Inverting Generative Adversarial Renderer for Face Reconstruction  
- Supplementary Material -

Jingtan Piao¹, Keqiang Sun¹, Quan Wang²,³, Kwan-Yee Lin¹,²*, Hongsheng Li¹,⁴*
¹CUHK-SenseTime Joint Laboratory, The Chinese University of Hong Kong  
²SenseTime Research and Tetras.AI  
³Shanghai AI Laboratory  
⁴School of CST, Xidian University
{1155116308, kqsun}@link.cuhk.edu.hk, {wangquan, linjunyi}@sensetime.com, hsli@ee.cuhk.edu.hk

The content of our supplementary material is organized as follows.

1. More results of face geometry reconstruction on other datasets.

2. Additional ablation studies and detailed analysis of components in our framework.

3. More qualitative evaluation of face image editing.

1. More Results of Face Reconstruction.

Our newly proposed face geometry reconstruction method achieves state-of-the-art results, not only on MICC as shown in the main paper, but on several other benchmarks, i.e. AFLW-3D[7], BJUT3D[1], as well. Due to the limit of the space, we present the results in this supplementary material. Note that no ground-truth mesh is used in our method. However, as shown in Table 1, the mean error of our method is 2.92 on AFLW-3D, which outperforms the second-best semi-supervised method by 29.13%, and is competitive with state-of-the-art fully supervised methods.

Since the images in BJUT3D are all generated by rendering from 3D meshes, the error value does not reflect the modeling effect of algorithms. We show qualitative results in Figure 1. It is worth noting that the reconstructed geometry is modeling the identity and expression of the image so vividly that we could recognize the identity of a person and his micro-expressions even by only viewing the corresponding face mesh.

Experiments on all these datasets indicate that, equipped with the novel Generative Adversarial Renderer and the novel Render Inverting method for initialization, our method has a state-of-the-art performance on face geometry reconstruction.

<table>
<thead>
<tr>
<th>Method</th>
<th>Supervision</th>
<th>NME</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DDFA [6]</td>
<td>Full</td>
<td>5.42</td>
</tr>
<tr>
<td>3DSTN [2]</td>
<td>Full</td>
<td>4.49</td>
</tr>
<tr>
<td>Uncertain [4]</td>
<td>Full</td>
<td>2.81</td>
</tr>
<tr>
<td>Ours</td>
<td>Semi</td>
<td>2.92</td>
</tr>
</tbody>
</table>

Table 1. Reconstruction errors of NME landmarks positions on AFLW2000-3D.


Choices of the Conditions. Previous works tend to employ 3DMM parameters or depth as the conditioning inputs. However, the 3DMM parameters are not an ideal option to condition the generation process of a generative adversarial network, due to its abstractness. And the depth might have ambiguity, since the same geometry might correspond to totally diverse depth, considering the translation along the depth axis. We found that the face normal map is a more effective form of condition for our GAR. As shown in Figure 2 (a), with the face normal maps as inputs, the loss is minimized more stably and faster, and the network could converge to a better optimum.

Effect of the Normal Injection Module (NIM). As shown in Figure 2 (b), compared with the simple concatenation of the normal map into the feature maps, the proposed NIM is effective in further minimizing the loss value, which demonstrates the effectiveness of the proposed NIM. Moreover, equipped with the NIM, the loss curve falls steeper, and the training converges faster.

3. Qualitative Evaluation on Face Image Editing

We mentioned in the main paper that our method is capable of editing faces. Actually, attributed to the novel Generative Adversarial Render, as well as the accurate face geometry reconstruction method, we can edit face images vividly.
Figure 1. Example face reconstruction results from AFLW dataset and BJUT dataset.

Figure 2. (a) Loss of generator with different conditions. (b) Loss of generator with different injection method.
In Figure 3, we show more editing results.

In the first two rows, the pose is set to turn along Euler Angles pitch and yaw. (roll is an in-plane rotation and can be easily applied in image wrapping). The faces pose in the rendered images gradually change while the identity and the expression maintain unchanged. In the last two rows, we present the resulting images of editing facial expressions. Thanks to the accurate control of the normal maps, we can generate various face images with exaggerated expressions.

References


