Carson Humulock

2/2/2023

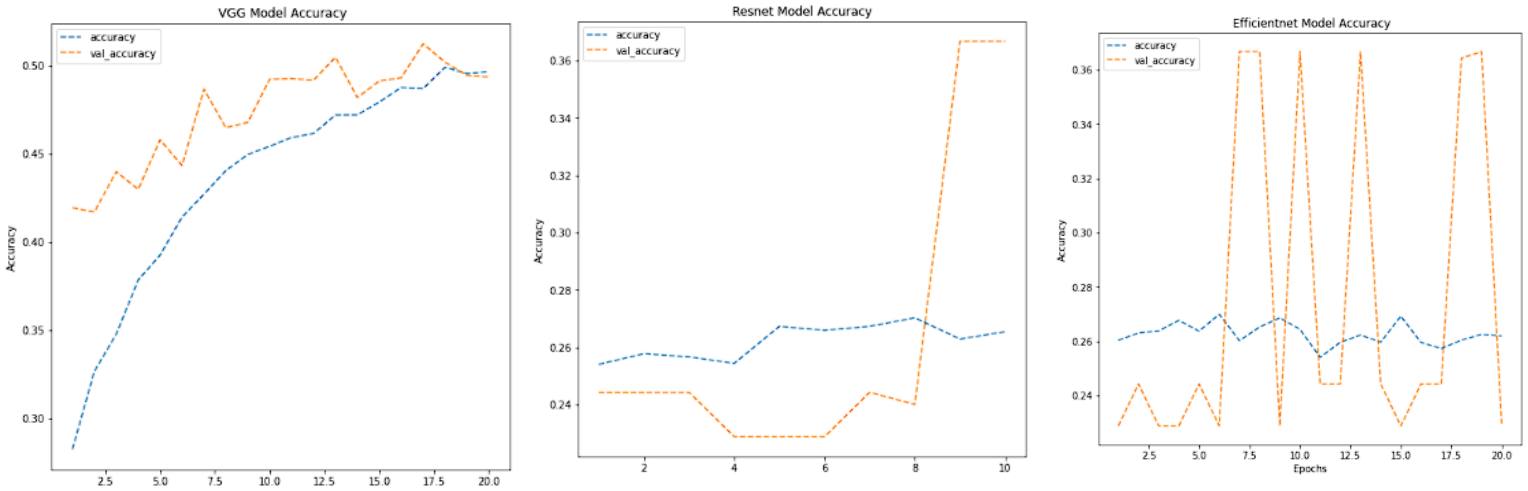
Milestone 2

Refined insights:

After running the transfer learning models, not many new observations were uncovered but many observations from milestone 1 were reaffirmed. All of the models struggled to distinguish the difference between neutral and sad. These two were misclassified as one another in nearly every model. The features of the two categories both include partially hidden pupils with dark pixels around the eyes, straight across mouth, and relaxed eyebrows and cheeks. surprised face performed the best over all on all of the models, which makes sense as it is the most distinct with raised or arched eyebrows, generally the mouth is open, pupils are very visible, and cheeks are raised.

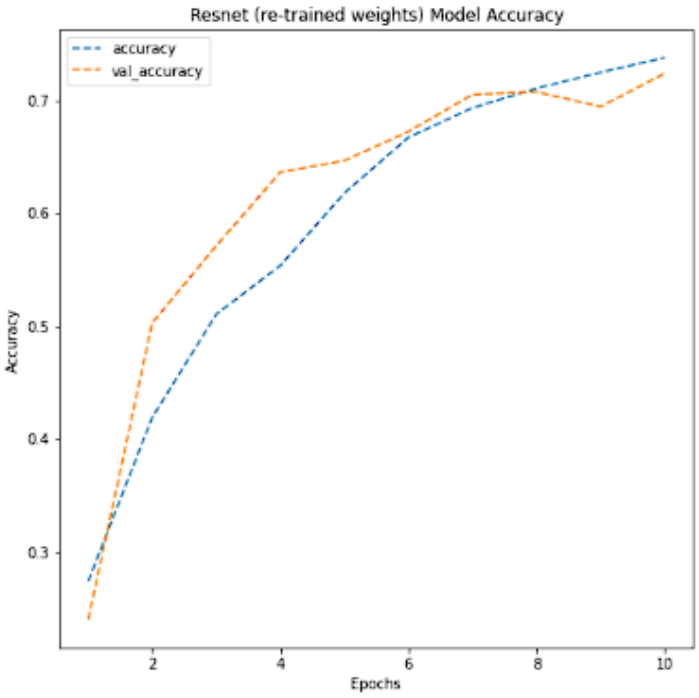
Comparison of the models:

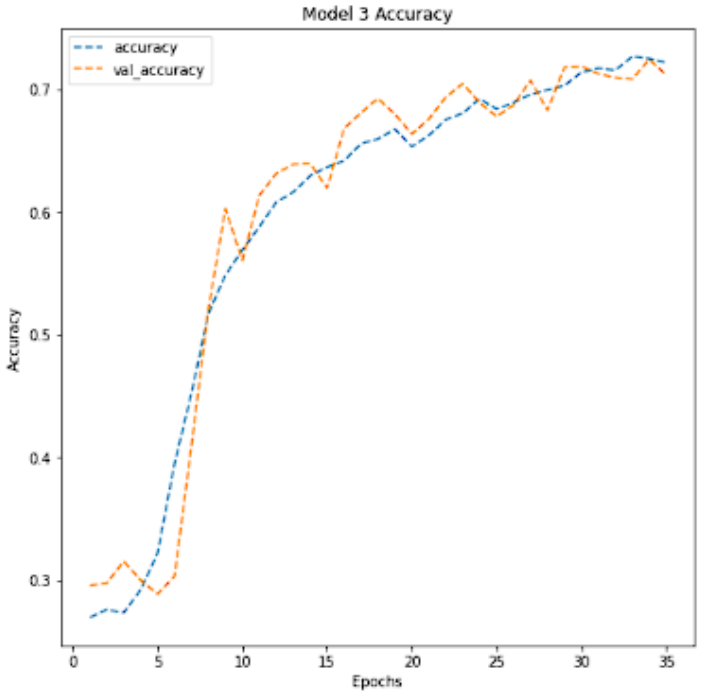
Unfortunately, the transfer learning models overall performed very poorly.



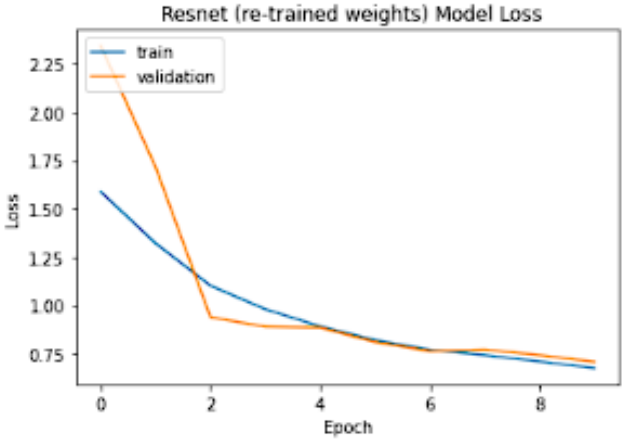
The VGG16 model performed the best out of the three transfer models, although it still did not give satisfactory results. In fact, the VGG model performed worse than the initial CNN models. Its accuracy for both validation and training data reached almost 50%. While both the Resnet model and efficient net model would get stuck at exactly 36.7% accuracy consistently. Even with added layers, different activation functions, different learning rates, and different optimization methods, the models remained with these funky charts that topped out at 36.7% accuracy. I believe these models performed poorly because they were not trained on human faces specifically. They are object detection models, so the weights from these models don't transfer over well to facial emotional detection. A more careful consideration of the transfer model, for example using VGG face, a model trained for face detection,

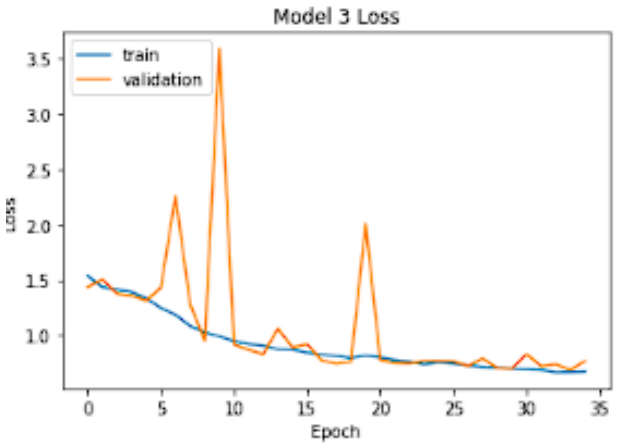
could have yielded better results. Now let's compare the models that worked well, Resnet with retrained weights, and the third CNN Model.





These two models performed fairly Similarly with both reaching accuracies of about 71%. This is 10% better than the first two CNN models and meaningfully higher than the transfer learning models. There is little to no over fitting in these models. The models are generalizing the training data fairly well.

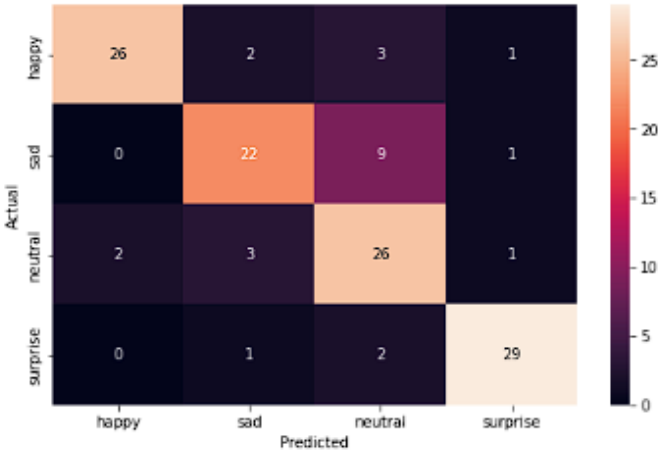




The loss charts show that model 3 reached a slightly better loss value of 0.7 while the Resnet model loss stayed around 0.75. The loss curve for the validation data in the retrained Resnet model is much smoother than the model 3. After reducing the learning rate, and removing some of the dropout layers, these spikes became significantly smaller but I could not get rid of them. I believe this shows that the Resnet retrained model is again performing better than model 3.

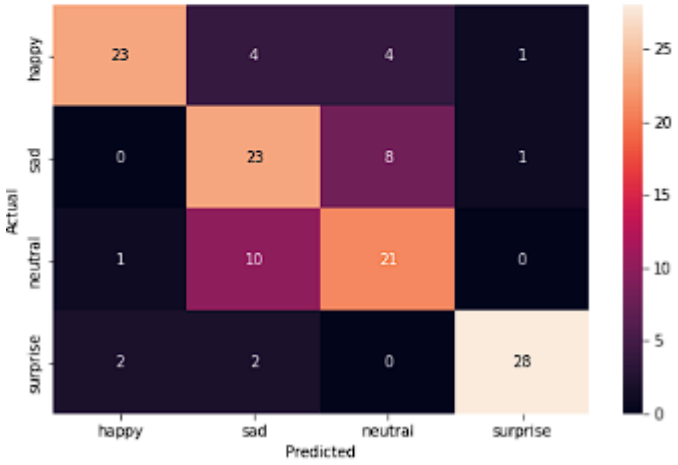
Resnet (re-trained weights)

	precision	recall	f1-score	support
0	0.93	0.81	0.87	32
1	0.79	0.69	0.73	32
2	0.65	0.81	0.72	32
3	0.91	0.91	0.91	32
accuracy			0.80	128
macro avg	0.82	0.80	0.81	128
weighted avg	0.82	0.80	0.81	128



CNN model 3

	precision	recall	f1-score	support
0	0.88	0.72	0.79	32
1	0.59	0.72	0.65	32
2	0.64	0.66	0.65	32
3	0.93	0.88	0.90	32
accuracy			0.74	128
macro avg	0.76	0.74	0.75	128
eighted avg	0.76	0.74	0.75	128



The confusion matrix shows us that the Resnet with re-trained weights properly classifies neutral faces significantly better than any model has so far at 81% recall. On the downside it does not do much better with the sad category, misclassifying it nearly 33% of the time, about the same as the other models that performed well on the training data. It also does notably better classifying happy faces. Surprise remains about 90% correct which is consistent among the models that have worked well. The precision on the Resnet model with retrained weights is better in every category except for surprised, and even there it is only 0.01 lower. The Recall is better for the Resnet model in 3/4 categories with the exception being sad b again the difference is negligible. F1 scores also come out 1 – 8% better in this retrained Resnet model than in CNN model 3. The accuracy on the Resnet model is also 6% better than the CNN model 3.

Final solution design:

The model that I chose as the proposed solution was the Resnet model with retrained weights. This is the best solution by nearly all metrics. The training and validation accuracy came out to be around 71%. This is better than our initial CNN models and about the same as model 3. The reason I chose this Resnet model over the CNN model 3 is because of its performance on the test data, mainly the neutral categorization. Neutral and sad were consistently misclassified as one another on all models, even the ones with the highest accuracies. On this retrained Resnet model Neutral was only misclassified 19% of the time vs 35% misclassification on the second best model. One downside to this model is that on the test data it did do slightly worse in the sad category but the major improvement in the neutral category makes this negligible. As stated before, the F1 score, recall, and precision are better

in the retrained Resnet model. In conclusion the metrics and results dictate that the Resnet model is the best model for classifying human facial emotion.