AFAR: A Deep Learning Based Tool for Automated Facial Affect Recognition

Itir Onal Ertugrul[1], László A. Jeni[1], Wanqiao Ding[2] and Jeffrey F. Cohn[2]
[1] Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA
[2] Department of Psychology, University of Pittsburgh, PA, USA

I. MOTIVATION

Automated facial affect recognition is crucial to multiple domains (e.g., health, education, entertainment). Commercial tools are available but costly and of unknown validity. Open-source ones [1] lack user-friendly GUI for use by non-programmers. For both types, evidence of domain transfer and options for retraining for use in new domains typically are lacking.

Deep approaches have two key advantages. They typically outperform shallow ones for facial affect recognition [2], [3], [4]. And pre-trained models provided by deep approaches can be fine tuned with new datasets to optimize performance.

We demo AFAR[1]: an open-source, deep-learning based, user-friendly tool for automated facial affect recognition.

II. SOFTWARE DESCRIPTION

AFAR consists of a pipeline having four components: (i) face tracking, ii) face registration, (iii) action unit (AU) detection and (iv) visualization as shown in Fig. 1. AFAR was written in the widely available MATLAB programming language and runs on Windows, Linux and OS platforms. It has been used to assess treatment response to deep brain stimulation (DBS), [5], investigate cross-domain generalizability [6], and explore facial affect in social interactions.

The first component is a real-time face tracking software, that can estimate a dense set of facial landmarks. This component outputs frame-level tracked landmark positions, 3D head orientation, and a video of tracked landmarks and head orientation overlaid on the input video.

The second component is a face registration software, that accomplishes dense 3D registration from 2D videos and images without requiring person-specific training [7]. Input to the second component is the input video and tracked landmark positions while the outputs include normalized face video and frame-level extracted features (SIFT, HoG).

The third component performs multi-label AU detection using the normalized face video as input. The current version is a 3-layered Convolutional Neural Network (CNN) architecture [5] that was trained using Extended BP4D+ on Keras platform. Input to this architecture is 200×200 pixel images of faces with 80 pixels interocular distance. This component outputs detection probabilities of 12 AUs: 1, 2, 4, 6, 7, 10, 12, 14, 15, 17, 23 and 24. The pretrained model is in AFAR so that users can fine tune it with their own datasets.

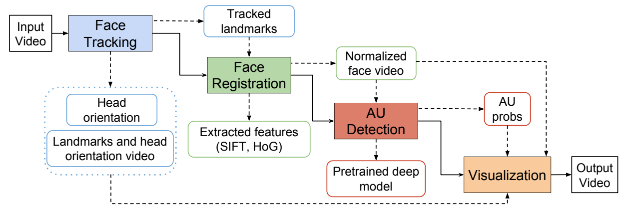

Fig. 1: Pipeline of the AFAR tool.

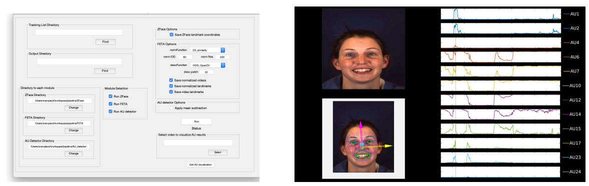

Fig. 2: GUI and Visualization

The fourth component is a visualization software (see Fig. 2) that allows users to visualize the outputs of the first three components. Users can visualize the input video, tracked landmarks, normalized face video, and detection probabilities of AUs simultaneously. AU probability signals are displayed, and current time is denoted with a moving line. This visualization allows for easy tracking of changes in the AU probabilities throughout the video.

III. ACKNOWLEDGMENTS

This research was supported in part by NIH awards NS100549 and MH096951.

REFERENCES

[1] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *FG*. IEEE, 2018, pp. 59–66.
[2] W.-S. Chu, F. De la Torre, and J. F. Cohn, "Learning spatial and temporal cues for multi-label facial action unit detection," in *FG*. IEEE, 2017, pp. 25–32.
[3] I. O. Ertugrul, L. A. Jeni, and J. F. Cohn, "Facscaps: Pose-independent facial action coding with capsules," in *CVPRW*, 2018, pp. 2211–2139.
[4] W. Li, F. Abtahi, Z. Zhu, and L. Yin, "Eac-net: Deep nets with enhancing and cropping for facial action unit detection," *TPAMI*, 2018.
[5] J. F. Cohn, L. A. Jeni, I. Onal Ertugrul, D. Malone, M. S. Okun, D. Borton, and W. K. Goodman, "Automated affect detection in deep brain stimulation for obsessive-compulsive disorder: A pilot study," in *ICMI*. ACM, 2018.
[6] I. Onal Ertugrul, J. F. Cohn, L. A. Jeni, Z. Zhang, L. Yin, and Q. Ji, "Cross-domain au detection: Domains, learning approaches, and measures," in *FG*. IEEE, 2019.
[7] L. A. Jeni, J. F. Cohn, and T. Kanade, "Dense 3d face alignment from 2d video for real-time use," *Image and Vision Computing*, vol. 58, pp. 13–24, 2017.

[1] `http://www.jeffcohn.net/resources/AFAR/`

978-1-7281-0089-0/19/$31.00 ©2019 IEEE