Supplemental File:
Holo-Relighting: Controllable Volumetric Portrait Relighting from a Single Image

Yiqun Mei[1] Yu Zeng[1†] He Zhang[2†] Zhixin Shu[2†] Xuaner Zhang[2] Sai Bi[2]
Jianming Zhang[2] HyunJoon Jung[2] Vishal M. Patel[1]
[1]Johns Hopkins University [2]Adobe Inc.

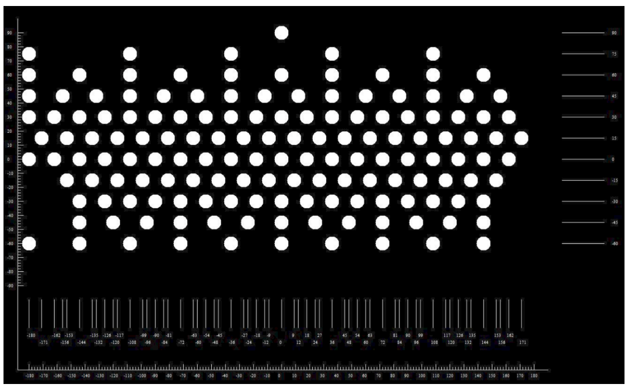

Figure 1. Illustration of the light distribution in our light stage. The image is presented in the panoramic format, where the horizontal axis represents longitude, and the vertical axis represents latitude.

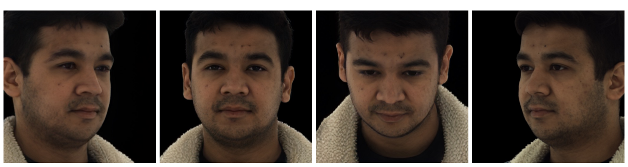

Figure 2. Examples of captured four views using our light stage.

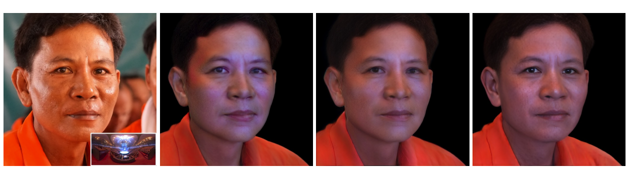

Figure 3. Visual comparison for ablation study. Either single-view inversion or removing shading transfer (ST) results in blurry lighting effects and harms image quality.

1. Video Demonstration

We encourage readers to view the provided supplemental video for a better demonstration of the controllability and relighting quality of *Holo-Relighting*.

[†]The second authors

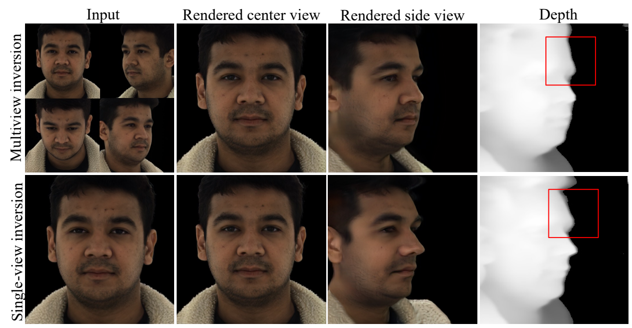

Figure 4. Illustration of multi-view regularization. Single-view inversion fails to reconstruct the actual geometry due to the depth ambiguity (see side view depth). Using multi-view inversion relieves this issue and provides a more accurate geometry.

Figure 5. Additional free-view relighting comparison with the state-of-the-art parametric face-based method PN-Relighting [12]. Due to the limited expressiveness of their parametric face model, [12] fails to synthesize realistic lighting effects, hairs and mouth interiors, and creates "holes" in the regions that are invisible from the input images.

2. Implementation Details

Network Architecture. Both the albedo net and the normal net have a simple U-shaped [10] structure with three down-sampling layers and three up-sampling layers. Both

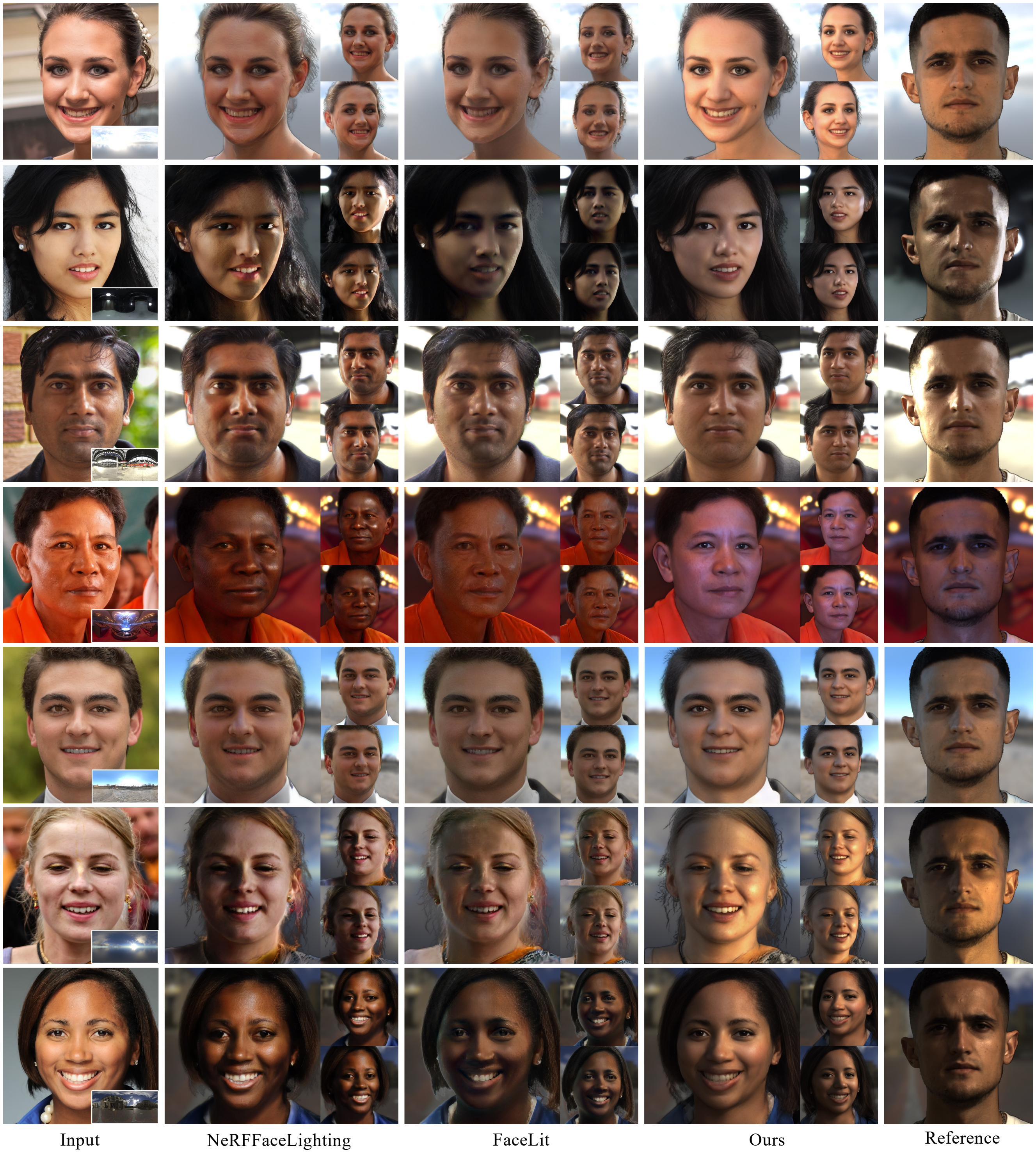

Figure 6. Visual comparison on free-view relighting. We compare our method with NeRFFaceLighting [6] and FaceLit [9].

networks have 64-128-256-512-256-128-64 hidden channels. For the relighting net, each resolution stage (*i.e.* the relighting block) contains one convolution layer, one residual block and one transposed convolution layer for upsampling. The output channel number matches the channel number of corresponding style block in the triplane generator.

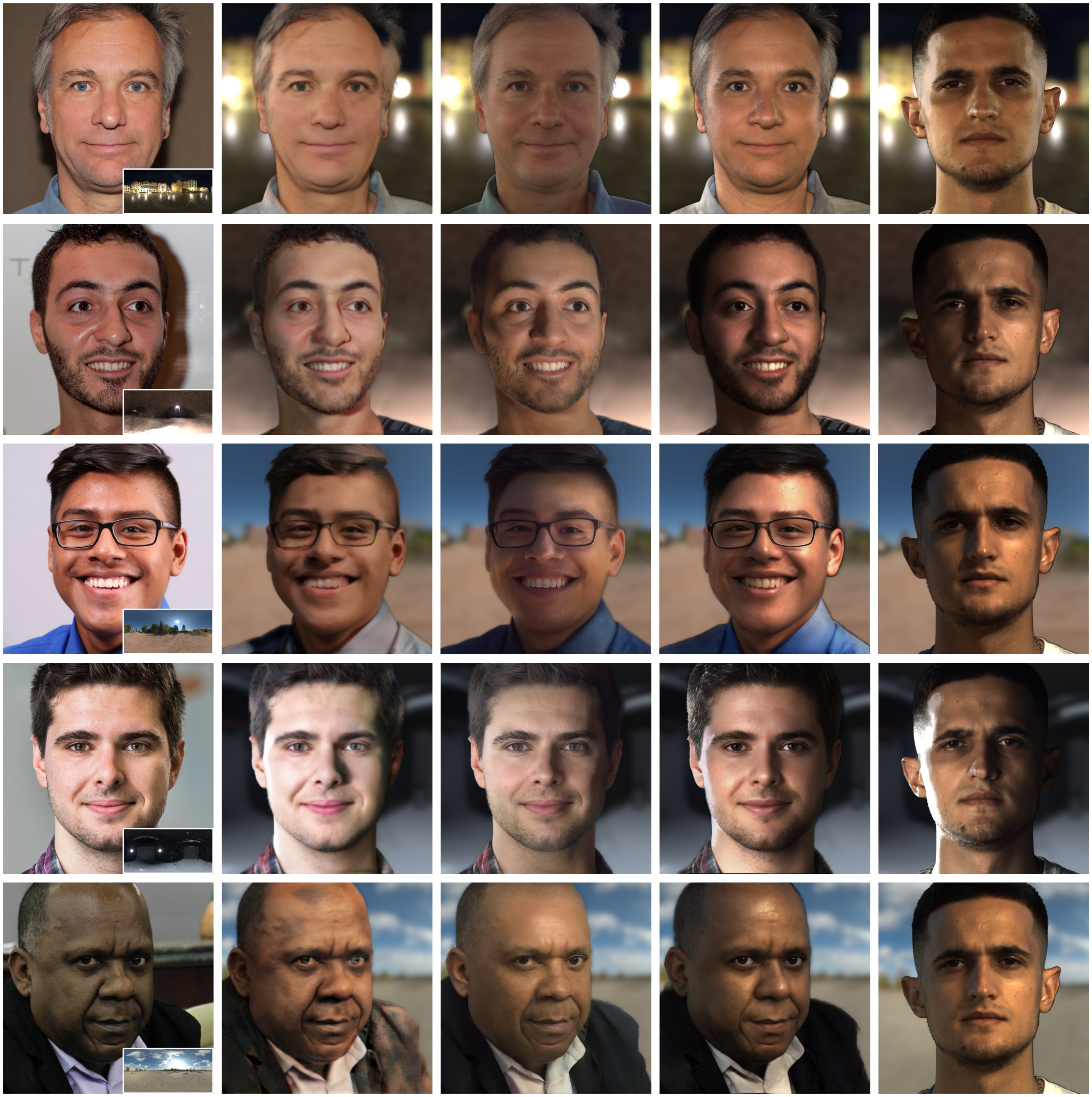

Figure 7. Visual comparison on 2D portrait relighting. We compare our method with SIPRW [13] and Total Relighting [8] (TR).

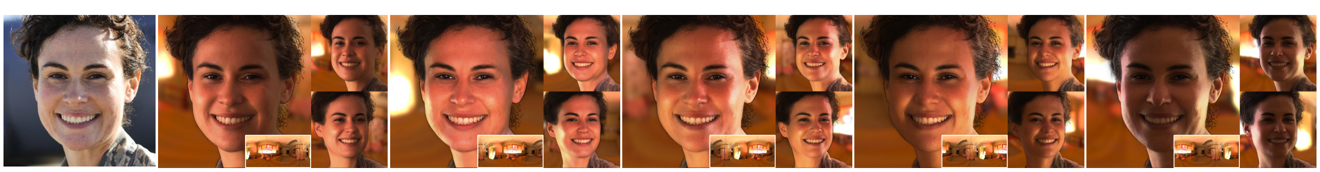

Figure 8. Illustration of lighting control using dynamic illumination. Our method can produce consistent lighting effects with a rotating lighting environment.

Input (a) coarse-only (b) fine-only (c) coarse-to-fine

Figure 9. Ablation study on coarse-to-fine relighting. Without the coarse-to-fine design, the model fails to capture either local effects (*e.g.* highlights) or the global shading distribution ((a) & (b)).

Inference and Training Details. We separately train the delighting stage and relighting stage. During both inference and training, we use the off-the-self estimators [4] and [5] to obtain camera parameters and head pose respectively. Training our method takes about 3 days on 8 NVIDIA A100 GPUs.

Light Stage Setup and Rendering Details. We train *Holo-Relighting* using data rendered by the light stage captures. Specifically, we use a light stage similar to [7]. The rig has a diameter of 3.6 meters, and is equipped with 160 programmable LED lights and four frontal-view cameras. Figure 1 shows the distribution of lights. We use the MER2-502-79U3M high-speed camera to capture subjects' reflectance field at 5 megapixel resolution with an exposure time of 20ms. We crop head regions from the captured OLAT sequences and resize them to the resolution of 512×512 for data rendering. Examples of the captured views (after cropping and resizing) are shown in Figure 2. We approximate the albedo using a lighting normalized image captured under a flat illumination following [7, 8]. Similar to TR [8], we pair an environment map with an OLAT sequence to get a relit image via image-based relighting [3]. Then, we apply shading transfer to obtain a pseudo ground truth for supervising the relighting net.

Shading Transfer Details. We create pseudo ground-truth images using shading transfer to train the relighting net. The pseudo ground-truth images are used in all terms of $\mathcal{L}_{relit}$. Relighting with real ground-truth images leads to blurry results, because the input to the relighting stage is not the ground-truth albedo, instead it is the ground-truth albedo's inversion, whose details are not perfectly aligned with real ground truth images. To alleviate this, we create a pseudo ground-truth image by transferring the shading effect to the inverted albedo, allowing pixel-wise alignment with the input in training. When transfering the shading, we add a small $\epsilon = 1e-4$ to the denominator to ensure numerical stability.

3. Visual Results for Ablation Study

We report visual results for ablation study in Figure 3. Either using single-view inversion or removing shading transfer (ST) harms relighting quality and leads to blurry results. We provide an additional illustration of the multi-view regularization using depth maps in Figure 4. While both single-view inversion and multi-view inversion can reconstruct the input view well, the geometry from the single-view inversion fails to reflect the true geometry of the subject due to the depth ambiguity. This is illustrated by the rendered side view from the inverted latent code. Inaccurate geometry impedes the network from learning to perform relighting by using geometry clues, and results in blurry lighting effects as shown in Figure 3. Multi-view regularization effectively alleviates this problem.

Here we also investigate the coarse-to-fine design of the relighting net. We compare it with "coarse-only" and "fine-only" feature injection and the results are shown in Figure 9. Without the coarse-to-fine design, the model fails to capture either local effects (*e.g.* highlights) or the global shading distribution. Thus coarse-to-fine design is crucial to handle both global & local effects well (Figure 9 (c)).

4. More Visual Results

We conduct additional visual comparison with the state-of-the-art parametric-face-based method PN-Relighting [12] on free-view relighting. As their code for free-view relighting is not released, we acquire results from their authors. As shown in Figure 5, due to the limited expressiveness of the parametric face model, their method fails to synthesize realistic lighting effects, hairs and mouth interiors and also produces black "holes" in the areas that are not visible from the input views.

In Figure 6, we report more visual comparsion with NeRFFaceLighting [6] and FaceLit [9] on free-view relighting.

In Figure 7, we provide extra qualitative comparsion on 2D portrait relighting. Here we also compare our method with SIPRW [13]. Results for SIPRW [13] and TR [8] are acquired from their authors, as their code is not available.

5. Lighting Control: Dynamic Illumination

We have shown our method can robustly control lighting in Figure 6 & 7. Here we further demonstrate *Holo-Relighting* also handles *dynamic illumination*, *i.e.* a rotating lighting environment around the subject. As shown in Figure 8, our method stably produces realistic shading and specular highlights across frames. *Holo-Relighting* also renders plausible rim lighting around the contour of the face, as shown in the last column of Figure 8.

6. Limitations

Holo-Relighting leverages the pretrained EG3D [2] and GAN inversion [14] to extract 3D information from a 2D input image. It is thus challenging to apply our current implementation to upper-body portrait relighting, which is beyond the capability of EG3D. Extending our method to more specialized human generative models [15, 19] could be an interesting direction for future work. Moreover, GAN inversion [14] might hallucinate inaccurate details such as the freckles shown in Figure 3 in the main paper. We use shading transfer to alleviate this issue when preparing for the training data. However, this problem still remains at inference time. Our method also suffers from the common limitations of GAN inversion-based image editing, where the imperfect inversion might lead to identity shift and losing some details. The inversion might also induce some inconsistency upon tiny geometry details (*e.g*. hairs) that causes flickers when rendering to videos. Developing more advanced inversion techniques [1, 17, 18] could be a potential solution to explore for future work.

For relighting, we demonstrate that our method is generally robust to diverse lighting conditions and control signals. However, similar to previous approaches [7, 8, 11], *Holo-Relighting* can only generate lighting effects that are represented in the training data. More complicated lighting effects such as foreign shadow do not exist in the light stage training data and thus cannot be produced. Further, as our method learns to approximate the light transport from data rather than adhering to physical constraints, we found some challenging cases (*e.g*. view-dependent effects) may not be perfectly handled. In addition, as shown in Figure 7 (third row, fourth column), our method fails to render eyeglasses glares as such effect only accounts for a minor portion of the overall loss function. A possible solution is to add an explicit supervision on eyeglasses in a way similar to [16].

References

[1] Ananta R. Bhattarai, Matthias Nießner, and Artem Sevastopolsky. Triplanenet: An encoder for eg3d inversion. In *WACV*, 2024.

[2] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *CVPR*, pages 16123–16133, 2022.

[3] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 145–156, 2000.

[4] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *CVPRW*, 2019.

[5] Yao Feng, Haiwen Feng, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *ACM Transactions on Graphics (ToG)*, 40(4): 1–13, 2021.

[6] 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 TOG*, 42(3):1–18, 2023.

[7] Yiqun Mei, He Zhang, Xuaner Zhang, Jianming Zhang, Zhixin Shu, Yilin Wang, Zijun Wei, Shi Yan, HyunJoon Jung, and Vishal M Patel. Lightpainter: Interactive portrait relighting with freehand scribble. In *CVPR*, pages 195–205, 2023.

[8] Rohit Pandey, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. Total relighting: learning to relight portraits for background replacement. *ACM TOG*, 40(4):1–21, 2021.

[9] Anurag Ranjan, Kwang Moo Yi, Jen-Hao Rick Chang, and Oncel Tuzel. Facelit: Neural 3d relightable faces. In *CVPR*, pages 8619–8628, 2023.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. pages 234–241. Springer, 2015.

[11] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul E Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM TOG*, 38(4):79–1, 2019.

[12] Youjia Wang, Kai He, Taotao Zhou, Kaixin Yao, Nianyi Li, Lan Xu, and Jingyi Yu. Free-view face relighting using a hybrid parametric neural model on a small-olat dataset. *IJCV*, 131(4):1002–1021, 2023.

[13] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. Single image portrait relighting via explicit multiple reflectance channel modeling. *ACM TOG*, 39(6):1–13, 2020.

[14] Jiaxin Xie, Hao Ouyang, Jingtan Piao, Chenyang Lei, and Qifeng Chen. High-fidelity 3d gan inversion by pseudo-multi-view optimization. In *CVPR*, pages 321–331, 2023.

[15] Zhuoqian Yang, Shikai Li, Wayne Wu, and Bo Dai. 3dhumangan: 3d-aware human image generation with 3d pose mapping. In *ICCV*, pages 23008–23019, 2023.

[16] Yu-Ying Yeh, Koki Nagano, Sameh Khamis, Jan Kautz, Ming-Yu Liu, and Ting-Chun Wang. Learning to relight portrait images via a virtual light stage and synthetic-to-real adaptation. *ACM TOG*, 2022.

[17] Fei Yin, Yong Zhang, Xuan Wang, Tengfei Wang, Xiaoyu Li, Yuan Gong, Yanbo Fan, Xiaodong Cun, Ying Shan, Cengiz Oztireli, et al. 3d gan inversion with facial symmetry prior. In *CVPR*, pages 342–351, 2023.

[18] Ziyang Yuan, Yiming Zhu, Yu Li, Hongyu Liu, and Chun Yuan. Make encoder great again in 3d gan inversion through geometry and occlusion-aware encoding. In *ICCV*, pages 2437–2447, 2023.

[19] Jichao Zhang, Enver Sangineto, Hao Tang, Aliaksandr Siarohin, Zhun Zhong, Nicu Sebe, and Wei Wang. 3d-aware semantic-guided generative model for human synthesis. In *ECCV*, pages 339–356. Springer, 2022.