Analysis and Modification of Encoded Facial Expression Data

by Colin Sullivan

Goal

The goal of this project is to use auto encoders to break down and analyze facial expression data with the aim of gleaning useful information from the distribution of data points in the latent space. For example, I would hope to discover some correlation or grouping of similar emotions in the image data (e.g. happy, sad, angry, etc). If found, I hope to use this information to "shift" encoded images' latent spaces toward any of the discovered emotional centers with the hopeful outcome being a shift in emotion of the initial encoded facial expression.

The inputs of this project would be a few thousand low-pixel (32x32 or 64x64) images of my face, relatively centered, and on a neutral back drop. The auto encoder should simply output the same images, but, more importantly, it will place the input data into a compressed latent space which can be broken down further for analysis with PCA or other compression strategies.

This projects could potentially provide insights on the distribution of facial expression that would prove useful for tasks such as the compression of facial imagery or facial identification. Further, if the latter half of the project is successful (emotional shifting of facial expressions), it could be used in video chats or conferences to improve productivity and collaboration — I believe there are some psych studies that back this up.

Previous and Related Works

This project will require some form of auto encoder, so some research on the various forms and their implementations will be necessary (e.g. VAE, b-VAE, etc). There are a variety of readings online which explain and build off of these structures. Further, there is a good amount of research that has been done on emotion detection in face images (e.g. **https://arxiv.org/pdf/1801.08329.pdf**). This research could provide valuable insight on what kinds of structures work best for encoding and interpreting facial expression data.

Dataset

As stated previously, the data should consist of several thousand low-pixel images of my face. This data can be gather fairly quickly by taking snapshots of my face during an hour long conversation. If one snapshot is taken every second, it will take less than an hour to generate 3,000 images. While this data will obviously not be sampled entirely independently, the one second delay should hopefully be enough to reduce correlation between consecutive snapshots. Also, since the data will only contain images of my face during one session, the model will not generalize very well, but this is not a concern — the goal of this model is to analyze encoded facial expressions, so limiting the data to one person will hopefully allow the model to focus only on variation in expression as opposed to face shape, eye color, hair, etc.

Methodology and Experiments

As stated previously, I will likely use some form of auto encoder to compress the image data and PCA or some other strategy to visualize/analyze the compressed data. Depending on the results there, it might be helpful to run K-means or fit a mixed Gaussian model to the encoded data, after which I'll use a simple method (possible just linear shifting) to shift encoded images closer to some of these emotional center points.

Evaluation

The results of this experiment can be evaluated based on whether or not useful information is gleaned from the analysis of the auto encoder's latent space. To evaluate this specifically, I might label the images in which I'm smiling, for example, to see if those images are close together in the latent space. The emotional shifting part of this experiment can be evaluated based on how realistic the shifted images look and how well they match up with expectations. Potential plots and images could include but are not limited to training loss by iteration, closeness of compressed smile images by iteration, k-groups by loss, randomly generated images, and new images shifted.