Using Deep Learning and Computer Vision for Interpreting Facial Expressions and Emotional States

Adrian Del Bosque, Kevin Jackson
adrian.delbosque01@utrgv.edu
kevin.jackson01@utrgv.edu

September 30, 2019

Summary of the Proposal

In recent years research has been conducted using machine learning and computer vision to map the geometry of the human face for biometric data and facial recognition. In this project, we will be exploring the use of deep learning algorithms alongside computer vision frameworks to detect human emotions based on images. Over the semester we hope to implement a machine learning algorithm that using some form of computer vision to rank emotions based on facial features.

Background

The facial action coding system (FACS) defined by Ekman and Friesen in 1978, is a system used to characterize facial expressions of human emotions. Changes in the facial landmarks (FLs) the ends of the eyebrows, bridge of the nose, eyes, and points of the mouth by action units (AUs) can be used to determine the expression of an emotion.

Conventional FER systems use geometric features, appearance features or a mix between the two. The geometric approach utilizes a feature vector based on facial components in image sequences using multi-class AdaBoost. Where as appearance features are extracted from the global facial region and recognized using stepwise linear discriminant analysis. The overall recognition performance for Conventional FER systems average around 59% - 70% accuracy for still frame images whereas Deep Learning based FERs average between 69% - 77%.

The highest average Deep Learning FER uses a hybrid approach by utilizing the spatial image characteristics using CNN and spatial temporal feature learned using LSTM. Even though Deep Learning FER shows great success there are still limitations keeping it from a higher success rate such as computing power, datasets and sold algorithm theory.

Goal and Objectives

Our goal is to find a way to detect human emotion using facial expressions. Based on prior readings, existing algorithms have found a success rate average of at least 63% using the standard model and 72% using deep-learning models. Therefore, using the results of previous models we hope to get at least a 60% success rate on the learning model.

Data and Methods

For our training data, we will be using some reference data found on Kaggle.com that was used in the facial recognition challenge. The training set data consists of 28,709 48x48 pixel grayscale images of faces each showing different facial expressions. The included CSV file contains two columns, an emotion and a pixel column. The pixel column contains a code ranging from 0 to 6 that is inclusive for the emotion shown in the picture, and the "pixel" column contains a string surrounded in quotes for each image. The testing data set contains 3,589 images along with another final test set that also contains 3,589 examples.

To train the model we will be using a CNN and focus our training data on three parts of the face: eyebrows, eyes and the wrinkling of the nose. These features will be used because they convey the most emotion, but are also difficult to manipulate unless in certain situations where some form of emotion is involved. Due to this nature, a scoring system will be given on each image based on the extreme representation of that emotion.

References

Towards Data Science, Priya Dwivedi, accessed September 30, 2019
https://towardsdatascience.com/face-detection-recognition-and-emotion-detection

Algorithmia, February 28, 2018, accessed September 30, 2019
https://blog.algorithmia.com/introduction-to-emotion-recognition

NCBI, Byoung Chul Ko, Jan 30, 2018, accessed September 30, 2019
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5856145/B72-sensors-18-00401