Supplementary Material for “Unpaired Multi-domain Attribute Translation of 3D Facial Shapes with a Square and Symmetric Geometric Map”

1. Details for Geometric Mapping

We clip the frontal part (see Figure 1) of the original 3D template on the FaceScape dataset [1], considering that the variations of facial attributes (expression, age, and gender) are manifest only in the frontal parts in current 3D scans. The resulting resolution of the clipped face is $10,857$ vertices. The size of the geometric map (**Gmap**) is designed to be $128 \times 128 \times 3$, being a trade-off of computational efficiency and representation accuracy. The resolution for the Gmap ($128 \times 128 = 16384$) is also on par with the resolution of the clipped template, thereby being sufficient to represent the details of 3D facial shapes.

In addition, the details of the Gmap are shown in Figure 2. It preserves the adjacency relationship of all vertices on the original 3D mesh in a **local least-square** sense while being **square and symmetric**. It also makes the mapping from 3D mesh to 2D geometric map to be one-to-one for each vertex, avoiding triangle flipping which is correlated to interpolation errors. We can see there is seldom triangle flipping even in the most difficult regions, *i.e.* inner mouth and eye surroundings.

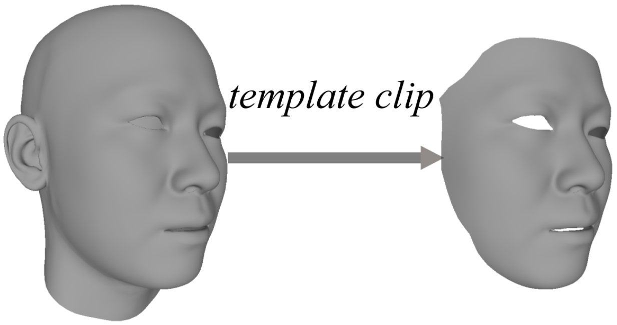

Figure 1. Illustration for the clipped template.

Sampling the locations between a 3D shape and its representation on the 2D Gmap (on an image grid) involves bi-directional mappings between the 3D shape $\mathcal{V}$ and its vertex locations on the 2D image grid. The mappings are as follows (also refer to Figure 3).

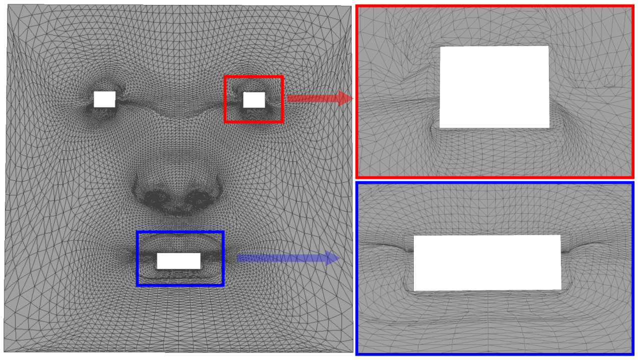

Figure 2. The visualized details of the geometric map. Please zoom in to view the structures.

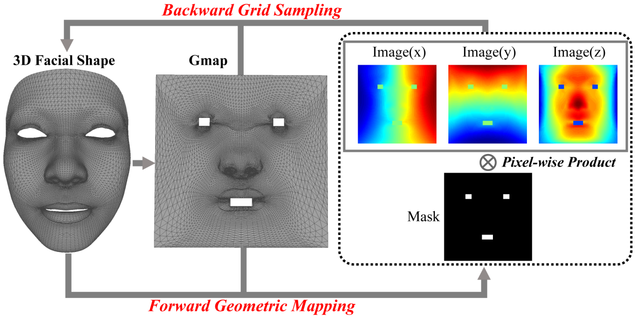

Figure 3. Sampling the locations between a 3D facial shape and its representation on the Gmap.

- **Forward mapping** is computed by *barycentric interpolation*. Suppose a pixel $P = \{x, y\}$ on the image grid lies inside a triangle $\Delta P_1P_2P_3$ indexed by $i_1i_2i_3$ on the geometric map, then the shape coding $I_{x,y}$ is

  $$I_{x,y} = w_1 v_{i_1} + w_2 v_{i_2} + w_3 v_{i_3} (v_{i_1}, v_{i_2}, v_{i_3} \in \mathcal{V}), \quad (1)$$

  where

  $$\begin{aligned} w_1 &= (\overrightarrow{PP_2} \times \overrightarrow{PP_3})/(\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}), \\ w_2 &= (\overrightarrow{PP_3} \times \overrightarrow{PP_1})/(\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}), \\ w_3 &= (\overrightarrow{PP_1} \times \overrightarrow{PP_2})/(\overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3}), \end{aligned} \quad (2)$$

  and × is the outer-product between two 2D vectors.

Index	Type	Kernel	Stride	Output	Others	Appended Loss
1	Input shape	-	-	3×10857	-	-
2	Geometric mapping	-	-	$3 \times 128 \times 128$	Mask Area set to 0	-
3	Label cat.	-	-	$26 \times 128 \times 128$	-	-
4	Conv.	7×7	1×1	$64 \times 128 \times 128$	IN+ReLU	-
	Conv.	4×4	2×2	$128 \times 64 \times 64$	IN+ReLU	-
	Conv.	4×4	2×2	$256 \times 32 \times 32$	IN+ReLU	-
5	Residual blocks $\times 6$	-	-	$256 \times 32 \times 32$	-	-
6	DeConv.	4×4	2×2	$128 \times 64 \times 64$	IN+ReLU	-
	DeConv.	4×4	2×2	$64 \times 128 \times 128$	IN+ReLU	-
7	Conv.	7×7	1×1	$3 \times 128 \times 128$	Mask Area set to 0	Symmetric loss
8	Bilinear grid sampling	-	-	3×10857	Mask	Cycle&Reconstruction loss

Table 1. The architecture of the generator network. The padding size of each layer is determined to be compatible with the input and output feature sizes. "IN" denotes for instance normalization operation. "ReLU" denotes for rectified linear unit activation.

Index	Type	Kernel	Stride	Output	Others	Appended Loss
1	Input shape	-	-	3×10857	-	-
2	Geometric mapping	-	-	$3 \times 128 \times 128$	Mask Area set to 0	-
3	Conv.	4×4	2×2	$64 \times 64 \times 64$	LReLU	-
	Conv.	4×4	2×2	$128 \times 32 \times 32$	LReLU	Conv. + Cat. + Adv. loss
	Conv.	4×4	2×2	$256 \times 16 \times 16$	LReLU	-
	Conv.	4×4	2×2	$512 \times 8 \times 8$	LReLU	Conv. + Cat. + Adv. loss
	Conv.	4×4	2×2	$1024 \times 4 \times 4$	LReLU	-
	Conv.	4×4	2×2	$2048 \times 2 \times 2$	LReLU	Conv. + Cat. + Adv. loss
4	Conv.&Output for class	2×2	1×1	$23 \times 1 \times 1$	-	Classification loss

Table 2. The architecture of the discriminator network. The padding size of each layer is determined to be compatible with the input and output feature sizes. "LReLU" denotes leaky ReLU activation. The Conv. operations listed in the appended loss denote convolutions to feature size 1 by 3×3 kernels and 1×1 strides. The output features for the 3 pyramid layers are then flattened and concatenated to fed into the adversarial loss.

- **Backward mapping** is computed by *bilinear grid sampling* from the image grid to the geometric map.

2. Detailed Network Architecture

The detailed architectures for the generator and the discriminator of the proposed adversarial learning framework are elaborated in Table 1 and Table 2 for ease of reproductively, respectively. **We will also release our code.**

3. Additional Qualitative Results

We have mentioned in the main manuscript that our method is also capable of translating gender and age. Figure 4 shows additional results for some authorized samples in the test set of FaceScape [1]. In addition, the propose method supports both continous (*e.g.* the expressions and genders) and discrete attribute labels (*e.g.* the ages). We also suggest that the fractional labels for expressions can be acquired by linear interpolations on the output directly. Figure 5 shows continous variations from neural to certain expressions. Therefore, our proposed method is capable of generating realistic shapes with different attributes given an input 3D facial shape.

References

[1] Haotian Yang, Hao Zhu, Yanru Wang, Mingkai Huang, Qiu Shen, Ruigang Yang, and Xun Cao. Facescape: a large-scale high quality 3d face dataset and detailed riggable 3d face prediction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 601–610, 2020. 1, 2

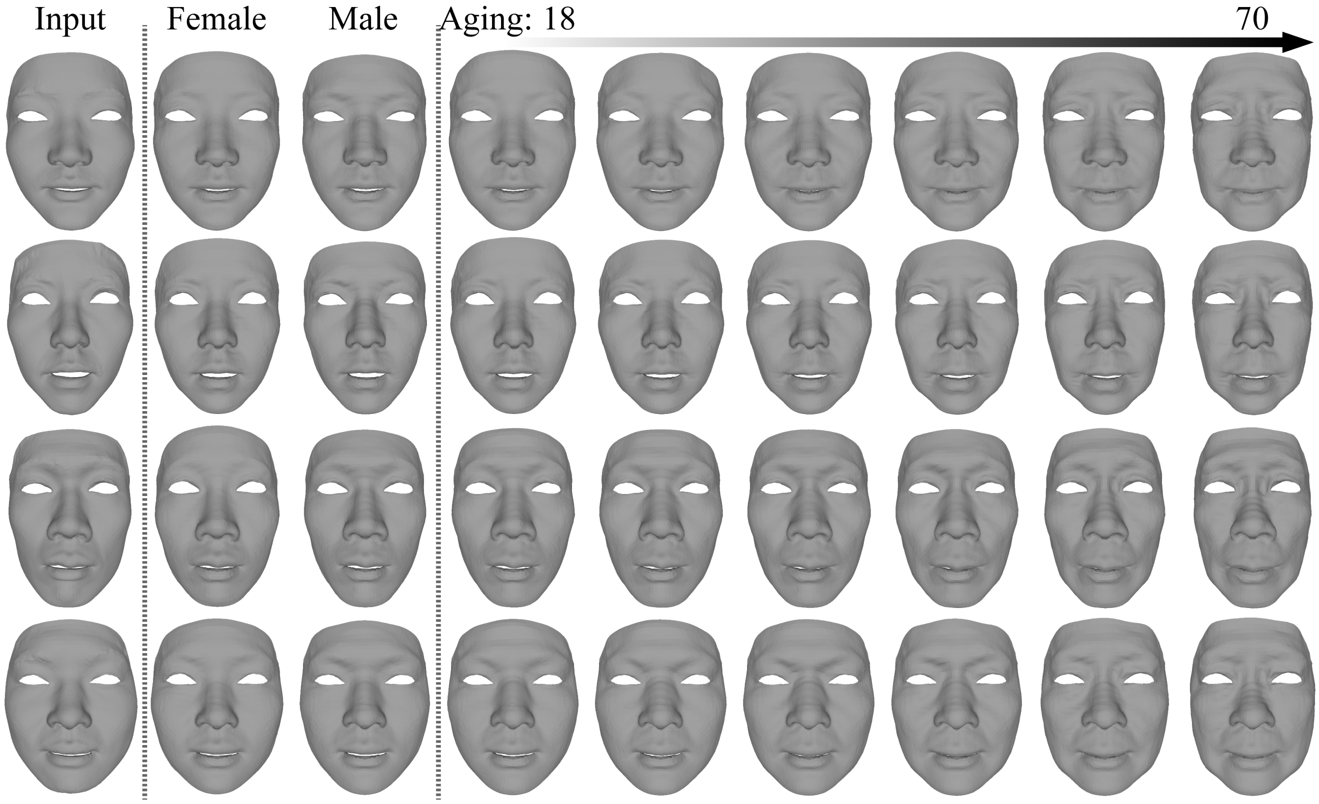

Figure 4. Some additional results for gender and age translations.

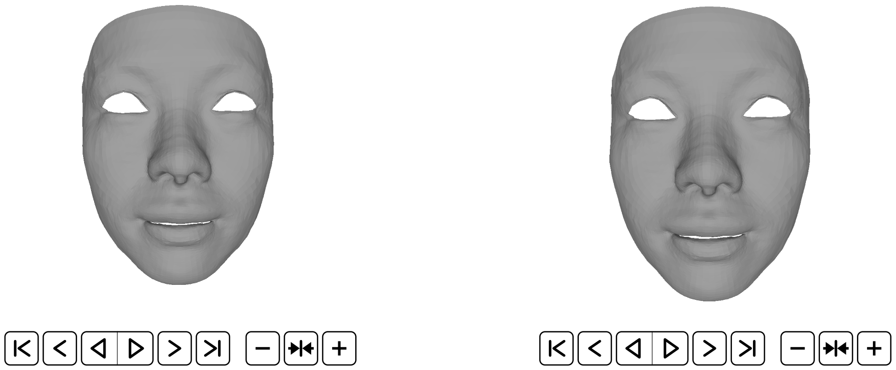

Figure 5. Two examples for continous variations of the expressions.