Supplementary Material:
GANFIT: Generative Adversarial Network Fitting for High Fidelity 3D Face Reconstruction

Baris Gecer[1,2], Stylianos Ploumpis[1,2], Irene Kotsia[3], and Stefanos Zafeiriou[1,2]
[1]Imperial College London
[2]FaceSoft.io
[3]University of Middlesex
{b.gecer, s.ploumpis, s.zafeiriou}@imperial.ac.uk , drkotsia@gmail.com

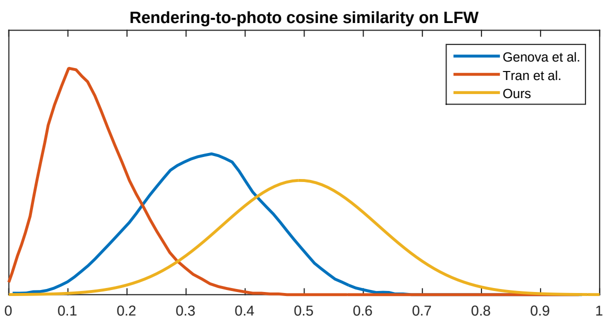


Figure 1: Cosine similarity distributions of rendered and real images LFW based on activations at the embedding layer of VGG-Face network[3]. Our method achieves more than 0.5 similarity on average which [1] has 0.35 average similarity and [6] 0.16 average similarity. Camera and lighting parameters are fixed for all renderings.

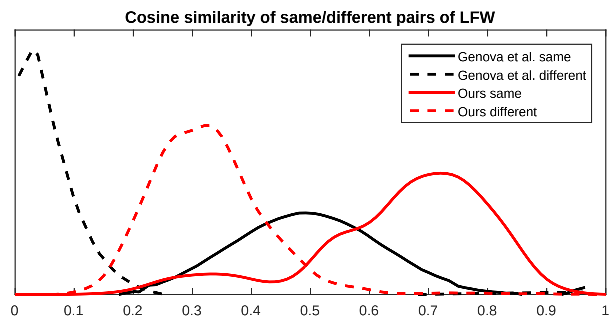


Figure 2: Our method successfully preserve identity so that distribution of cosine similarity of same/different pairs is separable by thresholding. Camera and lighting parameters are fixed for all renderings.

1. Experiments on LFW

In order to evaluate identity preservation capacity of the proposed method, we run two face recognition experiments on Labelled Faces in the Wild (LFW) dataset [2]. Following [1], we feed real LFW images and rendered images of their 3D reconstruction by our method to a pretrained face recognition network, namely VGG-Face[3]. We then compute the activations at the embedding layer and measure cosine similarity between 1) real and rendered images and 2) renderings of same/different pairs.

In Fig. 1 and 2, we have quantitatively showed that our method is better at identity preservation and photorealism (i.e., as the pretrained network is trained by real images) than other state-of-the-art deep 3D face reconstruction approaches [1, 6].

2. More Qualitative Results

Figures 3, 4, 5, and 6 illustrate the reconstructions of our method under different settings in comparison to the other state-of-the-art methods. Please see figure captions for detailed explanation.

References

[1] 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, 2

[2] Gary B. Huang, Marwan Mattar, Honglak Lee, and Erik Learned-Miller. Learning to align from scratch. In *NIPS*, 2012. 1

[3] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015. 1

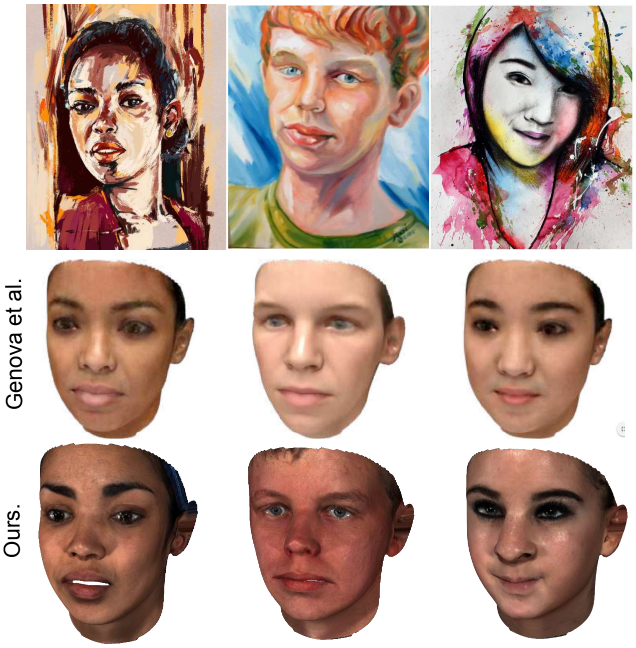


Figure 3: Our results on BAM dataset[7] compared to [1]. Our method is robust to many image deformations and even capable of recovering identities from paintings thanks to strong identity features.

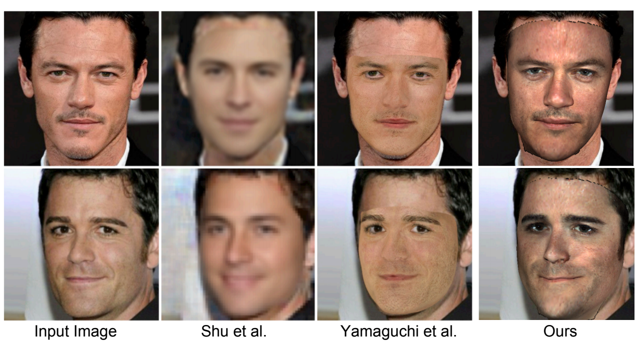


Figure 4: Qualitative comparison with [8, 5] by overlaying the reconstructions on the input images. Our method can generate high fidelity texture with accurate shape, camera and illumination fitting.

[4] Shunsuke Saito, Lingyu Wei, Liwen Hu, Koki Nagano, and Hao Li. Photorealistic facial texture inference using deep neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, volume 3, 2017. 3

[5] Zhixin Shu, Ersin Yumer, Sunil Hadap, Kalyan Sunkavalli, Eli Shechtman, and Dimitris Samaras. Neural face editing with intrinsic image disentangling. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pages 5444–5453. IEEE, 2017. 2

[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 *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pages 1493–1502. IEEE, 2017. 1

[7] Michael J Wilber, Chen Fang, Hailin Jin, Aaron Hertzmann, John Collomosse, and Serge J Belongie. Bam! the behance artistic media dataset for recognition beyond photography. In *ICCV*, pages 1211–1220, 2017. 2

[8] Shuco Yamaguchi, Shunsuke Saito, Koki Nagano, Yajie Zhao, Weikai Chen, Kyle Olszewski, Shigeo Morishima, and Hao Li. High-fidelity facial reflectance and geometry inference from an unconstrained image. *ACM Transactions on Graphics (TOG)*, 37(4):162, 2018. 2

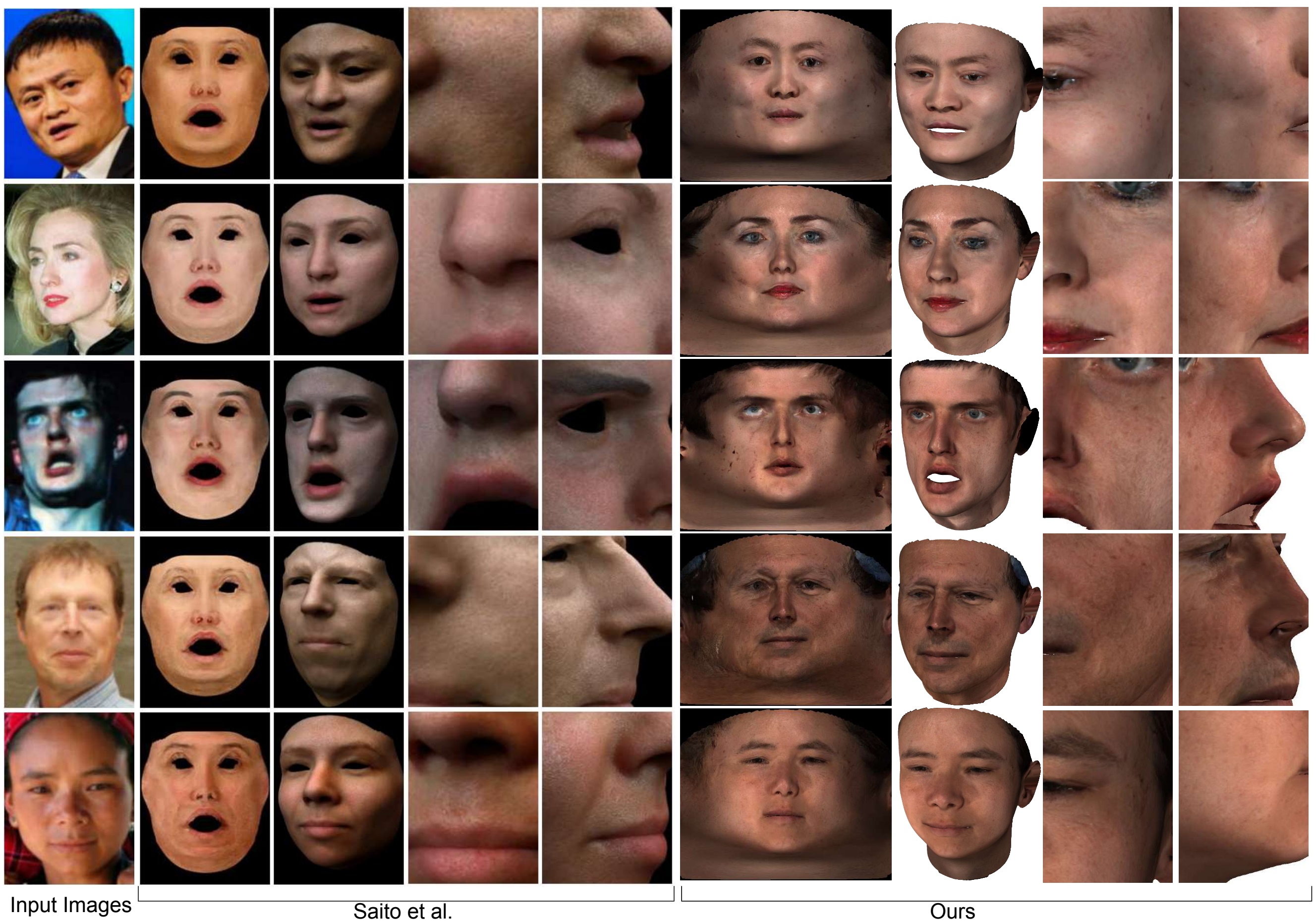
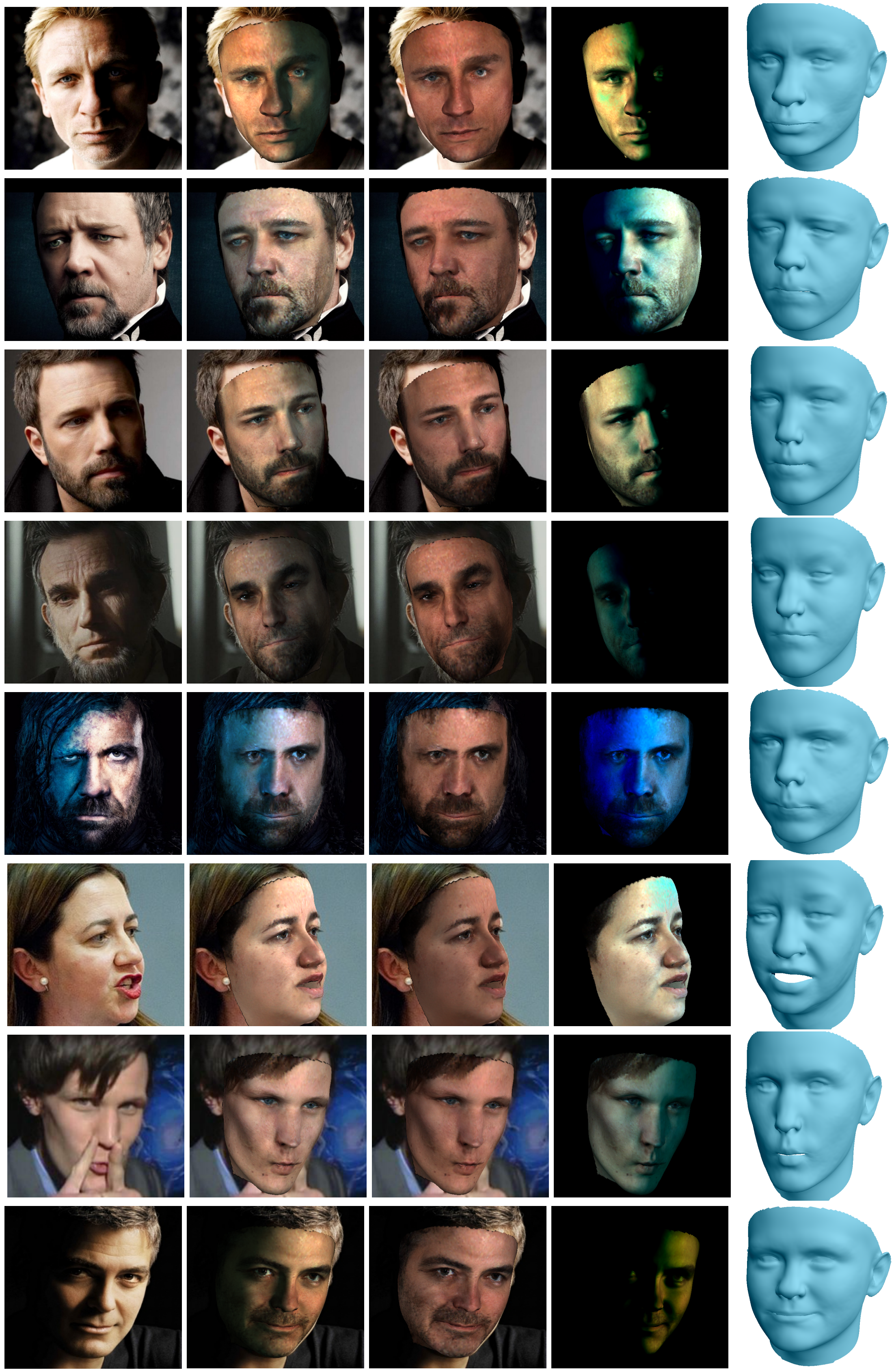


Figure 5: Qualitative comparison with [4] by means of texture maps, whole and partial face renderings. Please note that while our method does not require any particular renderer for special effects, e.g., lighting, [4] produce these renderings with a commercial renderer called Arnold.



(a) $\mathbf{I^0}$ (b) $\mathbf{I}^{\mathcal{R}}$ (c) $\mathbf{I}^{\mathcal{R}}_{\mathbf{alb.}}$ (d) $\mathbf{I}^{\mathcal{R}} - \mathbf{I}^{\mathcal{R}}_{\mathbf{alb.}}$ (e) $\mathbf{S}$

Figure 6: Results under more challenging conditions, *i.e*. strong illuminations, self-occlusions and facial hair. (a) Input image, (b) Estimated fitting overlayyed including illumination estimation, (c) Overlayyed fitting without illumination, (d) Pixel-wise intensity difference of (b) to (c), (e) Estimated shape mesh