

Supplementary Material – FaceLit: Neural 3D Relightable Faces

Anurag Ranjan¹ Kwang Moo Yi² Jen-Hao Rick Chang¹ Oncel Tuzel¹
¹Apple ²The University of British Columbia

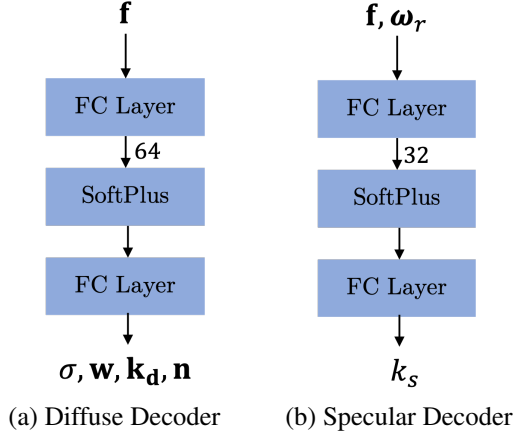


Figure 1. Architecture of Diffuse Decoder and Specular Decoder

A. Network Architecture

We show the architecture of the diffuse Decoder and the specular decoder as described in Section 3 in Fig. 1. We follow a feed-forward architecture with fully connected layers and softplus activation [4]. The fully connected layers have a hidden dimension of 64 for diffuse decoder and 32 for specular decoder.

B. Training Time

Our training time is slightly higher than EG3D due to addition of illumination model. EG3D takes 4 days, FaceLit-d takes 4.5 days and FaceLit-f takes 4.8 days on 8 NVIDIA A100 GPUs to train at a neural rendering resolution of 64^3 .

C. Additional Qualitative Results

Albedo. In Fig. 2, we show the visualization of albedo images of the corresponding generated samples. The albedo images are obtained by rendering the images before the superresolution module under constant illumination.

Uncurated Samples. In Fig. 3, we show uncurated samples generated by our model with a truncation factor of $\psi = 1.0$. In Fig. 4, we show uncurated samples generated by our model with a truncation factor of $\psi = 0.7$. In both case, we

use the FaceLit-F model trained on the FFHQ dataset.

Latent Space Interpolations. In Fig. 5, we show interpolation across generated samples from our model. In each row, we interpolate the latent code linearly and observe that the generated samples vary smoothly and generalize to wide variety of faces, even with accessories such as eyeglasses.

Videos. For visuals, please open the `README.html`. The webpage contains all the video results. The webpage is tested to run on a MacOS and Ubuntu and supports Safari, Chrome, and Firefox browsers. All videos are encoded with h264 encoding and the raw files are present in the data directory.

D. Additional Quantitative Results

We compare our method with VoLux-GAN [3] and ShadeGAN [2] by evaluating the face identity consistency of the generated samples by measuring the cosine similarity between the samples using Arcface [1]. In Tab. 1, we compute the cosine similarity of the face rendered with yaw $\in [-0.5, -0.25, 0.25, 0.5]$ radians with yaw = 0. The methods are trained on different datasets therefore it does not reflect a direct comparison. However, it shows that our method preserves identity similar to the baseline methods.

In Tab. 2, we evaluate the identity similarity under different lighting conditions as compared to the rendered albedo of the face. Similar to previous evaluation, the methods are trained on different datasets and have different models of illumination. ShadeGAN uses point light sources, and the face consistency is measured under different point light sources and the albedo. In VoLux-GAN, the face consistency is measured under different environment maps and the albedo. In our case, we measure it using 3 different spherical harmonics lights and the albedo. We observe that the face similarity is preserved under different lighting conditions.

References

- [1] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on*

		Dataset	Yaw changes in radians			
			-0.5	-0.25	0.25	0.5
ShadeGAN [2]	CelebA		0.481	0.751	0.763	0.500
VoLux-GAN [3]	CelebA		0.606	0.774	0.800	0.599
FaceLit-D	FFHQ		0.615	0.817	0.781	0.596
FaceLit-F	FFHQ		0.581	0.795	0.783	0.626

Table 1. Face consistency metrics with yaw changes.

	Dataset	Illumination Type	Setting		
			1	2	3
ShadeGAN [2]	CelebA	Point Source	0.581	0.649	0.666
VoLux-GAN [3]	CelebA	Environment Map	0.760	0.890	0.808
FaceLit-D	FFHQ	Spherical Harmonics	0.837	0.911	0.811
FaceLit-F	FFHQ	Spherical Harmonics	0.839	0.906	0.815

Table 2. Face consistency metrics with illumination changes.

computer vision and pattern recognition, pages 4690–4699, 2019. 1

- [2] Xingang Pan, Xudong Xu, Chen Change Loy, Christian Theobalt, and Bo Dai. A shading-guided generative implicit model for shape-accurate 3d-aware image synthesis. *Advances in Neural Information Processing Systems*, 34:20002–20013, 2021. 1, 2
- [3] Feitong Tan, Sean Fanello, Abhimitra Meka, Sergio Orts-Escolano, Danhang Tang, Rohit Pandey, Jonathan Taylor, Ping Tan, and Yinda Zhang. Volux-gan: A generative model for 3d face synthesis with hdri relighting. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–9, 2022. 1, 2
- [4] Huizhen Zhao, Fuxian Liu, Longyue Li, and Chang Luo. A novel softplus linear unit for deep convolutional neural networks. *Applied Intelligence*, 48(7):1707–1720, 2018. 1



Figure 2. **Albedo Visualization:** Generated images on top with their corresponding albedo images at the bottom.



Figure 3. Uncurated FFHQ samples at $\psi = 1.0$



Figure 4. Uncurated FFHQ samples at $\psi = 0.7$



Figure 5. **Latent Space Interpolation:** In each row, we smoothly vary the latent code from left to right.