Supplementary material for the paper: Sparse to Dense Dynamic 3D Facial Expression Generation

Naima Otberdout
Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
naima.otberdout@univ-lille.fr

Claudio Ferrari
Deparment of Architecture and Engineering
University of Parma, Italy
claudio.ferrari2@unipr.it

Mohamed Daoudi
IMT Nord Europe, Institut Mines-Télécom, Univ. Lille, Centre for Digital Systems, F-59000 Lille, France
Univ. Lille, CNRS, Centrale Lille, Institut Mines-Télécom, UMR 9189 CRIStAL, F-59000 Lille, France
mohamed.daoudi@imt-nord-europe.fr

Stefano Berretti
Media Integration ad Communication Center
University of Florence, Italy
stefano.berretti@unifi.it

Alberto Del Bimbo
Media Integration ad Communication Center
University of Florence, Italy
alberto.delbimbo@unifi.it

1. Landmarks Configuration

In Figure 1 we show, for three different expressions, the configuration of landmarks used to guide the generation of the facial expression.

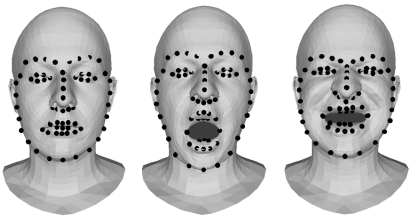

Figure 1. Landmarks configuration used to guide our model.

2. Logarithm and Exponential Maps

In order to map the SRVF data forth and back to a tangent space of $\mathcal{C}$, we use the logarithm $\log_p(.)$ and the exponential $\exp_p(.)$ maps defined in a given point p by,

$$\begin{aligned} \log_p(q) &= \frac{d_{\mathcal{C}}(q,p)}{sin(d_{\mathcal{C}}(q,p))}(q - cos(d_{\mathcal{C}}(q,p))p), \\ \exp_p(s) &= cos(\|s\|)p + sin(\|s\|)\frac{s}{\|s\|}, \end{aligned} \tag{1}$$

where $d_{\mathcal{C}}(q,p) = \cos^{-1}(\langle q,p\rangle)$ is the distance between q and p in $\mathcal{C}$.

3. Architecture of S2D-Dec

The architecture adopted for S2D-Dec is based on the architecture proposed in [1]. S2D-Dec takes as input the displacements of 68 landmarks illustrated in Figure 1. The architecture includes a fully connected layer of size 2688, five spiral convolution layers of 64, 32, 32, 16 and 3 filters. Each spiral convolution layer is followed by an up-sampling by a factor of 4.

4. Ablation Study

In this section, we report a visual comparison between reconstructions obtained with the standard L1 loss and our proposed weighted L1. Figure 2 clearly shows the effect of

our introduced weighting scheme that allows for improved expression modeling.

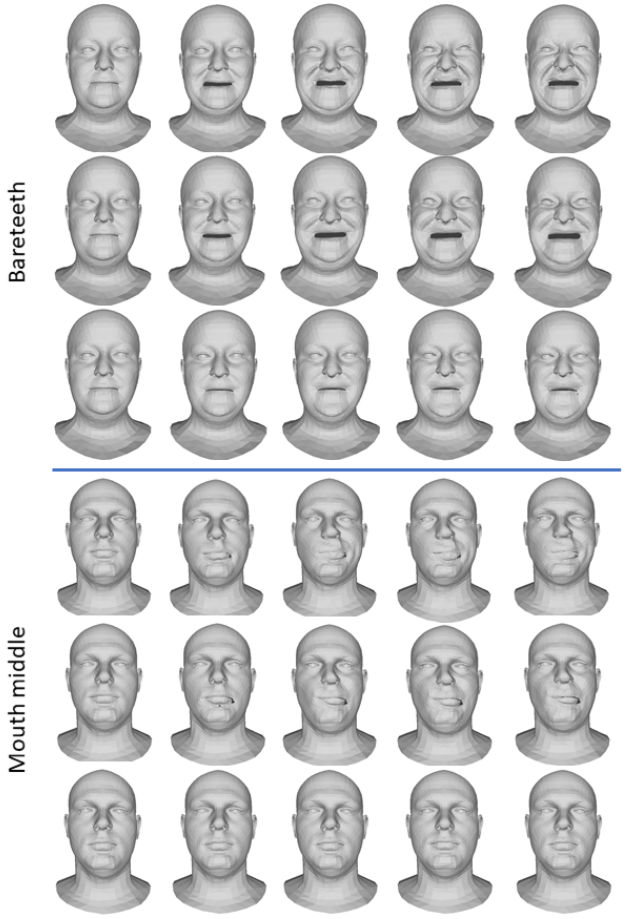

Figure 2. Ablation study: qualitative comparison between ground truth (first row) our model with (second row) and without (last row) weighted loss.

References

[1] Giorgos Bouritsas, Sergiy Bokhnyak, Stylianos Ploumpis, Stefanos Zafeiriou, and Michael Bronstein. Neural 3D morphable models: Spiral convolutional networks for 3D shape representation learning and generation. In *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, pages 7212–7221, 2019. 1