Deciphering emotions using convolutional neural networks on video data
Elizabeth Tran, Michael B. Mayhew, Alan D. Kaplan
Lawrence Livermore National Laboratory

BACKGROUND
In recent years, there has been a push for automatic facial recognition and to interpret facial expressions using deep learning in computer vision tasks. Here, we are leveraging these methods in order to create continuous emotion predictions.

Emotion Recognition
Emotional recognition has been focused on using categorical models to group emotions into discrete categories, but that method does not capture all expressible human emotions, especially with microexpressions. Microexpressions are brief facial expressions that occur when a person is deliberately or unconsciously concealing an emotion.

Universal Facial Emotions
There is strong evidence for universal facial expression of seven emotions—anger, disgust, fear, happy, sad, surprise, and neutral.

METHODS
Categorical emotion prediction to continuous emotion prediction
Using Kaggle’s Facial Expression Recognition Challenge, we trained a convolutional neural network to classify human faces into discrete emotion categories. With four categories (angry, happy, sad, neutral), we were able to achieve a test accuracy of 68.4%. With Kaggle’s weights, we were able to continue to experiment with the model using the RECOLA’s dataset (a multimodal dataset combining ECG, EDA, audio, video recordings, and annotations ranking multiple emotional characteristics). Through different architectures and hyperparameters, such as quadrant pooling and fine-tuning, we were able to assess the overall performance of neural networks to recognize emotion from videos.

CNN Architecture

Table 1. Six raters (3 males, 3 females) rated each participant’s arousal and valence ranging from [-1, 1]. The above scores show only the participant’s valence score.

Universal Facial Emotions
There’s strong evidence for universal facial expression of seven emotions—anger, disgust, fear, happy, sad, surprise, and neutral.

RESULTS

Table 2. Performance comparison between our model and other methods. In our experiment, we focused on predicting the valence score using the video modality of the RECOLA dataset. Unlike the other methods, where they averaged the six rater’s scores, we decided to keep each rater separate in order to see the variability.

CONCLUSION & FUTURE WORK
Not a big network, like our original Kaggle model, is required to perform such accuracy for emotion recognition. However, our model finds certain emotions, such as angry, neutral, and sad, similar and tends to mistake them for one another.

Human perception is extremely tuned to small configurations and shape changes. We hope to improve our algorithm to emulate this capacity of precise detection of faces, and facial features in order to bridge the gap between continuous and categorical emotion recognition. Having emotionally aware algorithms will improve our understanding of human cognition and behavior.

REFERENCES