Assignment 1 - CAP 5516

Michael Fielder
University of Central Florida
College of Engineering and Computer Science
Department of Computer Science
MichaelFielder@knights.ucf.edu

The code for this assignment can be accessed by clicking this blue link: Google Colab Link.

1. Models and Hyperparameters

This assignment contains two models. Both of these models are constructed using ResNet-18 as their base. The major difference between these two models is that one has been initialized using random weights and the other was pre-trained on the ImageNet dataset. In this paper, *Model 1* is the model with randomly initialized weights and *Model 2* is the one that is pre=trained. Another thing to note is that the fully-connected layer at the end of this program has been modified to only classify 2 classes instead of 1000. This is because the dataset only denotes whether an X-Ray does or does not contain Pneumonia. A Softmax layer was also added at the end of the model to make comparisons easier.

In terms of hyperparameters, both networks were trained for 25 epochs, with a batch size of 32. Both used Cross-Entropy Loss and the Adam optimizer. A learning rate of 0.001 and a weight decay of 0.0001 were used.

Several image preprocessing techniques were also applied to the dataset including resizing the images to 256 x 256, center cropping them to as size of 224 x 224, and finally converting each image to a tensor so that they can be interpreted by the models.

Figure 1. Training loss for both models

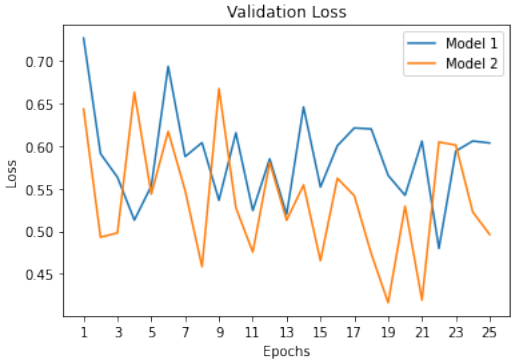

Figure 2. Validation loss for both models

2. Training and Validation Loss

The pre-trained Model 2 possessed a lower loss compared to Model 1 during the entire training process. This can most likely be attributed to the fact that Model 2 utilized transfer learning for its weights. It most likely learned general properties about images which made it better in a different application. **Figure 1** shows the compared losses between Models 1 and 2 during training.

Validation loss was inconclusive due to the fact that the validation dataset only contains 16 images. This makes the loss more erratic during evaluation. Usually, the testing set would be smaller than the validation set but in this case, it's the opposite. To make this dataset better, it would be beneficial to change the number of validation images. **Figure 2** shows the erratic results of the validation loss for Model 1 and Model 2.

3. Testing Classification Accuracy

Figure 3 shows the total accuracy and class accuracy for both Model 1 and 2. As you can see, Model 2 gets a higher accuracy in every category. It seems to show that transfer learning positively improves the performance of a model

Category	Model 1	Model 2
Normal	61.96%	64.52%
Pneumonia	98.46%	98.46%
Total	84.77%	85.73%

Figure 3. Accuracies for both models

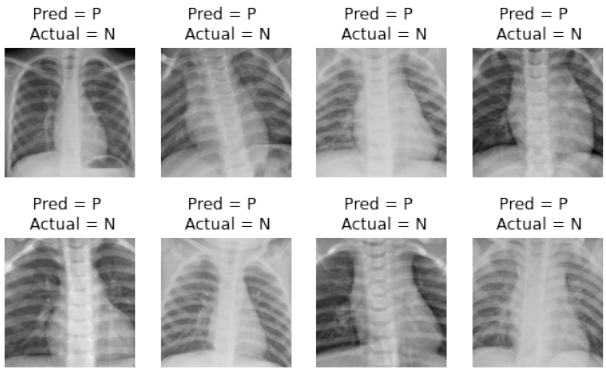

Figure 4. Validation loss for both models

when used. Some other things to note are that both models seem to favor pneumonia images over normal ones. This could be due to an imbalance in the data which could be remedied using image synthesis. I unfortunately could not do this due to realizing it too late in the process.

4. Failure Cases

Images that were misclassified by both models is shown in **Figure 4**. In general, both models misclassified similar images. As show in the previous section, most of these happened to be Normal images. In the case of this dataset, it's also interesting to note that most of the images are similar in orientation.

5. Conclusion

The biggest change that could be made is to add more synthetic data to give the model more examples of the Normal class. This would involved applying image filters that would simulate real-world x-ray conditions.