Interpreting Emotion Classification Using Temporal Convolutional Models

Authors: Manasi Bharat Gund - manasigund22@gmail.com Abhiram Ravi Bharadwaj - raviabhiram@yahoo.co.in Dr. Ifeoma Nwogu - ion@cs.rit.edu

INTRODUCTION

This study proposes a temporal convolutional model for emotion classification using facial landmarks.

- Hypothesis: Changes in facial expression are best recognized with movement (temporal modeling)
- Image based ConvNets provide good result. So is temporal information even important?
- Video based ConvNets tend to be more computationally heavy.
- Solution: **T-ConvNet** uses facial landmarks (thus, temporal modeling) and less computation (ignores appearance of person)

DATA AND PREPROCESSING

The **CK+ dataset**: (Training, Validation)

- 593 videos of various subjects going from neutral to a class of emotion
- Classes are: Anger, Contempt, Happy, Sad, Disgust, Surprise, Fear
- Model was trained primarily on CK+
- Data was preprocessed to grayscale and 20 frames per video

The **SAMM dataset**: (Testing)

- Humans showing same 7 emotions with different intensities throughout the video.
- Emotions are macro and micro (subtler than CK+), thus overfitting can be detected.

FACIAL LANDMARKS, AUs and EMOTIONS

Emotions activate certain Action Units (AU), or landmarks.

- Happiness: AU6, AU12 (Cheek Raiser, Lip Corner Puller)
- Surprise: AU1, AU2, AU5, AU26 (Inner Brow Raiser, Outer Brow Raiser, Upper Lid Raiser, Jaw Drop)

Some AUs shown in Figure 2.

The motion of landmarks over video can be learned to predict the emotion.

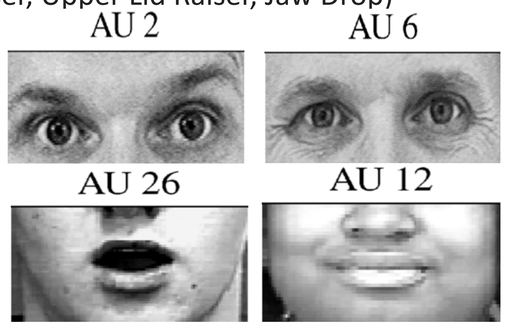


Figure 1. Examples of AUs

TEMPORAL CONVOLUTIONAL NETWORK

- Input: 3D tensor of shape $\{2 \times 20 \times 68\}$ where 20 is num of frames, 68 is num of landmarks and 2 channels for (x, y)-coordinates of each landmark.
- Model architecture: multiple blocks of convolutional layer, ReLU layer, Batch normalization, and Pooling layer. Shown in Figure 2.
- Filters convolved in temporal and partially spatial space

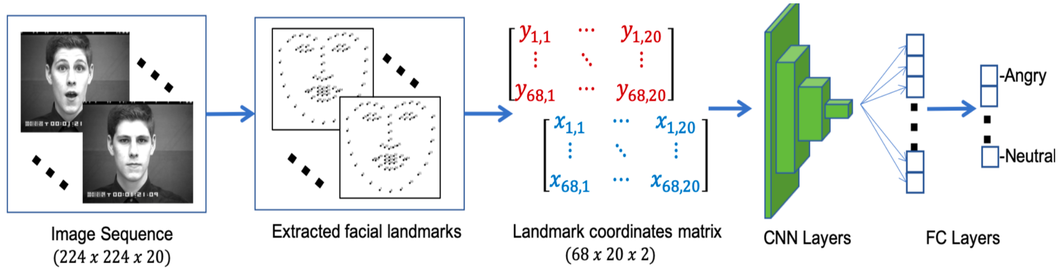
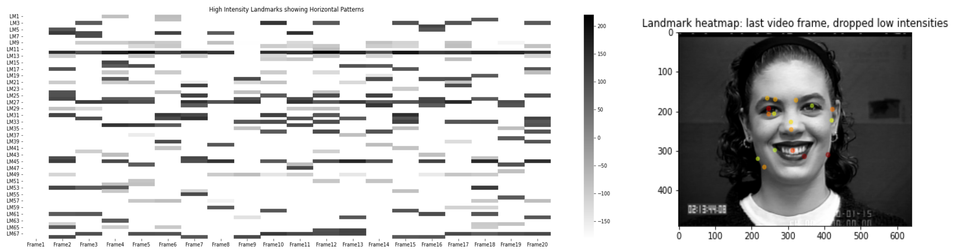
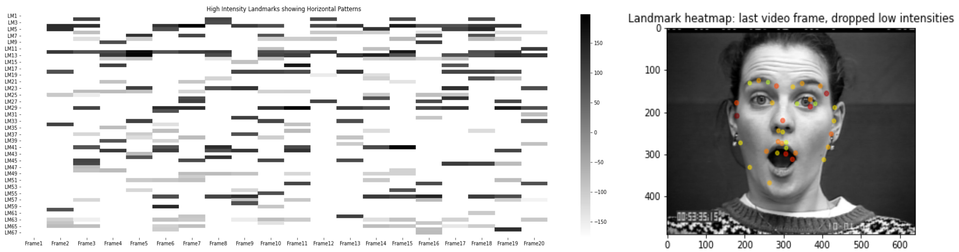


Figure 2. T-ConvNet Architecture

RESULTS AND CONCLUSION



(a) Predicted class: Happiness



(b) Predicted class: Surprise

Figure 3. Results. L: the highlighted landmarks throughout the video, R: Sum of activated landmarks on the last frame of the video

- Model predicted emotions with the accuracy of **99.6%** on CK+ and 41% on SAMM (more than x5 times F1 score than baseline)
- **Horizontal patterns** in Figure 3(Left) **show important landmarks** responsible for prediction. intensity of the same projected on the last video frame on the right.
- Predictions show correlation between the highlighted landmarks and the action units.
- Figure 3(a) shows prominent landmarks are associated with AU6 and AU12 (AUs associated with happiness).
- Figure 3(b) shows prominent landmarks on the eyebrows, open mouth. AUs associated are AU1, AU2, AU26 (Action units that indicate surprise)
- Colorful dots indicate heatmap: white (lowest intensity) to red (highest)

RIT | B. Thomas Golisano
College of COMPUTING AND INFORMATION SCIENCES

REFERENCES

Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, andI. Matthews, “The extended cohn - kanade dataset”
C. H. Yap, C. Kendrick, and M. H. Yap, “Samm long videos: Aspontaneous facial micro- and macro-expressions dataset”
D. Y. Liliana, “Emotion recognition from facial expression using deepconvolutional neural network”
F. Khan, “Facial expression recognition using facial landmark detectionand feature extraction via neural networks”