ALPhA: Week 3

Jose Cruz and Andrew Hoyle

Goals for the past week

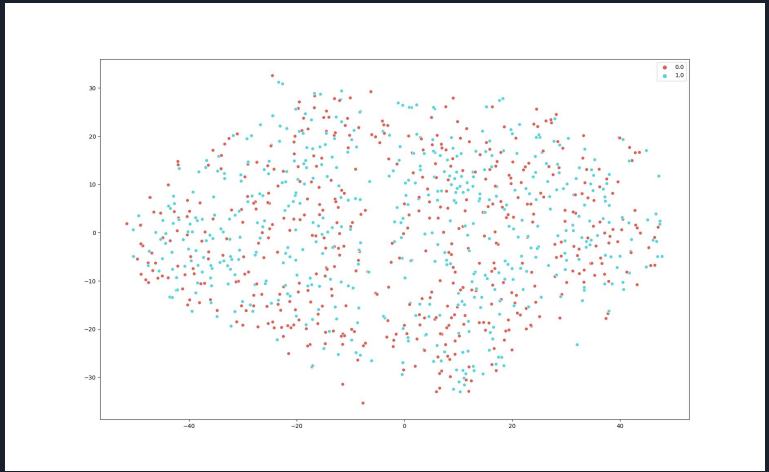
- Run our data through the Logistic Regression Algorithm from scikit-learn
- Use T-SNE to visualize our data
- Add new layers to our VGG16 base model and train the model

Logistic Regression Algorithm

- Horrible Results
  - Training Accuracy: ~70%
  - Validation Accuracy: ~20%
    - Worse than guessing

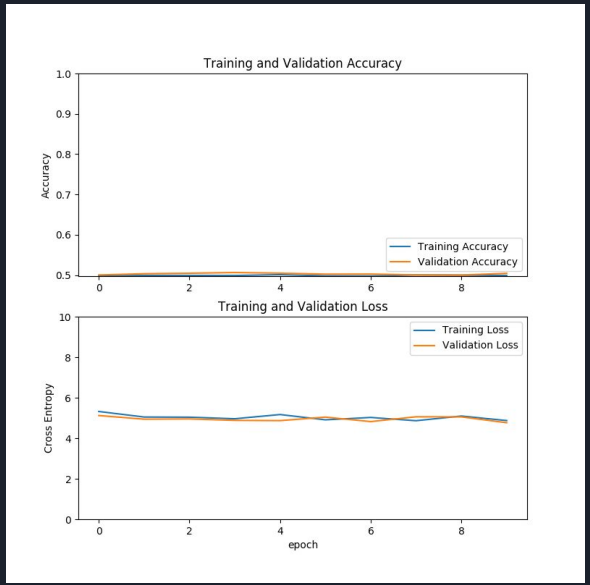
T-SNE Data Visualization

- Visualized the vectors from our base model using T-SNE
  - No distinguishable difference between positive and negative data samples



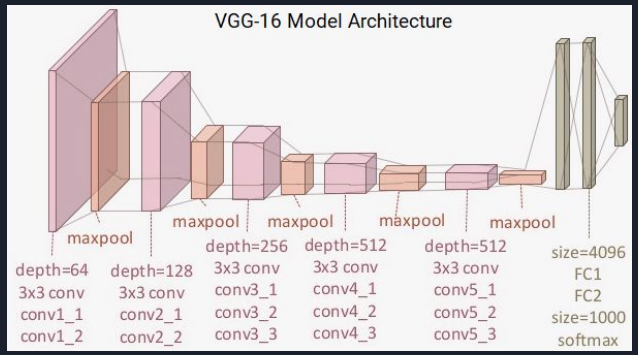
Convnet Classification Layer

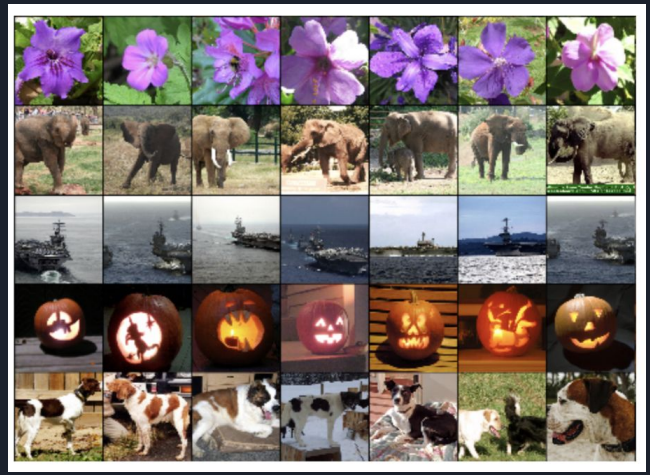
- Added an extra layer onto our VGG16 base model
- Trained on roughly 27000 images, Validated on roughly 9000 images
  - 10 epochs, batch size of 32



Reasons for our results

- Differences between VGG16 training images (ImageNet) and our data
- Lack of info in our images
- VGG16 Architecture





Next Steps

- Use fewer layers of VGG16 and assess accuracy
- Maybe build a Convnet from scratch
- Reformat Data and reconsider problem