Contributions: 1. Clearly state everyone's contributions in your team, not only your own contribution. 1 point out of 10 will be deducted if this part is missing on your report.

Alan Leon: Code, Testing Model on Outside Images, Report

Model Description and Methodology: 2. Please describe your model in detail and what schemes you have tried to improve the performance.

I conducted trial and error to finally arrive at a model with high prediction accuracy not only on LOOCV, but also on images outside the dataset. For my finalized model, I created a preprocess function that takes the image by using facial detection in order to crop all the images to a set size (224,224), this is to eliminate the influence of noise in the prediction process of the model later on. Then I created a VGG model with pre-trained weights from 'imagenet'. Then I froze the pretrained layer, flattened the output layer to 1 dimension, and added a dense layer with softmax activation for the number of classes.

Moreover, during the model compilation the model is trained using LOOCV but also with the augmentation images. However, the augmentation images are never a part of the validation set only for training and these 9 augmentations are applied to every single image for training using an apply augmentations function, so for each training fold the with number validation images being 180 which is the size of the Celebrities Dataset divided 10 for 10 number of training images are ((1800-180)+(9*(1800-180))=16200) where (9*(1800-180)) are augmented images.The list of the performed augmentations for this model include:

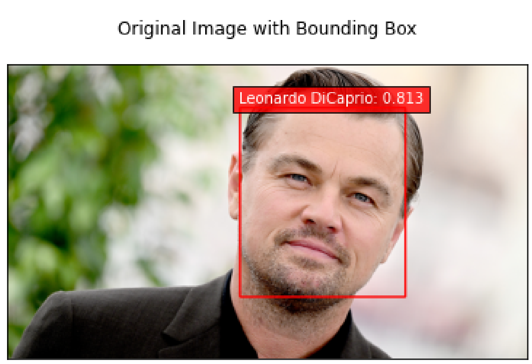
1. Applying Grayscale
2. Applying Darken (Shadow effect)
3. Increasing Contrast
4. Increasing Saturation
5. Horizontal Flip
6. Horizontal Flip and Grayscale
7. Horizontal Flip and Darken
8. Horizontal Flip and Contrast Increase
9. Horizontal Flip and Saturation Increase

The reason I utilized these augmentation techniques was I found that these augmentations to be optimal after trial and error as they did not harshly modify the images to the point where they are unrecognizable to the model, but they are just enough to enhance the prediction accuracy of the model. The reason I only used horizontal flip for combination augmentations was not only for image recognizability reasons, but because I had found it the best way to manipulate the orientation of the faces in the images within our dataset in contrast to methods like shearing which were more of detriment than an enhancement.

Here is an image displaying a varying range of augmentations I had tried for previously run models:

Lastly, to validate on images outside the dataset, the image is processed using previous functions and is used as a primary input to the model and the same processed image with histogram equalization is applied as a secondary input. Then the highest average probability from these two inputs are considered as the predicted label (this technique is known as ensembling even though there are not two separate models).

Here is an image showing an example of what the Predicted Probability Labels Look Like:



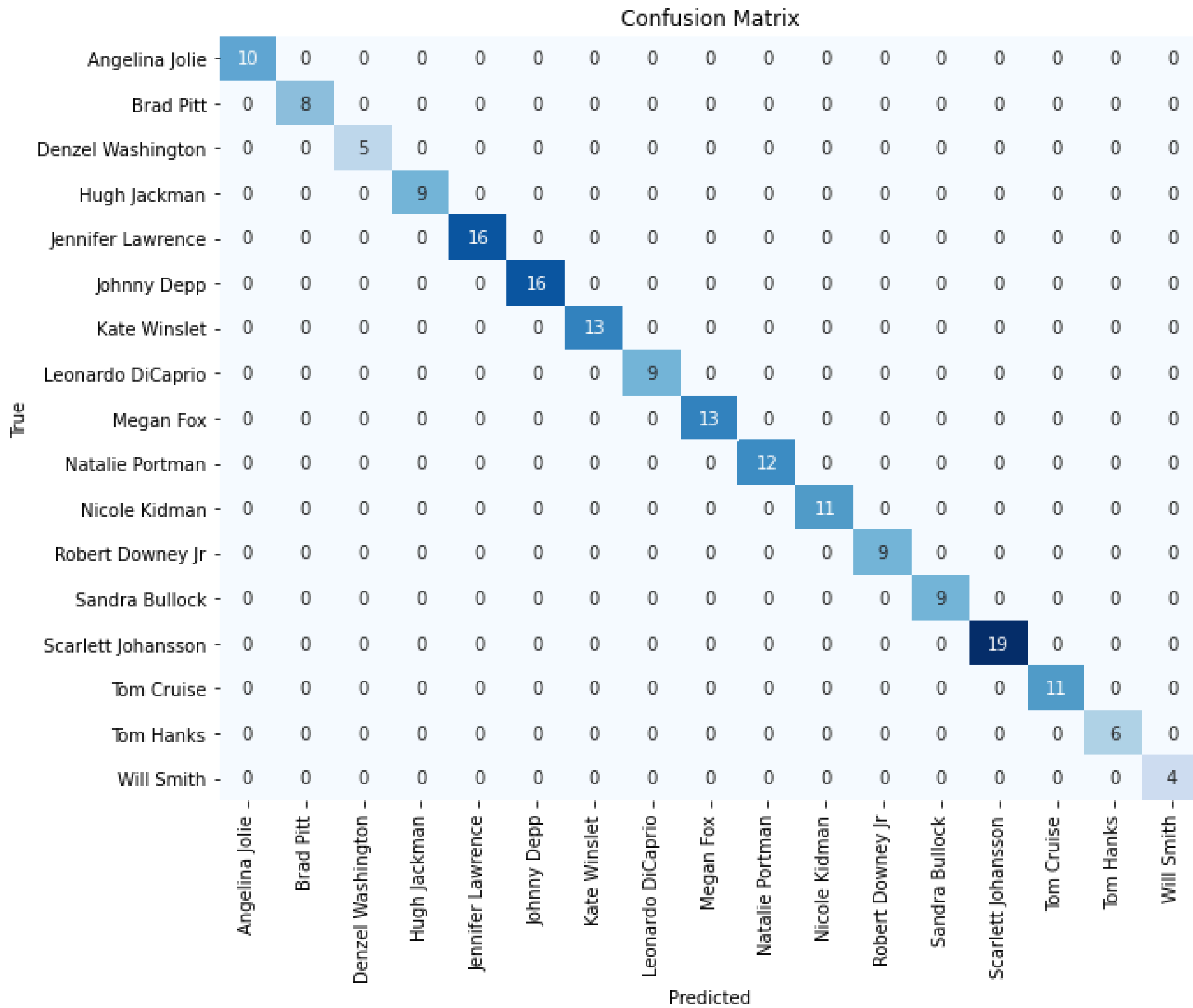

```
Ensemble Probability for Angelina Jolie: 7.277238864844549e-07
Ensemble Probability for Brad Pitt: 0.0809137225151062
Ensemble Probability for Denzel Washington: 1.307296861341456e-05
Ensemble Probability for Hugh Jackman: 0.07461114227771759
Ensemble Probability for Jennifer Lawrence: 2.783345962598105e-06
Ensemble Probability for Johnny Depp: 1.9214370695408434e-05
Ensemble Probability for Kate Winslet: 2.4467324237775756e-06
Ensemble Probability for Leonardo DiCaprio: 0.8133200407028198
Ensemble Probability for Megan Fox: 2.7386017009689567e-08
Ensemble Probability for Natalie Portman: 5.7594451874365404e-08
Ensemble Probability for Nicole Kidman: 2.363402563787531e-05
Ensemble Probability for Robert Downey Jr: 0.00010580607340671122
Ensemble Probability for Sandra Bullock: 9.026915326337814e-10
Ensemble Probability for Scarlett Johansson: 6.0369919083314016e-05
Ensemble Probability for Tom Cruise: 0.0007151822210289538
Ensemble Probability for Tom Hanks: 0.0032557977829128504
Ensemble Probability for Will Smith: 0.026955926790833473
```

LOOCV accuracy: 3. Report the face recognition accuracy using LOOCV scheme on the Celebrity dataset.

```
Training on fold 1/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 339s 668ms/step - loss: 0.4543 - accuracy: 0.8853 - val_loss: 0.8187 -
val_accuracy: 0.7833
Training on fold 2/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 360s 709ms/step - loss: 0.0492 - accuracy: 0.9925 - val_loss: 0.1212 -
val_accuracy: 0.9611
Training on fold 3/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 360s 710ms/step - loss: 0.0102 - accuracy: 0.9996 - val_loss: 0.0235 -
val_accuracy: 0.9944
Training on fold 4/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 371s 731ms/step - loss: 0.0048 - accuracy: 0.9999 - val_loss: 0.0097 -
val_accuracy: 1.0000
Training on fold 5/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 366s 722ms/step - loss: 0.0031 - accuracy: 1.0000 - val_loss: 0.0043 -
val_accuracy: 1.0000
Training on fold 6/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 347s 684ms/step - loss: 0.0021 - accuracy: 1.0000 - val_loss: 0.0039 -
val_accuracy: 1.0000
Training on fold 7/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 349s 688ms/step - loss: 0.0016 - accuracy: 1.0000 - val_loss: 0.0037 -
val_accuracy: 1.0000
Training on fold 8/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 355s 700ms/step - loss: 0.0012 - accuracy: 1.0000 - val_loss: 0.0027 -
val_accuracy: 1.0000
Training on fold 9/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 354s 699ms/step - loss: 8.8345e-04 - accuracy: 1.0000 - val_loss: 0.00
20 - val_accuracy: 1.0000
Training on fold 10/10
Number of training images: 16200
Number of validation images: 180
507/507 [==============================] - 353s 696ms/step - loss: 6.7553e-04 - accuracy: 1.0000 - val_loss: 0.00
17 - val_accuracy: 1.0000
```

Average Validation Accuracy: 0.97388

Confusion Matrix: 4. Draw a confusion matrix, and discuss who's faces can be easily confused with others.



From trial and error I learned that the model can distinguish very minor facial features for prediction. Here are some of the commonly confused faces I noticed and the reasoning I believe behind them.

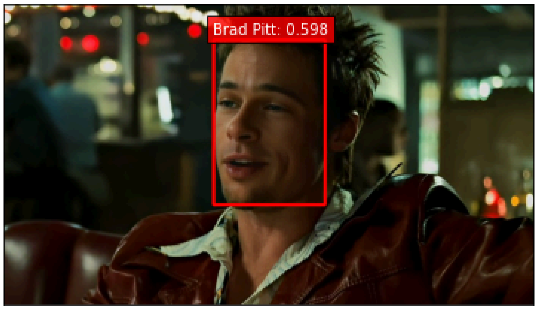
Brad Pitt image incorrectly predicted Sandra Bullock, potential reason (Nose Shape)

Angelina Jolie image incorrectly predicted Megan Fox, potential reason (Face Shape, Makeup Style)

Brad Pitt image incorrectly predicted Denzel Washington, potential reason (Oval Face, Image Quality, Brightness)

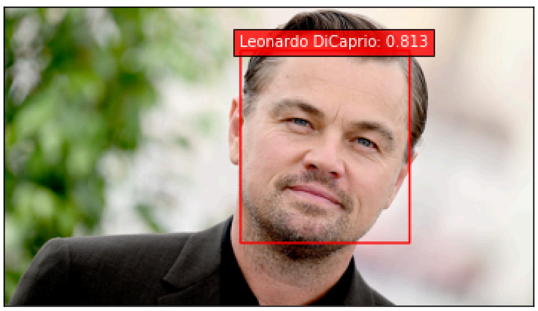
Final Results on Images Outside of Dataset With Ensemble (7/7 accuracy = 100%)

Original Image with Bounding Box



Processed Face

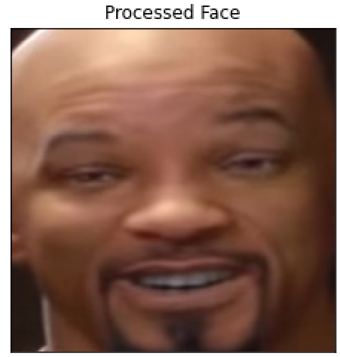
Original Image with Bounding Box



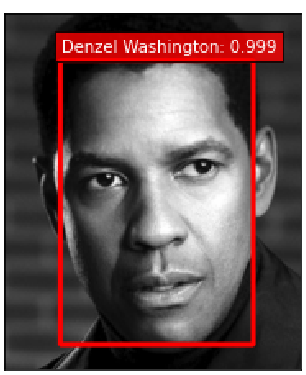
Processed Face

Original Image with Bounding Box

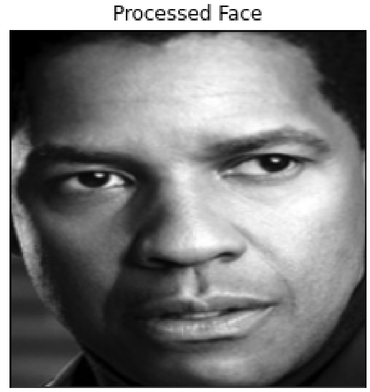
Processed Face



Original Image with Bounding Box



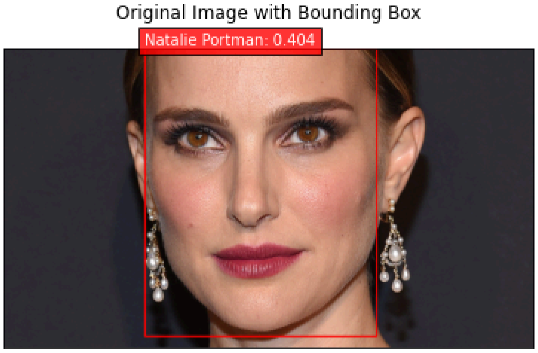
Processed Face



Original Image with Bounding Box

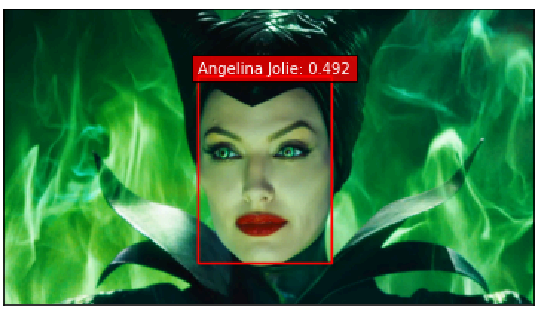
Processed Face

Original Image with Bounding Box

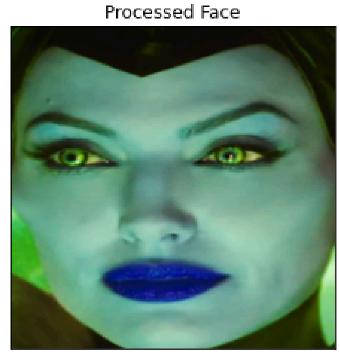


Processed Face

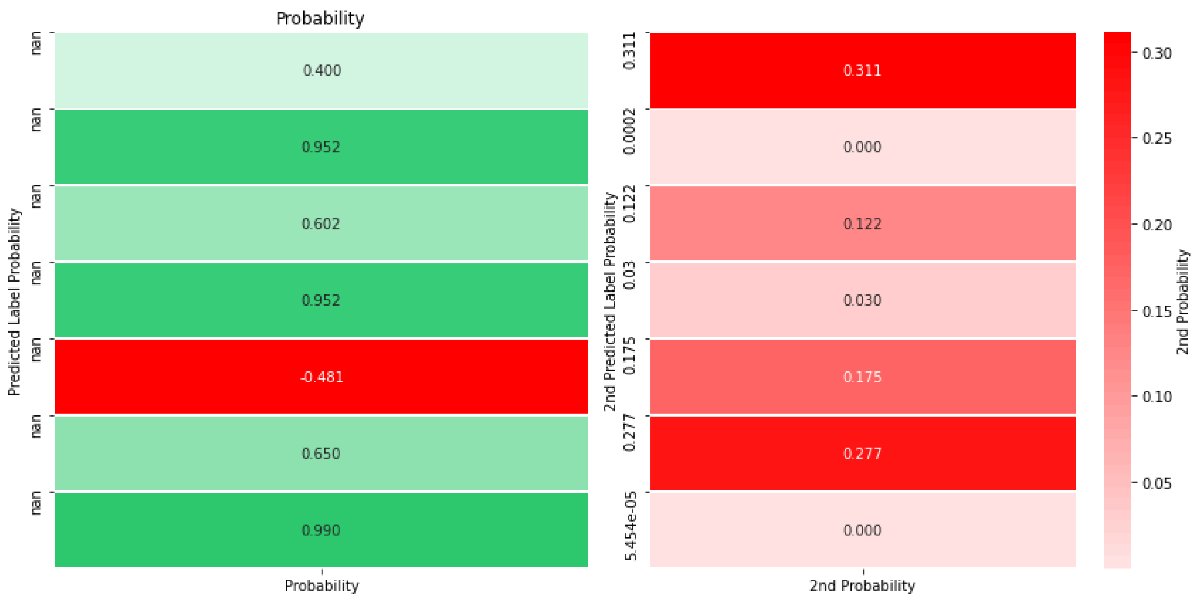
Original Image with Bounding Box



Processed Face

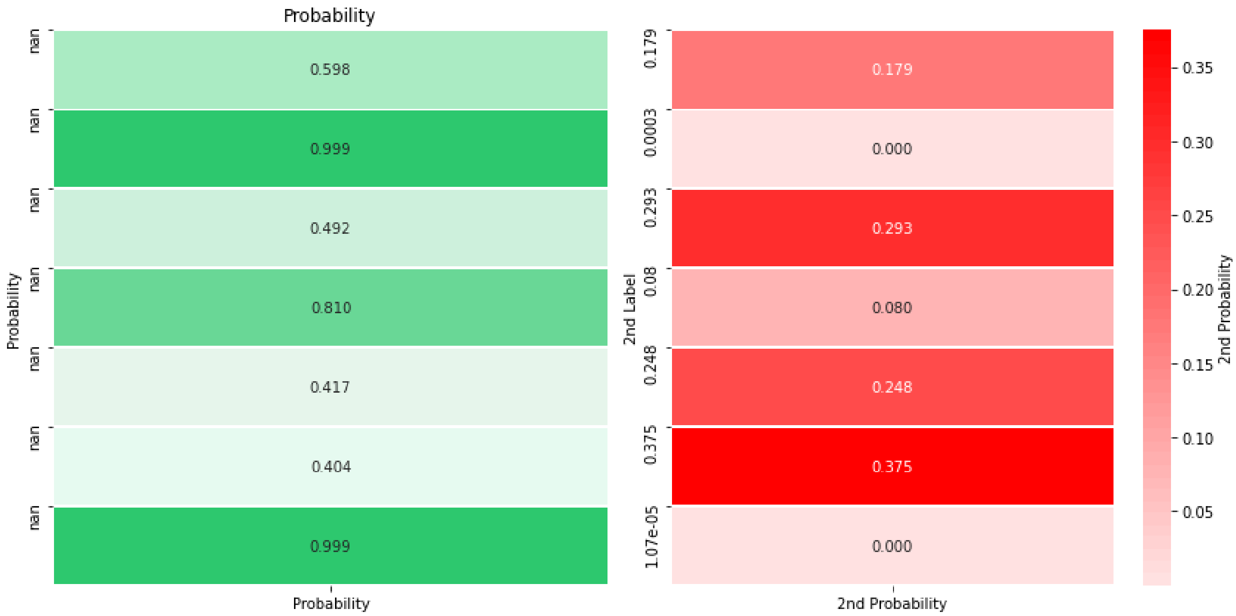


Heatmaps for Predicted Labels Without Ensemble



For the Model without ensemble most of the Predicted Labels are correct except for the image pitt_bald_test.jpg. It is Predicted as Denzel Washinton with a probability of 0.481 hence the red representation in the first heatmap for the incorrect prediction. So the model without an ensemble has 6/7 accuracy for the 7 test images. Similarly, the second heatmap revolves around the 2nd Highest Predicted Labels Probabilities. Which is relatively low for the most part with an average of 0.13 Probability for the Second Highest Predicted Label, which is good considering we want a high Prediction Probability for the correct label and low ones for the rest of the labels. It is also important to note the 2nd Highest Probability Predicted Label for the incorrectly predicted image is not Brad Pitt but actually the Robert Downey Jr. label, meaning the correct label is not predicted in the first two highest probabilities but only the third. This is what was indicative of using ensemble combination rather than training a whole new model.

Heatmaps for Predicted Labels With Ensemble



For the Model without ensemble all of the Predicted Labels are correct. So the model without an ensemble has 7/7 accuracy for the 7 test images. Similarly, the second heatmap revolves around the 2nd Highest Predicted Labels Probabilities. Which is a little higher than the model without ensemble an average of 0.167 Probability for the Second Highest Predicted Label, this is still good since it is relatively low. It is important to note that the ensemble portion without histogram equalization as a processing step of the testing image classifies the pitt_bald_test.jpg incorrectly but the other portion of the ensemble with histogram equalization as a preprocessing step of the test image correctly classifies this image but incorrectly classifies two other images that the without histogram equalization portion classified correctly consisting of portaman_test.jpg and jolie_hair_eyes_test.jpg. Averaging the probability results of these two ensembled models (6/7) and (5/7) is what allows for the (7/7) accuracy to be achieved.

Conclusion

To conclude the finalized model is robust on images inside and outside the dataset. Nevertheless, there are some limitations, I did not use many images outside of the dataset that include obscurities of the eye except one, it may be hard to validate on images outside the dataset where the subject is wearing sunglasses or makeup. If I had more time on this project, I would use GAN's in my training like BeautyGAN or CirleGAN to account for these potential scenarios and in order to increase the robustness of the model. Ultimately, this model is a good foundation for simple purposes like a singular subject input images without obscurities. But this model needs to be developed for more advanced purposes/scenarios such as facial recognition with obscurement, makeup, recognition of future images (images when subjects are much older), recognition past images (images where the subjects are much younger), facial recognition with many other subjects in shot, and facial recognition in the frames video.