Assignment 1: Computer Vision

Reflections

Exercise 1

The Py-Feat detector has several stages. First it finds the faces in an image, and then it uses this information to extract other forms of data (AU activations, expressed emotion, ...).
Based on the visualization you produced, do you agree with all its predictions? What seems to confuse the system when it fails? Are there any cases that would be tricky for a human observer?

The model is not always accurate with its predictions, as illustrated in Figure 1 and Figure 2. In some cases, the system incorrectly detects faces where there are none, and in other cases, it fails to detect faces altogether. For instance, Figure 1 shows where the face is missed, likely due to the person only showing their profile, while Figure 2 displays a case where the model wrongly identifies a face in a background.

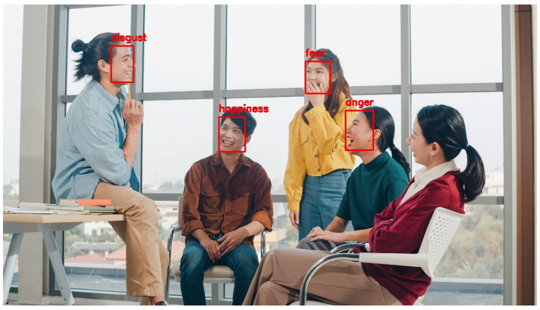

Figure 1: Model failing to identify someone's face.

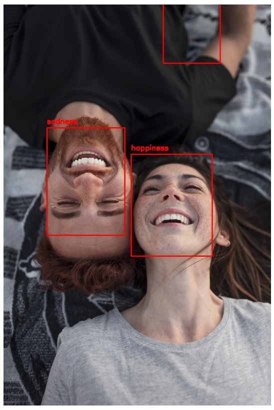

Figure 2: Model wrongly identifying a face in the image where there isn't one.

What seems to confuse the system sometimes is not being able to properly see the whole face, e.g., only the profile of the person is visible, the person is wearing glasses or something that obstructs the view of the whole face. In such situations, the model struggles to extract reliable features for emotion detection, leading to incorrect predictions.
Additionally, certain Action Units (AUs) are ambiguous and can be associated with multiple emotions. This ambiguity can make it difficult for the model to correctly differentiate between emotions, as these AUs are characteristic of multiple facial expressions. For example, a person expressing anger may also have their lips apart (AU25), similar to someone displaying surprise. This overlap creates confusion for the model, as the AU could point to more than one possible emotion.
Another aspect that affects the model's performance is the image quality. Poor resolution, blurring, or bad lighting can confuse the model's ability to accurately detect facial features and emotions.

Regarding tricky cases for human observers, they also face difficulties in interpreting emotions, especially when certain features are less expressive or obscured. For instance, in cultural contexts, emotional expressions vary widely across different regions and communities. Some cultures barely show any facial expression and thus make it hard to infer someone's emotion. It can also be the case that they display a mix of emotions in their faces, so it is difficult to classify which is the dominant one.
Finally, when only parts of the face are visible due to obstructions (such as hats, glasses, or hands partially covering the face), the interpretation of the emotion becomes more challenging.

Exercise 2

Based on the analysis of the AU data you have performed, suppose you need to choose a subset of the AUs as inputs for a predictive algorithm. Which AUs would you choose, and why? What's the problem with using too many features?

Using too many features in a predictive algorithm can lead to several issues, including overfitting, where the model becomes too personalized to the training data and performs poorly on unseen data. Dimensionality can also make it harder to identify meaningful patterns, requiring significantly more data to achieve reliable predictions. Having many features can also decrease the interpretability of the model. Therefore, it is essential to balance the inclusion of differentiating features with the risk of overfitting, and the ability to interpret the model's decisions.

So when selecting a subset of AUs for a predictive algorithm, as a first step, I think it is important to prioritize those that have the largest absolute differences of mean between the positive and negative conditions, as these AUs are likely to characterize more a defined emotional valence. In contrast, AUs with small absolute differences may introduce ambiguity, making it harder for the model to distinguish between emotions. Such AUs should be excluded if they don't contribute clear distinctions, as they could introduce noise and reduce model accuracy.
Next, it's important to consider the correlation between AUs. Highly correlated AUs provide redundant information, which can limit the model's effectiveness. To address this, if possible, I would perform a correlation analysis to identify pairs of AUs that are closely related. In these cases, one of the correlated AUs can often be removed without loosing too much information.