Text and Image Guided 3D Avatar Generation and Manipulation

Zehranaz Canfes* M. Furkan Atasoy* Alara Dirik* Pinar Yanardag
Boğaziçi University
Istanbul, Turkey
{zehranaz.canfes.2022, muhammed.atasoy.2022}@alumni.boun.edu.tr
{alaradirik, yanardag.pinar}@gmail.com

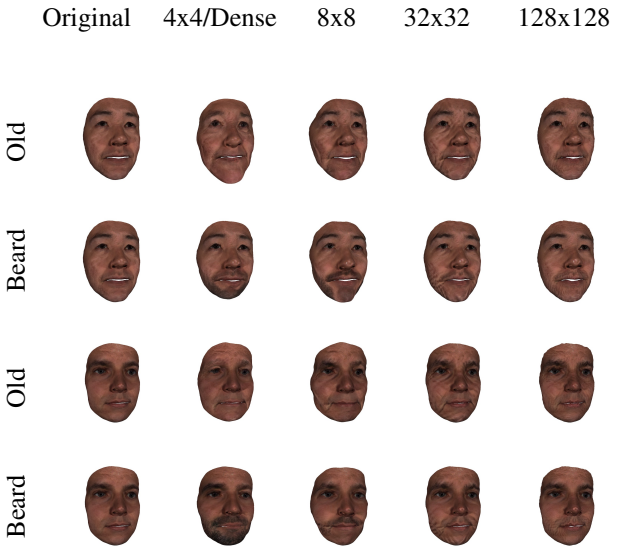

Figure 1. The comparison of manipulations on different layers for two different 3D faces. First two column show the 'beard' and 'old' manipulations on one 3D face and the second column show the results for the same manipulations on another 3D face.

A. Ablation Study

In this section, we perform ablation studies on the effects of identity loss and layer selection for latent space manipulation.

A.1. Effect of Layer Selection

We perform our manipulations on the 44/Dense layer of TBGAN, the layer that provides the best results in terms of identity preservation and meaningful manipulations. The comparison of our method on different layers can be found in Figure 1. We show that the manipulations on other layers give defected results with undesirable artifacts, so that the results deviate from the desired text prompt.

*Denotes equal contribution.

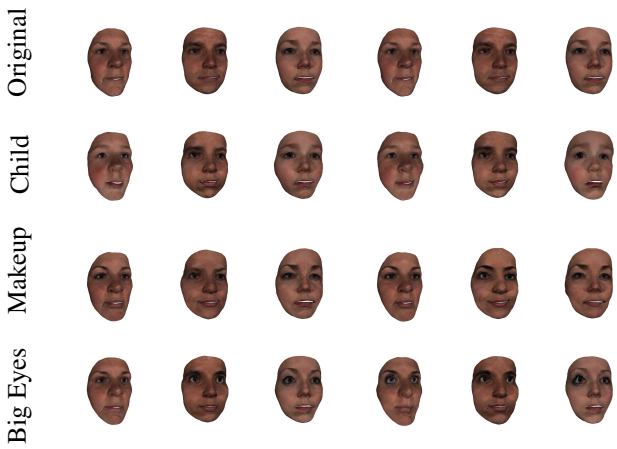

Figure 2. Results with and without ID loss.

A.2. Effect of Identity Loss

Our method uses ArcFace, a large-scale pre-trained face recognition network, to compute identity loss $\mathcal{L}_{\text{ID}}$ and enforce identity preservation during manipulation. We perform an ablation study with different target texts describing emotion-, shape-, and texture-related changes to demonstrate the effect of $\mathcal{L}_{\text{ID}}$ on the manipulation results, and present the results in Figure 2. For the identity loss experiments, we simply set $\mathcal{L}_{\text{ID}} = 0$ and leave the other hyperparameters the same. As can be seen in Figure 2, identity loss is crucial for preserving the identity of the input, and omitting it leads to manipulation results that are significantly different from the input.

B. Sentence Templates for Prompt Engineering

Our method uses 74 sentence templates. The list of templates we use for augmentation can be found in Table 1.

‘a bad photo of a’	‘a sculpture of a’
‘a photo of the hard to see’	‘a low resolution photo of the’
‘a rendering of a’	‘graffiti of a’
‘a bad photo of the’	‘a cropped photo of the’
‘a photo of a hard to see’	‘a bright photo of a’
‘a photo of a clean’	‘a photo of a dirty’
‘a dark photo of the’	‘a drawing of a’
‘a photo of my’	‘the plastic’
‘a photo of the cool’	‘a close-up photo of a’
‘a painting of the’	‘a painting of a’
‘a pixelated photo of the’	‘a sculpture of the’
‘a bright photo of the’	‘a cropped photo of a’
‘a plastic’	‘a photo of the dirty’
‘a blurry photo of the’	‘a photo of the’
‘a good photo of the’	‘a rendering of the’
‘a in a video game.’	‘a photo of one’
‘a doodle of a’	‘a close-up photo of the’
‘a photo of a’	‘the in a video game.’
‘a sketch of a’	‘a face of the’
‘a doodle of the’	‘a low resolution photo of a’
‘the toy’	‘a rendition of the’
‘a photo of the clean’	‘a photo of a large’
‘a rendition of a’	‘a photo of a nice’
‘a photo of a weird’	‘a blurry photo of a’
‘a cartoon’	‘art of a’
‘a sketch of the’	‘a pixelated photo of a’
‘itap of the’	‘a good photo of a’
‘a plushie’	‘a photo of the nice’
‘a photo of the small’	‘a photo of the weird’
’the cartoon’	‘art of the’
‘a drawing of the’	‘a photo of the large’
‘the plushie’	‘a dark photo of a’
‘itap of a’	‘graffiti of the’
‘a toy’	‘itap of my’
‘a photo of a cool’	‘a photo of a small’
‘a 3d object of the’	‘a 3d object of a’
‘a 3d face of a’	‘a 3d face of the’

Table 1. List of templates that our method uses for augmentation. The input text prompt is added to the end of each sentence template.