Inverting Generative Adversarial Renderer for Face Reconstruction - Supplementary Material -

Jingtan Piao[1], Keqiang Sun[1], Quan Wang[2,3], Kwan-Yee Lin[1,2*], Hongsheng Li[1,4*]
[1]CUHK-SenseTime Joint Laboratory, The Chinese University of Hong Kong
[2]SenseTime Research and Tetras.AI [3]Shanghai AI Laboratory [4]School 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.

*K. Lin and H. Li are the co-corresponding authors.

Method	Supervision	NME
3DDFA [6]	Full	5.42
3DSTN [2]	Full	4.49
PRN [3]	Full	3.62
Uncertain [4]	Full	2.81
Nonlinear [5]	Semi	4.12
Ours	**Semi**	**2.92**

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

2. More Ablation Studies.

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 divers 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.

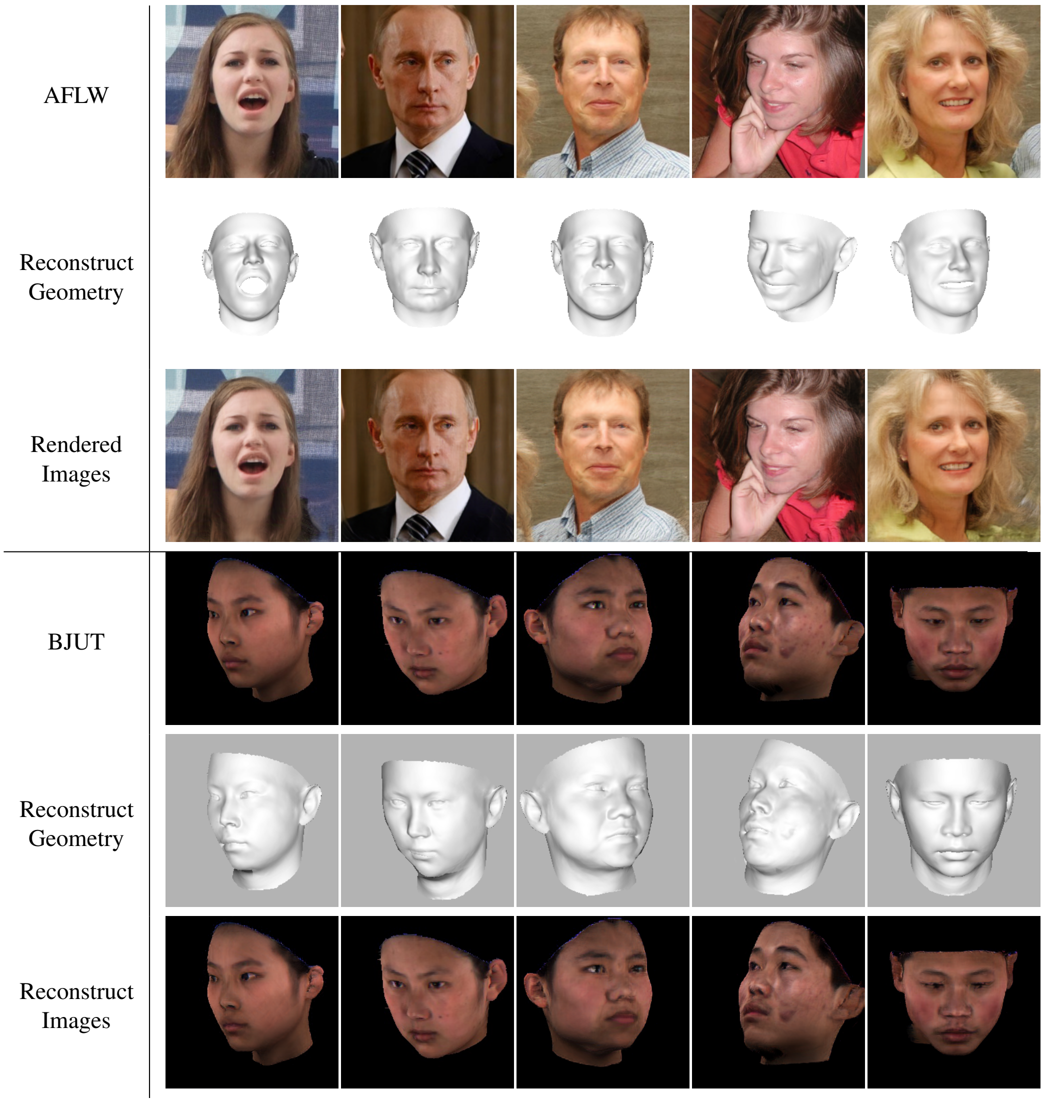
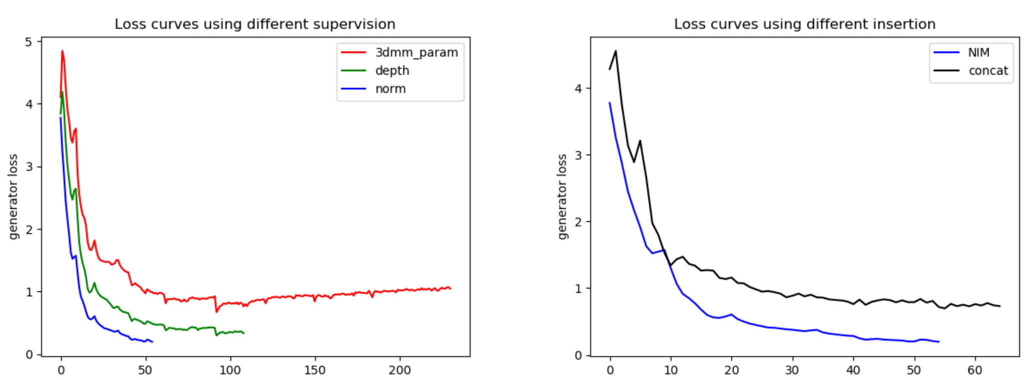


Figure 1. Example face reconstruction results from AFLW dataset and BJUT dataset.



(a) Loss of generator with different conditions.

(b) Loss of generator with different injection method.

Figure 2. name of the figure

Figure 3. Example face reconstruction results from AFLW dataset and BJUT dataset.

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

[1] Yin Baocai, Sun Yanfeng, Wang Chengzhang, and Ge Yun. Bjut-3d large scale 3d face database and information processing. *Journal of Computer Research and Development*, 46(6):1009, 2009. 1

[2] James Booth, Epameinondas Antonakos, Stylianos Ploumpis, George Trigeorgis, Yannis Panagakis, and Stefanos Zafeiriou. 3d face morphable models" in-the-wild". In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5464–5473. IEEE, 2017. 1

[3] Kyle Genova, Forrester Cole, Aaron Maschinot, Aaron Sarna, Daniel Vlasic, and William T Freeman. Unsupervised training for 3d morphable model regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8377–8386, 2018. 1

[4] Gun-Hee Lee and Seong-Whan Lee. Uncertainty-aware mesh decoder for high fidelity 3d face reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6100–6109, 2020. 1

[5] Luan Tran, Feng Liu, and Xiaoming Liu. Towards high-fidelity nonlinear 3d face morphable model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1126–1135, 2019. 1

[6] Anh Tuan Tran, Tal Hassner, Iacopo Masi, and Gérard Medioni. Regressing robust and discriminative 3d morphable models with a very deep neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5163–5172, 2017. 1

[7] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li. Face alignment across large poses: A 3d solution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 146–155, 2016. 1