Analyzing and Improving the Skin Tone Consistency and Bias in Implicit 3D Relightable Face Generators (Supplementary Document)

Libing Zeng
Texas A&M University
College Station, USA
libingzeng@tamu.edu

Nima Khademi Kalantari
Texas A&M University
College Station, USA
nimak@tamu.edu

Abstract

In the supplementary document, we further compare our method against NeRFFaceLighting [3] visually and numerically. Specifically, we first evaluate the geometry and relighting quality of the albedo and relit images of our model obtained by fine-tuning EG3D [1] like NeRFFaceLighting [3] using our strategy. We then compare our method without any fine-tuning against NeRFFaceLighting [3]. Our approach consistently produces improved visual and numerical results across all comparisons.

1. Geometry and Relighting Quality Evaluation

We compare our generated albedo and relit images from different yaw and pitch angles against NeRFFaceLighting in Fig. 1. Both approaches generate reasonable albedo images from different views with high geometric consistency. Our relit images, however, are considerably better than NeRFFaceLighting. In particular, NeRFFaceLighting creates unrealistic shading on the teeth and nose. This demonstrates that our strategy not only enhances skin tone consistency and reduces bias, but also improves the quality of relighting. We attribute this to the fact that, in addition to handling lighting directions, NeRFFaceLighting also needs to manage the magnitude. In contrast, our generator focuses solely on the direction and produces relit images with varying illumination magnitudes through scaling.

Furthermore, we numerically assess the geometric and relighting consistency of the generated images. Specifically, we randomly sample 100 latent vectors and relight each with a random lighting. Subsequently, we render each sample from multiple viewpoints with varying yaw angles: -0.5, -0.25, 0 (frontal), 0.25 , and 0.5 radians. We then compute identity similarity scores between yaw-posed and frontal views using ArcFace [2], a face recognition algorithm. We show the average score of both the relit and albedo images for each method in Table 1. Note that we separate EG3D from the remaining approaches since it is not relightable. As seen, our method produces the best results among the relightable approaches and is closest to our backbone EG3D,

Figure 1. Qualitative comparison against NeRFFaceLighting [Jiang et al.]. We show generated albedo images from various poses on the left, while the corresponding relit images are shown on the right. Both methods produce high quality albedo images, but our method produces significantly better relit results.

Table 1. Quantitative comparison against the other approaches in terms of identity similarity.

	Relit image identity similarity ↑					Albedo image identity similarity ↑				
	-0.5 rad	-0.25 rad	0 rad	0.25 rad	0.5 rad	-0.5 rad	-0.25 rad	0 rad	0.25 rad	0.5 rad
EG3D	0.6632	0.8527	-	0.8512	0.6815	-	-	-	-	-
NeRFFaceLighting	0.5987	0.8219	-	0.8263	0.6064	0.6402	0.8601	-	0.8603	0.6069
NeRFFaceLighting-DECA	0.5733	0.8076	-	0.7910	0.5580	0.6033	0.8584	-	0.8452	0.6116
SH-Norm	0.6087	0.8304	-	0.8299	0.6090	0.6456	0.8739	-	0.8602	0.6213
Ours	**0.6303**	**0.8442**	-	**0.8320**	**0.6230**	**0.7115**	**0.8809**	-	**0.8661**	**0.6565**

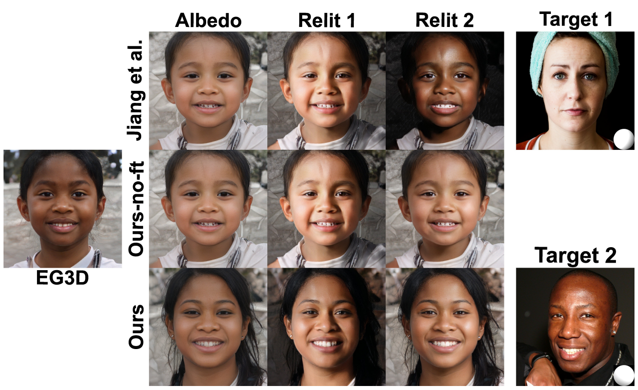

Figure 2. On the top, we show two relit images produced by NeRFFaceLighting [Jiang et al.] [3], using the lighting extracted from images of individuals with fair and dark skin tones (shown on the right). As seen, NeRFFaceLighting produces relit images with inconsistent skin tones. Additionally, when distilling the EG3D triplane, NeRFFaceLighting tends to produce albedo maps that are biased towards lighter skin colors. Our method mitigates this bias and improves the consistency of the skin tone in relit images. Note that even though we use the same latent vector to generate the results with EG3D, NeRFFaceLighting, and ours, there are variation in the images as the backbone EG3D network is fine-tuned separately in NeRFFaceLighting and ours. Our method can also be directly applied to NeRFFaceLighting without fine-tuning, as demonstrated in the second row.

demonstrating that our relit images are realistic.

2. Further Comparisons against NeRFFace-Lighting

To clearly illustrate the difference between applying our approach directly to NeRFFaceLighting without fine-tuning and fine-tuning EG3D [1] using our approach, we replicate Fig. 1 from the main paper and add an additional row of results without fine-tuning. As shown in the second row of Fig. 2, our approach, when applied without fine-tuning, does not affect the albedo image or the relit image when using lighting from a fair-skinned face. However, when the albedo image is relit using lighting from a dark-skinned face, our result maintains greater skin tone consistency compared to NeRFFaceLighting. This demonstrates that our approach can remove the skin tone bias in NeRFFaceLighting simply by processing the lighting conditions, indicating that the bias originates from the lighting itself.

References

[1] Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein. Efficient geometry-aware 3D generative adversarial networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2

[2] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4690–4699, 2019. 1

[3] Kaiwen Jiang, Shu-Yu Chen, Hongbo Fu, and Lin Gao. Nerffacelighting: Implicit and disentangled face lighting representation leveraging generative prior in neural radiance fields. *ACM Transactions on Graphics (TOG)*, 42, 2023. 1, 2