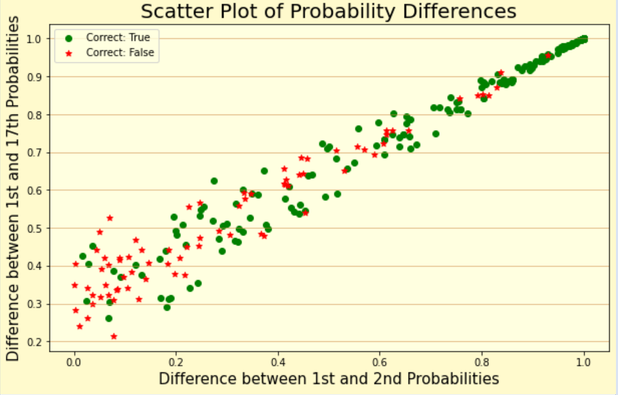
Analytically Determining Optimal Methods for Facial Recognition Models

By: Alan Leon

Introduction

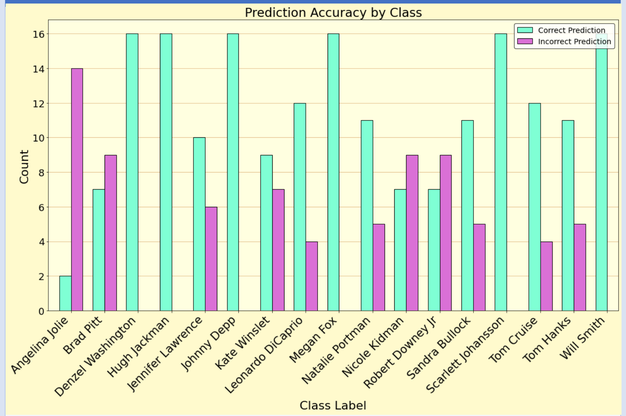
Utilized modelling techniques fall into categories: Color, Ensemble, Augmentations, & Regularization. Top left, right, & center, represent how model will be trained with/out Color, Ensemble, and Augmentations, respectively. Regularization cannot be visualized but it is a technique used to prevent overfitting. All combinations of these 4 techniques (2^4=16) used to predict on 17 different classes(people), totaling 272 predictions to analyze.

High Level View



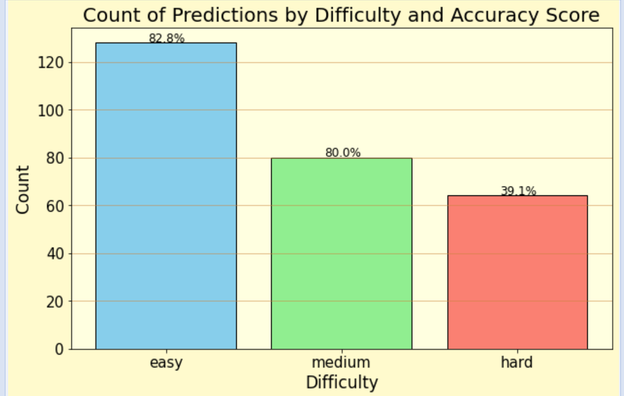
This is a scatterplot showing difference between 1st and 2nd Highest prediction probabilities against difference between highest and lowest prediction probabilities, Not only does this show the relationship between these two items, but also overall model performance, some of the models are overfitting demonstrated by the medium density of red stars ranging from (0.5,0.5) upwards. This is because the model is predicting incorrect values (stars) with high certainty meaning the model for these observations are not generalizing. Although, this visualization does not pinpoint the problematic models.

Class Labels



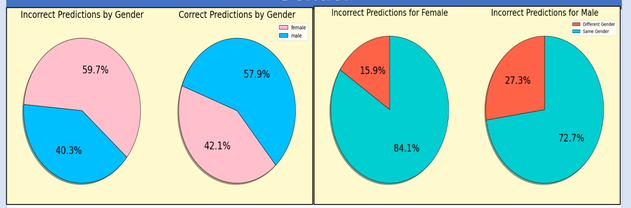
Classification of each label (person) is shown above. This is significant as if there are higher number of correct predictions than incorrect for all labels we can just use an ensemble of all model types to achieve a 100% prediction accuracy on the validation images. Since that is not the case as can be seen by labels such as Angelina, Brad, Nicole, and Robert. It is important to determine which model types result in these poor predictions, so if necessary for prediction all the high quality models can be combined for high accuracy ensemble prediction.

Image Difficulty



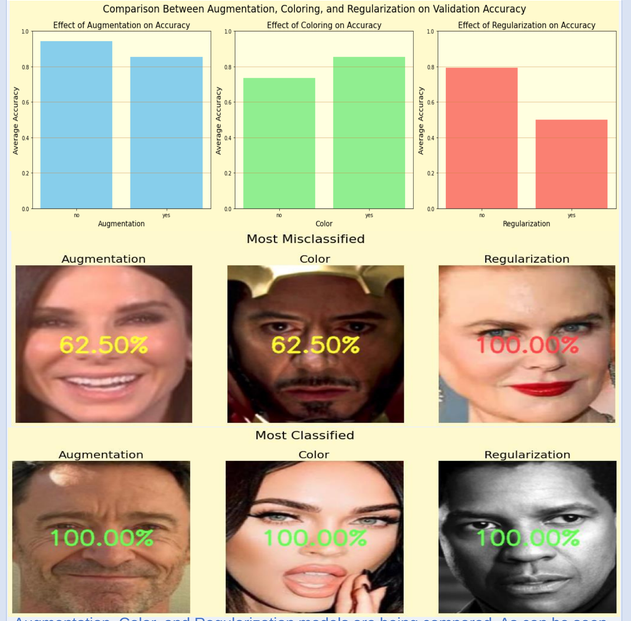
Images are labeled into difficulty groups consisting of easy, medium, & hard depending on angle, lighting, facial obstructions, and quality of the image. This plot displays the count of images present within each of these categories and the respective accuracy scores within each difficulty level annotated at the top of each bar. In short, the hard category has the least image count and significantly lower overall accuracy in contrast to easy/medium difficulty image types.

Gender



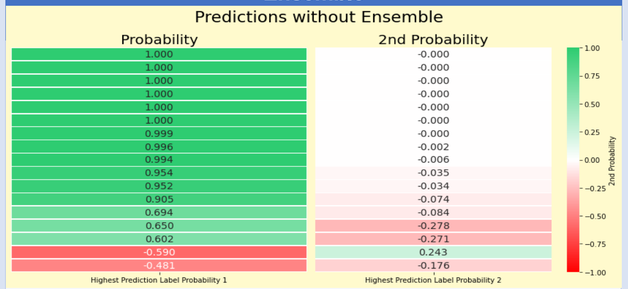
The purpose of the visualizations above are to determine whether gender of class label influences prediction accuracy. First with the plot on the left was made to deduce whether one gender class is causing majority of misclassifications which is found to be female classes, after reviewing the images this could be due to the presence of makeup in the images specifically smokey eye shadow making it difficult for the model distinguish between faces. The right plot conveys the false predictions are mostly the same gender, so adding a gender classifier will result in minor improvements, but in the future a makeup GAN could be used for training.

Comparing Model Types



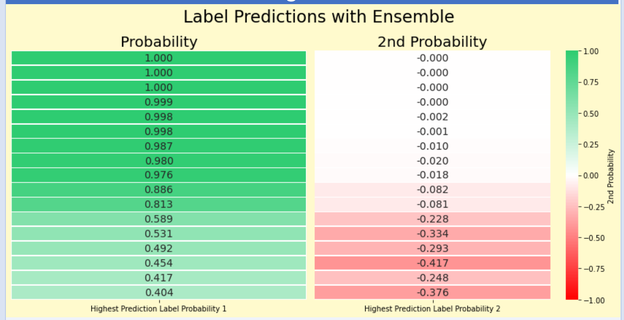
Augmentation, Color, and Regularization models are being compared. As can be seen by the plot on the top there are only minor differences when Color and Augmentation are used in models versus when they are not. However, when regularization is removed from models there is a significant boost in model prediction performance. Additionally, for the most misclassified labels, regularization model has more instances of misclassification shown by bottom figure. Minimal difference for top correct classifications.

Ensemble



So far, all model types have been evaluated except ensemble, so this section will focus on it. The figure above depicts the best model without ensemble. It classifies most of the class labels correctly with only 2/17 misclassifications (incorrect predictions in red). However, those mistakes have a relatively high probability, meaning the model is slightly overfit. The positive is that the 2nd probability for one of the misclassifications is the correct label (right plot bottom 2nd green), so the model is close but needs tuning.

Visualizing Final Results



Moreover, this figure illustrates the results of the same model from the figure above but with Ensemble incorporated (other factors remain unchanged). The specific method used was transforming the image using histogram equalization to predict on the images twice and combining the probabilities from each separate prediction (refer to image on top right of introduction section). From the results we can see that all the classes are correctly predicted but with a wide range of confidence ranging from 40% to near 100%. For future improvements of the optimal model a more representative accuracy score of model's predictions is desired.