Carson Humulock

1/26/23

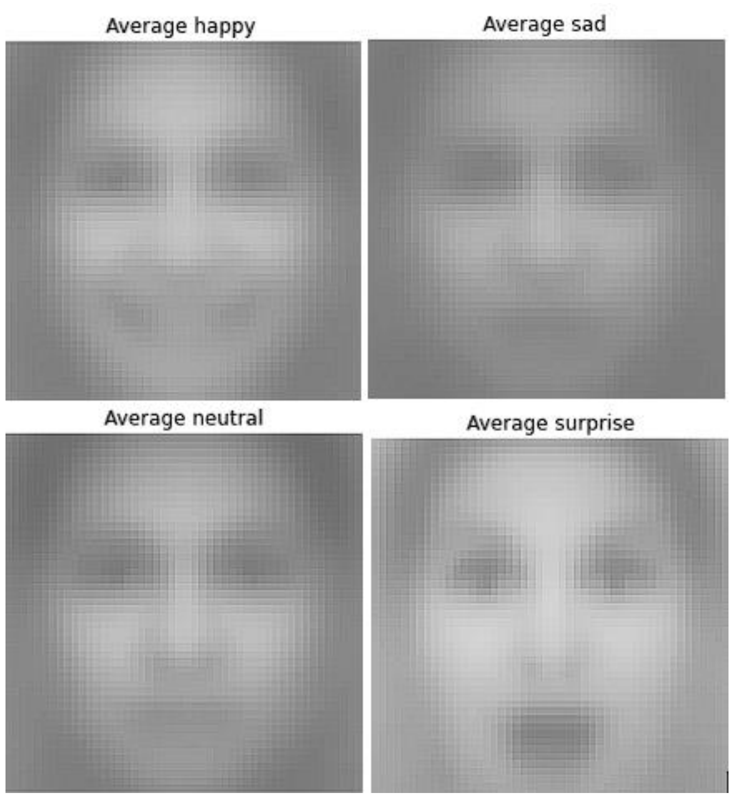
Milestone 1

Problem Definition:

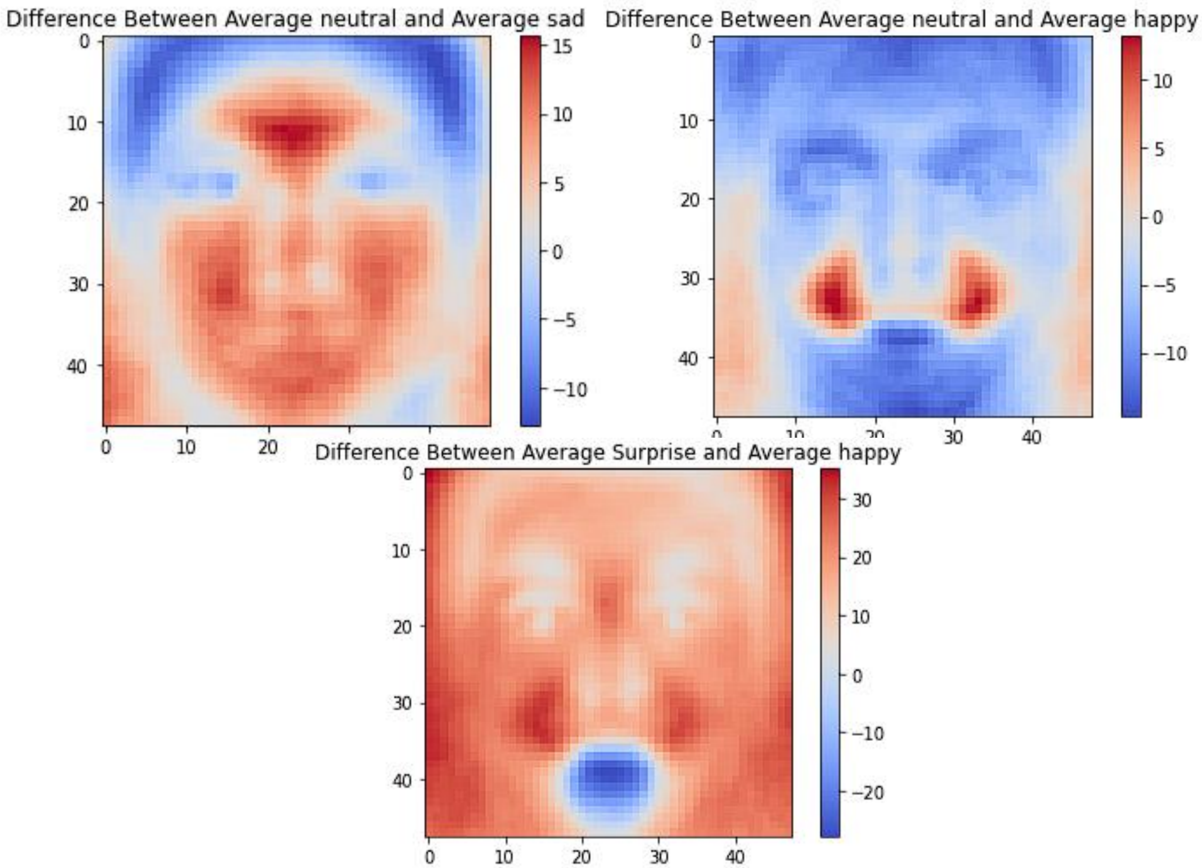
Recognizing emotions on one's face is a crucial part of human interaction. Emotional recognition is a multifaceted division of artificial intelligence. Its use cases include aiding in medical evaluation by monitoring patient emotions, assessing customer satisfaction and reactions to products or advertisements, providing a more personalized education experience for students, and overall creating a more pleasing interaction between humans and computers. The objective of this assignment is to train a neural network model to accurately categorize pictures of people's faces into one of four categories. The categories are, happy, sad, neutral, and surprised. The key question I am trying to answer is what model architecture can produce the best results to classify the emotions displayed on the face. The best model should accurately and efficiently classify the images. Similar to the key questions, the problem that needs to be solved is what is the best way to classify the facial emotion images.

Data exploration:

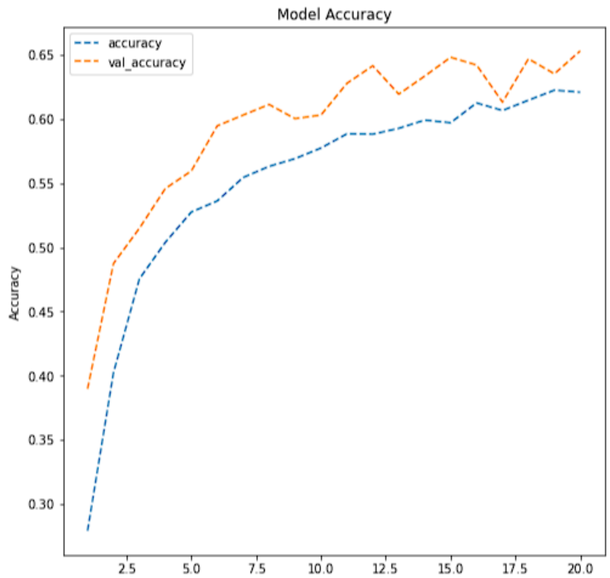
The data to train this CNN consists of pictures of human faces displaying happy, sad, neutral, and surprised faces. There are about 4000 images of each emotion excluding surprised, of which there are about 3200 images. Below is the average image of each category:

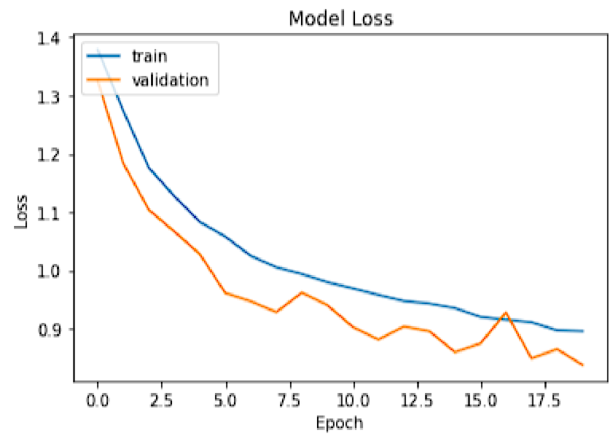


From this one can see that the average happy face has slightly raised eyebrows and cheeks as well as upturned lips and eyes fairly opened. The average sad face has lowered eyebrows, the pupils are likely not as visible, and the mouth is held in a straight across or lowered position. The average neutral face has somewhat visible pupils, relaxed eyebrows and relaxed mouth. Finally, the average surprise face has very raised eyebrows, fully visible pupils and an open or rounded mouth. The sad and neutral faces are fairly similar, the biggest difference is the eyes.



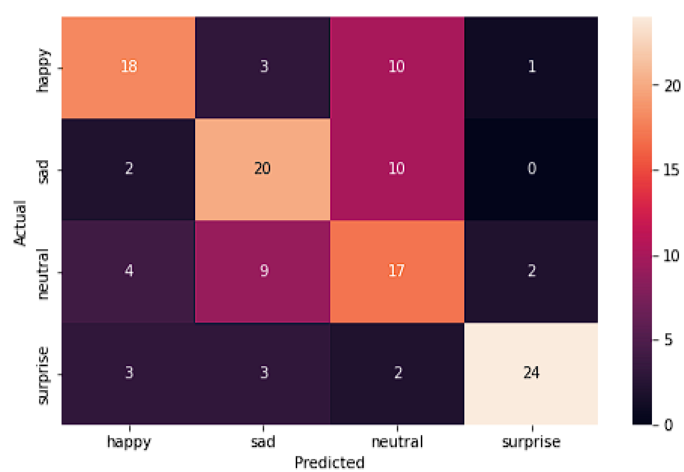
When contrasting a few of the average faces it is confirmed that the average neutral and average sad face are fairly similar differing most around the eyes. neutral and happy differ greatly around the mouth with the smile being very pronounced. Between average surprised and happy the surprised mouth differs greatly as well as the distinct cheek difference that is shown when a person is smiling. The data was already split into training, validation, and test data. Each containing folders categorized by the emotion. Random selections of the images are flipped horizontally, brightness shifted, and given a shear transformation. The images are also randomly shuffled. This image augmentation will help prevent the model from over fitting the training data.





The results from the first CNN model, which fit the data better, show that the validation data accuracy hit about 65%. This is an okay accuracy but can definitely get better. The model did generalize the data well. Our model loss shows that the model is learning fairly well as once it gets close to 20 epochs, the loss is around 0.7.

	precision	recall	f1-score	support
0	0.67	0.56	0.61	32
1	0.57	0.62	0.60	32
2	0.44	0.53	0.48	32
3	0.89	0.75	0.81	32
accuracy			0.62	128
macro avg	0.64	0.62	0.62	128
weighted avg	0.64	0.62	0.62	128



Looking at the confusion matrix the model struggled with accurately predicting happy, sad, and neutral. It often confused happy faces with neutral faces as well as sad faces with neutral faces and vice versa. One interesting result is the surprise face accuracy is the highest at 75% even though it was the least represented in the dataset. The average surprised face was also very different than the others so this could be why the model classified it more accurately.

Proposed approach

The results from the first two CNNs were okay but they can definitely be better. One way to better the model could be to add more training data. In this case it would not be too difficult to find and label new images, although it could be very time consuming. After looking online, I found that the data set used is the fer-2013. it contains 30,000 images so there are potentially more labeled facial images that could be used. I believe the accuracy could also be increased by finding a happy medium between model 1, which was the better model, and model 2 which was far too complex. The number of hidden layers and filters in model 2 was far greater that of model 1. I believe transfer learning would be a better solution. Using a pre-trained model would be much easier and less time consuming than attempting to find more images and label them. Using both pre-trained parameters and weights could make the model more efficient and accurate. Using transfer learning to create the potential solution design, I would need to find a pre-trained model, add some classification layers, and tweak the classification layers until they give a satisfactory result. To measure if a result is successful, when comparing this new technique to the previous technique one must look at a few variables. The first being Training accuracy vs the validation accuracy. This will tell you if the model is over fitting or giving accurate predictions only for the training data. The next is the accuracy on the test data. Another measure is the loss which is the prediction error of the model. The loss will also tell you if the model is over fitting if the test data curve is below the validation learning curve. Loss can also show if you if your learning rate is too high or too low. This will show if your model is correctly classifying images that were not used to train it. Finally, the F1 score. The F1 score is a measure of precision and recall.