Supplementary Material for Deep Face Normalization

1 ADDITIONAL RESULTS AND EVALUATIONS

In Fig.1, we show an additional quantitative comparison with [Fried et al. 2016] using the optical flow numerical score described in their paper. We run perspective undistortion of each method on the entire CMDP dataset [Burgos-Artizzu et al. 2014], rigidly align the predicted image and ground truth, and measured optical flow displacements as a pixel alignment error for all the photos of all the subjects. In Fig.1, we report the median of the optical flow distances for all the photos comparing ours with the input image (undistorted) and [Fried et al. 2016]. Boxes are 25th to 75th percentile and the orange line is a median of medians. As can be seen, the error is reduced after the perspective undistortion with both methods, but ours gives the better results than [Fried et al. 2016]. In Table 1, we report the value of the median of medians (orange line).

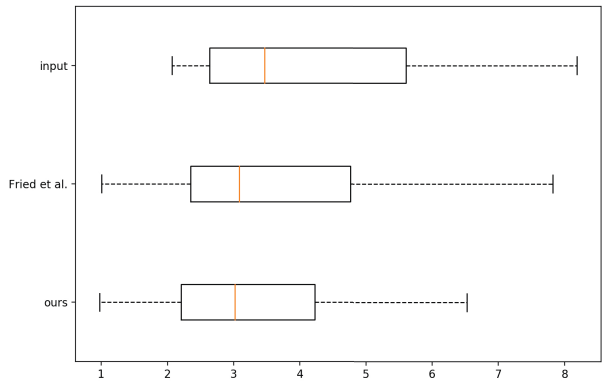

Fig. 1. Additional numerical score comparison to [Fried et al. 2016]. We compute the median value of the optical flow pixel alignment error between the undistorted image and ground truth over the entire CMDP dataset.

Table 1. Median of medians, corresponding to the red lines in Fig.1. Ours gives the best score.

Input	Fried et al.	Ours
3.467065	3.085133	**3.019704**

Fig. 2 shows the difference of the generated avatars before and after lighting normalization in CG scenes. Our method enables faithful relighting under novel illumination conditions. The CG scenes are rendered in the Unity3D game engine using an HDR light probe.

Fig. [Li et al. 2014] shows comparison with the state of the art intrinsic image decomposition technique [Li et al. 2014]. While our result reveals plausible skin color of the subject, the previous method fails to recover meaningful skin color.

Author's address:

Fig. 2. Relighting in a CG scene.The avatar without illumination normalization bakes the original lighting into the texture, which yields incorrect skin tones. Original image (bottom) is courtesy of Vadim Pacev.

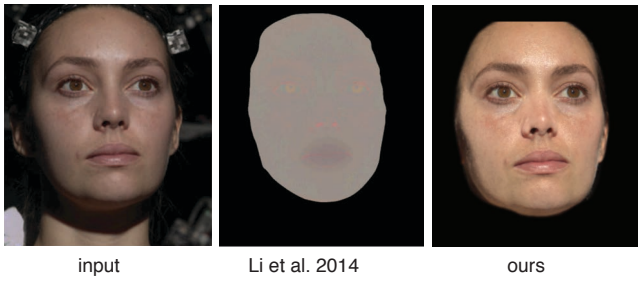

Fig. 3. Comparison with a data-driven intrinsic decomposition method [Li et al. 2014] (produced by original authors).

Fig. 4. Numerical ground truth evaluation of our lighting normalization method using synthetic lighting inputs (top row). Our results show consistent skin color that is close to the ground truth (second row). The pixel difference error to the ground truth is visualized in the heat map in the third row.

In Fig.4, we show quantitative ground truth evaluation of our lighting normalization method using synthetic illuminations (top row). The pixel difference error visalized in heat map (third row) confirms that our method predicts skin tone that is close to the ground truth (second row).

In Fig.5, we show additional lighting consistency evaluations under varying lighting conditions. Even under extreme lighting

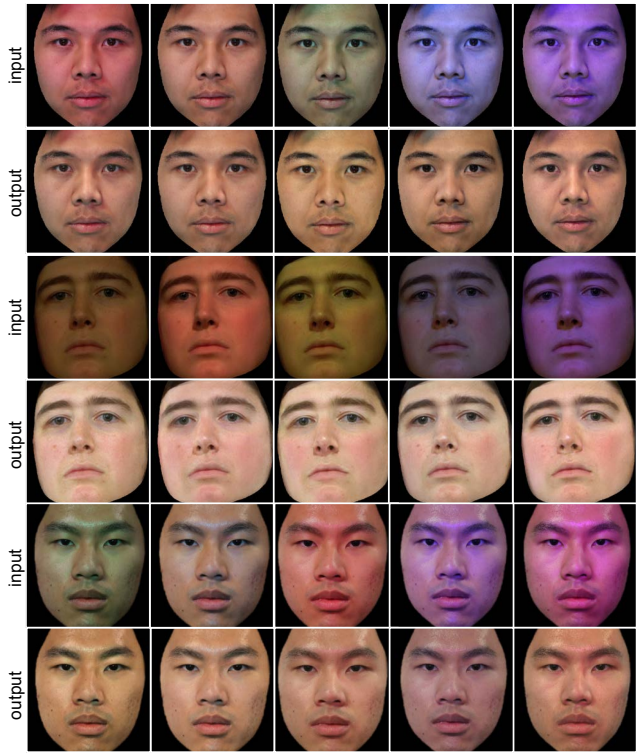

Fig. 5. Lighting consistency. Top rows show input photos with varying lighting conditions and the bottom rows show lighting normalized photos.

conditions, our method can produce faithful skin colors that are appropriate to different races (see Fig.21 for a different instance).

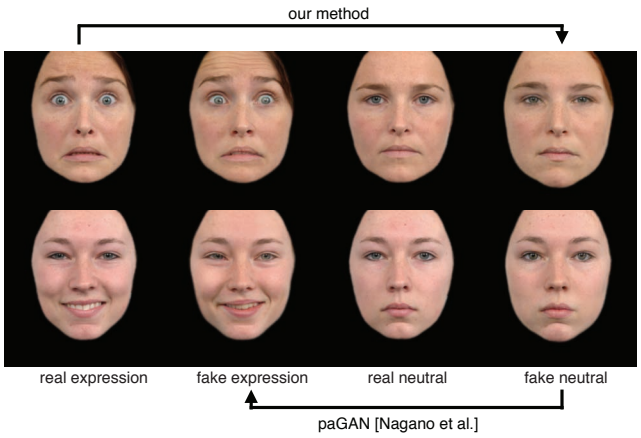

Fig. 6. Cycle consistency of our work using paGAN [Nagano et al. 2018]. Our expression neutralization behaves like the inverse operation of paGAN.

In Fig. 6, we evaluate the cycle consistency of our method. Given a portrait with expression (first column), we apply our expression neutralization method to synthesize a neutral face (forth column). The neutral face is passed to paGAN [Nagano et al. 2018] to synthesize the expression back to get the fake expression image (second column). Our method successfully keeps the consistency between the real and generated images with expression.

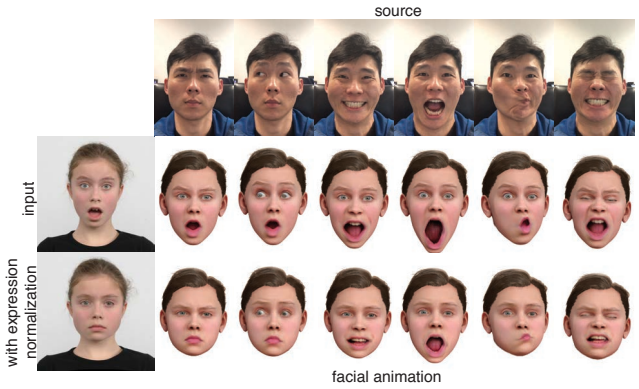

Fig. 7. Comparison of facial animation results with and without expression normalization. The expression of the input image is baked into the animation when normalization is not performed.

In Fig. 7, we demonstrate the importance of expression neutralization on a facial animation application. Without expression neutralization, the mouth open and eye wide expressions are baked in the avatar geometry, and cannot produce consistent expressions as the source sequence.

We also evaluate the robustness of our pose frontalization technique using challenging in the wild examples (Fig. 8). Although the given inputs are low resolution and captured under a poor lighting condition, our method can handle them and even produce reasonable results with decent resolutions.

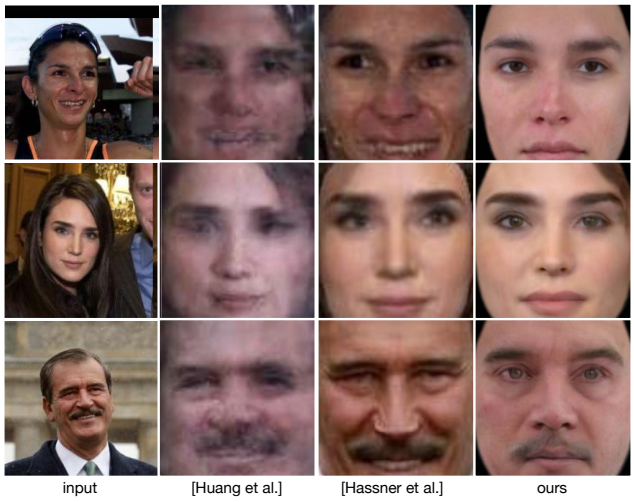

Fig. 8. Our results compared to previous work on pose normalization.

In Fig.11, we show additional results and corresponding 3D avatars with and without normalization. The figure demonstrates that our

method can handle a wide variety of ethnicities, ages, and skin types with varying cameras, lighting conditions, and expressions.

In Fig.9, we show samples of test images used for user study for perspective normalization, lighting normalization, and expression normalization. As described in the PDF, users are asked to pick a ground truth normalized image, given the unnormalized input (first, fourth columns), ground truth (second, fifth columns), and output from our network (third, sixth columns). For each fake image, we show a corresponding fooling rate (0.5 is completely random guess) on the lower right corner.

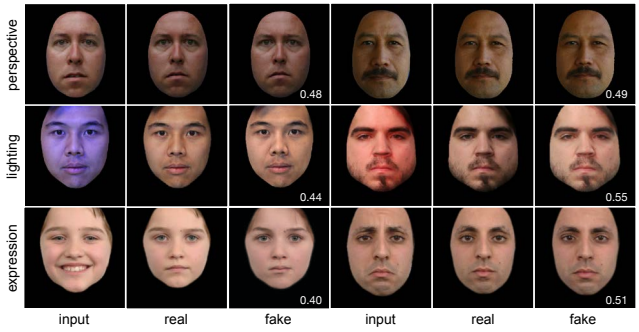

Fig. 9. Test samples used for user study for perspective normalization, lighting normalization, and expression normalization.

2 IMPLEMENTATION DETAILS

Perspective Undistortion. As described in the paper, we applied random brightness, contrast, and blurring to the grayscale input image during the perspective undistortion network training. For the brightness and contrast, we employed the color jitter transform in the Pytorch library with a value range of 0.3 for both brightness and contrast. For the blurring, we used gaussian blurring with a kernel randomly drawn from 0 to 11.

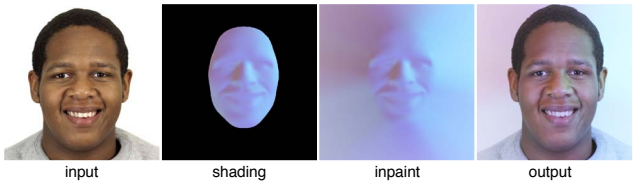

Fig. 10. A step by step process to create synthetic illuminations.

Even Light Illumination. In Fig.10, we show a step by step process to create illumination variations for our light normalization network training. Using the extracted 3D face proxy, we first render the face geometry using a lighting pattern of choice (directional lighting or an HDR environment) to create a face shading image (second column). To illuminate the background and hair, we perform image inpainting [Telea 2004] with an inpainting radius of 3 to propagate the shading image to the image background (third column). Finally the inpainted shading image is multiplied to the input portrait in the studio lighting to produce the relit portrait (fourth column). Fig.10 shows an example relighitng using an HDR environment map.

For the light normalization training, we used directional lighting and an HDR environment lighting to simulate a wide variety of natural lighting conditions using the GLSL shader described in the PDF. For directional lighting, we randomly placed one directional light on the front hemisphere of the face and created the rendering. For the HDR environment, we collected around 300 HDR light probes from a public database [Greg Zaal 2019]. At the rendering time, we picked a random HDR environment and produced the face rendering using spherical harmonics lighting [Ramamoorthi and Hanrahan 2001]. To further increase the variation of the training data, we applied the color jitter in Pytorch with a brightness range of 0.2 and contrast range of 0.15.

REFERENCES

Xavier P. Burgos-Artizzu, Matteo Ruggero Ronchi, and Pietro Perona. 2014. Distance Estimation of an Unknown Person from a Portrait. In *ECCV*. Springer International Publishing, Cham, 313–327.

Ohad Fried, Eli Shechtman, Dan B Goldman, and Adam Finkelstein. 2016. Perspective-aware Manipulation of Portrait Photos. *ACM Trans. Graph.* (July 2016).

Greg Zaal. 2019. https://hdrihaven.com.

Chen Li, Kun Zhou, and Stephen Lin. 2014. Intrinsic Face Image Decomposition with Human Face Priors. In *ECCV*. 218–233.

Koki Nagano, Jaewoo Seo, Jun Xing, Lingyu Wei, Zimo Li, Shunsuke Saito, Aviral Agarwal, Jens Fursund, and Hao Li. 2018. paGAN: Real-time Avatars Using Dynamic Textures. *ACM Trans. Graph.* 37, 6, Article 258 (Dec. 2018), 12 pages.

Ravi Ramamoorthi and Pat Hanrahan. 2001. An efficient representation for irradiance environment maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 497–500.

Alexandru Telea. 2004. An Image Inpainting Technique Based on the Fast Marching Method. *J. Graphics, GPU, and Game Tools* 9, 1 (2004), 23–34. http://dblp.uni-trier.de/db/journals/jgtools/jgtools9.html#Telea04

Fig. 11. First column: input photo. Second to fourth column: portrait with individual normalization component applied, i.e., perspective normalization, perspective+lighting normalization, and full normalization. The fifth and sixth columns show an avatar generated from an original input without normalization and from a fully normalized picture. From top to bottom, original images are courtesy of Michael Beserra, NWABR, The Pentecostals of OC, SPÄÜ Presse und Kommunikation, Brett Morrison, and Jelle