Mood Estimation Based on Facial Expressions and Postures

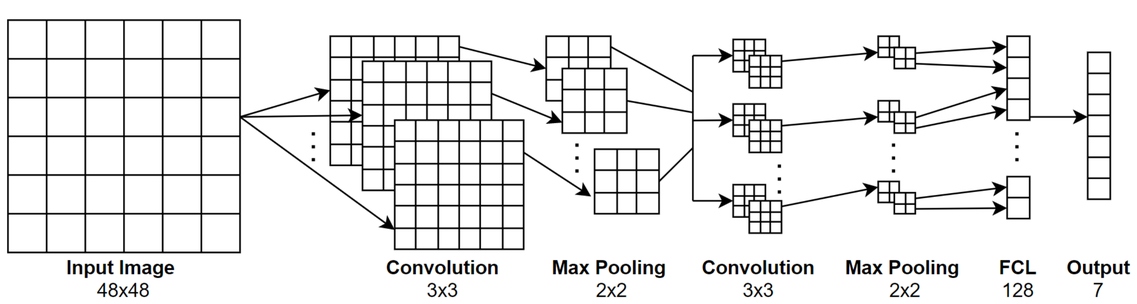
Daniel Canedo and António J. R. Neves
IEETA – Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro
E-mails: {danielduartecanedo, an}@ua.pt

1. Introduction

This poster presents a mood estimation algorithm based on facial expressions and postures using Computer Vision and Deep Learning. This algorithm consists in two well-known modalities within Computer Vision: facial expression recognition and pose estimation. Such algorithm can be useful in a wide range of applications that may benefit from feedback regarding the mood of a user. A specific application that estimates the mood of a speaker during a speech was used for testing the developed software.

2. Facial Expressions

Since real time performance was one of the goals of this work, a simple convolutional neural network (CNN) was designed for facial expression recognition. The CK+ dataset [1] was used for training. Before the training step, the dataset was pre-processed by applying rotation correction, cropping, intensity normalization, histogram equalization and smoothing, respectively. The proposed CNN achieved 93% validation accuracy:



3. Postures

Pose Estimation

PoseNet [2] – Vision model that is able to estimate the pose of a person by estimating where are the key body joints.

Implementation – The MobileNet version was implemented to cope with real time performance.

Expansiveness

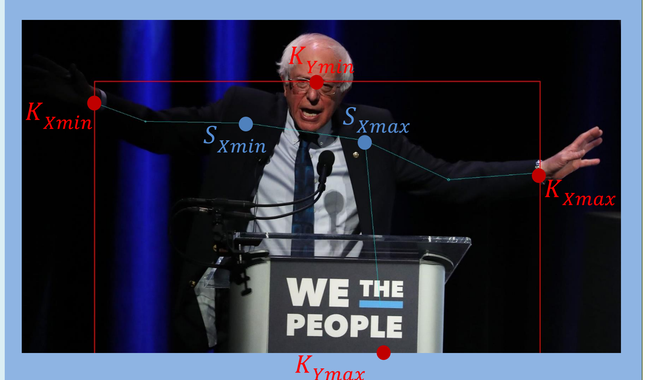
Idea - Having an expansive body posture is often correlated with dominance, power and confidence [3].

Expansiveness calculation - Ratio between the occupied area [4] and the minimum area that the speaker could be occupying:

$$A_{min} = |K_{Ymax} - E_{Ymin}| \times |S_{Xmax} - S_{Xmin}|$$

$$A_{current} = |K_{Ymax} - K_{Ymin}| \times |K_{Xmax} - K_{Xmin}|$$

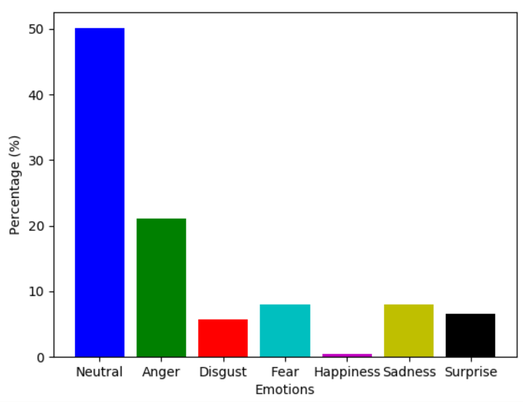
$$A_{ratio} = \frac{A_{current}}{A_{min}} \quad , \quad 1 \leq A_{ratio} \leq 5$$



4. Results and Discussion

A 1-minute segment of a speech given by Professor António J. R. Neves in TEDxAveiro 2019 was used for testing the developed software. When processing the segment, it was observed that the facial motion of the speaker when he was giving the speech, mainly mouth movement and head pose variation, contributed to some false positives. During the whole segment, the speaker presented a neutral expression, however the facial expression recognition model only detected that expression 50% of the times.

False positives. From left to right: anger, disgust, fear, happiness, sadness and surprise.



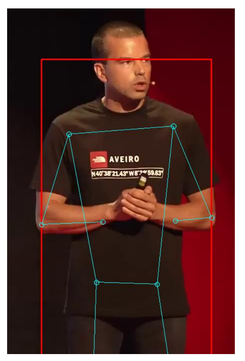
Emotion distribution for the whole segment.

The segment was successfully processed with PoseNet and the expansiveness was calculated using the A_{ratio} equation. The calculated expansiveness was 1.59.

Category	Value
Negative (anger, disgust, fear, sadness)	1
Neutral (neutral, surprise)	3
Positive (happiness)	5

Fusion	Label
1	Anxious
3	Comfortable
5	Confident

Fusion approach.



Processed frame from the segment.

The mood of the speaker was then calculated using the following equation:

$$\boldsymbol{Mood} = \frac{Expansiveness + (Negative + Neutral \times 3 + Positive \times 5)}{2}$$

The estimated mood was 1.87, which is somewhere between anxious and comfortable. This value is reasonable since the speaker revealed that he was nervous and anxious about the speech, but at the same time he was comfortable since he is an expert on the topic.

5. Conclusion and Further Work

This work presented a mood estimation algorithm based on facial expressions and postures using Computer Vision and Deep Learning. Using the keypoints of the body pose estimation, it was possible to estimate the expansiveness of a speaker, and using a simple CNN network trained with the CK+ dataset, it was possible to recognize his facial expressions. Using these two modalities, the mood of a speaker was estimated. As for future work, it is necessary to improve the facial expression recognition model in uncontrolled environments, as well as adding more relevant modalities, such as tone of voice and movement. The developed software also needs extensive testing to validate its accuracy.

6. References

[1] Lucey, Patrick, et al. "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression." 2010 ieee computer society conference on computer vision and pattern recognition-workshops. IEEE, 2010.

[2] Oved, D., I. Alvarado, and A. Gallo. "Real-time human pose estimation in the browser with TensorFlow. js." TensorFlow Medium, May (2018).

[3] Carney, Dana R., Amy JC Cuddy, and Andy J. Yap. "Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance." Psychological science 21.10 (2010): 1363-1368.

[4] Barros, Fábio, et al. "Understanding Public Speakers' Performance: First Contributions to Support a Computational Approach." International Conference on Image Analysis and Recognition. Springer, Cham, 2020.

7. Acknowledgements

Supported by the Integrated Programme of SR&TD SOCA (Ref. CENTRO-01-0145-FEDER-000010), co-funded by Centro 2020 program, Portugal 2020, European Union, through the European Regional Development Fund.