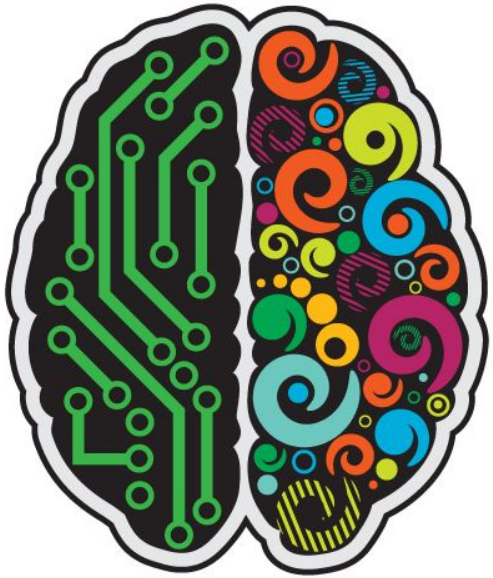
Facial Expression Recognition

Business Case

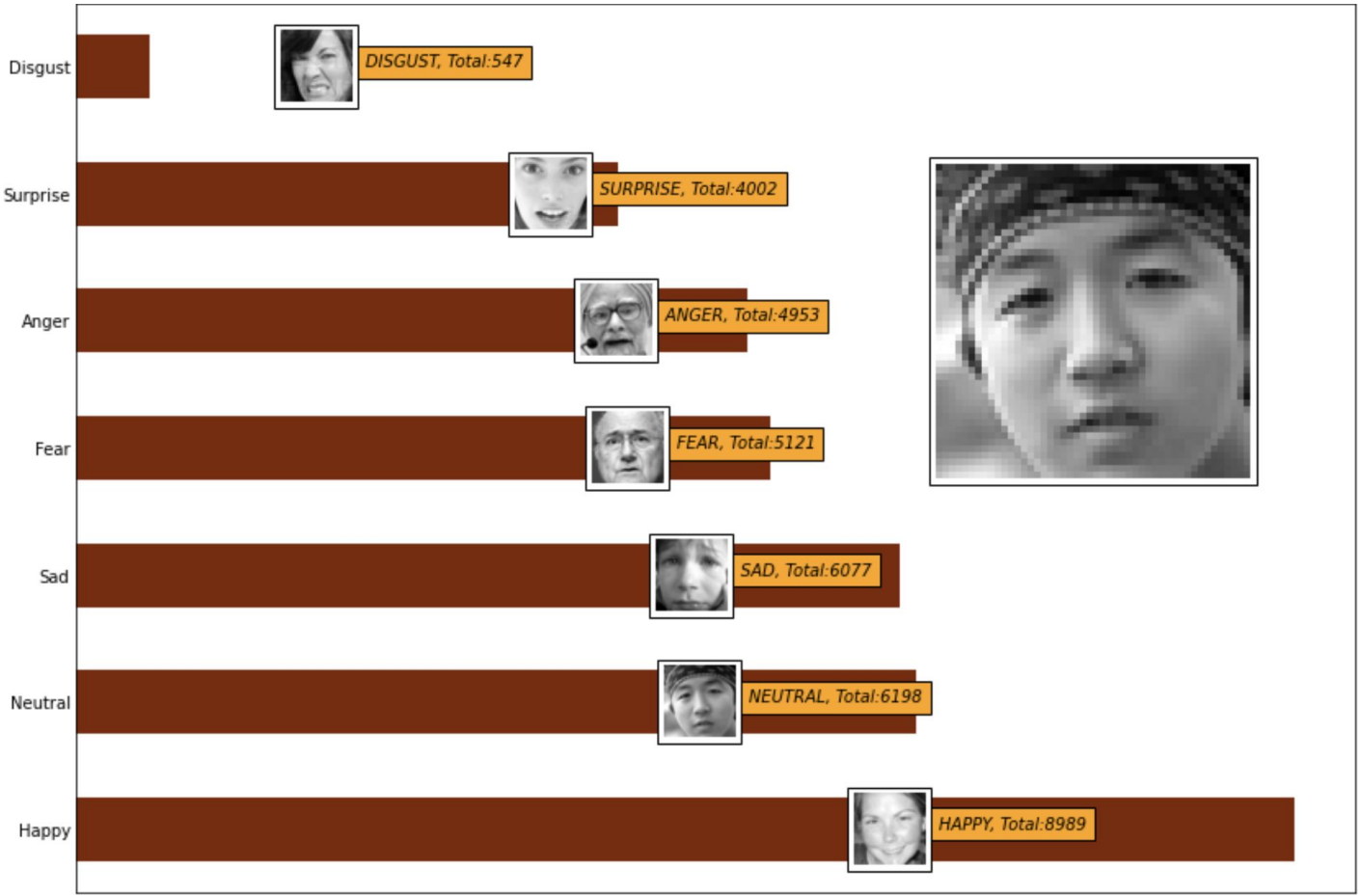
Helping those who suffer from Social-Emotional Agnosia (or the inability to interpret meaning from non verbal cues.



Why is This Important?

Experts say that 70-93% of communication is non-verbal meaning that the vast majority of information being conveyed in a conversation resides in visual cues such as body language or facial expressions. As a result, people who suffer from Social-Emotional Agnosia often have to learn to identify these social cues on their own which you can imagine is not as easy as it sounds. Using object detection and image classification techniques, a real-time classification of at least some of these nonverbal cues can be provided to aid those who struggle with these skills.

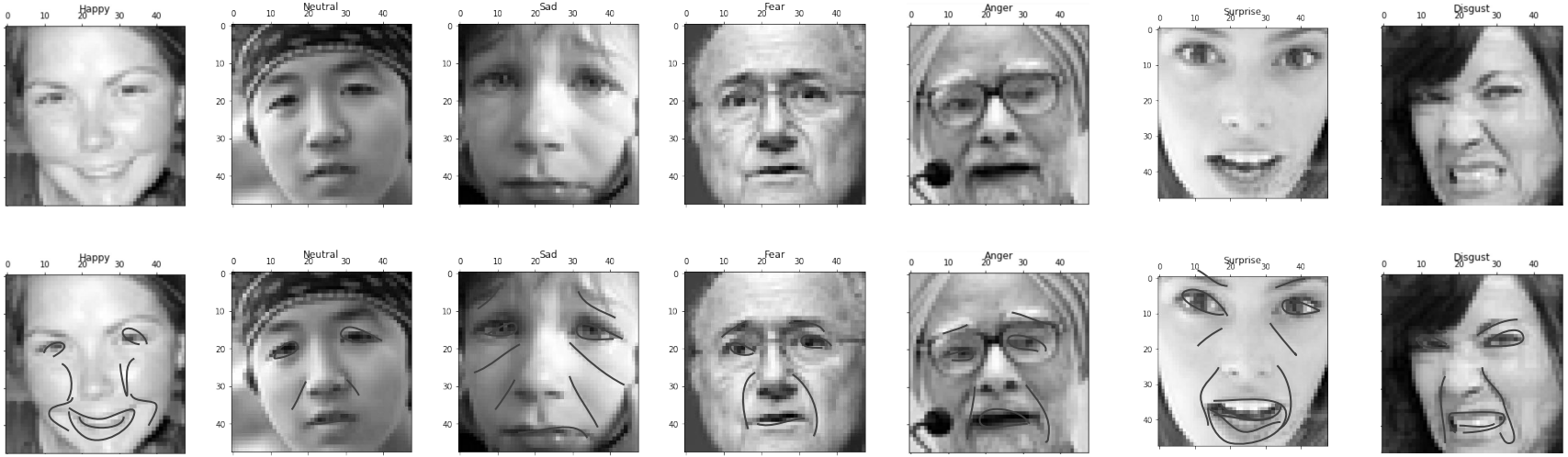
Class Imbalance



Data Cleaning

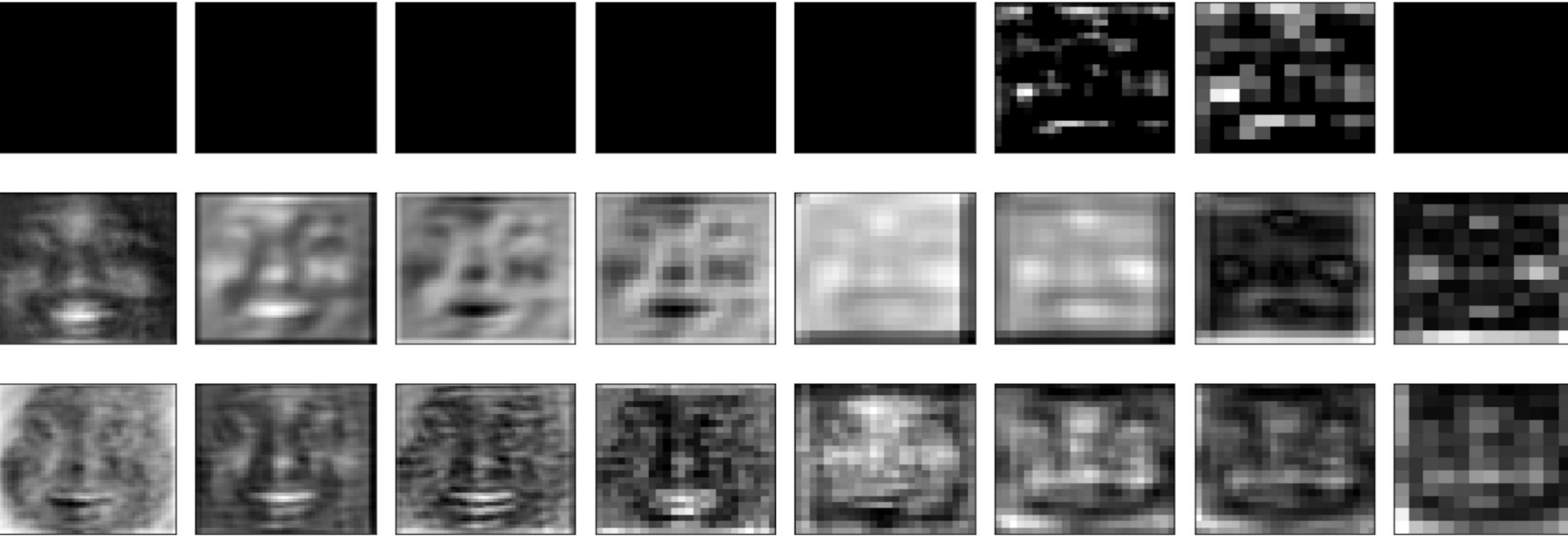
The data came in a csv where the image pixels were a string in the first column. Data cleaning consisted of parsing this dataframe and reformatting the pixels into the original 48x48 grayscale pictures and making train and validation tensors.

What Patterns Should The Model Recognize?



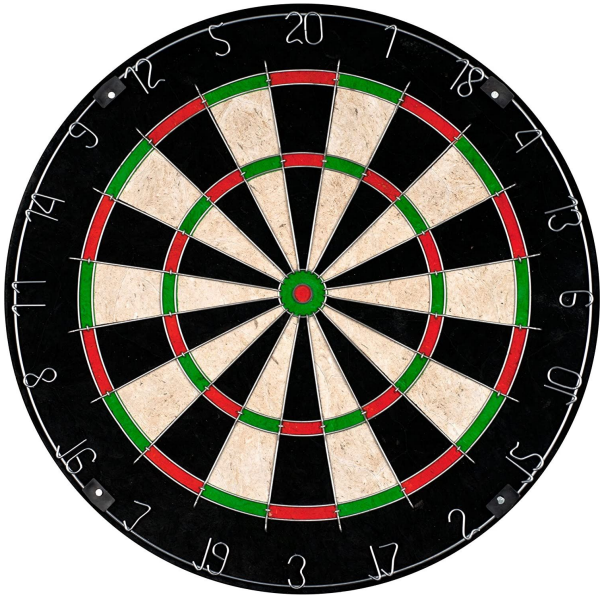
Modeling Process

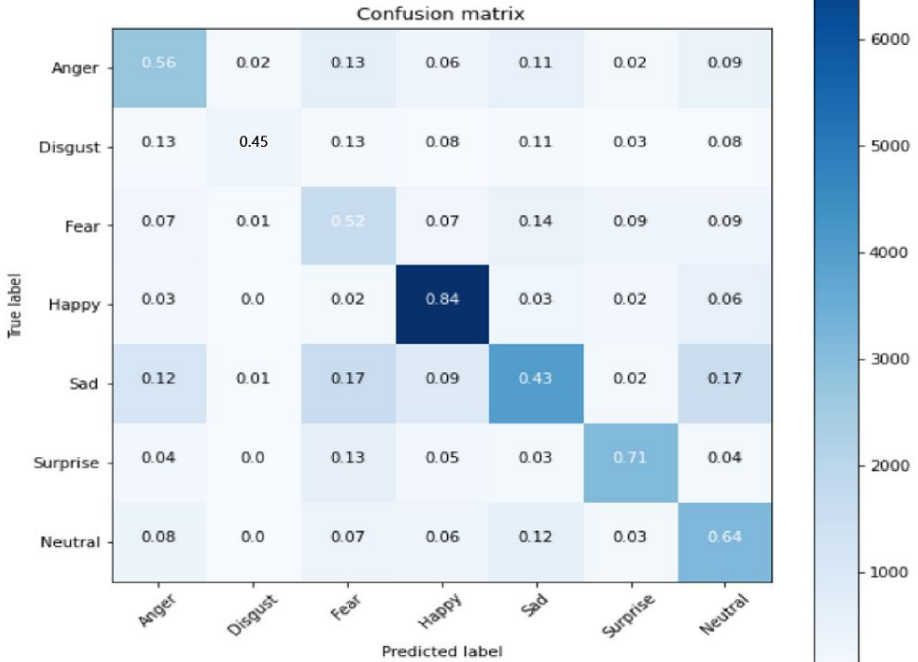
During the modelling process a wide range of models were created and analyzed on how each was coming to its conclusions. Through activation layer visualizations, the sequence conv layers that led to the best feature extraction was identified and dropout layers were added to combat overfitting.

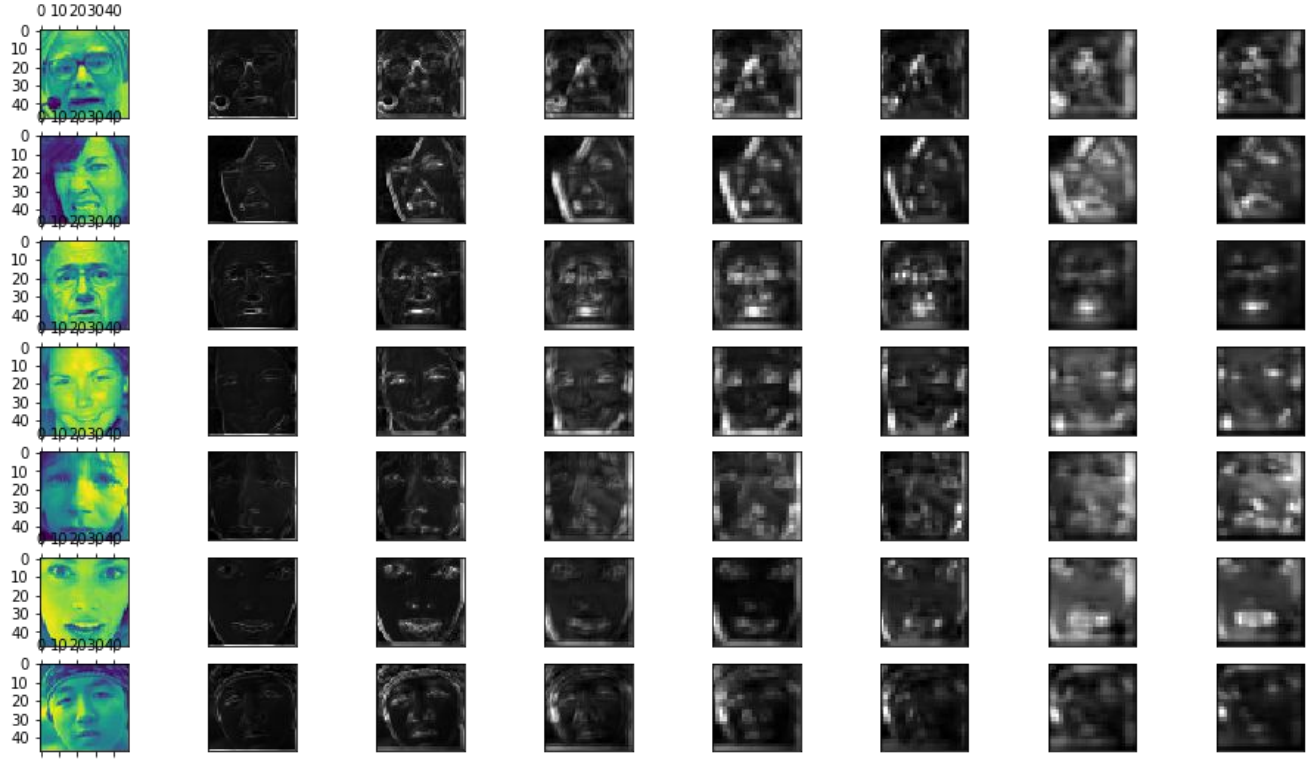


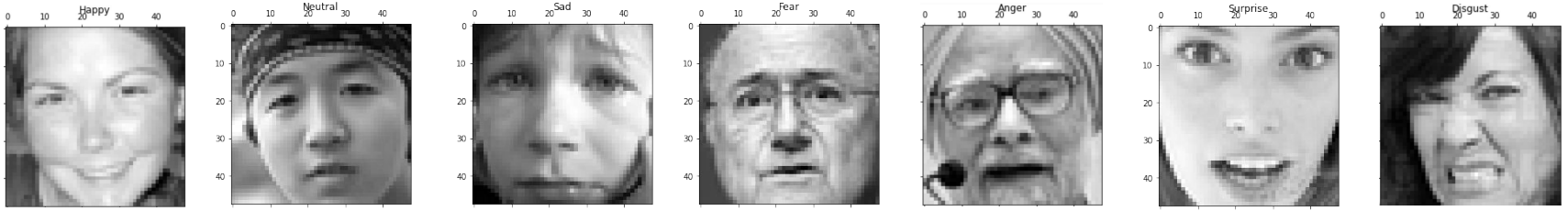
Evaluation Metrics

As the goal of this project is the be as generally accurate as possible and there is no preferred class, the evaluation metric used was an f1 score. The final model had an f1 of 61.2% on the validation data and 65% on the training data.



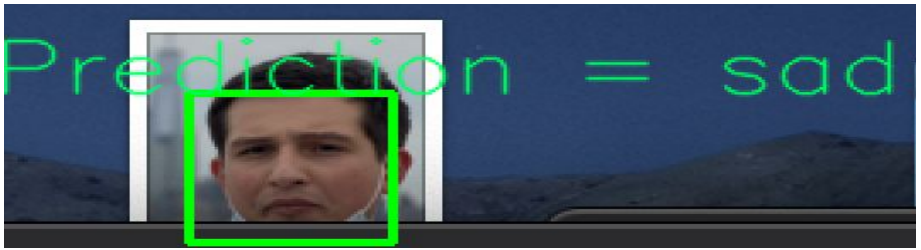





Happy
Neutral
Sad
Fear
Anger
Surprise
Disgust

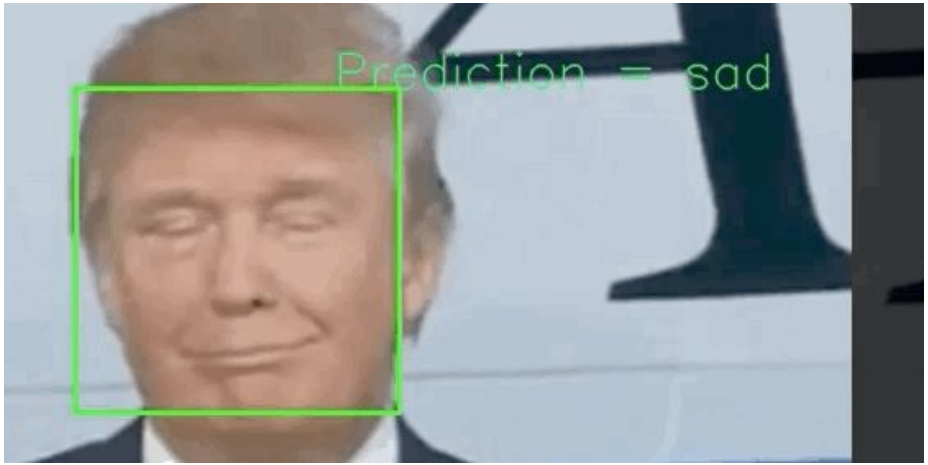
Real-time application

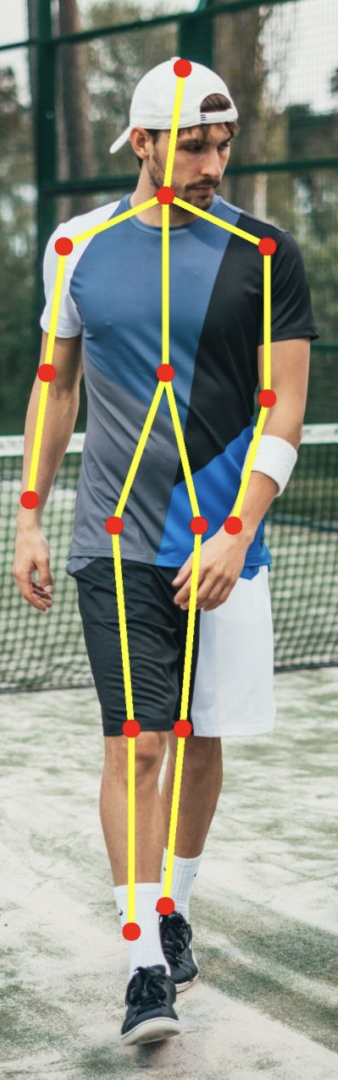
As a result of the models being trained on 48x48 images of faces, any image that is to be fed into a model has to be downscaled and converted to black and white. Also, because the models were trained on images with a single face, if the image inputted into a model does not have a face or has multiple faces, there is no way the model can make a meaningful prediction.



Facial Recognition

Using a Haar Cascade classifier and opencv, the coordinates of probable faces were obtained. These coordinates were then used to crop faces. The cropped faces were then downscaled to end up with a picture closer to what the training data was. This makes the model more accurate and allows for the prediction of multiple people at a time.





Pose Estimation

While facial expression is a major part of non-verbal communication, body language can be used in conjunction with an accurate classifier to add contextual information and therefore, further increase overall modeling accuracy.

Next Steps

- Improvements to the Haar Cascade Classifier can be made. It frequently struggles to find faces that are tilted more than 30 degrees and it often interprets stripes as faces.
- A new supplemental dataset can be constructed in which the classes are changed to positive and negative using pose estimation. This data can be used to train an even more accurate model.
- Higher resolution pictures may provide more accurate results. The model currently has trouble with tasks such as finding lips and often misinterprets parts of an image due to low resolution.
- A depth map would be an advantageous addition. Knowing the distance for each predicted body part can prove to be very useful.

Thank You

Any questions?