Pacific Graphics 2023
R. Chaine, Z. Deng, and M. H. Kim
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 42 (2023), Number 7

A Perceptual Shape Loss for Monocular 3D Face Reconstruction - Supplementary Material

C. Otto[1,2], P. Chandran[1], G. Zoss[1], M. Gross[1,2], P. Gotardo*[1], D. Bradley[1]

[1]DisneyResearch|Studios, Switzerland
[2]ETH Zürich, Switzerland

1. Architecture Details

Figure 1 shows more details of our convolutional network $\mathcal{D}$ [GAA*17, ACB17], which takes the aligned RGB image and gray-shaded render with noise background as a four channel input. Output is a real-valued 1-dimensional scalar $\mathcal{S}$, representing the perceptual shape loss score.

2. Implementation Details

The results for the EMOCA_V2 [DBB22] version that we compare to in the paper are based on their *EMOCA_v2_lr_mse_20* model checkpoint, which uses the SPECTRE [FRPP*22] lip reading loss.

As SPECTRE [FRPP*22] itself is a video-based method, we create small 10 frame videos consisting out of the same single-image frame for running their reconstruction. We process the video with chunk size 10 and extract the last valid reconstruction from the first chunk.

3. Additional Results

We present additional optimization results using our perceptual shape loss in Figure 2. The results are generated following the same optimization schedule as mentioned in the main paper. For all parameter updates we use the AdamW optimizer [LH19] with a learning rate of 0.005.
Additional inference-based qualitative comparisons with related state-of-the-art work are presented in Figure 3.

4. Data of Human Subjects

For all of the personal data that is shown in the main paper and in this supplementary material, we have obtained the consent of the respective individuals.

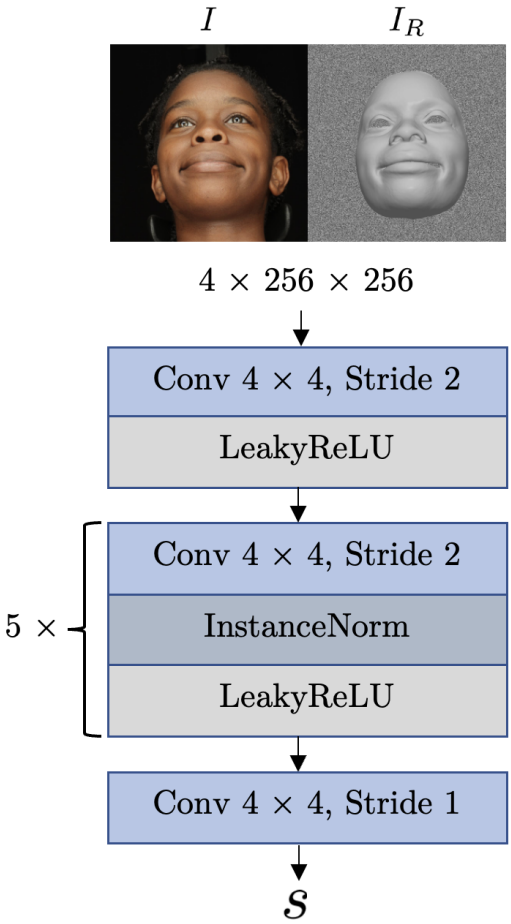

Figure 1: *Architecture details for our convolutional network $\mathcal{D}$ [GAA*17, ACB17]. The LeakyReLU activations use a negative slope of* 0.2.

References

[ACB17] ARJOVSKY M., CHINTALA S., BOTTOU L.: Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning* (06–11 Aug 2017), Precup D., Teh

*Now at Google

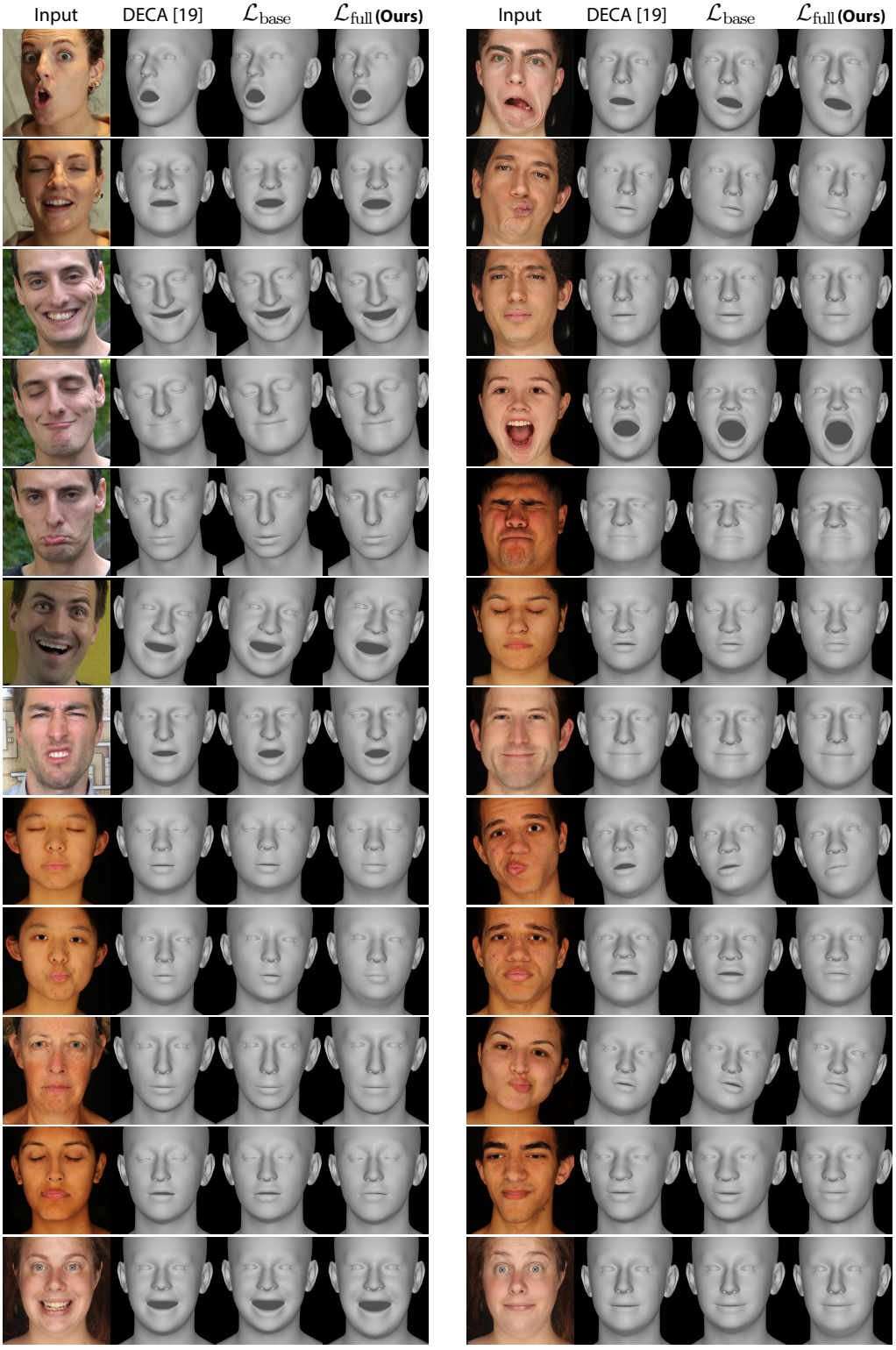

Figure 2: *We show additional qualitative optimization results for face reconstructions on in-the-wild data and on our held-out validation set. The optimization is initialized by the DECA [FFBB21] parameters. We show the results for the traditional losses in the column ($\mathcal{L}_{base}$) and the results including our proposed perceptual shape loss in column ($\mathcal{L}_{full}$).*

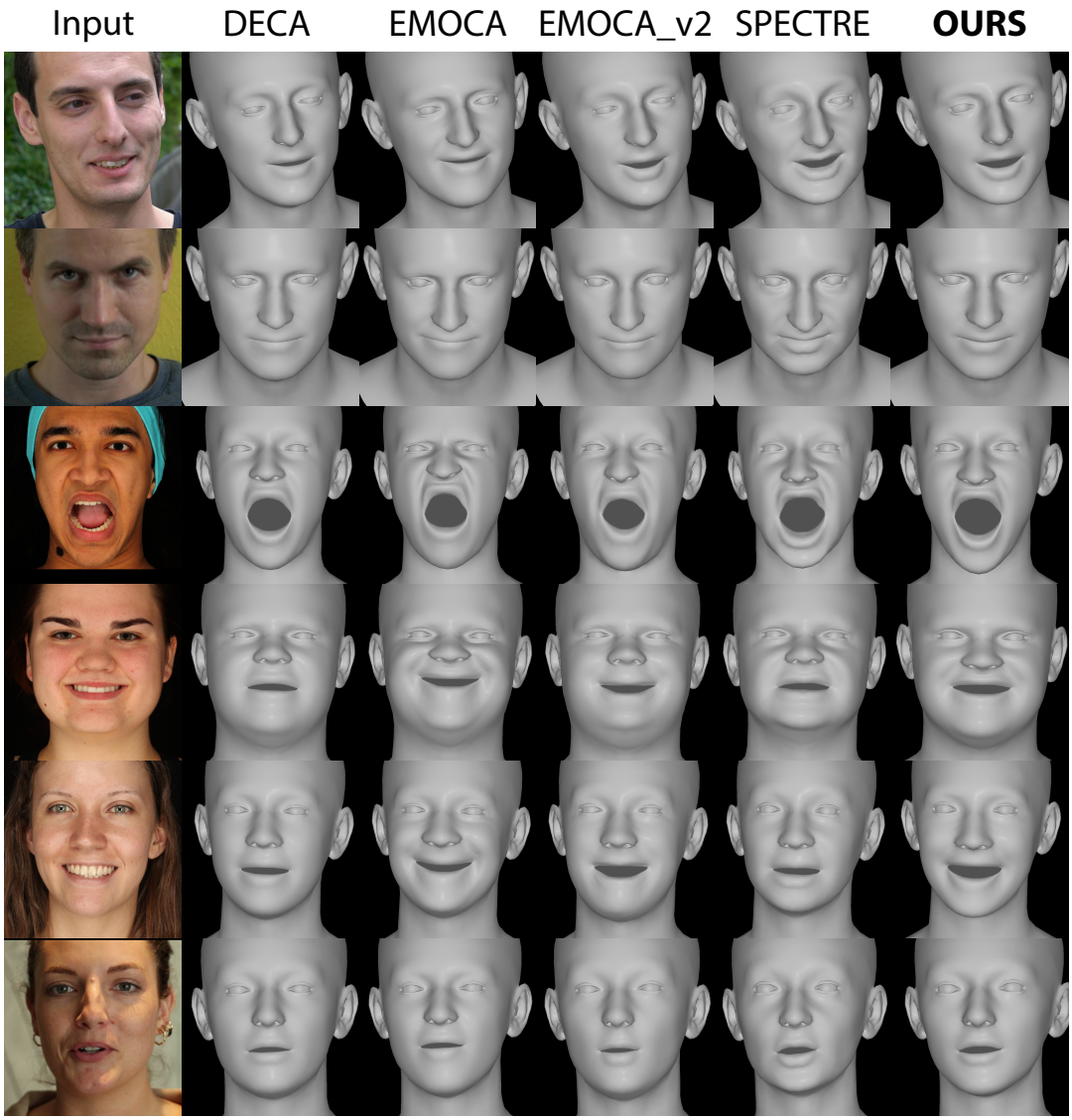

Figure 3: *Here we show more qualitative inference-based results for face reconstructions, comparing the initial DECA [FFBB21] fit with our new method PSL - $\mathcal{L}_{full}$ and related state-of-the-art methods EMOCA [DBB22], EMOCA_V2 [DBB22] and SPECTRE [FRPP*22].*

Y. W., (Eds.), vol. 70 of *Proceedings of Machine Learning Research*, PMLR, pp. 214–223. 1

[DBB22] DANECEK R., BLACK M. J., BOLKART T.: EMOCA: Emotion driven monocular face capture and animation. In *Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 20311–20322. 1, 3

[FFBB21] FENG Y., FENG H., BLACK M. J., BOLKART T.: Learning an animatable detailed 3D face model from in-the-wild images. vol. 40. 2, 3

[FRPP*22] FILNTISIS P. P., RETSINAS G., PARAPERAS-PAPANTONIOU F., KATSAMANIS A., ROUSSOS A., MARAGOS P.: Visual speech-aware perceptual 3d facial expression reconstruction from videos. 1, 3

[GAA*17] GULRAJANI I., AHMED F., ARJOVSKY M., DUMOULIN V., COURVILLE A. C.: Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems* (2017), Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R., (Eds.), vol. 30, Curran Associates, Inc. 1

[LH19] LOSHCHILOV I., HUTTER F.: Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019* (2019). 1