Anthony Long
Affective Computing
CAP 5627

# Project 2 Questions

1. With both 128 and 64, the architecture classified pain the most. This can be seen from the confusion matrices for 64 and 128 below:

   64x64:
   ```
   [[1002  302]
    [ 622  411]]
   ```
   128x128:
   ```
   [[981 323]
    [303 730]]
   ```
   .

   The rows retain count sums since they represent the actual numbers of each class, however upon inspecting the cells, we can see that correct predicion counts were higher for pain in the 64, but higher for no pain in 128. This could possibly be due to no pain being more accuractely predicted with more distinct features, whereas pain can be classified quite well in either, which is why correct predictions for pain in both are very close (1002 vs 981).

2. When comparing scores against each other in both datasets, they make sense. When looking at recognition systems, it is common to use multiple metrics since accuracy by itself can give an incomplete image of the success of a system. Other metrics allows the researcher to distinguish between things such as overfitting or memorization of data. When inspecting the accuracy of both models, loss is correspondingly lower with a higher accuracy, while precision, recall and F1 scores all increase in direct proportion with accuracy.

   64:
   ```
   ACCURACY: 0.6046212911605835
   LOSS: 2.3359804153442383
   PRECISION SCORE: 0.576437587657784
   RECALL SCORE: 0.3978702807357212
   F1 SCORE: 0.47079037800687284
   ```
   128:
   ```
   ACCURACY: 0.7321352362632751
   LOSS: 0.9229076504707336
   PRECISION SCORE: 0.6932573599240266
   RECALL SCORE: 0.7066795740561471
   F1 SCORE: 0.6999041227229147
   ```

3. Based on my highest accuracy score of ~%73.2 from the 128x128 model, I do not think it makes sense to classify pain based on images alone. Images are a very easily manipulatable form of data and can be tampered with by either the subject or someone else. Furthermore, according to [1], various methods of emotion recognition can reach around 75% accuracy through use of unimodal classifiers. However, in almost all cases, using multimodal classifiers always improves accuracy for detection. For this reason, other classifiers such as physiological ones explored in project 1 should be combined with image recognition to improve accuracy of the emotion classifier.

4. The highest accuracy for project 1 was 88.3%, compared to the highest accuracy of project 2 which was 73.2%. I believe this is an unfair comparison, since project 1 used data fusion, which allowed for the models to make more “fully” informed classifications,

whereas project 2’s models only had image data to go off of. As mentioned previously, in almost all cases of classification, multimodal improves accuracies, as seen in comparing project 1 and project  2 accuracies.

5. Pain recognition systems should strive for higher recall rather then precision. Precision is a measure of how many selected items are relevant, or a measure of relevant data out of correctly classified data. Recall is a measure of how many relevant items are selected, or how many relevant items out of items classified in the same class. In certain context where correct classification is more important than simply identifying things that match, the researcher would value correct classification. For example, in determining if a child is in pain, it would be much more useful to determine not only the relevant data (if the child is in pain), but also specifically what children are correctly classified as in pain relative to all correctly classified children (pain and no pain).

6. I believe using multimodal classification would be necessary for stress identification. As we discussed in class and observed in some homework assignments, stress can be more difficult to identify than other emotions, because not only do people express stress on a wider scale than other emotions typically, but stress is typically not as distinguishable by facial features. For this reason, I think that it would be necessary to utilize physiological data to change our system to a stress recognition system. This would vary in difficultly, and would require either feature, score or decision level fusion in order to classify, and depending on the fusion selected might require a different algorithm altogether for classification. This would make it more difficult in terms of time since the code would need to be substantially reworked. As stated, since can be harder to identify with images, I think it would be more difficult to recognize stress over pain through use of facial images only.

Full data for 64 x 64:

```
ACCURACY: 0.6046212911605835
LOSS: 2.3359804153442383
PRECISION SCORE: 0.576437587657784
RECALL SCORE: 0.3978702807357212
F1 SCORE: 0.47079037800687284
[[1002  302]
 [ 622  411]]

D:\Affective-Computing-Projects\Project 2>
```

Full data for 128 x 128:

```
ACCURACY: 0.7321352362632751
LOSS: 0.9229076504707336
PRECISION SCORE: 0.6932573599240266
RECALL SCORE: 0.7066795740561471
F1 SCORE: 0.6999041227229147
[[981 323]
 [303 730]]

D:\Affective-Computing-Projects\Project 2>
```

References

[1] Dupré, Damien, et al. “A Performance Comparison of Eight Commercially Available Automatic Classifiers for Facial Affect Recognition.” PLOS ONE, Public Library of Science, journals.plos.org/plosone/article?id=10.1371/journal