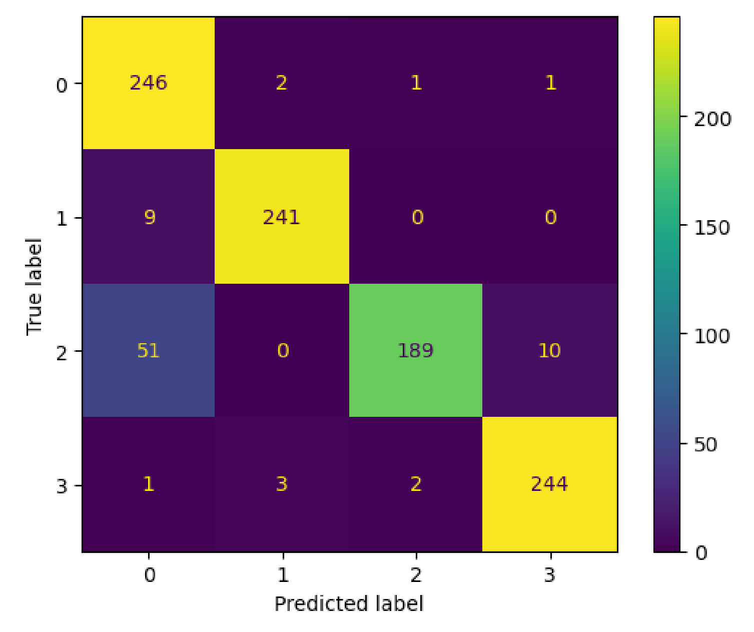
Test Metrics

- With a balanced test set on which to evaluate the models, it is clear that the ResNet50v2 model out-performed the others across all of the metrics.
- Accuracy greater than .96

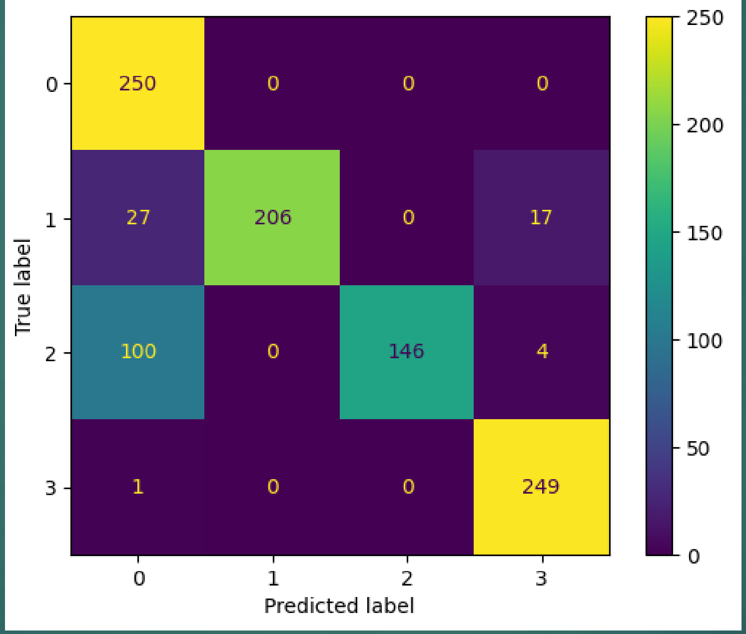
Metric	VGG-Style	ResNet-Style	ResNet50v2
Accuracy	0.92	0.851	0.963
Precision	0.931	0.896	0.968
Recall	0.92	0.851	0.963
F Score	0.919	0.849	0.963

Confusion Matrix

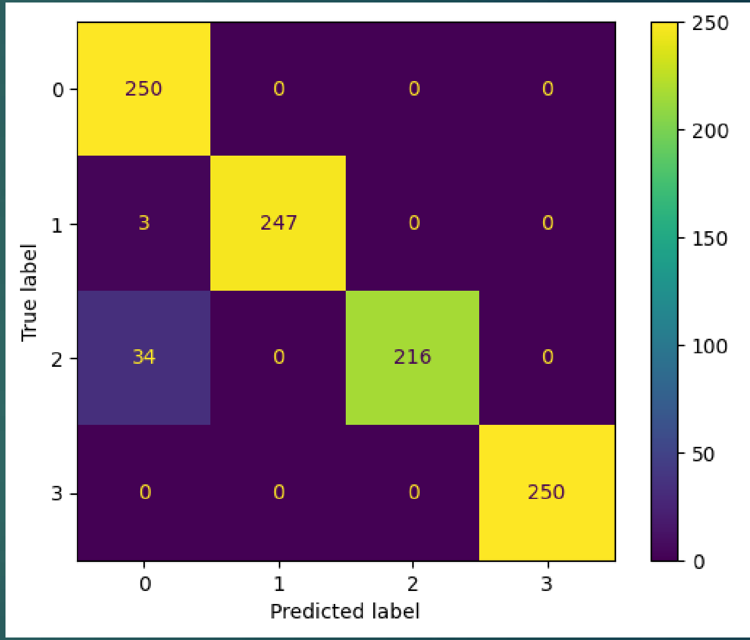
VGG-Style



ResNet-Style



ResNet50v2 (Pre-Trained)



CNV: 0, DME: 1, DRUSEN: 2, NORMAL: 3

Discussion

- Future iterations of this project could attempt to oversample data from minority classes.
- Validation data alone is not enough to gauge the performance of a model. My validation data mirrored the makeup of my training data. Therefore, for those classes where there were many images, a greater number were also in the validation data. This skewed the results.
- Another place for improvement is a better strategy for saving models. I attempted to checkpoint models, and this worked only partially. Models were lost or corrupted.
- Pre-trained models can perform better on this task with fewer resources expended.

Conclusion

- Using two models trained from scratch as well as fine-tuned pre-trained model, ResNet50v2, it is clear that the pre-trained model performed better using fewer same resources.
- Models trained on other image recognition tasks can be successfully adapted to other classification projects.
- An accuracy of over .96 on this task is possible. Further gains can be made through optimization and longer training.
- There may be space in the doctor’s office for trained neural networks in diagnosing pathologies.