applied sciences

MDPI

Supplementary Materials: Viewpoint Robustness of Automated Facial Action Unit Detection Systems

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

1. Supplementary Information

Automatic Facial Action Detection System

AFAR

This study used the AFAR toolbox, which is an open-source, deep-learning based, user-friendly tool for automated facial movement detection. The details are described in Ertugrul et al. (2020). The target AUs were as follows: 1, 2, 4, 6, 7, 10, 12, 14, 15, 17, 23, and 24.

The other three systems (FaceReader, OpenFace, and Py-feat) are described in the main text; the AUs assessed were similar, and 11 overlapped: 1, 2, 4, 6, 7, 10, 12, 14, 15, 17, and 23.

Results and Discussions

To evaluate the robustness of the automatic AU detection systems to facial angle variation, we calculated biserial correlations between manual FACS coding and machine output. In the ANOVA, angle was the within-group factor (0°, 15°, 30°, and 45°) and machine type was the between-group factor (AFAR, FaceReader, OpenFace, and Py-Feat). The number of samples differed among angles because FaceReader and the AFAR toolbox sometimes failed to fit or compute facial data (0°: 72 samples, 15°: 68 samples, 30°: 49 samples, and 45°: 11 samples).

The ANOVA (Figure 1A) showed a main effect of system ($F(3, 588) = 35.41$, partial eta = 0.15, $p < 0.001$). Multiple comparisons using Shaffer's modified sequentially rejective Bonferroni procedure indicated that the correlation coefficients for OpenFace and FaceReader were higher than for the other two systems ($t > 2.67$, $p < 0.02$). The correlation coefficient for AFAR was higher than that for Py-Feat ($t(196) = 5.91$, $p < 0.001$, Hedge's $g = 0.84$, 95% CI: 0.55, 1.13). However, there was no main effect of angle ($F(3, 196) = 1.65$, partial eta = 0.02, $p = 0.18$).

Overall, there was a significant interaction effect between machine and angle ($F(9, 588) = 2.04$, partial eta = 0.03, $p = 0.03$). There were significant interaction effects for AFAR and OpenFace ($F > 2.83$, partial eta > 0.04, $p < 0.04$) but not for FaceReader or Py-feat ($F < 1.70$, partial eta < 0.03, $p > 0.17$). However, for both AFAR and OpenFace, there was no significant main effect of angle ($t < 2.62$, $p < 0.07$).

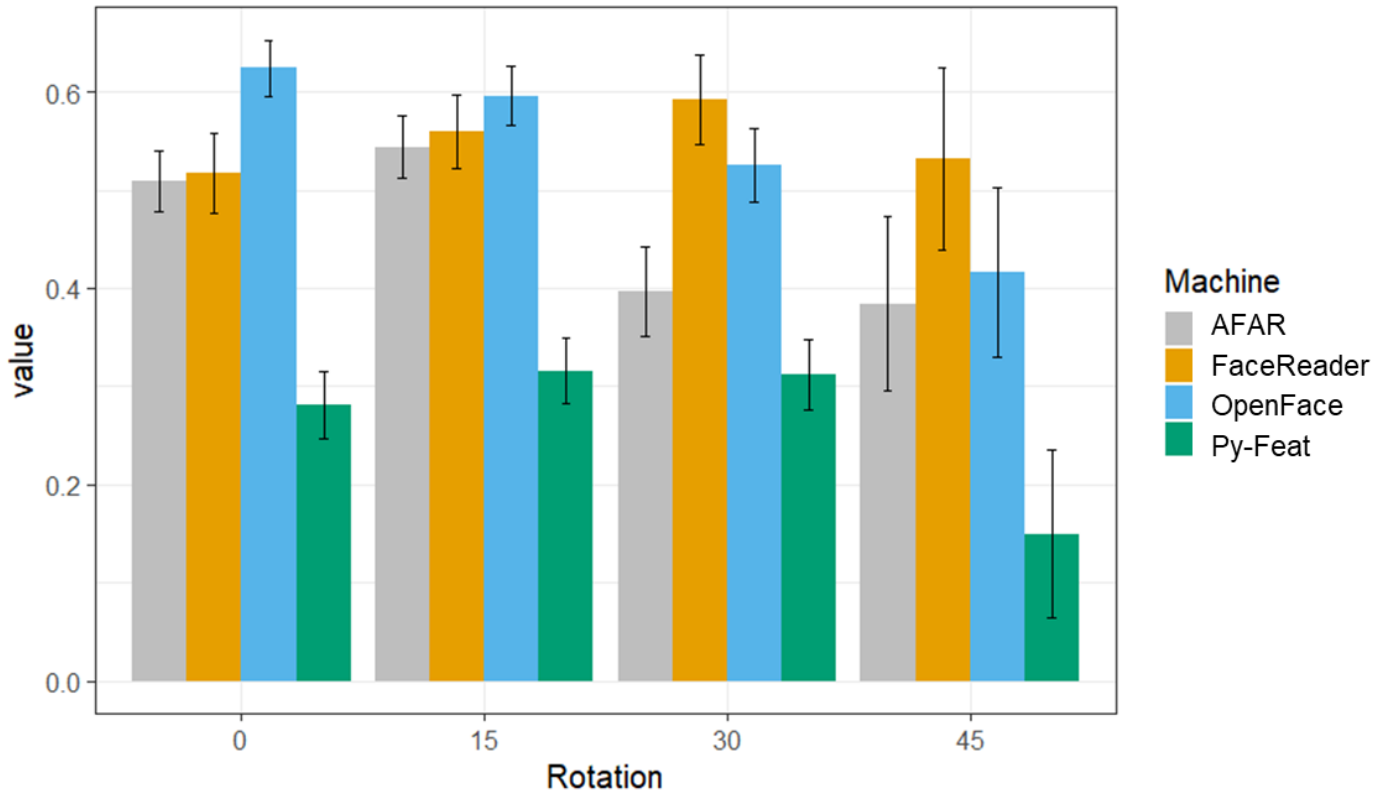

Figure S1. Average biserial correlation values between manual coded AU and predicted AUs for all angles and machines. Error bars represent standard errors.

References

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

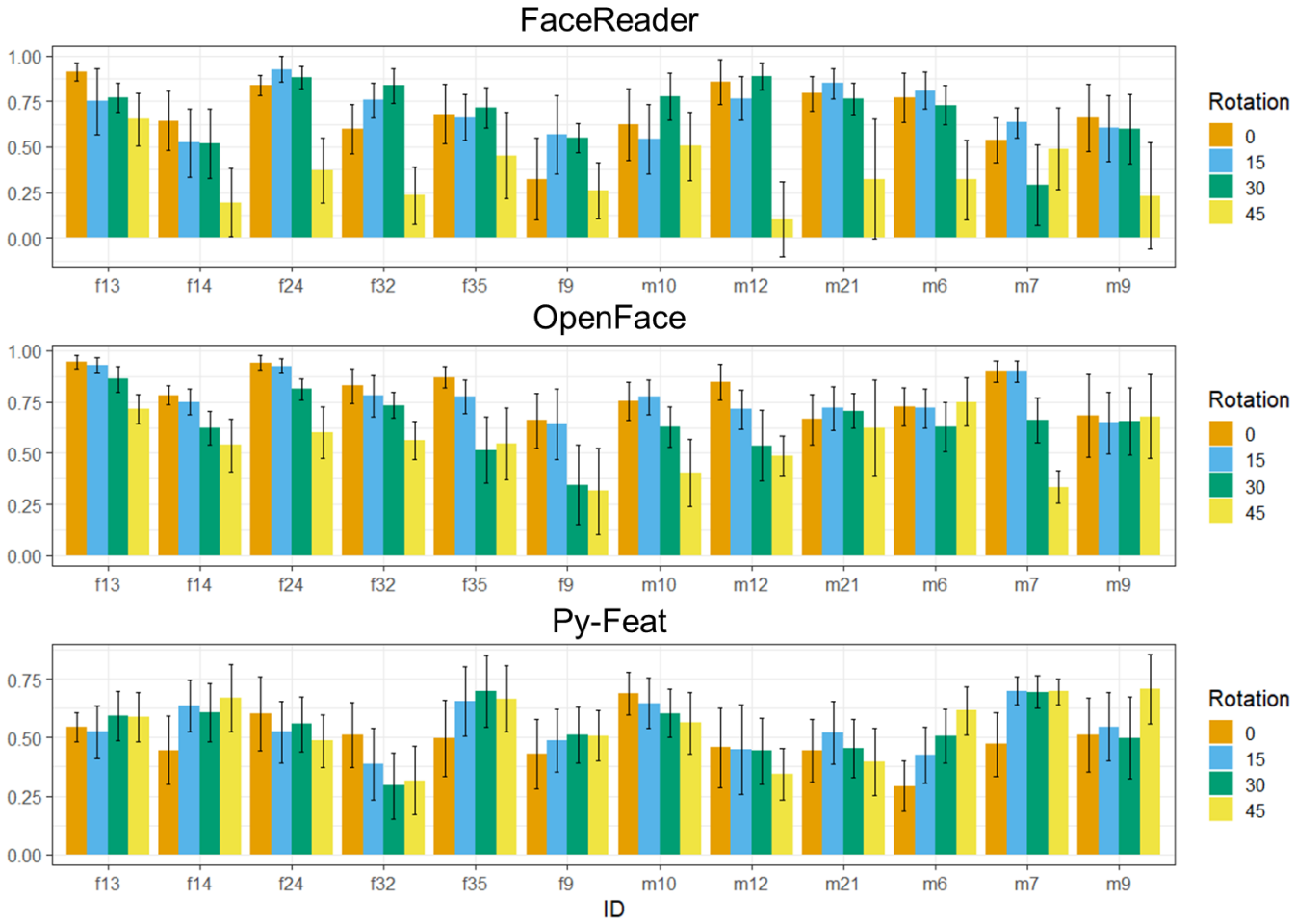

Figure S2. Average biserial correlation values between manually coded AU and predicted AUs for all angles, machines, and participants. f, female; m, male. Error bars represent standard errors.