What is the problem being addressed?

The lack of the ability for modern A.I. to recognize subtle facial expressions.

Why does technology need to understand facial expressions?

Everyday humans act on social cues, whether it's looking at a driver when at a stop sign or understanding that someone was surprised when an individual entered a room. For technology to advance, there needs to be a way for it to understand and respond to human expression.

How does this specific research contribute?

One important expression, as identified in "The Dictionary of Body Language: A Field Guide to Human Behavior", is that of raised eyebrows. This research focuses on this expression because it is fairly distinct, common, and has fairly strong emotional meaning.

This research is designed to test how difficult it is to accurately identify eyebrow raises. Knowing the difficulty of one expression will give guidance to the overall difficulty of measuring any facial expression.

What is an overview of the general approach?

The general structure is as follows:

1. Receive video data from an outside source, and split the video into image frames.
2. Using the **dlib library**, compute the **68 facial landmarks** for each frame.
3. Compute two features for every frame:
    1. A single-frame-eyebrow-raise-score based of the 68 facial landmarks.
    2. A mouth-openness score based of the 68 facial landmarks.
4. For every frame; aggregate the features of the previous 9 frames to create a 10x2 matrix.
5. Feed the 10x2 matrices into an **SVM** trained to predict a boolean classification of whether of not the eyebrows are raised.

What data was used to train the SVM?

A script was created that randomly explored, downloaded, and filtered YouTube video based on whether or not they had a face present in many of their frames. The resulting videos acted as a sufficiently random sample of faces that were capable of being recognized by the 68 facial landmarks.

All videos were stored in .mp4 format, and were scaled to a resolution of 1920×1080.

Out of the random sample, 9 video clips were chosen for training. All of them contain only one person in frame. Every 10th frame of the 9 videos was hand labelled for a total

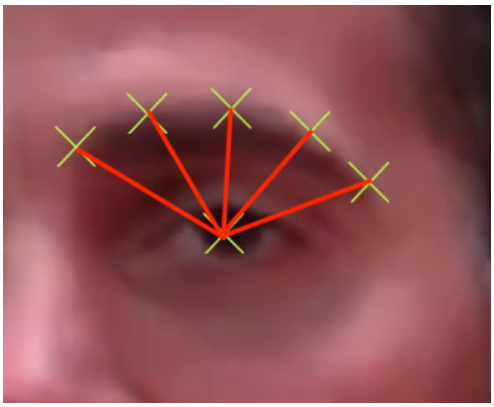
of 148 labelled frames. Each of those frames was labelled with a score from 0 to 100. The number reflects the answer to the question “out of 100 people, how many would claim the person in the image has their eyebrows raised”.

Durning training time this 0-to-100 score is converted into a boolean value by picking a threshold, such as 50. In which case, every value under 50 is classified as false, and every value over 50 is classified as true.

How are the features derived?

The single-frame-eyebrow-raise-score is computed in a few steps.
1. Find the center of the eye, based on the 68 facial landmarks.

2. Find the distance between the center of the eye and each eyebrow landmark.



3. Find the mean distance for each eye.

4. Divide the mean distance for each eye by the width of the eye to standardize the value across different facial structures.
5. Combine the value for left and right eye.

The mouth-openness score is more straightforward. Their are 3 landmarks on the inside-top of the mouth and 3 landmarks on the inside-bottom of the mouth. The distance between each of these vertical pairs was measured and averaged (mean) to create the mouth score.

How were edge cases handled?

Some frames do not contain faces, which results in missing scores for those frames. The next available frame with a face is used as a substitute, for the case of a missing frame. This is designed to emulate a kind of 0-padding.

Sometimes random background elements are recognized as faces by dlib. This was addressed by having a minimum face height of 200 pixels, which sufficiently filtered out this noise.

Multiple faces was not an issue since all videos were chosen to only contain one person.

How well did the model perform?

Running a 6-fold cross validation on the training data resulted in a accuracy score of 66%. While this is better than random, inspecting the videos and labels themselves revealed that the model is simply predicting “True” for every frame, including videos that were never used in cross validation.

This process was repeated with different numbers of look-back frames, e.g. a 3x2 matrix instead of a 10x2 matrix. However, this did not substantially change the validation score.

Why did the model underperform, how can it be improved?

This performance is certainly a case of under-fitting. One of the causes of this is likely due to the data itself not having a equal number of examples for each classification.

This hypothesis could be tested by simply duplicating some of the negative examples to create a balanced set for verification and training. A better solution would be to label more data until the classes are equal.

There are several other areas could also be affecting performance.

- The parameters of the SVM have not been tuned for this model, which could likely address under-fitting.
- The single-frame-eyebrow-raise-score could be an unreliable measure due to dividing by the width of the eyes. When a person turns their face left or right, the measurable width of the eye artificially changes. This could be tested and addressed in a number of ways.

- The dataset of 9 videos and 148 frames is quite small, it could be the case that more labelled data was needed to accurately find a boundary.
- Eyebrow raises often are notable because of their change from a previous state. However, the SVM is only being provided with the static values, not the change in value over time. This could be placing an additional strain on learning.
- The features for the SVM are not normalized, making for additional strain on learning.
- Rather than looking at the previous 9 frames, the model could be given a logarithmic look back, such as the 2nd-most-recent frame, 4th-most-recent frame, 8th, 16th, 32nd, etc to get a better sample for change over time.

Conclusion

The case for accurately measuring eyebrow raises still requires further investigation. There is evidence that is it possible to measure, however there is yet to be an implementation for doing so.