PyAFAR: Python-based Automated Facial Action Recognition library for use in Infants and Adults

Saurabh Hinduja[1], Itir Onal Ertugrul[2], Maneesh Bilalpur[3], Daniel S. Messinger[4], and Jeffrey F. Cohn[1]

[1] Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

[2] Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

[3] Intelligent Systems Program, University of Pittsburgh, Pittsburgh, USA

[4] Department of Psychology, University of Miami, Miami, FL, USA

***Abstract*—PyAFAR[1] is a Python-based, open-source facial action unit detection library for use with adults and infants. Convolutional neural networks were trained on BP4D+ for adults and MIAMI and CLOCK databases for infants. In adults, AU occurrence and intensity detection are enabled for 12 action units (AU). The AU chosen were selected on the criterion that they have base rates greater than 5% in BP4D+. Action unit intensity estimation is enabled for 5 of these AU. In infants, AU occurrence is enabled for 9 action units that are involved in expression of positive and negative affect. For both adults and infants, facial landmark and head pose tracking are enabled as well. For adults, multiple persons within a video may be tracked. The library is developed for ease of use. The models are available for fine-tuning and further training. PyAFAR may be easily incorporated into user Python code.**

I. Introduction

Facial action (AU) recognition has gained importance in multiple domains (e.g., marketing, education, and mental health). While there has been extensive work in detecting AUs [2], [3], [14], most are experimental and unavailable for general use. Open-source ones [1], [4] lack ease of use, fail to allow for retraining on new datasets, and lack robustness to faces of infants and young children [11]. We propose a new Python library, PyAFAR, with the following novelties:

1) Modules for adults and for infants
2) Platform independent, open source, easy to use library
3) Individualized tracking for multiple persons
4) Pre-trained models available for fine tuning
5) Customizable output

II. Software Description

PyAFAR was developed with ease of use in mind. It is fully developed in Python using other open-source libraries and is compatible with Windows, Linux, and MacOS platforms. The framework has three major components (i) Face detection and individualized tracking, (ii) face normalization and identification, and (iii) Action Unit detection.

The first component is face detection and individualized tracking. For face detection and dense 3D facial landmark estimation, Mediapipe [9] library is used. Facenet [13] enables tracking individuals even if they leave and reenter the scene. This component outputs frame level tracked landmarks and 3D head orientation calculated using the PnP perspective method [12].

The second component is face normalization. For face normalization we use dlib [8] library. Inputs to the second component are input video and tracked landmark positions while output is normalized face with background subtraction using a convex hull.

The third component performs AU detection and intensity estimation using the normalized frames as input. Detection modules are available for adults and infants. Both adult and infant modules are based on ResNet50 [7] that are pre-trained on ImageNet using tensorflow. The adult module is fine-tuned using BP4D+ [15]. The infant module is fine-tuned using MIAMI [6] and CLOCK [10]. Due to the recent advancements that raise privacy concerns, we release the encrypted infant module.

The output of the framework is customizable; it can be directly imported into user code and CSV and JSON files.

III. Results

Within database performance for AU detection was evaluated using 3-fold cross-validation in BP4D+ (adults), CLOCK and MIAMI (infants). (Tables I and II, respectively). For all AUs, PyAFAR achieved AUC values greater than 0.8 and 0.78 for adult and infant models, respectively.

Within database performance for AU intensity estimation was evaluated in BP4D+ in adults (Table III). ICC, Kendall Tau, and Spearman Correlation values are high for AU that have high base rates and less so for those that have low base rates. Cross-database performance and comparisons with OpenFace and the previous version of AFAR was done in GFT (Tables IV and V respectively). PyAFAR achieved the best PA on 7 of 12 AUs for cross-domain AU detection and achieved the best ICC for all AUs for cross-domain AU intensity estimation.

IV. Discussions

PyAFAR provides state-of-the-art AU detection in both adults and infants and intensity estimation in adults. In common with previous findings, performance for AU that have

Development of PyAFAR was supported in part by NIH awards R01MH096951, R01GM105004, and UH3NS100549.

[1]Code Available on: https://github.com/AffectAnalysisGroup/PyAFAR

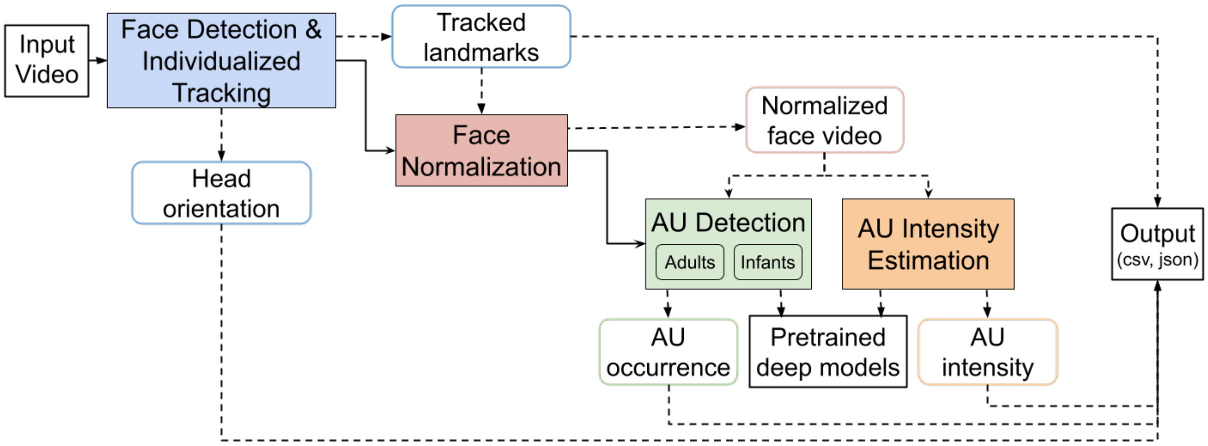

Fig. 1. Pipeline of PyAFAR library

TABLE I
WITHIN-DOMAIN DETECTION IN ADULTS IN BP4D+ [15]

AUs	Base Rates	AUC	PA	NA
1	0.09	0.84	0.47	0.94
2	0.11	0.81	0.38	0.94
4	0.03	0.82	0.42	0.97
6	0.32	0.94	0.86	0.85
7	0.41	0.90	0.89	0.75
10	0.40	0.94	0.92	0.82
12	0.32	0.95	0.90	0.85
14	0.11	0.80	0.82	0.66
15	0.09	0.85	0.46	0.94
17	0.29	0.84	0.48	0.92
23	0.22	0.85	0.57	0.90
24	0.12	0.90	0.27	0.99
Avg		0.87	0.61	0.88

TABLE II
WITHIN-DOMAIN OCCURRENCE DETECTION IN INFANTS IN CLOCK [10] AND MIAMI [10]

AUs	Base Rates	AUC	PA	NA
1	0.26	0.78	0.54	0.83
2	0.20	0.78	0.44	0.84
3	0.23	0.80	0.61	0.90
4	0.11	0.89	0.63	0.93
6	0.33	0.93	0.80	0.91
9	0.07	0.94	0.60	0.96
12	0.22	0.96	0.78	0.93
20	0.18	0.91	0.67	0.91
28	0.08	0.91	0.49	0.96
Avg		0.88	0.62	0.91

low base rates remains less than would be wished. Imbalanced data remain a stubborn challenge for PyAFAR as well as for OpenFace and other approaches. Poor performance for AU 14 may reflect changes in AU annotation in BP4D+ training data.

We release the codes and pretrained models for AU detection and intensity estimation modules in adults. They allow users to further fine-tune our models with their own databases for downstream tasks. For infants, AU detection modules are encrypted to maintain confidentiality of the participants as recent work [5] suggests that low resolution reconstruction of training inputs is possible.

TABLE III
WITHIN-DOMAIN INTENSITY ESTIMATION FOR ADULTS IN BP4D+ [15]

AU	ICC	Kendall Tau	Spear Corr
6	0.86	0.70	0.77
10	0.89	0.72	0.80
12	0.89	0.73	0.80
14	0.52	0.34	0.34
17	0.61	0.40	0.40
Avg	0.75	0.58	0.62

TABLE IV
CROSS-DATABASE COMPARISON FOR OCCURRENCE IN ADULTS IN GFT

AUs	PA			NA		
	Openface	AFAR	PyAFAR	Openface	AFAR	PyAFAR
1	0.18	0.11	**0.19**	0.88	**0.97**	0.96
2	**0.35**	0.03	0.25	0.85	**0.93**	**0.93**
4	0.09	**0.14**	0.10	0.79	0.86	**0.92**
6	0.62	0.63	**0.69**	0.76	0.79	**0.89**
7	0.60	0.68	**0.69**	**0.66**	**0.66**	0.63
10	0.56	0.62	**0.65**	0.74	0.80	**0.83**
12	0.62	0.67	**0.69**	0.70	0.85	**0.89**
14	**0.07**	**0.07**	**0.07**	0.43	**0.61**	0.57
15	0.23	**0.27**	0.25	0.81	0.93	**0.95**
17	0.46	0.28	**0.49**	0.78	**0.80**	0.77
23	0.32	**0.36**	0.32	0.70	**0.82**	0.72
24	-	**0.19**	0.16	-	**0.93**	0.86
Avg	0.37	0.35	**0.40**	0.74	**0.82**	**0.82**

TABLE V
CROSS-DATABASE ICC FOR AU INTENSITY IN ADULTS IN GFT

AU	Openface	PyAFAR
6	0.49	0.56
10	0.32	0.57
12	0.65	0.72
14	−0.01	0.00

REFERENCES

[1] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE, 2018.

[2] W. Chu et al. Learning spatial and temporal cues for multi-label facial action unit detection. In *Face and Gesture Recognition*, 2017.

[3] I. Ertugrul et al. Cross-domain au detection: Domains, learning approaches, and measures. *FG*, 2019.

[4] I. O. Ertugrul, L. A. Jeni, W. Ding, and J. F. Cohn. Afar: A deep learning based tool for automated facial affect recognition. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–1. IEEE, 2019.

[5] N. Haim, G. Vardi, G. Yehudai, O. Shamir, and M. Irani. Reconstructing training data from trained neural networks. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 22911–22924. Curran Associates, Inc., 2022.

[6] Z. Hammal, J. F. Cohn, and D. S. Messinger. Head movement dynamics during play and perturbed mother-infant interaction. *IEEE transactions on affective computing*, 6(4):361–370, 2015.

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[8] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[9] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann. Mediapipe: A framework for perceiving and processing reality. In *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019*, 2019.

[10] D. V. Luquetti, M. L. Speltz, E. R. Wallace, B. Siebold, B. R. Collett, A. F. Drake, A. L. Johns, K. A. Kapp-Simon, S. L. Kinter, B. G. Leroux, et al. Methods and challenges in a cohort study of infants and toddlers with craniofacial microsomia: the clock study. *The Cleft palate-craniofacial journal*, 56(7):877–889, 2019.

[11] I. Onal Ertugrul, Y. A. Ahn, M. Bilalpur, D. S. Messinger, M. L. Speltz, and J. F. Cohn. Infant afar: Automated facial action recognition in infants. *Behavior Research Methods*, pages 1–12, 2022.

[12] F. Rocca, M. Mancas, and B. Gosselin. Head pose estimation by perspective-n-point solution based on 2d markerless face tracking. In *International Conference on Intelligent Technologies for Interactive Entertainment*, pages 67–76. Springer, 2014.

[13] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015.

[14] Z. Shao, Z. Liu, J. Cai, and L. Ma. Jâa-net: Joint facial action unit detection and face alignment via adaptive attention. *International Journal of Computer Vision*, 2020.

[15] Z. Zheng et al. Multimodal spontaneous emotion corpus for human behavior analysis. In *Computer Vision and Pattern Recognition*, pages 3438–3446, 2016.