Retinal disorders detection based on Optical Coherence Tomography (OCT) images using ResNet

Tanxin Qiao

Background:

The development of neural networks has enabled image classification to a new level. Through layers of filters, convolutions or normalizations, images are abstracted to varying sizes of feature maps that eventually give the desired classification outcomes[1]. From image to direct labels, neural networks reduce the need for traditional image processing steps, such as segmentation or calibration. Hence, the application of neural networks on medical images may provide quick, reliable disease diagnosis with little cost. Potentially, it could facilitate a series of clinical practices, provide accessible health instructions where medical resources are limited and fuel the coming of age of digital healthcare.

Retinal disease detection is an important challenge in computer aided diagnosis (CAD) for medical applications. Retinal diseases could easily cause blurred vision or even total blindness. Most of them cause detectable disorders in the retina, a thin layer of tissue on the inside back wall of one's eye. Optical coherence tomography (OCT) is a common clinical procedure in ophthalmology used to inspect retinal conditions and diagnose retinal diseases. It is a non-invasive imaging technique that uses a long-wavelength, broad-bandwidth light source to illuminate the retina and assess the light reflected from retinal tissue[2]. A typical OCT image of normal retina is shown in Fig. 1, displaying smooth layers with no breaks or loss of layer continuity[3].

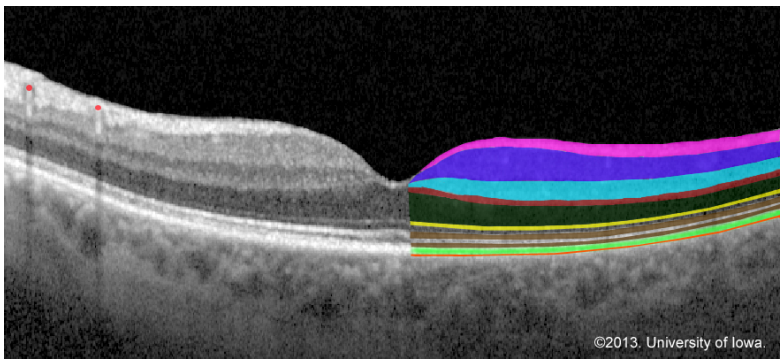

Fig 1. Optical coherence tomography (OCT) image of the retina with the layers colored for clarification on the right[4]

From top to bottom are the internal limiting membrane, nerve fiber layer, ganglion cell layer (pink), inner plexiform layer (purple), inner nuclear layer (turquoise), outer plexiform layer (red), outer nuclear layer (cell bodies of the photoreceptors; dark green), external/outer limiting membrane (yellow), and the photoreceptor layer (internal and external segments; brown). The retinal pigment epithelium is

immediately underneath the retina (bright green). Bruch's membrane (orange) is a thin membrane that separates the retinal pigment epithelium from the underlying highly vascular choroid. The retinal vasculature (red dots upper left) cast shadows on the OCT.

Some of the retinal diseases (concerned in this report) are: choroidal neovascularization (CNV), diabetic macular edema (DME) and drusen. Choroidal neovascularization can be roughly deemed as the generation of new blood vessels in the choroid layer of the eye. The causes of CNV include age-related macular degeneration (AMD), extreme myopia and ocular histoplasmosis, etc. It may lead to loss of central vision. The symptoms of CNV are usually represented as neovascular membrane (red arrow) and retinal pigment epithelium detachment (yellow arrow) in OCT images (Fig. 2a).

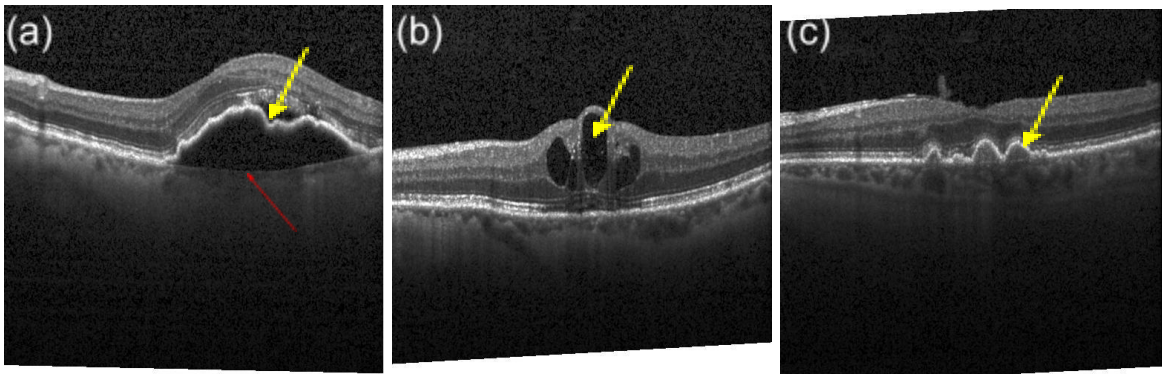

Fig 2. OCT images of different kinds of retinal disorders
(a): CNV (b): DME (c): DRUSEN
Yellow and red arrows indicate lesion positions.

Diabetic macular edema is one of the most common complications of diabetes. It is caused by the accumulation of fluid in the macula and may induce irreversible blindness[5]. In OCT images, DME often displays as hollows (Fig. 2b yellow arrow) above the retinal pigment epithelium (RPE) layer. Drusen are yellow deposits made up of lipids and proteins under the retina. Reasons for drusen are generally associated with aging and early AMD. Drusen appear in OCT images as little waves (Fig. 2c, yellow arrow) on the RPE layer.

Residual network (ResNet) is a convolutional neural network model that was proposed in 2015 by researchers at Microsoft Research. Addressing the degradation problem (a quick reduction in training accuracy) brought by stacking more layers in the models, it surpassed all other models in the ImageNet 2015 competition. ResNet features shortcut connections, where a previous layer is connected directly to a new layer, skipping some middle layers. The structure formed by a skip connection is called a residual block. ResNet consists of various numbers of different residual blocks and thus has multiple structures (ResNet18, ResNet50, ResNet152, etc.). Enabling significantly deeper networks, ResNet is suitable for conducting complex tasks such as image classification,

and therefore it would be a good practice for the retinal disorders diagnosis task.

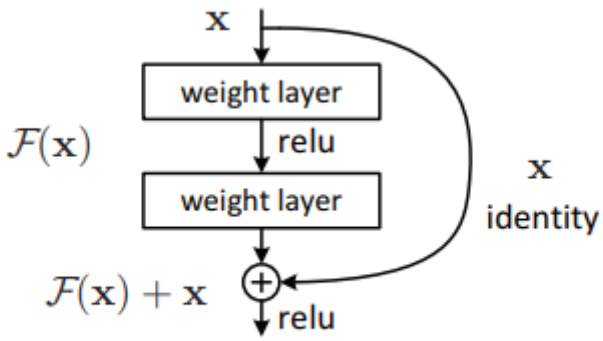

Fig 3. An illustration of a residual block[6]

Computational problem:

The computational problem can be formulated as:
Construct and train a ResNet model that can take the OCT images as input and output the labels of the images, ranging from CNV, DME, DRUSEN and NORMAL, with a high accuracy.

Method:

1. Dataset

The images used for training and testing are from a labeled OCT dataset[7]. Images were collected from adult patients 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 between July 1, 2013 and March 1, 2017[8]. The dataset contains a training set of 83,484 high resolution gray-scale images in JPEG format (CNV: 37,205, DME: 11,348, DRUSEN: 8,616, NORMAL: 26,315) and a testing set of 1,000 images (250 each class).
2. The images were all resized to 224 × 224 pixels and read as numpy arrays of size (224, 224, 3) with the cv2 package as the initial input for the model.

2. Models

Three kinds of ResNet models were implemented in this project: Resnet18, ResNet34 and ResNet50.
For all three types of models, they all went through first:

Conv1: a 7x7 kernel, stride=2, with a feature map size of 64, followed by a batch normalization and a ReLU function. The output size is (64,112,112).
MaxPool: a 3x3 maxpool layer, stride=2, padding=1. The output size is (64,56,56).

And then they went through different numbers of four kinds of residual blocks, after which the output size is (2048,7,7). Then they all went through:
AveragePool: a 7x7 average pool layer, stride=1. The output size is (m,1,1) and then flattened into m. The value of m for ResNet18 and ResNet34 is 512 while 2048 for ResNet50.
FC: a fully-connected layer of (2048 - 4) where 4 is the number of classes.

The residual blocks for ResNet18 and ResNet34 are the same.
Block1:
Input size=(64,56,56)
3x3 convolution kernel - batch normalization - ReLU;
3x3 convolution kernel - batch normalization - (+input) - ReLU;
Output size=(64,56,56)
Block2:
Input size=(64,56,56)
3x3 convolution kernel - batch normalization - ReLU;
3x3 convolution kernel - batch normalization - (+input) - ReLU;
Output size=(128,28,28)
Block3:
Input size=(128,28,28)
3x3 convolution kernel - batch normalization - ReLU;
3x3 convolution kernel - batch normalization - (+input) - ReLU;
Output size=(256,14,14)
Block4:
Input size=(256,14,14)
3x3 convolution kernel - batch normalization - ReLU;
3x3 convolution kernel - batch normalization - (+input) - ReLU;
Output size=(512,7,7)
ResNet18 went through 2x*Block1*, 2x*Block2*, 2x*Block3*, 2x*Block4*. ResNet34 went through 3x*Block1*, 4x*Block2*, 6x*Block3*, 3x*Block4*.

The residual blocks for ResNet50 differentiate from those of ResNet18 and ResNet34 in that each residual block is a bottleneck block, which employs a 1x1 kernel at the beginning and in the end to reduce the number of parameters and matrix multiplications.
Block1:
Input size=(64,56,56)

1x1 convolution kernel - batch normalization - ReLU;
3x3 convolution kernel - batch normalization - ReLU;
1x1 convolution kernel - batch normalization - (+input) - ReLU;
Output size=(256,56,56)
Block2:
Input size=(256,56,56)
1x1 convolution kernel - batch normalization - ReLU;
3x3 convolution kernel - batch normalization - ReLU;
1x1 convolution kernel - batch normalization - (+input) - ReLU;
Output size=(512,28,28)
Block3:
Input size=(512,28,28)
1x1 convolution kernel - batch normalization - ReLU;
3x3 convolution kernel - batch normalization - ReLU;
1x1 convolution kernel - batch normalization - (+input) - ReLU;
Output size=(1024,14,14)
Block4:
Input size=(1024,14,14)
1x1 convolution kernel - batch normalization - ReLU;
3x3 convolution kernel - batch normalization - ReLU;
1x1 convolution kernel - batch normalization - (+input) - ReLU;
Output size=(2048,7,7)
ResNet50 went through 3x*Block1*, 4x*Block2*, 6x*Block3*, 3x*Block4*.

3. Training

Adam optimization algorithm is used as the optimizer. CosineAnnealing is used as the scheduler for adjusting learning rates. Gradients (losses) are scaled to prevent vanishing gradients. CrossEntropy is the loss function used. Different learning rates and weight decay values were tested and compared to obtain the best results.

Results:

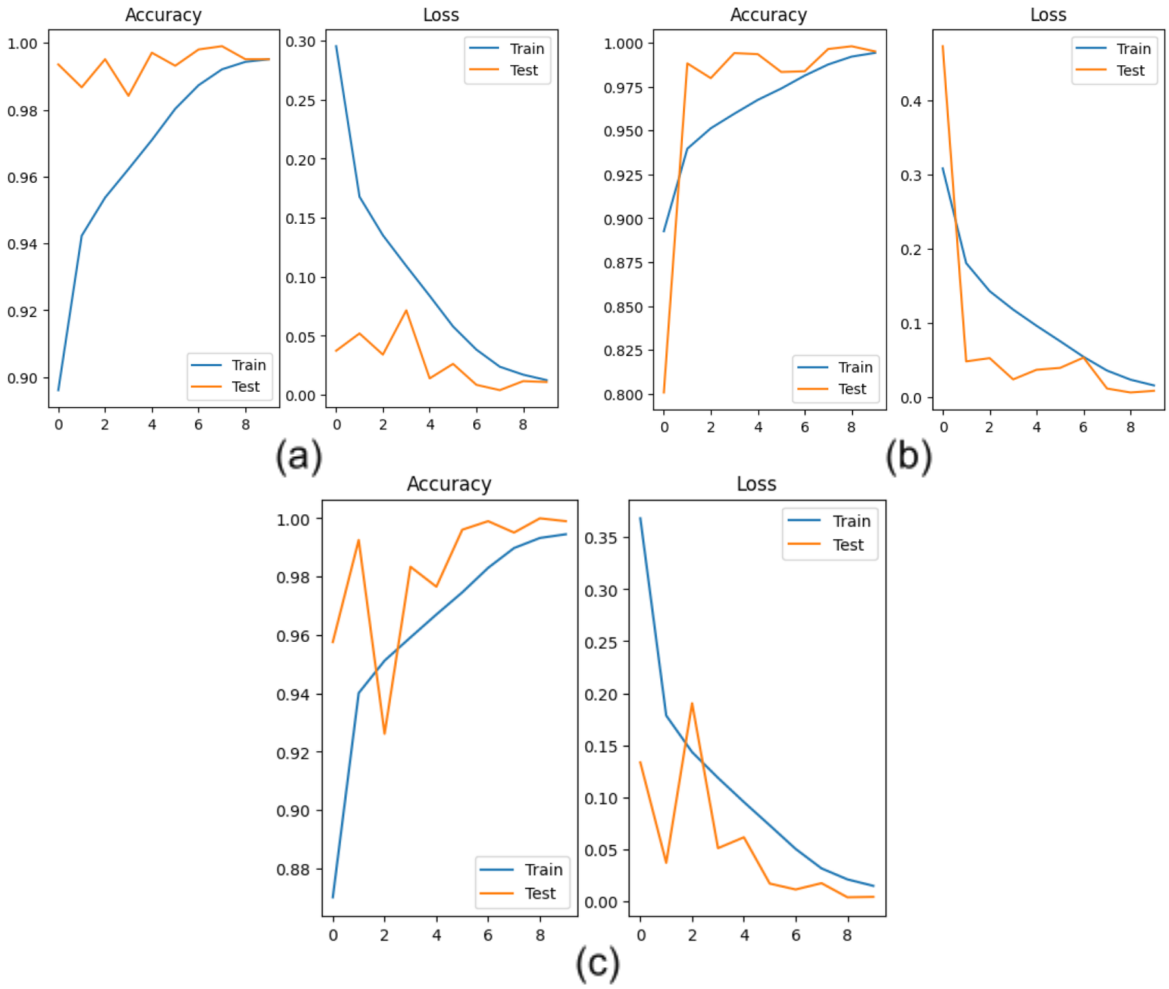

Fig 4. Accuracy and Loss of OCT image classification using different ResNet structures
(a): ResNet18 (b): ResNet34 (c): ResNet50
All models were trained for 10 epochs (x-axis). Learning rate=1e-4. Weight decay=5e-4

As is shown in Fig. 4, all three models showed a test accuracy rate of over 80% in the first epoch and reached over 99% in the end. The accuracy and loss trends of the training set were similar across all three models. For the testing set, the performance of ResNet18 is the most stable, never dropping below 98% of accuracy. ResNet50 may have extracted abundant irrelevant information from the images at epoch 2, resulting in a drop in accuracy. The stacking of layers seemed unnecessary for this task. The following experiments were thus all performed with ResNet18, saving computing resources.

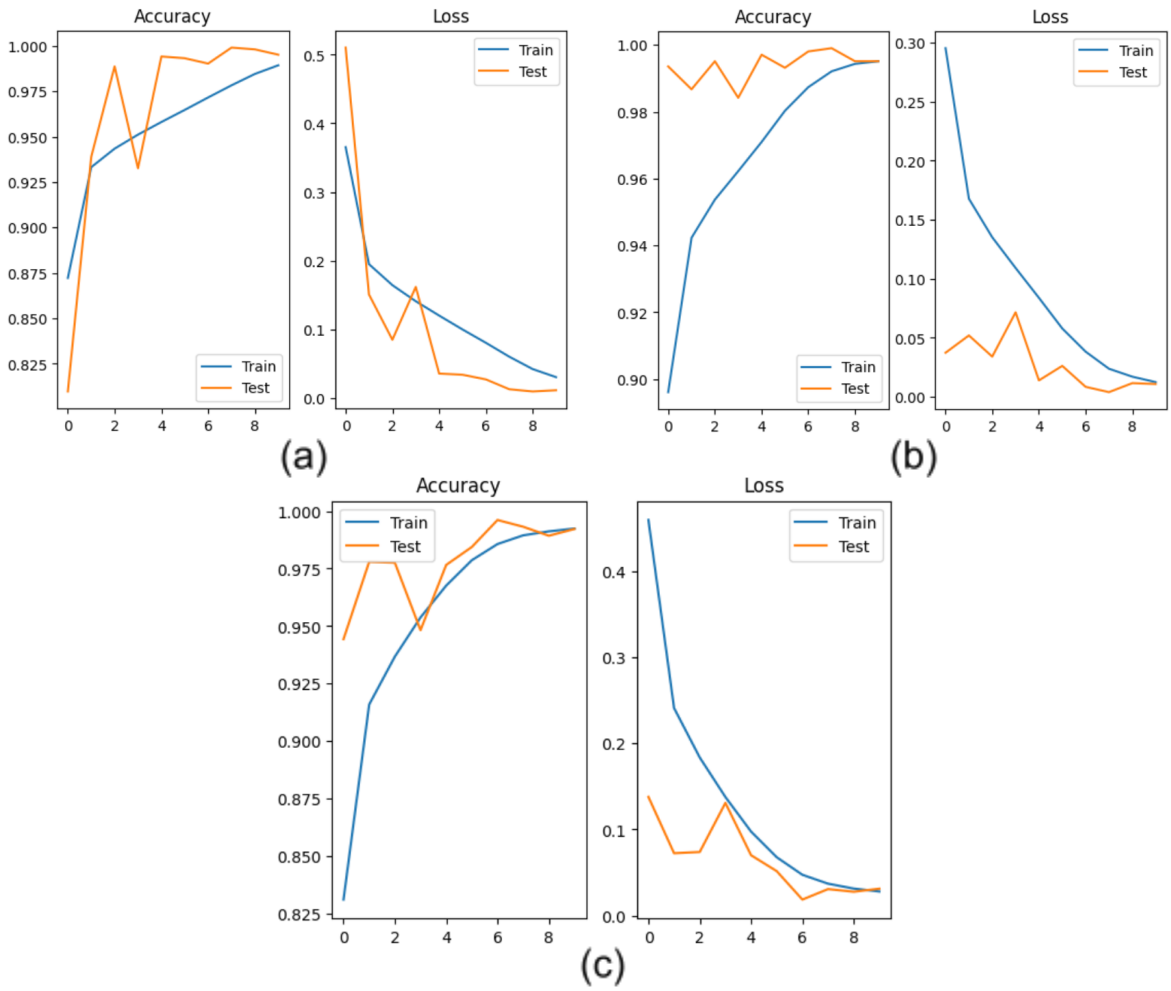

Fig 5. Accuracy and Loss of OCT image classification using different learning rate
(a): 1e-3 (b): 1e-4 (c): 1e-5
All three models were ResNet18, trained for 10 epochs (x-axis). Weight decay=5e-4.

As is shown in Fig. 5, all three models showed a test accuracy rate of over 80% in the first epoch and reached over 98% in the end. Learning rate determines the length of a 'step' taken when altering the parameters based on the results of back propagation. It is therefore reasonable for the model with a bigger learning rate to demonstrate a more fluctuating result (Fig. 5a).

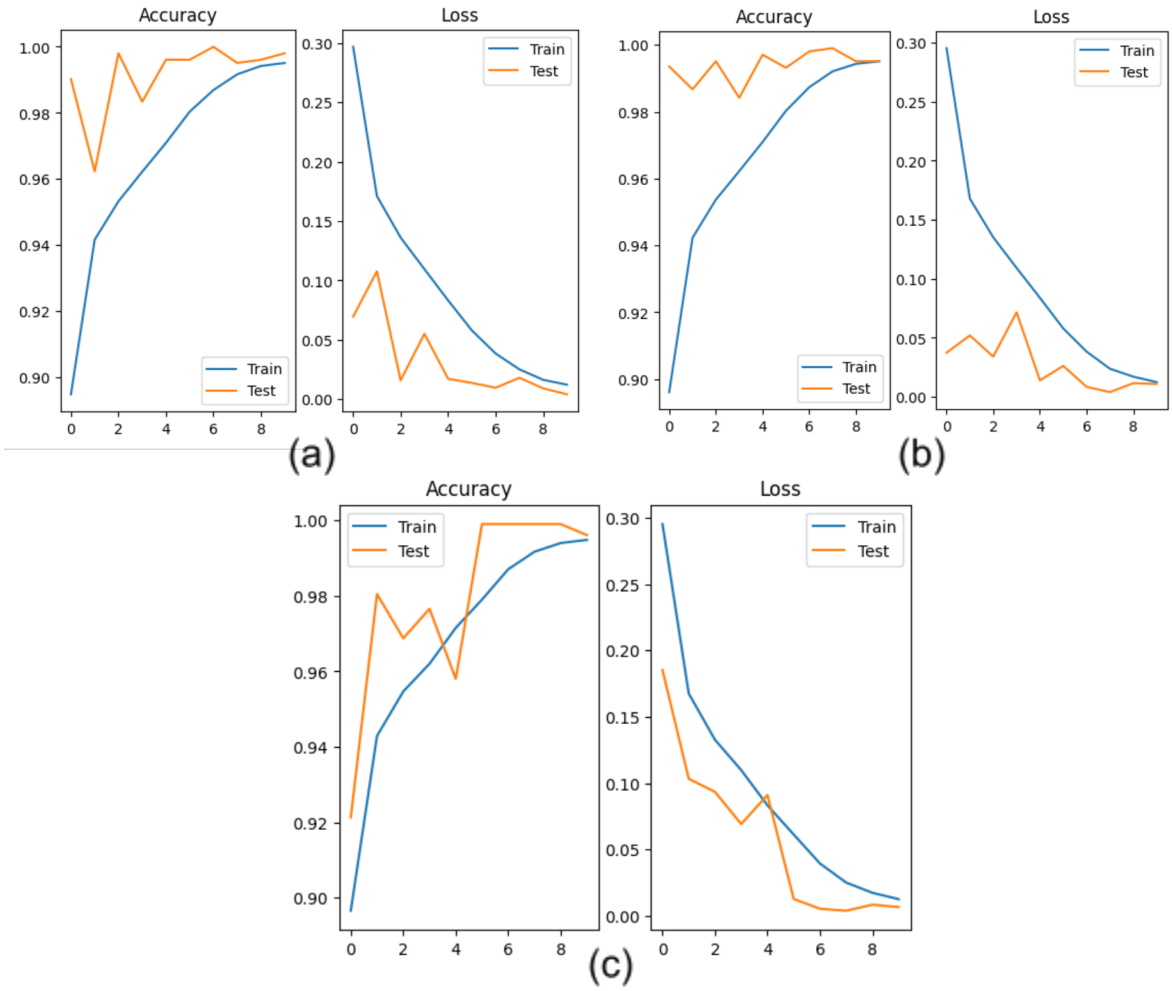

Fig 6. Accuracy and Loss of OCT image classification using different weight decay
(a): 5e-3 (b): 5e-4 (c): 5e-5
All three models were ResNet18, trained for 10 epochs (x-axis). Learning rate=1e-4.

As is shown in Fig. 6, all three models showed a test accuracy rate of over 90% in the first epoch and reached over 99% in the end. Weight decay, also known as L2 regularization, is the regularization factor which is applied to the weights to penalize the absolute values of the weights. If the absolute values of the weights are huge, the outputs may be way off because of some tiny noises in the input. Weight decay is thus applied to ensure the stability of the model. It is not surprising to see from Fig. 6c that the model with the smallest weight decay appeared to be the most variable.

Discussion:

1. Mistakes made by the models

Confusion matrices were not drawn because each model achieved around 99% accuracy and made only one or two (no more than ten) mistakes. I did collect the

mistakes made by all the models and found an interesting fact that almost all mistakes were mistaking DRUSEN as CNV (Fig. 7). There are two possible explanations for the prevalence of this mistake. For one, CNV and DRUSEN do share some similarity in that they both involve the distortion of the RPE layer. It is less likely to mistake these for DME because DME would generally have intact RPE layers. The main difference between CNV and DRUSEN is that, CNV usually causes only one but bigger distortion in the RPE layer while DRUSEN would appear as several smaller bumps. When the small bumps of DRUSEN are close enough that they are somehow infused together (Fig.7, right), it may resemble CNV to some extent. Secondly and perhaps more importantly, the reason may lie in the disproportionate distribution of different labels in the training set. CNV has the most amount of samples (37,205) while DRUSEN has the least amount of samples (8,616). With the tilted instances of training samples, it is actually foreseeable that the model would output more CNV labels and less DRUSEN labels. This issue may be addressed by adjusting training samples so that the distribution is more balanced.

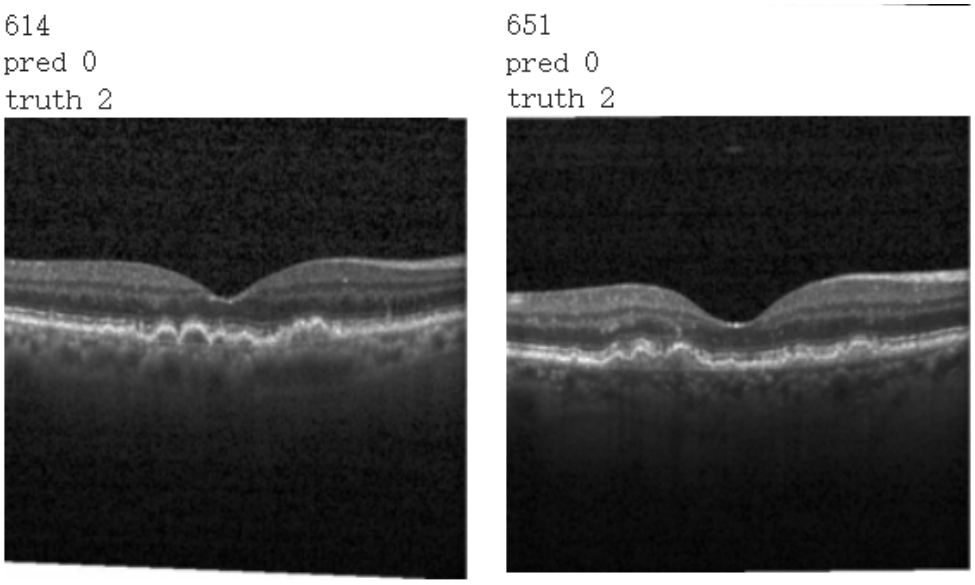

Fig 7. Examples of the classification mistakes made by ResNet models

Above each picture is three lines of information of the specific instance. From top to bottom is the index of the picture, the predicted label made by the model and the actual label of the picture. 0,1,2,3 corresponds to CNV, DME, DRUSEN and NORMAL respectively.

2. Additional thoughts

Despite the pleasing results achieved by ResNet models on this dataset, the capability of deep learning models may not stand up to all other tasks. In fact, I used ResNet models on another dataset (from the same dataset source). The task was to examine chest X-ray images and tell whether there was pneumonia or not. The seemingly easier 2-class classification task did not receive desirable outcomes. As is shown in Fig. 8, although an accuracy of around 85% could be obtained, the test accuracy was not increasing with the number of epochs but fluctuated vastly. That would mean that the

learning of the model was not in the right direction at all. Different ResNet structures and hyperparameters were tried but none of them displayed reliable results. For one thing, ResNet models may not be suitable for this task. There isn't a one-model-for-all for different tasks, or even different datasets. For every task, the choice of model and tuning of parameters would still be challenging. For two, this task may actually be harder compared to the 4-class retinal disease classification task. The boundary between pneumonia and normal lungs may be not contrastive in X-ray images whereas the differences between CNV, DME, DRUSEN and NORMAL are quite distinctive. As a complete ophthalmology outsider, it took me around 1 hour to learn how to differentiate the four types (although I could not guarantee a 99% accuracy).

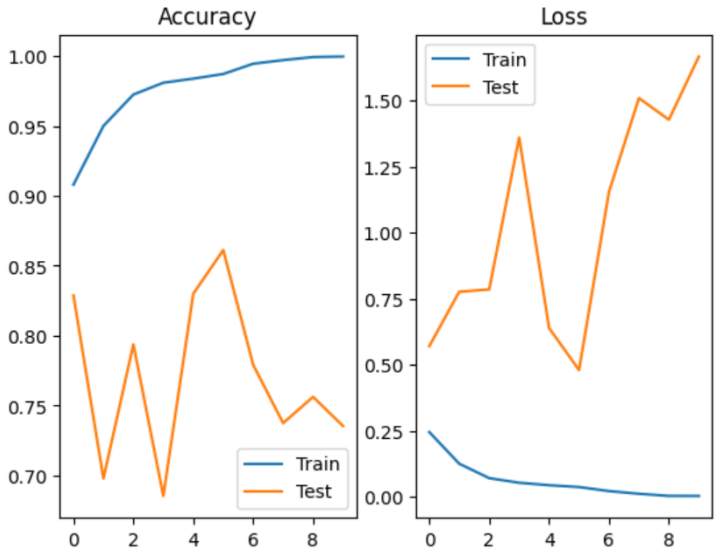

Fig 8. Accuracy and Loss of X-ray image classification using ResNet

References:

[1]Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5):1122–1131, 2018.

[2]https://www.aao.org/young-ophthalmologists/yo-info/article/oct-how-it-works-and-when-to-use-it

[3]Abhishek Vahadane, Ameya Joshi, Kiran Madan, and Tathagato Rai Dastidar. Detection of diabetic macular edema in optical coherence tomography scans using patch based deep learning. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pages 1427–1430. IEEE, 2018.

[4]http://webeye.ophth.uiowa.edu/eyeforum/tutorials/retinal-detachment-med-students/index.htm
[5]Wang Z, Zhang W, Sun Y, Yao M, Yan B. Detection of Diabetic Macular Edema in Optical Coherence Tomography Image Using an Improved Level Set Algorithm. Biomed Res Int. 2020 Apr 30;2020:6974215. doi: 10.1155/2020/6974215. PMID: 32420362; PMCID: PMC7210525.
[6]K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
[7]Kermany, D., Zhang, K., Goldbaum, M.: Labeled optical coherence tomography OCT and chest x-ray images for classification. Mendeley data (2018). https://data.mendeley.com/datasets/rscbjbr9sj/3
[8]Chetoui, M., Akhloufi, M.A. (2020). Deep Retinal Diseases Detection and Explainability Using OCT Images. In: Campilho, A., Karray, F., Wang, Z. (eds) Image Analysis and Recognition. ICIAR 2020. Lecture Notes in Computer Science(), vol 12132. Springer, Cham. https://doi.org/10.1007/978-3-030-50516-5_31
[9]Tasnim, Nowshin & Hasan, Mahmudul & Islam, Ishrak. (2019). Comparisonal study of Deep Learning approaches on Retinal OCT Image.
[10]He, K., Zhang, X., Ren, S., Sun, J. (2016). Identity Mappings in Deep Residual Networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9908. Springer, Cham. https://doi.org/10.1007/978-3-319-46493-0_38
[11]Saman Sotoudeh-Paima, Ata Jodeiri, Fedra Hajizadeh, Hamid Soltanian-Zadeh, Multi-scale convolutional neural network for automated AMD classification using retinal OCT images, Computers in Biology and Medicine, Volume 144, 2022, 105368, ISSN 0010-4825. https://doi.org/10.1016/j.compbiomed.2022.105368.
[12]Little, Karis & Ma, Jacey & Yang, Nan & Chen, Mei & Xu, Heping. (2018). Myofibroblasts in macular fibrosis secondary to neovascular age-related macular degeneration - the potential sources and molecular cues for their recruitment and activation. EBioMedicine. 38. 10.1016/j.ebiom.2018.11.029.
[13]M. Goldbaum et al., "Automated diagnosis and image understanding with object extraction, object classification, and inferencing in retinal images," Proceedings of 3rd IEEE International Conference on Image Processing, Lausanne, Switzerland, 1996, pp. 695-698 vol.3, doi: 10.1109/ICIP.1996.560760.
[14]Yankui Sun, Shan Li, Zhongyang Sun, "Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning," J. Biomed. Opt. 22(1) 016012 (20 January 2017) https://doi.org/10.1117/1.JBO.22.1.016012
[15]Abirami, M.S., Vennila, B., Suganthi, K. et al. Detection of Choroidal Neovascularization (CNV) in Retina OCT Images Using VGG16 and DenseNet CNN.

Wireless Pers Commun 127, 2569–2583 (2022).
https://doi.org/10.1007/s11277-021-09086-8