

Supplementary File: LightPainter: Interactive Portrait Relighting with Freehand Scribble

Yiqun Mei¹ He Zhang² Xuaner Zhang² Jianming Zhang² Zhixin Shu² Yilin Wang²
Zijun Wei² Shi Yan² HyunJoon Jung² Vishal M. Patel¹

¹Johns Hopkins University ²Adobe Inc.

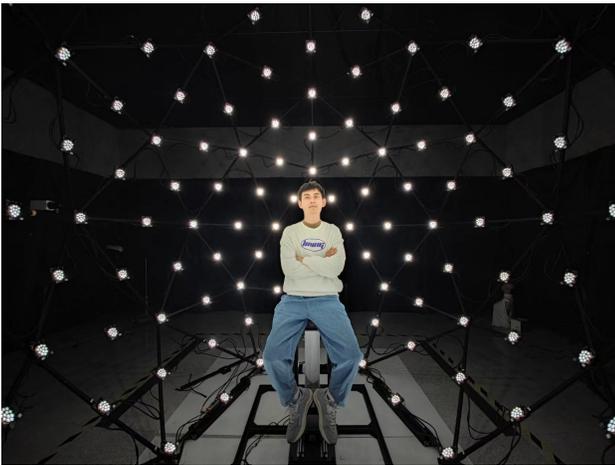


Figure 1. An illustration of our light stage system.

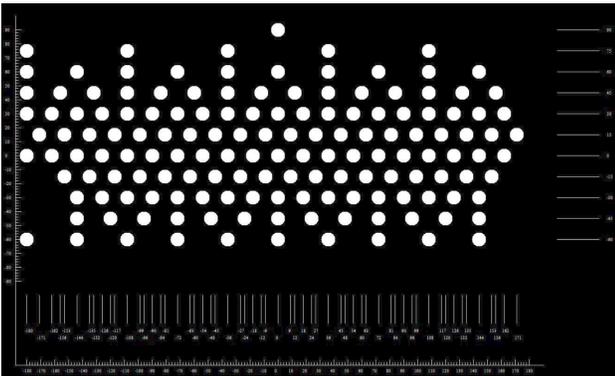


Figure 2. The LED light distribution of our light stage system. The image is shown in the panoramic format with x -axis denoting the longitude and y -axis denoting the latitude.

1. Interactive Application

The interactive application of LightPainter implements drawing tools for users to draw and edit their scribbles with ease. We provide a [recorded video](#) of using LightPainter to perform in-the-wild portrait relighting to demonstrate our interface and workflow.

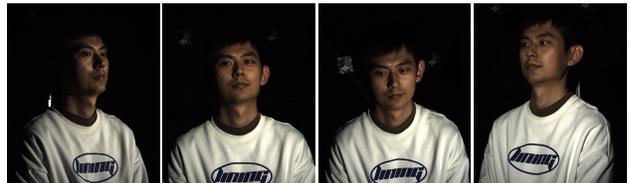


Figure 3. Examples of the captured OLAT images from 4 frontal-view cameras



Figure 4. Examples of skin-tone control. Given a source image (left), LightPainter can predict the albedo that respects user-specified skin color (bottom left).



Figure 5. Limitations of our method. While novice users can easily transfer the lighting patterns from a target image by simple scribbling, the scribble-based interface does not allow professional users to specify the exact shape/boundary of shadows. In LightPainter, these details are handled by the network to tolerate imperfect scribbles from novices. In addition, as shown in the relit image, our network also failed to produce non-Lambertian reflections in the subject's eyes.

2. Implementation Details

Unconditional Albedo Net. We also train an unconditional albedo net using the same architecture, which takes only the portrait image and its normal as input. If the skin color is not provided, our backend will automatically switch



Figure 6. Examples of relighting in-the-wild portraits from user scribbles



Figure 7. More visual results on images with eyeglasses, light hair color and three color lighting.

to this unconditional network for inference.

Network Architecture. In LightPainter, both Albedo Net and Shading Net adopt a U-Shaped structure. Specifically, Albedo Net is a standard UNet [3] and Shading Net has one additional non-local layer [5] and dilated convolution [7] in the bottleneck of both the encoder (shading completion branch) and the decoder. For both networks, we use 3 up-sampling layers in the encoder and 3 down-sampling layers in the decoder, resulting in 64-128-256-512 and 512-256-128-64 hidden channels respectively. All latent features in the bottleneck have 512 channels. The up-sampling operation is implemented using transposed convolution with a kernel size of 6×6 . The dilated convolutions have a dilation rate of 7×7 . And the kernel size for standard convolutions is set to 3×3 .

More Scribble Simulation Details. To compute shading maps, we follow Total Relighting [2] and use the prefilter-

ing technique for environment-map-based Phong shading. We pre-compute the diffuse and specular irradiance maps for each environment map by filtering it with cosine lobe functions corresponding to Lambertian and Phong specular BRDF (where the Phong exponent is set to 32). This bakes the *ambient* light term into the diffuse term. The *diffuse* and *specular* shading of the subject are obtained by looking up into these irradiance maps according to the normal or reflection directions, respectively. The final shading is: $0.85 \times \text{diffuse} + 0.15 \times \text{specular}$.

We provide more details about the quantization process in our scribble simulation algorithm. As discussed, we first convert the obtained shading into *Lab* space and quantize the luminance into multiple bins. We empirically set the number of bins to 25 to ensure the simulated scribbles are sufficiently coarse in terms of intensity so that can mimic causal user input. However, quantization inherently limits the network to be trained on only those 25 predefined bin levels and may result in poor generalization for other scrib-



Figure 8. Visual comparisons on environment-map-based portrait relighting. We compare our method with Total Relighting [2] (TR) and SIPR-W [6].

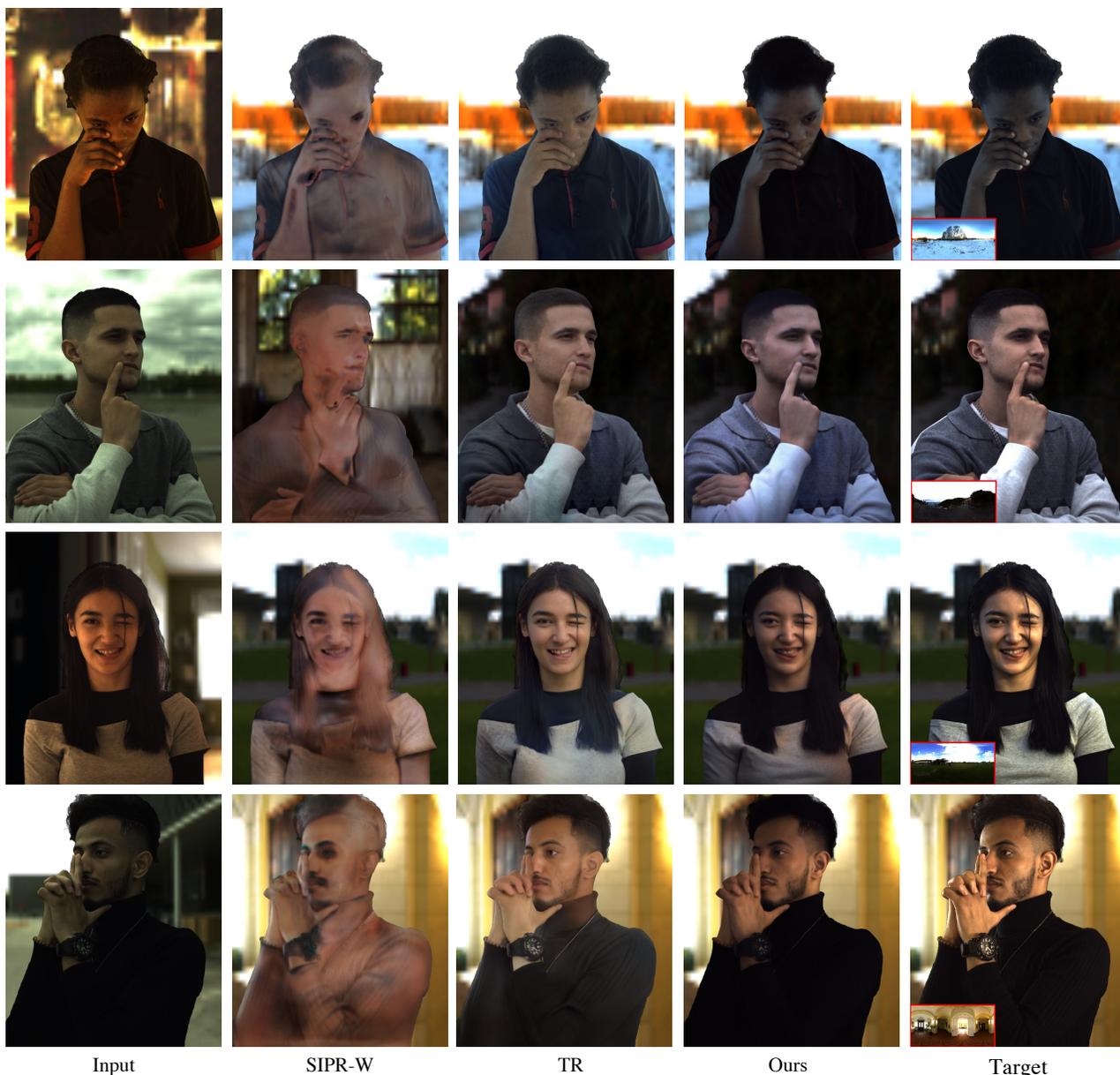
ble intensities. To address this, we propose to shift the bin levels by a random value p at each training step, where p is drawn from 0 to the bin width. With the random shift, the simulated scribbles can cover the full range of input intensity.

Light Stage Configurations. As discussed, we rely on the light stage [1] to prepare our training data. Specifically, we construct a light stage illumination system with a diameter of 3.6m and 160 LED lights. The system structure and the distribution of lights are illustrated in Figure 1 & 2. We

use the MER2-502-79U3M high-speed camera to record the reflectance field of the subject at 5 megapixel resolution and exposure time of 20ms. We provide some examples of the captured OLAT images from 4 frontal camera views in Figure 3.

3. More Study on Skin-tone Control

As discussed, LightPainter can generate an albedo image that respects the user-specified skin color. We demonstrate its effectiveness in terms of user control in Figure 4. As one can see, our approach can predict realistic albedo images



Input

SIPR-W

TR

Ours

Target

Figure 9. Visual comparisons on environment-map-based portrait relighting. We compare our method with Total Relighting [2] (TR) and SIPR-W [6].

that match the user’s intent.

4. Limitations

LightPainter is not without limitations. First, similar to previous relighting methods [2,4,8,9], our method is trained on the data rendered with image-based relighting [1]. This inherently limits the expressiveness of the model to the lighting patterns that can be represented by the environment maps, indicating it cannot handle lighting effects and cast shadows caused by occlusions. Second, similar to Total Relighting [2], it fails to generate specular reflections in the

eyes of the subjects (see Figure 5). This is because the eye regions only contribute to a small portion of the overall loss function. One possible solution is to add explicit supervision on the eye regions and we leave this extension for future work. Third, in a few examples, we found that some fine details of the hair region are smoothed out. We suspect it may be due to simply fine-tuning the off-the-shelf normal net is not sufficient to estimate the fine and dense geometry of hair strands, and it is also difficult to learn sophisticated light transport in hairs without explicit supervision. Finally, as discussed, by design our method is robust

to noisy and incomplete scribbles from novices, where the plausible completion of the lighting effects, such as sharpness of highlights and hardness of shadow boundary, are left to the network. Therefore, as shown in Figure 5, although our method is good at producing realistic light patterns that match the user’s desires, the scribble-based interface does not allow professional users to specify the lighting details, for example, the exact shape and boundary of shadows. We believe extending the current scribble-based interface to support more fine-grained control for professional users will be an interesting direction for future exploration.

5. More Visual Results

In Figure 6 & 7, we provide more examples of in-the-wild portrait relighting to better demonstrate the effectiveness of LightPainter. In Figure 8 & 9, we also provided additional qualitative comparisons with state-of-the-art methods [2, 6] on environment-map-based portrait relighting. Note that we did not compare SIPR-W [6] on this task in the main paper as it yields relit results with strong artifacts (because SIPR-W [6] is only trained for face images). These results are also included here.

References

- [1] 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.
- [2] 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 Transactions on Graphics*, 40(4):1–21, 2021.
- [3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [4] 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 Transactions on Graphics*, 38(4):79–1, 2019.
- [5] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7794–7803, 2018.
- [6] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. Single image portrait relighting via explicit multiple reflectance channel modeling. *ACM Transactions on Graphics*, 39(6):1–13, 2020.
- [7] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In Yoshua Bengio and Yann LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.
- [8] Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. Neural video portrait relighting in real-time via consistency modeling. In *IEEE International Conference on Computer Vision*, pages 802–812, 2021.
- [9] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. Neural light transport for relighting and view synthesis. *ACM Transactions on Graphics*, 40(1):1–17, 2021.