Interpreting Emotion Classification Using Temporal Convolutional Models

•••

Authors:
Manasi Bharat Gund
Abhiram Ravi Bharadwaj
Dr. Ifeoma Nwogu

The Team

Manasi Bharat Gund
manasigund22@gmail.com

Abhiram Ravi Bharadwaj
raviabhiram@yahoo.co.in

Dr. Ifeoma Nwogu
ion@cs.rit.edu

Motivation

- Hypothesis:
  - Changes in facial expressions are best recognized with movement.
- Problem:
  - Image based models provide good results.
  - Computational cost with video based convnets.
- Solution:
  - Build two models which use temporal data:
    - Stacked Convolutional Network (SCN)
    - Temporal Convolutional Network (TCN)

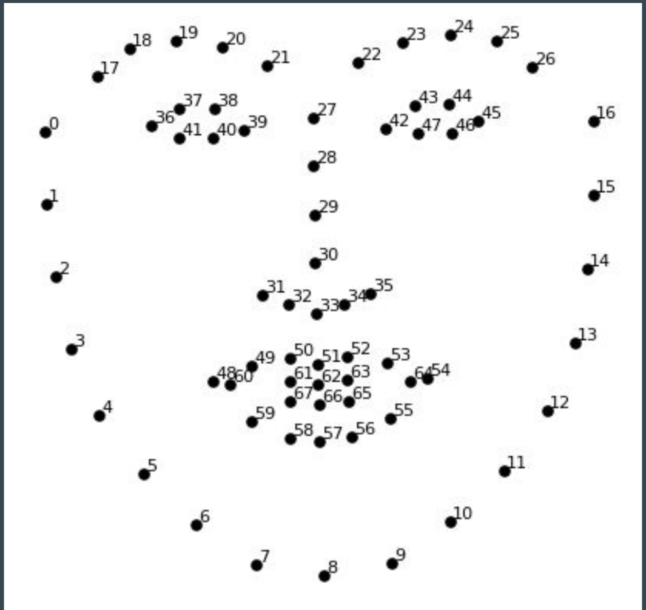

Fig 1. *Facial Landmarks*

Datasets

- CK+:
  - 593 videos of subjects going from neutral emotion to extreme emotion.
  - Classes: Anger, Contempt, Happy, Sad, Disgust, Surprise and Fear.
  - Data was preprocessed to grayscale and standardized to 20 frames per video.
- SAMM:
  - Videos classified into the same classes as with CK+.
  - Includes both macro and micro (extremely subtle) expressions.

Background

- Action Units:
  - Group of correlated facial landmarks.
  - Group of AUs determine emotion and every emotion triggers certain AUs.
  - Motion of landmarks during the video can be leaned to predict emotion.

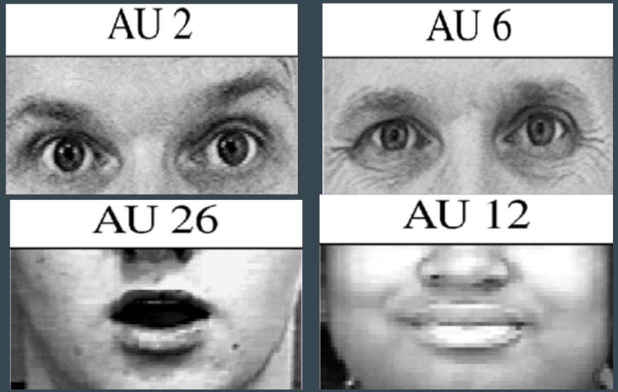

Fig 2. *Action Units for 2 emotions. (L) Surprise - AU2 and AU26. (R) Happiness - AU6 and AU12*

Steps - Preprocessing

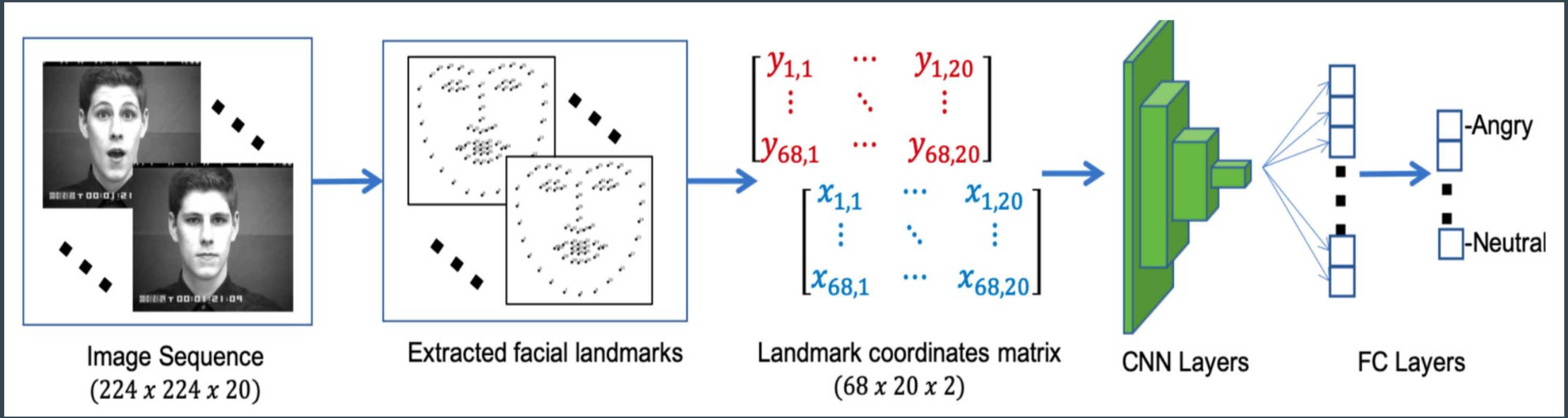

Fig 3. *Visual of the steps in the training process of the TCN.*

- Standardise to 20 frames.
- Extract facial landmarks for each image (to be consumed by the TCN).

Steps - Model

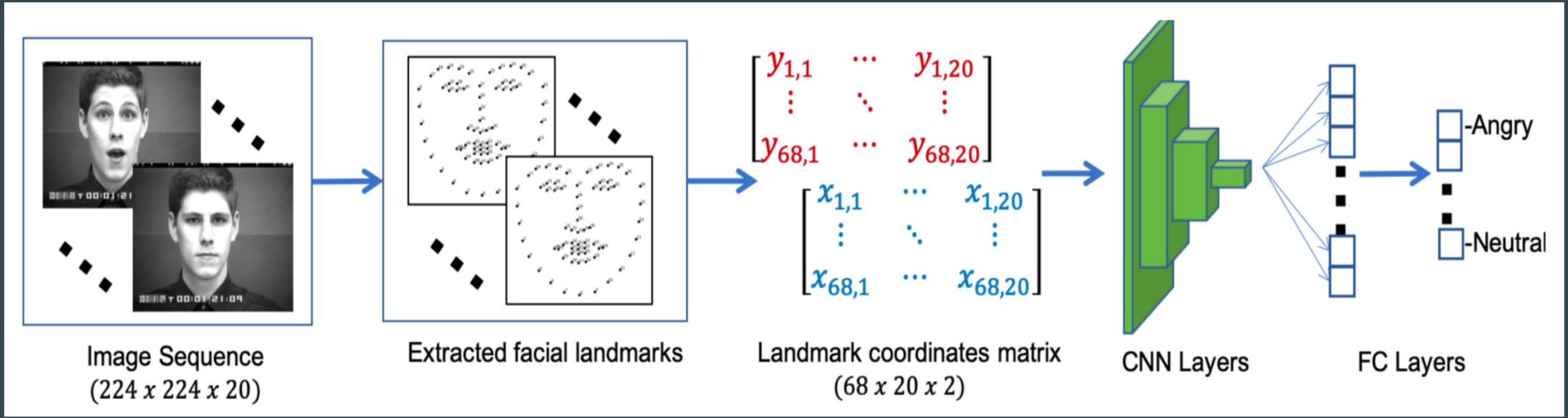

Fig 3. Visual of the steps in the training process of the TCN.

- **Input** - 3D tensor of shape { 2 x 20 x 68 }.
- **Architecture** - Multiple conv layers (ReLU activated).
- **Output** - Emotion class.

Steps - Interpretability (TCN)

- Horizontal patterns (Fig 4, left) show important landmarks responsible for prediction.
- Corresponding landmarks are highlighted (Fig 4, right)

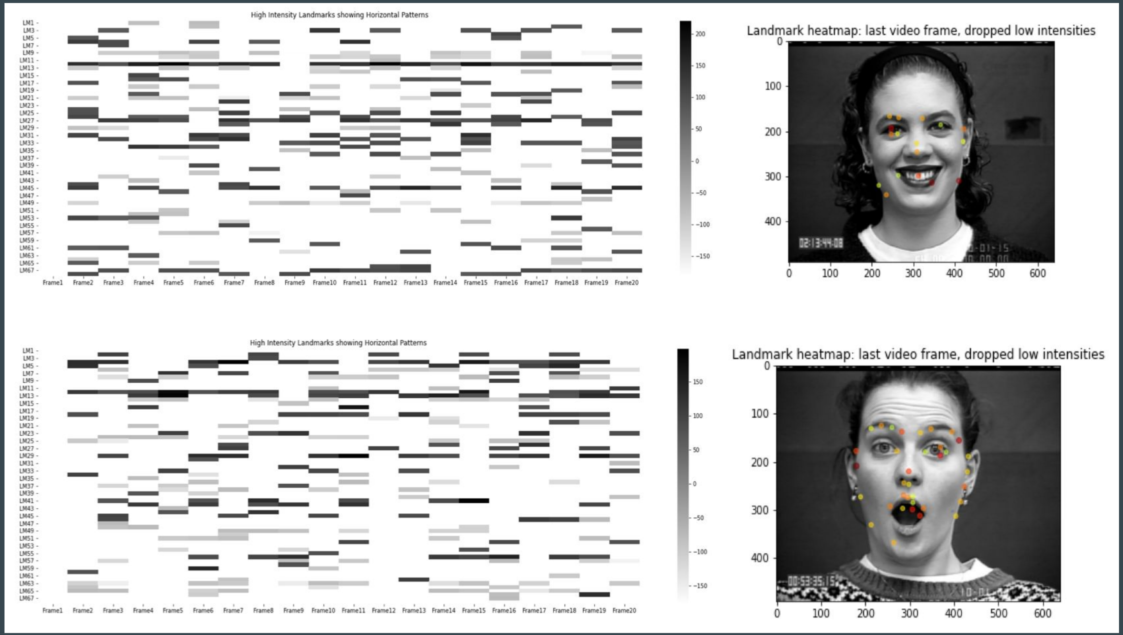

Fig 4. *Interpretation of the learned weights. (Top) Happy. (Bottom) Surprise. On the right, dots show the intensity of the landmarks with white being low and red being high.*

Results & Conclusion

- Prediction accuracy of 99.6% on CK+ by TCN.
- Both SCN and TCN outperform the baseline on the SAMM dataset.
- Predictions show correlation between highlighted landmarks and AUs

Model	CK+	SAMM	
	Accuracy(%)	TP	F1-score
FAN	**99.7**	-	-
DeepEmotion	99.3	-	-
DeepConv	92.8	-	-
TimeConvNets	97.9	-	-
Baseline MEGC	-	22	0.06
SCN	95.7	30	0.16
TCN	**99.6**	**41**	**0.33**

Fig 5. *Result comparison with existing models.*

Thank You