# Supplementary Material for
# Synthetic Expressions are Better Than Real for Learning to Detect Facial Actions

Koichiro Niinuma[1], Itir Onal Ertugrul[2], Jeffrey F Cohn[3], László A Jeni[4]
[1]Fujitsu Laboratories of America, Pittsburgh, PA, USA, [2]Tilburg University, Tilburg, The Netherlands
[3]University of Pittsburgh, Pittsburgh, PA, USA, [4]Carnegie Mellon University, Pittsburgh, PA, USA
kniinuma@fujitsu.com, i.onal@uvt.nl, jeffcohn@pitt.edu, laszlojeni@cmu.edu

## Abstract

*This document accompanies the paper: "Synthetic Expressions are Better Than Real for Learning to Detect Facial Actions". We provide detailed results of our experiments.*

## S1 Experiments

### S1.1 Experiments for models trained on combined datasets

In the paper, we report results for models trained on single datasets. In this supplementary material, we report results for models trained on combined datasets. Table S1 and S2 show the results for FERA Test partition and DISFA, respectively. The first, second and forth columns show the results for models trained on single datasets, and the third, fifth and sixth columns show results for models trained on combined datasets. As shown in the tables, combined datasets show slightly better results.

### S1.2 Comparison without and with idiosyncratic loss

Table S3 shows the results without and with idiosyncratic loss ($\lambda_{ids} = 1.0$). It does not improve the ICC, but we confirmed that it improves the image quality as mentioned in Sec. 4.4.

### S1.3 Synthetic MultiPIE images

Fig. S1 shows some examples of MultiPIE synthesized expressions without and with idiosyncratic loss.

Table S1. ICC comparison of models trained on single vs combined datasets for FERA 2017 Test partition. Training size is the number of images per intensity per AU used to train models.

| | FERA real | FERA synthetic | FERA real + FERA synthetic | MultiPIE synthetic | FERA real + MultiPIE synthetic | FERA synthetic + MultiPIE synthetic |
|---|---|---|---|---|---|---|
| Training size | 5,000 | 5,000 | 10,000 | 4,605 | 9,605 | 9,605 |
| AU01 | 0.343 | 0.381 | 0.446 | 0.311 | 0.315 | **0.457** |
| AU04 | 0.260 | 0.219 | 0.324 | 0.202 | **0.342** | 0.288 |
| AU06 | 0.751 | **0.804** | 0.794 | 0.786 | 0.784 | 0.801 |
| AU10 | **0.785** | 0.773 | 0.772 | 0.726 | 0.760 | 0.763 |
| AU12 | 0.806 | 0.795 | 0.800 | 0.792 | **0.807** | 0.801 |
| AU14 | 0.084 | **0.244** | 0.171 | 0.168 | 0.116 | 0.235 |
| AU17 | 0.391 | **0.461** | 0.433 | 0.365 | 0.408 | 0.436 |
| Mean | 0.489 | 0.525 | 0.534 | 0.479 | 0.505 | **0.540** |

Table S2. ICC comparison of models trained on single vs combined datasets for DISFA. Training size is the number of images per intensity per AU used to train models.

| | FERA real | FERA synthetic | FERA real + FERA synthetic | MultiPIE synthetic | FERA real + MultiPIE synthetic | FERA synthetic + MultiPIE synthetic |
|---|---|---|---|---|---|---|
| Training size | 5,000 | 5,000 | 10,000 | 4,605 | 9,605 | 9,605 |
| AU01 | 0.394 | 0.314 | 0.365 | 0.418 | **0.470** | 0.346 |
| AU04 | **0.634** | 0.400 | 0.571 | 0.541 | 0.544 | 0.402 |
| AU06 | 0.404 | 0.573 | 0.507 | 0.524 | 0.496 | **0.576** |
| AU12 | 0.750 | 0.748 | 0.723 | 0.698 | 0.749 | **0.762** |
| AU17 | 0.293 | 0.373 | 0.296 | 0.290 | 0.377 | **0.390** |
| Mean | 0.495 | 0.482 | 0.493 | 0.494 | **0.527** | 0.495 |

Table S3. ICC for intensity estimation without and with idiosyncratic loss on FERA 2017 Test partition.

| | Without idiosyncratic loss | With idiosyncratic loss |
|---|---|---|
| AU1 | **0.381** | 0.350 |
| AU4 | 0.219 | **0.302** |
| AU6 | **0.804** | 0.802 |
| AU10 | 0.773 | **0.780** |
| AU12 | **0.795** | 0.793 |
| AU14 | **0.244** | 0.181 |
| AU17 | **0.461** | 0.452 |
| Mean | **0.525** | 0.523 |

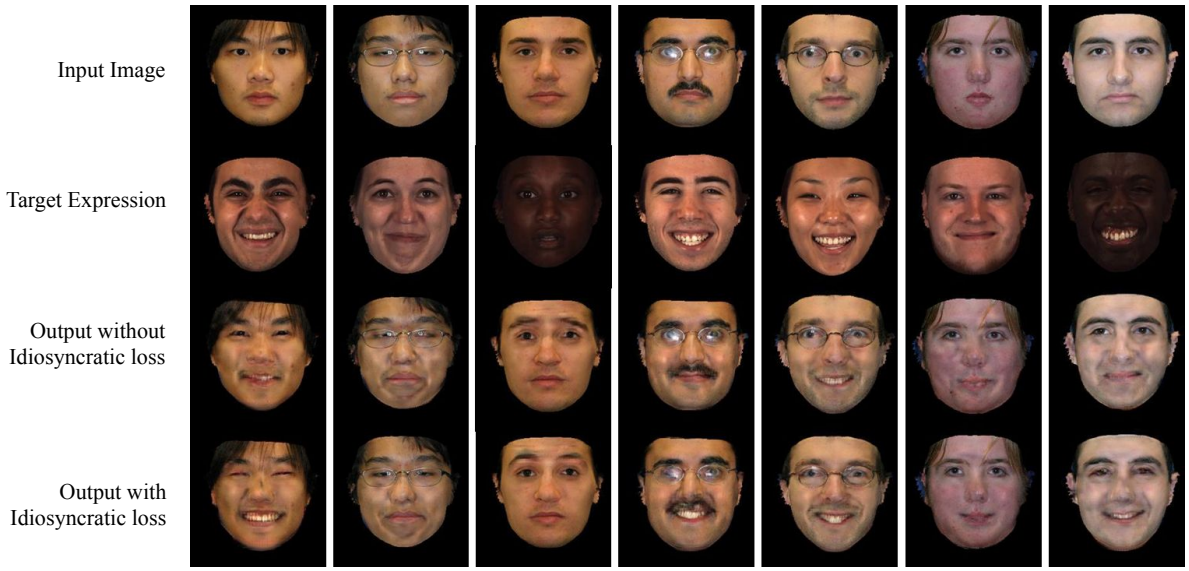


Figure S1. Comparison without and with idiosyncratic loss on MultiPIE.

Table S4. ICC comparison for GAN architectures.

| | GANimation | StarGAN | GANimation internal classifier |
|---|---|---|---|
| AU1 | **0.381** | 0.367 | 0.380 |
| AU4 | 0.219 | **0.259** | 0.065 |
| AU6 | **0.804** | 0.793 | 0.712 |
| AU10 | 0.773 | **0.788** | 0.743 |
| AU12 | 0.795 | **0.807** | 0.793 |
| AU14 | **0.244** | 0.199 | 0.123 |
| AU17 | **0.461** | 0.451 | 0.364 |
| Mean | **0.525** | 0.523 | 0.454 |

Table S5. Cross domain ICC performance for UNBC Pain. (Synthetic training set → Real test set)

| | FERA→UNBC Pain | MultiPIE→UNBC Pain |
|---|---|---|
| AU04 | 0.130 | 0.149 |
| AU06 | 0.496 | 0.434 |
| AU10 | 0.034 | 0.038 |
| AU12 | 0.402 | 0.367 |
| Mean | 0.266 | 0.247 |

## S1.4 Comparison of GAN architectures

To examine the influence of GAN architectures, we coducted experiments for GANimation, StarGAN and GANimation internal classifier $D_{exp}$. Table S4 indicates that GANimation and StarGAN show almost the same performance, but GANimation internal classifier shows worse than the others.

## S1.5 Cross-domain experiments on UNBC Pain

Table S5 shows the cross domain results for UNBC Pain. Classifiers were trained either on synthetic expressions generated from FERA 2017 or synthetic expressions generated from MultiPIE, and they were tested on UNBC Pain.