Nicholas Newton, Tim Fitzpatrick

CV Part 3 Deliverable (2nd Update)

Project: Face Recognition

A. A short justification of the choice of classifier. For instance, if you selected SVM with RBF kernel, say why you think this classifier is good for your project. **(3 points).**

   For this project, we have decided to use Viola Jones for the face detection, and LBP to extract the features into a trained model. This model is used to identify images and discern if an image is in the training data. If the image is in the model, the model should be able to identify the image. Specifically, we are using LPBHFaceRecognizer as our recognizer which is optimized to detect facial features. The model created uses an 8x8 grid with 8 neighbors and a radius of 1 and a threshold of 1.7976931348623157e+308.

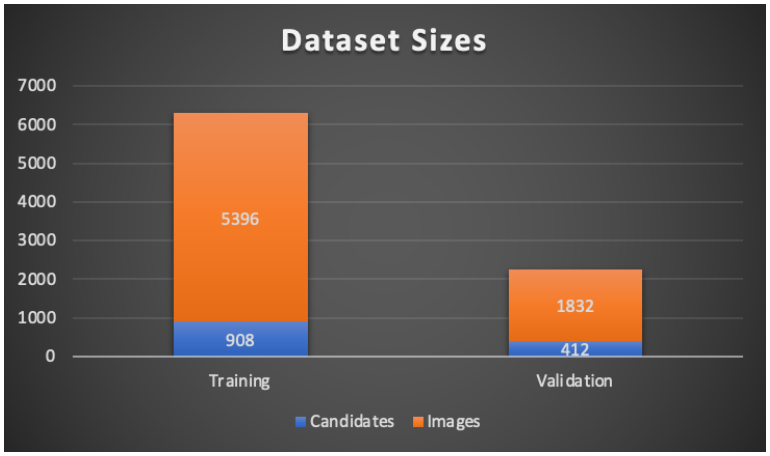
   Viola Jones is used for the face detection portion. It extracts the image of the face from the image and crops out the extra background, resizes, and converts it to grayscale. Once this is done, face detection can be used to find characteristics to differentiate faces. LBPH algorithm is optimized for our project to recognize different faces in a python environment that is not limited by the hardware of a local running laptop. The program should not need to utilize the power and processing capabilities on a platform like Google CoLabs.

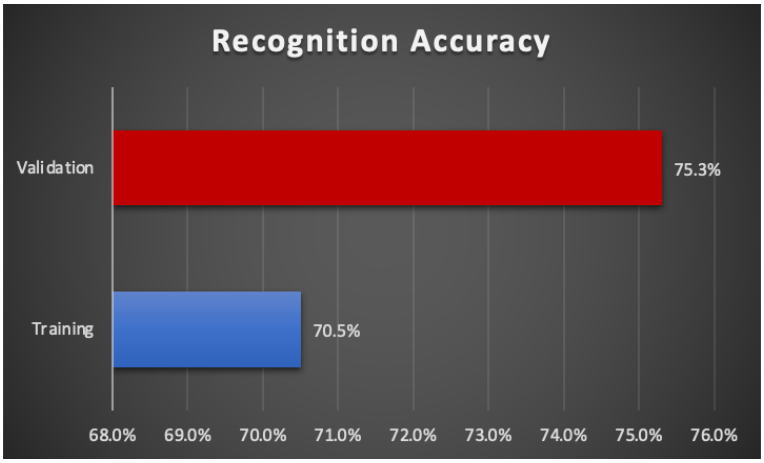
B. A classification accuracy achieved on the training and validation sets. That is, how many objects were classified correctly and how many of them were classified incorrectly (It is better to provide a percentage instead of numbers). **(3 points).**

   Our training data contained 5396 unique images. These were among 908 candidate sets. From our model, 70.5% of the images were identified correctly.

   Our validation data contained 1832 unique images. These were among 412 candidate sets. From our model, 75.3% of the images were identified correctly.

   Our data can be visualized with the graphs below:





C. A short commentary related to the observed accuracy and ideas for improvements. For instance, if you see almost perfect accuracy on the training set, and way worse on the validation set, what does it mean? Is it good? If not, what do you think you could do to improve the generalization capabilities of your solution? **(6 points)**

Based on our accuracy, we feel that we can move on to testing and training models on unknown data. For the next/final deliverable, we need to be able to take in image samples in real-time and be able to identify them correctly. Based on our results from training and validation data, we believe that there should be similar results on the unknown data. This is a good indicator overall for the program performance. To improve accuracy further we would likely need to run more intense models and compare image features using SVM.