Article

Viewpoint Robustness of Automated Facial Action Unit Detection Systems

Shushi Namba [1],* , **Wataru Sato** [1],* and **Sakiko Yoshikawa** [2]

1   Psychological Process Team, Guardian Robot Project, RIKEN, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

2   Faculty of Arts, Kyoto University of the Arts, 2-116 Uryuyama Kitashirakawa, Sakyo, Kyoto 606-8271, Japan; yoshikawa.sakiko.4n@kyoto-u.ac.jp

*   Correspondence: shushi.namba@riken.jp (S.N.); wataru.sato.ya@riken.jp (W.S.); Tel.: +81-0774-95-1360 (W.S.)

Abstract: Automatic facial action detection is important, but no previous studies have evaluated pre-trained models on the accuracy of facial action detection as the angle of the face changes from frontal to profile. Using static facial images obtained at various angles ($0°$, $15°$, $30°$, and $45°$), we investigated the performance of three automated facial action detection systems (FaceReader, OpenFace, and Py-feat). The overall performance was best for OpenFace, followed by FaceReader and Py-Feat. The performance of FaceReader significantly decreased at $45°$ compared to that at other angles, while the performance of Py-Feat did not differ among the four angles. The performance of OpenFace decreased as the target face turned sideways. Prediction accuracy and robustness to angle changes varied with the target facial components and action detection system.

Keywords: action unit; angle; automatic facial detection

1. Introduction

With the rise in affectivism [1], automatic facial action detection system is gaining widespread attention [2–5]. Given the importance of facial expressions for emotional communication [6,7], machines that can automatically detect facial movements are expected to be useful for studies of facial expressions and emotional communication. The most common index of facial movement is the action unit (AU), which is derived from the Facial Action Coding System (FACS) [8]. FACS describes facial movements in anatomical terms, while AU represents a single facial movement. For example, pulling of the lip corner and lowering of the eyebrows correspond to AUs 12 and 4, respectively. Automated facial action detection systems mainly provide outputs in terms of AUs.

As for automated facial action detection systems, the processing pipeline includes face detection, face registration, feature extraction, and classification [9]. A wide variety of algorithms have been proposed at each stage [2–5]. For example, Shao et al. [10] proposed a novel end-to-end deep learning framework for joint AU detection and face alignment, resulting in robust AU detection. Many facets of automated facial action detection systems (e.g., 3D [11] and multimodal information [12]) have been discussed in several review papers [2–5,13,14]. However, too few automated facial expression analysis systems are available for novice users who are motivated to study human facial expressions but are not experts in engineering techniques.

For such demands, the available options include commercial software [15] and limited open software (e.g., OpenFace [16], AFARtoolbox [17], and Py-Feat [18]), which usually have several pre-trained facial databases. Commercial tools are expensive, and open-source software often lacks user-friendly graphical user interfaces. To evaluate the software available to non-programmers with the best graphical user interface and performance for facial action detection, Namba et al. [19] compared the performances of three automatic facial action detection systems (AFAR toolbox, OpenFace, and FaceReader) using several dynamic

facial databases, including one derived from YouTube. OpenFace [16] was able to detect facial expressions more accurately than the other automatic facial action detection systems, including commercial software. Thus, among the many automated facial action detection systems available, OpenFace is easy for naïve researchers to use and performs well.

The identification of facial expressions in an unconstrained setting is limited by the commonly observed variability among facial poses. A major issue is the perception of facial movements from varying angles, including the facial profile. Human facial angles influence the psychological reactions of observers. A recent study reported that emotion recognition was more accurate for the frontal view of the face compared to the profile [20]. However, other studies reported that emotion recognition had similar accuracy between the profile and frontal views of the face [21,22]. Additionally, perception intensity [21], face recognition [23], impression evaluation [24], and threat perception [25] depend on the facial angle. Assuming that perceptions of facial expressions change with facial angle, automated facial action detection systems must show robustness to facial angles with respect to the detection of facial expressions to be useful for psychological experiments. Thus, it is important to systematically evaluate the robustness of these systems to different facial angles.

To the best of our knowledge, no previous study has compared the facial expression detection accuracy of pre-trained facial action detection systems as the facial angle changes from the frontal to profile view. The current study aimed to investigate the viewpoint robustness of four automated facial action detection systems: OpenFace, AFAR toolbox, Py-feat, and FaceReader. The former three are open software packages designed for use in academic research, whereas FaceReader is a commercial software package. The AFAR toolbox has a smaller number of estimable AUs than the other software packages. Furthermore, the AFAR toolbox did not reliably estimate AU intensity for more than half of the facial stimuli used in this study when the yaw angle was 45° (available N/all N = 16/72). To make full use of the available data, here, we report analysis results excluding data computed by AFAR toolbox. However, for their informational value, we also report the results that include AFAR toolbox data in Supplementary Materials (hence the variation in the amount of data for each angle) (Figure S1).

This study aimed to evaluate the robustness of automated facial action detection systems (FaceReader, OpenFace, and Py-feat) to varying camera poses (0°, 15°, 30°, and 45°) when calculating AUs for static facial images. These data were compared with those obtained via manual FACS coding. The exploratory nature of the study did not permit directional hypotheses to be formulated.

2. Materials and Methods

2.1. Dataset

Facial emotional expressions of 12 Japanese people (6 females and 6 males) were used as stimuli. Part of the data (i.e., analysis of data for the 0° condition) are reported elsewhere [26]. Each expresser was videotaped while expressing anger, disgust, fear, happiness, sadness, and surprise at angles of 0°, 15°, 30°, and 45° from the left using four video cameras (DSR-PD150, Sony, Tokyo, Japan). The expressers were instructed to "Imitate the facial expression in the photograph." As a model, we presented photographs of emotional expressions selected from Pictures of Facial Affect [27] (model: JJ). The video data were digitized and synchronized, and then a coder who was blind to the research objectives selected the apex frames of emotional expressions. The final dataset resulted in 288 images (i.e., 12 individuals × 6 emotions × 4 angles). Figure 1 shows representative photographs of an expresser. This study was approved by the Ethics Committee of the Primate Research Institute, Kyoto University, Japan, and conducted in accordance with the ethical guidelines of our institute and the Declaration of Helsinki. Manual FACS coding was performed by two certified coders. The dataset contained the following 20 AUs: 1 (inner brow raised), 2 (outer brow raised), 4 (brow lowered), 5 (upper lid raised), 6 (cheek raised), 7 (lid tightened), 9 (nose wrinkled), 10 (upper lip raised), 12 (lip corner pulled),

14 (dimple), 15 (lip corner depressed), 17 (chin raised), 18 (lip puckered), 20 (lip stretched), 23 (lip tightened), 24 (lip pressed), 25 (lips parted), 26 (jaw dropped), 27 (mouth stretched), and 43 (eyes closed). The data were binary (presence or absence of each action unit in an expression). Interrater reliability was confirmed (Cohen's kappa = 0.81).
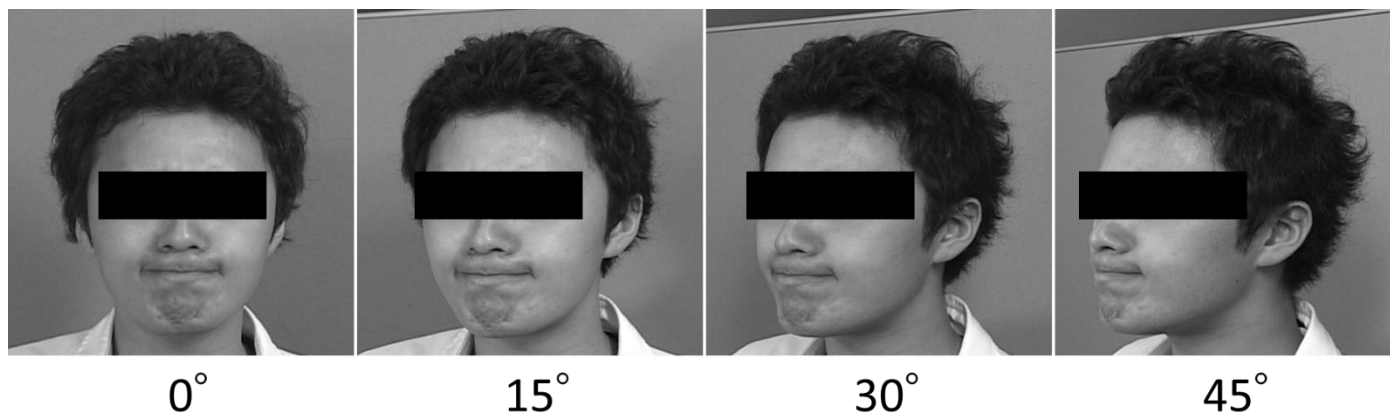

Figure 1. Example facial images exhibiting anger. The original images were in full color.

2.2. Automatic Facial Action Detection System

2.2.1. FaceReader

We used version 7 of FaceReader (Noldus Information Technology, Wageningen, The Netherlands). Although no information was available regarding the algorithms and training data of FaceReader, the documentation stated that it uses Active Appearance Models and facial shape as feature vectors, and a neural network learning model. Twenty AUs can be predicted (1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 18, 20, 23, 24, 25, 26, 27, and 43). Because the current study used facial expressions made by Japanese participants, we used a facial model trained on East Asian faces.

2.2.2. OpenFace

OpenFace was developed to detect facial landmarks, head pose, facial AUs, and eye gaze [16]. OpenFace detects facial landmarks using the Convolutional Experts Constrained Local Model (CE-CLM) [28] and dimensionally reduced histograms of oriented gradients as feature vectors. It uses a linear kernel support vector machine learning model. The following databases were used for model training: Bosphorus [29], BP4D [30], DISFA [31], FERA2011 [32], SEMAINE [33], UNBC [34], and CK+ [35]. Eighteen AUs were recognized (AUs 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28 (lip suck), and 45 (blink)). We used the default settings of the CE-CLM.

2.2.3. Py-Feat

Py-Feat [18] is an open-source software package that uses Python-based tools to detect facial expressions in images and videos. This toolbox uses several models, including state-of-the-art algorithms for detection of faces, facial landmarks, head poses, AUs, and emotions (https://py-feat.org/content/intro.html (accessed on 25 November 2021)). The current study used the default model settings: face detection = RetinaFace [36], facial landmark detection = MobileNets [37], and AU detection = Random Forest. The model was trained on histograms of oriented gradients extracted from BP4D, DISFA, CK+, UNBC, and AFF-wild2 [38–43]. Twenty AUs were recognized (AUs 1, 2, 4, 5, 6, 7, 9, 10, 11 (deep nasolabial fold), 12, 14, 15, 17, 20, 23, 24, 25, 26, 28, and 43).

2.3. Evaluation Metrics

To evaluate the detection algorithms, we calculated biserial correlations of AUs between manual and machine FACS coding according to angle. The manual coding data for the frontal view were considered the "gold standard" when assessing machine per-

formance. A 3 (between-group factor: machine system) × 4 (within-group factor: angle) mixed-design ANOVA was used for the analysis.

Area under the curve (AUC) values are superior to other indices (such as F1 scores) for comparison of the performance of automated facial action detection systems and were thus used here [19,44].

Statistical analyses were performed using R statistical software (version 4.0.3; R Development Core Team, Vienna, Austria) and the "pROC", "data.table", "psych", and "tidyverse" R packages [45–47].

3. Results

To evaluate the robustness of the automatic AU detection systems to changes in angle, biserial correlations between manual FACS coding and the machine output for each facial image were calculated, as stated above. In the ANOVA model, angle was the within-group factor (0°, 15°, 30°, and 45°) and machine type was the between-group factor (FaceReader, OpenFace, and Py-Feat), as also stated above. The number of images differed among angles because FaceReader sometimes failed to fit or compute facial data (0°: 72 images, 15°: 71 images, 30°: 69 images, and 45°: 60 images).

The ANOVA (Figure 2) showed a main effect of system ($F(2, 536) = 22.07$, $\eta^2_p = 0.08$, $p < 0.001$). Multiple comparisons using Shaffer's modified sequentially rejective Bonferroni procedure indicated that the correlation coefficient for OpenFace was higher than that for the other two systems ($ts > 3.46$, $ps < 0.001$). The correlation coefficient was also higher for FaceReader than Py-Feat ($t(268) = 2.93$, $p < 0.004$, *Hedge's g* = 0.25, 95% *CI* = 0.08–0.42).
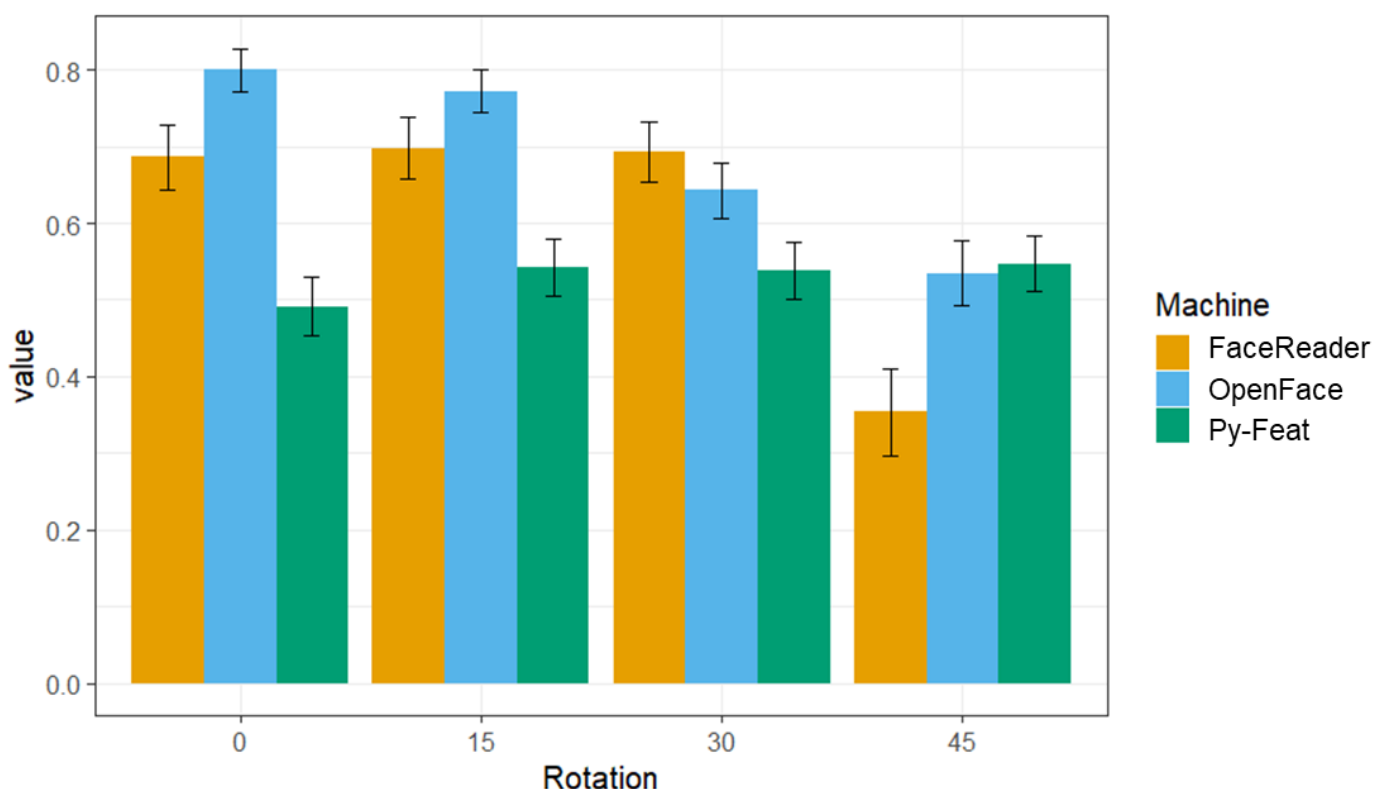

Figure 2. Average biserial correlation values between manually coded and machine-predicted action units (AUs) by angle. Error bars represent standard error.

Machine performance significantly differed by facial angle ($F(3, 268) = 9.82$, $\eta^2_p = 0.10$, $p < 0.001$). Performance at 45° was significantly lower than at other angles ($ts > 3.69$, $ps < 0.001$), while no significant performance difference was observed at any other angle ($ts < 1.23$, $ps > 0.66$).

There was also a significant interaction effect between machine and angle ($F(6, 536) = 8.37$, $\eta^2_p = 0.09$, $p < 0.001$). For OpenFace and FaceReader, there was a significant main effect of angle ($Fs > 12.51$, $\eta^2_ps > 0.12$, $ps < 0.001$) but not for Py-feat ($F(3, 268) = 0.48$, $\eta^2_p = 0.01$, $p = 0.70$). For OpenFace, there was no significant difference in performance between 0° and 15°, or between 30° and 45° ($ts < 2.17$, $ps > 0.06$). However, in the other comparisons, the correlation coefficient gradually decreased from 0° to 45° ($ts > 2.74$, $ps < 0.02$). For FaceReader, the performance at 45° was significantly lower than that at the other angles ($ts > 5.17$, $ps < 0.001$), but there were no other significant differences ($ts < 0.19$, $ps = 1.00$).

To confirm that the facial particularities of the individual being tested affected the results, the analysis was repeated with each participant removed. The result showed a significant effect similar to the above (main effect of system $Fs > 15.74$, $ps < 0.000$ and facial angle $Fs > 7.60$, $ps < 0.000$; interaction effect $Fs > 6.86$, $ps < 0.000$). The estimated performance accuracy for each individual's data is depicted visually in Supplementary Materials.

Figure 3 shows the average AUC values by AU. The receiver operating characteristic curves are available online (https://bit.ly/2ZKaT22 (accessed on 25 November 2021)). The AUC values for the facial angles varied among the AUs. A sharp reduction in AUC values was noted at 45° for FaceReader. The variation by angle was relatively low for AUs 7 (lid tightened), 12 (lip corner pulled), and 25 (lips parted), excluding at 45° for FaceReader. Regardless of angle, all three systems detected AUs 12 and 25 accurately but had difficulty in detecting AU 7. For AU 9 (nose wrinkled), OpenFace outperformed Py-feat and FaceReader. Py-feat had a significantly lower AUC value for AU 4 (brow lowered) and higher scores for AUs 5 (upper lid raised) and 24 (lip pressed) than the other two systems. Importantly, in contrast to OpenFace and FaceReader, AUs in the frontal view (0°) were not easily detected by Py-Feat. For Py-Feat, the AUC values for the frontal view were lower than those of the other systems for AUs 14 (dimpled), 15 (lip corner depressed), and 23 (lip tightened).
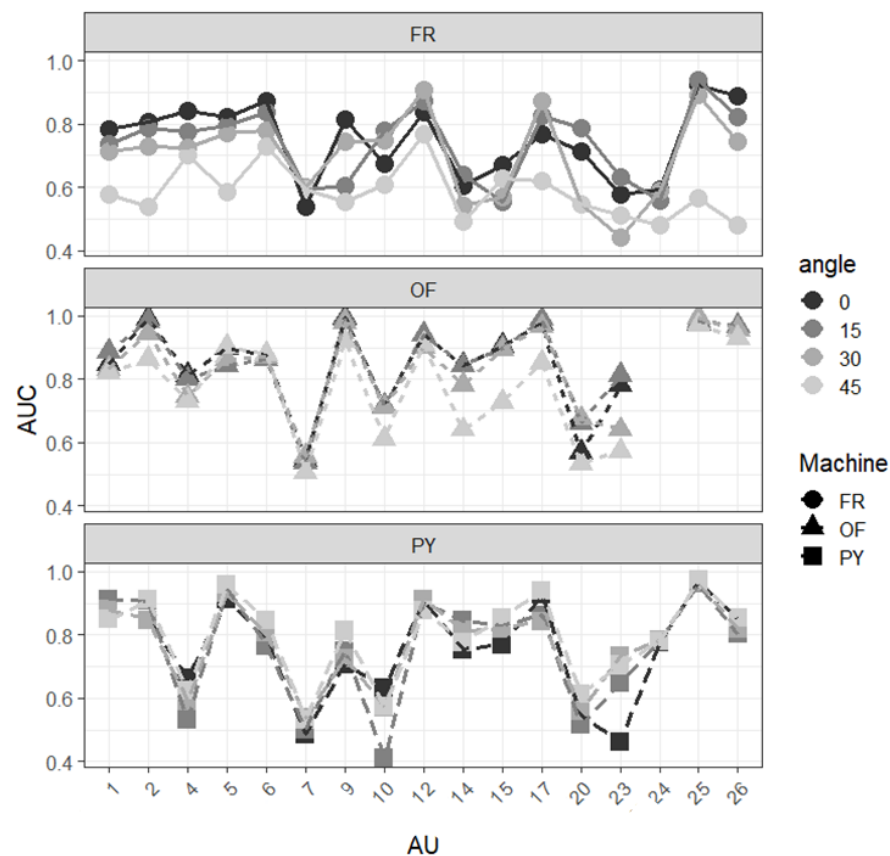

Figure 3. Average area under the curve (AUC) values for action units (AUs) predicted by the machines. FR: FaceReader; OF: OpenFace; PY: Py-Feat.

4. Discussion

We evaluated the robustness of three automated facial action detection systems to changes in facial angle by calculating the accuracy of predicted AUs. OpenFace had the best performance, followed by FaceReader and Py-Feat. The performance was significantly lower at 45°, and there was no significant performance difference among 0°, 15° and 30°. Importantly, the robustness to angle changes differed among the systems. FaceReader had markedly poorer performance at 45°. For OpenFace, performance was similar between 0° and 15°, and between 30° and 45°. However, the performance gradually decreased as the angle moved toward the profile view. Although the performance of Py-feat was poorer than the other two systems, it did not vary by angle.

FaceReader provided inaccurate results at 45°. The poor performance of FaceReader for recognizing dynamic facial expressions reported in a previous study [19] may be due to its vulnerability to angle changes. In this study, the AUC values were significantly lower at 45°, especially for AUs 1 (inner brow raised), 2 (outer brow raised), 5 (upper lid raised), 25 (lips parted), and 26 (jaw dropped). These AUs correspond to the prototypical surprised facial expression [8]. Therefore, FaceReader may not be suitable for evaluating the prototypical surprised expression, suggesting that future models should learn to recognize that expression from various angles.

OpenFace had the best AU detection performance among the systems evaluated in this study, consistent with the previous findings pertaining to dynamic facial expressions [19]. Among the pre-trained models, OpenFace had the best AU detection performance based on facial expression images. As shown in Figure 3, the AUC values for AUs 14 (dimpled), 15 (lip corner depressed), 17 (chin raised), and 23 (lip tightened) were lower at 45° compared to the frontal view. When using OpenFace, recognition of the lower part of the facial expression was likely to be affected by angle, whereas for static facial images, the AUC values were high for AUs 2 (outer brow raised), 9 (nose wrinkled), 12 (lip corner pulled), 17 (chin raised), and 25 (lips parted), even at 45°.

Although the overall performance of Py-Feat was not superior to that of FaceReader and OpenFace, it provided robust results at 0–45°. Therefore, the use of a large amount of "wild" data (i.e., Aff-Wild2 [38–43]), including various angles, for model pre-training may have been effective. The lower performance for the frontal view compared to the other angles for AU 23 may have been related to the properties of the training data. Importantly, despite the lower AU detection performance in the frontal view of Py-Feat compared to the other two systems, the detection of AU 5 (upper lid raised) was superior for Py-Feat (Figure 3). Therefore, Py-Feat may be particularly useful for evaluating eyelid opening in a static facial image.

Novice users should use a pre-trained automated facial action detection system; OpenFace may be the best choice. The performance of OpenFace is affected by facial angle, but it was still more robust to angle changes than FaceReader (Figure 2). Despite the inadequate performance of Py-Feat in the current study, it is still very promising. Py-Feat uses multiple models for face detection (e.g., MTCNN [48] and FaceBoxes [49]), AUs (e.g., JAANET [10] and DRML [50]), and facial landmarks (e.g., PFLD [51] and MobileFaceNet [52]). In this study, we used the default settings, which are the best-performing combination [18], although other combinations of pre-trained models may be used depending on the properties of the target facial expressions. Moreover, the settings of Py-Feat and OpenFace can be fine-tuned. The overall performance of Py-Feat can be improved by using a database of frontal view facial expressions as training data. Future research is required to improve its robustness to angle changes.

We investigated the performance of three systems in terms of identifying AUs from static facial images at various angles. Several limitations of this study must be mentioned. First, automated facial action detection systems may be affected by training data bias. Garcia et al. [53] reported that algorithms for deep face recognition show demographic bias. Dailey et al. [54] reported that facial expression detection systems performed differently depending on whether they were trained on Asian or Western facial expressions. To

overcome this problem, it is necessary to develop training and test datasets with a wide range of facial expressions. When developing an algorithm, the design is informed by the robustness to angle changes. Niimura et al. [55] evaluated robustness to pose variation, including pitch, of algorithms for "deep facial action coding" according to pre-training, feature alignment, model size, and optimizers. Many computer vision researchers have attempted to develop pose normalization techniques [56,57]. Li and Deng [3] recommended the use of the GAN network to overcome the effects of variations, including in head pose, when developing deep learning models to analyze facial expressions. We hope that future research will lead to the development of automated facial action detection systems accessible to novice users and non-programmers

5. Conclusions

This study showed that the overall recognition performance for static facial images, at various angles (0–45°), was best for OpenFace, followed by FaceReader and Py-Feat. The performance of FaceReader was significantly lower at 45° compared to the other angles, while Py-Feat's performance did not differ by angle. The performance of OpenFace decreased as the target faces turned sideways. Prediction accuracy and robustness to angle changes differed depending on the type of target facial components (i.e., AUs). Facial AU detection systems should be selected according to the particular purpose of the user.

References

1. Dukes, D.; Abrams, K.; Adolphs, R.; Ahmed, M.E.; Beatty, A.; Berridge, K.C.; Broomhall, S.; Brosch, T.; Campos, J.J.; Clay, Z.; et al. The rise of affectivism. *Nat. Hum. Behav.* **2021**, *5*, 816–820. [CrossRef]
2. Cohn, J.F.; Ertugrul, I.O.; Chu, W.S.; Girard, J.M.; Jeni, L.A.; Hammal, Z. Affective facial computing: Generalizability across domains. In *Multimodal Behavior Analysis in the Wild*; Alameda-Pineda, X., Ricci, E., Sebe, N., Eds.; Academic Press: Cambridge, MA, USA, 2019; pp. 407–441.
3. Li, S.; Deng, W. Deep facial expression recognition: A survey. *IEEE Trans. Affect. Comput.* **2020**. [CrossRef]
4. Martinez, B.; Valstar, M.F.; Jiang, B.; Pantic, M. Automatic analysis of facial actions: A survey. *IEEE Trans. Affect. Comput.* **2017**, *10*, 325–347. [CrossRef]
5. Zhi, R.; Liu, M.; Zhang, D. A comprehensive survey on automatic facial action unit analysis. *Vis. Comput.* **2020**, *36*, 1067–1093. [CrossRef]
6. Bavelas, J.; Chovil, N. Some pragmatic functions of conversational facial gestures. *Gesture* **2018**, *17*, 98–127. [CrossRef]
7. Van Kleef, G.A. The emerging view of emotion as social information. *Soc. Personal. Psychol. Compass* **2010**, *4*, 331–343. [CrossRef]
8. Ekman, P.; Friesen, W.V.; Hager, J.C. *Facial Action Coding System*, 2nd ed.; Research Nexus eBook: Salt Lake City, UT, USA, 2002.

9. Valstar, M. Automatic facial expression analysis. In *Understanding Facial Expressions in Communication*; Mandak, M.K., Awasthi, A., Eds.; Springer: New Delhi, India, 2015; pp. 143–172.
10. Shao, Z.; Liu, Z.; Cai, J.; Ma, L. JÂA-net: Joint facial action unit detection and face alignment via adaptive attention. *Int. J. Comput. Vis.* **2021**, *129*, 321–340. [CrossRef]
11. Zhang, Y.; Zhang, L.; Hossain, M.A. Adaptive 3D facial action intensity estimation and emotion recognition. *Expert Syst. Appl.* **2015**, *42*, 1446–1464. [CrossRef]
12. Meng, Z.; Han, S.; Liu, P.; Tong, Y. Improving speech related facial action unit recognition by audiovisual information fusion. *IEEE Trans. Cybern.* **2018**, *49*, 3293–3306. [CrossRef] [PubMed]
13. Liu, Y.; Zhang, X.; Zhou, J.; Li, X.; Li, Y.; Zhao, G.; Li, Y. Graph-based Facial Affect Analysis: A Review of Methods, Applications and Challenges. *arXiv* **2021**, arXiv:2103.15599.
14. Perusquía-Hernández, M. Are people happy when they smile?: Affective assessments based on automatic smile genuineness identification. *Emot. Stud.* **2021**, *6*, 57–71.
15. Dupré, D.; Krumhuber, E.G.; Küster, D.; McKeown, G.J. A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLoS ONE* **2020**, *15*, e0231968. [CrossRef] [PubMed]
16. Baltrušaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.P. OpenFace 2.0: Facial behavior analysis toolkit. In Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG), Xi'an, China, 15–19 May 2018; pp. 59–66.
17. Ertugrul, I.O.; Cohn, J.F.; Jeni, L.A.; Zhang, Z.; Yin, L.; Ji, Q. Crossing domains for AU coding: Perspectives, approaches, and measures. *IEEE Trans. Biom. Behav. Identity Sci.* **2020**, *2*, 158–171. [CrossRef] [PubMed]
18. Cheong, J.H.; Xie, T.; Byrne, S.; Chang, L.J. Py-Feat: Python Facial Expression Analysis Toolbox. *arXiv* **2021**, arXiv:2104.03509.
19. Namba, S.; Sato, W.; Osumi, M.; Shimokawa, K. Assessing Automated Facial Action Unit Detection Systems for Analyzing Cross-Domain Facial Expression Databases. *Sensors* **2021**, *21*, 4222. [CrossRef] [PubMed]
20. Surcinelli, P.; Andrei, F.; Montebarocci, O.; Grandi, S. Emotion Recognition of Facial Expressions Presented in Profile. *Psychol. Rep.* **2021**, 00332941211018403. [CrossRef]
21. Guo, K.; Shaw, H. Face in profile view reduces perceived facial expression intensity: An eye-tracking study. *Acta. Psychol.* **2015**, *155*, 19–28. [CrossRef]
22. Matsumoto, D.; Hwang, H.S. Judgments of facial expressions of emotion in profile. *Emotion* **2011**, *11*, 1223–1229. [CrossRef]
23. Hill, H.; Schyns, P.G.; Akamatsu, S. Information and viewpoint dependence in face recognition. *Cognition* **1997**, *62*, 201–222. [CrossRef]
24. Sutherland, C.A.; Young, A.W.; Rhodes, G. Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties. *Br. J. Psychol.* **2017**, *108*, 397–415. [CrossRef] [PubMed]
25. Hess, U.; Adams, R.B.; Kleck, R.E. Looking at you or looking elsewhere: The influence of head orientation on the signal value of emotional facial expressions. *Motiv. Emot.* **2007**, *31*, 137–144. [CrossRef]
26. Ekman, P.; Friesen, W.V. *Pictures of Facial Affect*; Consulting Psychologist: Palo Alto, CA, USA, 1976.
27. Sato, W.; Hyniewska, S.; Minemoto, K.; Yoshikawa, S. Facial expressions of basic emotions in Japanese laypeople. *Front. Psychol.* **2019**, *10*, 259. [CrossRef]
28. Zadeh, A.; Chong, L.Y.; Baltrusaitis, T.; Morency, L.P. Convolutional experts constrained local model for 3d facial landmark detection. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 2519–2528.
29. Savran, A.; Alyüz, N.; Dibeklio ğlu, H.; Çeliktutan, O.; Gökberk, B.; Sankur, B.; Akarun, L. Bosphorus database for 3D face analysis. In Proceedings of the European Workshop on Biometrics and Identity Management, Roskilde, Denmark, 7–8 May 2008; pp. 47–56.
30. Zhang, X.; Yin, L.; Cohn, J.F.; Canavan, S.; Reale, M.; Horowitz, A.; Liu, P.; Girard, J.M. Bp4d-spontaneous: A high-resolution spontaneous 3D dynamic facial expression database. *Image Vis. Comput.* **2014**, *32*, 692–706. [CrossRef]
31. Mavadati, S.M.; Mahoor, M.H.; Bartlett, K.; Trinh, P.; Cohn, J.F. DISFA: A spontaneous facial action intensity database. *IEEE Trans. Affect. Comput.* **2013**, *4*, 151–160. [CrossRef]
32. Valstar, M.F.; Jiang, B.; Mehu, M.; Pantic, M.; Scherer, K. The first facial expression recognition and analysis challenge. In Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, 21–25 March 2011; pp. 921–926.
33. McKeown, G.; Valstar, M.; Cowie, R.; Pantic, M.; Schroder, M. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Trans. Affect. Comput.* **2011**, *3*, 5–17. [CrossRef]
34. Lucey, P.; Cohn, J.F.; Prkachin, K.M.; Solomon, P.E.; Matthews, I. Painful data: The UNBC-McMaster shoulder pain expression archive database. In Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, 21–25 March 2011; pp. 57–64.
35. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.
36. Deng, J.; Guo, J.; Zhou, Y.; Yu, J.; Kotsia, I.; Zafeiriou, S. Retinaface: Single-stage dense face localisation in the wild. *arXiv* **2019**, arXiv:1905.00641.

37. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.

38. Kollias, D.; Nicolaou, M.A.; Kotsia, I.; Zhao, G.; Zafeiriou, S. Recognition of affect in the wild using deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 26–33.

39. Kollias, D.; Zafeiriou, S. Aff-wild2: Extending the Aff-wild database for affect recognition. *arXiv* **2018**, arXiv:1811.07770.

40. Kollias, D.; Zafeiriou, S. A multi-task learning & generation framework: Valence–arousal, action units & primary expressions. *arXiv* **2018**, arXiv:1811.07771.

41. Kollias, D.; Zafeiriou, S. Expression, affect, action unit recognition: Aff-wild2, multi-task learning and ArcFace. *arXiv* **2019**, arXiv:1910.04855.

42. Kollias, D.; Tzirakis, P.; Nicolaou, M.A.; Papaioannou, A.; Zhao, G.; Schuller, B.; Kotsia, I.; Zafeiriou, S. Deep affect prediction in-the-wild: Aff-wild database and challenge, deep architectures, and beyond. *Int. J. Comput. Vis.* **2019**, *127*, 907–929. [CrossRef]

43. Zafeiriou, S.; Kollias, D.; Nicolaou, M.A.; Papaioannou, A.; Zhao, G.; Kotsia, I. Aff-wild: Valence and arousal 'n-the-Wild' challenge. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 34–41.

44. Jeni, L.A.; Cohn, J.F.; De La Torre, F. Facing imbalanced data—Recommendations for the use of performance metrics. In Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction, Washington, DC, USA, 2–5 September 2015; pp. 245–251.

45. Dowle, M.; Srinivasan, A. Data.table: Extension of 'data.frame'. R Package, Version 1.14.2. 2021. Available online: https://CRAN.R-project.org/package=data.table (accessed on 6 October 2021).

46. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [CrossRef]

47. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.A.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]

48. Zhang, N.; Luo, J.; Gao, W. Research on Face Detection Technology Based on MTCNN. In Proceedings of the 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 25–27 September 2020; pp. 154–158.

49. Zhang, S.; Zhu, X.; Lei, Z.; Shi, H.; Wang, X.; Li, S.Z. FaceBoxes: A CPU real-time face detector with high accuracy. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 1–9.

50. Zhao, K.; Chu, W.S.; Zhang, H. Deep region and multi-label learning for facial action unit detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3391–3399.

51. Guo, X.; Li, S.; Yu, J.; Zhang, J.; Ma, J.; Ma, L.; Liu, W.; Ling, H. PFLD: A practical facial landmark detector. *arXiv* **2019**, arXiv:1902.10859.

52. Chen, S.; Liu, Y.; Gao, X.; Han, Z. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In Proceedings of the Chinese Conference on Biometric Recognition, Urumchi, China, 11–12 August 2018; pp. 428–438.

53. Garcia, R.V.; Wandzik, L.; Grabner, L.; Krueger, J. The harms of demographic bias in deep face recognition research. In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4–7 June 2019; pp. 1–6.

54. Dailey, M.N.; Joyce, C.; Lyons, M.J.; Kamachi, M.; Ishi, H.; Gyoba, J.; Cottrell, G.W. Evidence and a computational explanation of cultural differences in facial expression recognition. *Emotion* **2010**, *10*, 874–893. [CrossRef] [PubMed]

55. Niinuma, K.; Onal Ertugrul, I.; Cohn, J.F.; Jeni, L.A. Systematic Evaluation of Design Choices for Deep Facial Action Coding Across Pose. *Front. Comput. Sci.* **2021**, *3*, 27. [CrossRef]

56. Hassner, T.; Harel, S.; Paz, E.; Enbar, R. Effective face frontalization in unconstrained images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4295–4304.

57. Sagonas, C.; Panagakis, Y.; Zafeiriou, S.; Pantic, M. Robust statistical face frontalization. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3871–3879.