Computational Photography Lab.

Find the project web page here:
https://yaksoy.github.io/PhysicalRelighting/

Physically Controllable Relighting of Photographs Supplementary Material

CHRIS CAREAGA and YAĞIZ AKSOY, Simon Fraser University, Canada

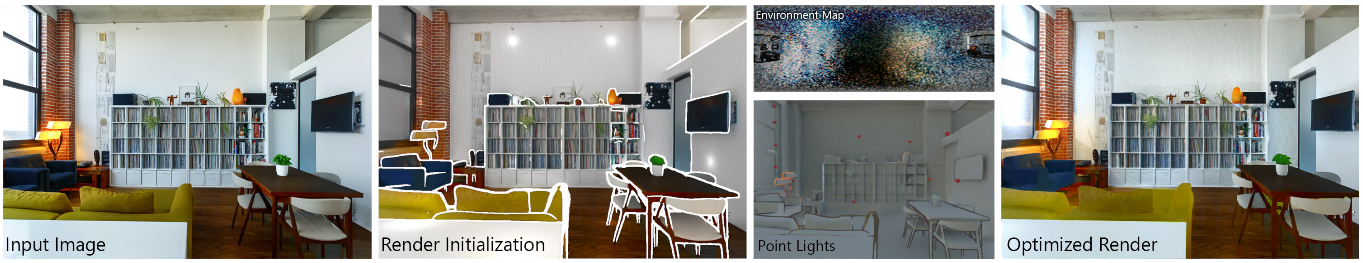

Fig. 1. Visualization of our lighting optimization process. For a given input image, we first extract the diffuse-only image using CID. We then initialize an environment map and a grid of point lights in Mitsuba. We optimize the values of both the environment map and the point lights. The third column shows the optimized environment map, and point light positions visualized as red spheres. The last column shows our rendered version of the scene after optimization. Image from Unsplash by Nathan Van Egmond

ACM Reference Format:
Chris Careaga and Yağız Aksoy. 2025. Physically Controllable Relighting of Photographs Supplementary Material. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '25), August 10–14, 2025, Vancouver, BC, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3721238.3730666

1 SUPPLEMENTARY RESULTS AND FIGURES

In Figure 1, we show an example scene with visualized lighting parameters from our differentiable optimization process. We provide example inputs and outputs for our network in Figures 2 and 3 and a set of training data pairs in Figure 4. Finally, we provide results from our two numerical experiments.

1.1 Quantitative Comparisons

In order to evaluate the efficacy of both our lighting optimization and neural renderer, we perform a quantitative evaluation using the BigTime dataset [Li and Snavely 2018] against two prior works, ICLight [Zhang et al. 2025] and RGB⇔X [Zeng et al. 2024]. The BigTime dataset consists of time-lapse sequences depicting various scenes under time-varying lighting conditions. This means that pairs of images can be sampled from each scene to represent a source and target relighting pair. For our experiment, we sampled 2 image pairs from each scene in the BigTime dataset, resulting in 424 pairs. For our method, we extract our PBR representation from the source image and run our lighting optimization to extract the lighting configuration of the target image. We then run the optimized scene representation through our neural renderer to compare it to the ground truth target image. For RGB⇔X, we feed the intrinsic components from the source image and the irradiance from the target image to their image generator. For ICLight, we input the source image and supply our optimized shading as the lighting condition. For both competing methods, we compute scale-invariant metrics between the target image and the generated output to account for any scale ambiguity. The results of the experiment are summarized in Table 1. We observe that RGB⇔X generally fails to reproduce the expected appearance of the scene from its intrinsic components, even when provided with the irradiance extracted from the target image. While ICLight can often apply the novel lighting condition, it typically fails to remove the original lighting from the input.

Table 1. We perform two numerical evaluation of our method: 1) a quantitative experiment using time-lapse data from the BigTime dataset, 2) a user study comparing subjective performance on environment map relighting. We evaluate our method against two diffusion-based relighting methods; RGB⇔X [Zeng et al. 2024] and ICLight [Zhang et al. 2025]

	BigTime		**User Study**	
Method	MSE ↓	SSIM ↑	Avg. Rank ↓	Rank 1 % ↑
RGB⇔X	0.0322	0.544	2.54	6.6
ICLight	0.0366	0.544	2.12	18.0
Ours	**0.0225**	**0.643**	**1.33**	**75.3**

1.2 User Study

For the user study, we collected a set of environment maps and in-the-wild photographs to create 20 different examples. We condition both competing methods using our rendered shading and provide a prompt describing both the scene and the illuminating environment. Each participant is asked to rank the 3 results from best to worst based on two criteria: 1) accuracy of lighting with respect to the environment map and 2) photorealism/naturalness of the result. Each of the twenty examples showed the input image,

SIGGRAPH Conference Papers '25, August 10–14, 2025, Vancouver, BC, Canada

ACM ISBN 979-8-4007-1540-2/2025/08…$15.00
https://doi.org/10.1145/3721238.3730666

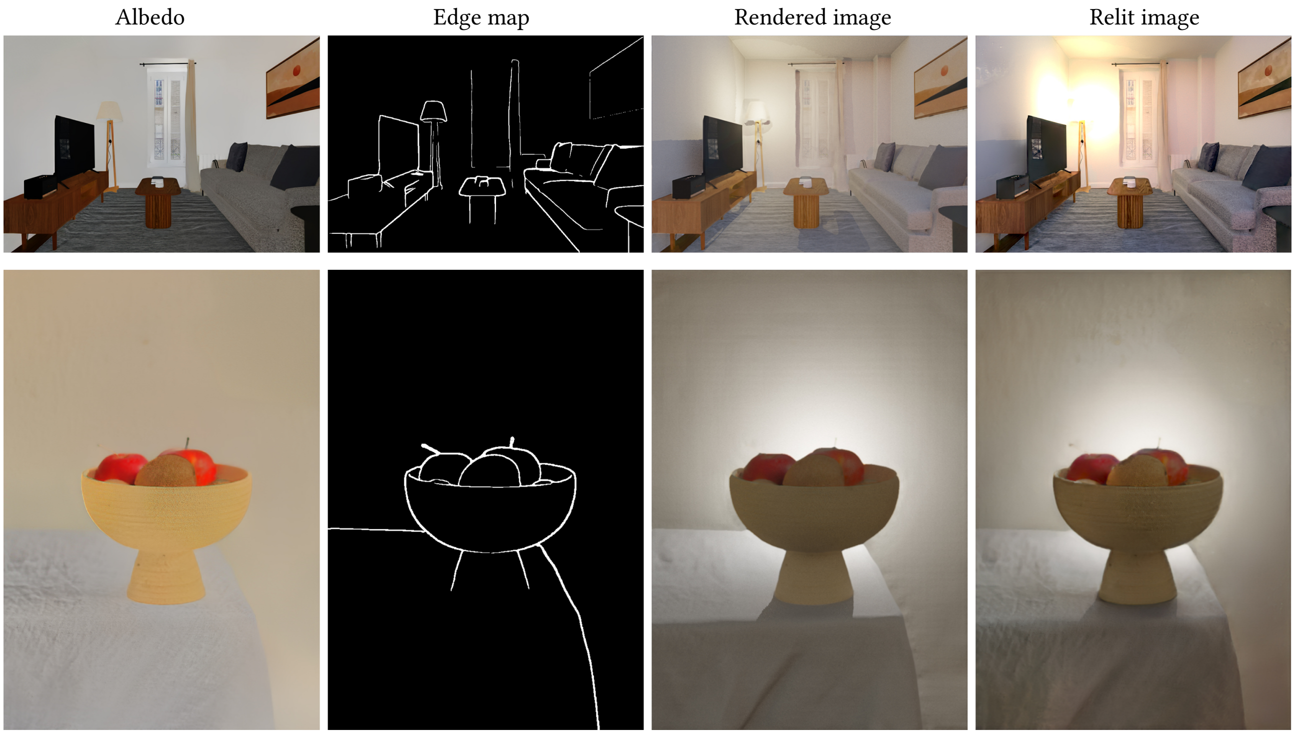

Fig. 2. Example inputs and estimated relighting from our neural render for in-the-wild examples
Images from Unsplash by Md. Shafaat Hossain (living room) and Suzanne Boureau (fruit)

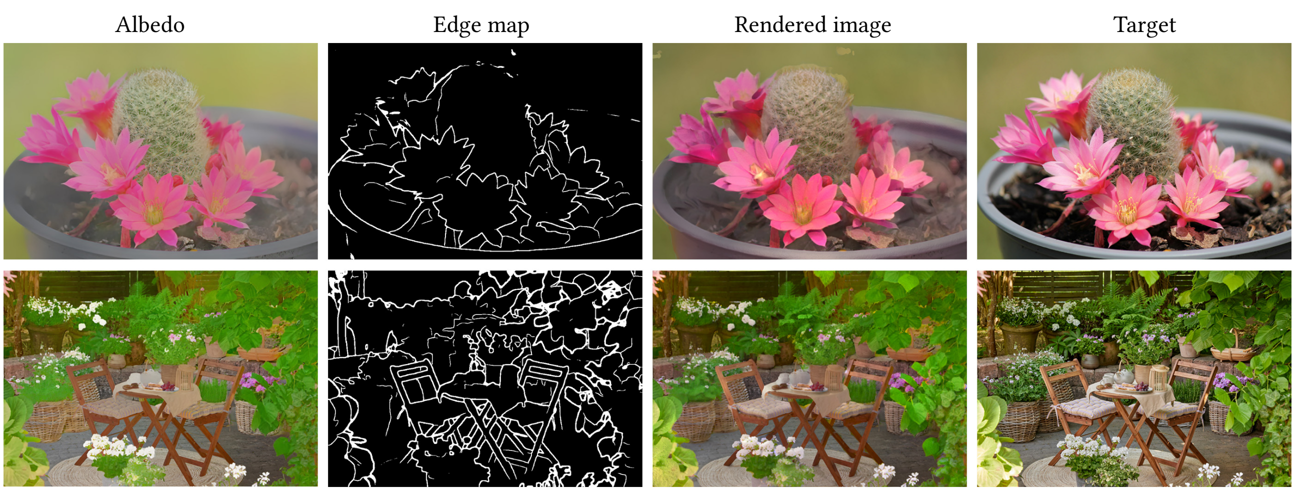

Fig. 3. Example training inputs and ground-truth targets. The input rendered image is created using our differentiable rendering optimization.
Images from Unsplash by Tyler Moore and Charlotte Cowell

the environment map, and the result of each method in a randomized order. We provide these image collages, as well as the method ordering in the supplementary material. We fielded responses from 45 participants. The average ranking for each method is given in Table 1. Among all responses, ours was ranked as the first choice 75% of the time, while ICLight was chosen as the top choice in 18% of comparisons.

REFERENCES

Zhengqi Li and Noah Snavely. 2018. Learning Intrinsic Image Decomposition from Watching the World. In *Computer Vision and Pattern Recognition (CVPR)*.

Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yiwei Hu Yannick Hold-Geoffroy, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. 2024. RGB↔X: Image Decomposition and Synthesis Using Material- and Lighting-aware Diffusion Models. In *Proc. SIGGRAPH*.

Fig. 4. Example ground truth and optimized rendered images. The left side represents the original image, and the right is the result of our lighting optimization. Images from Unsplash by Buğra Şıkel (pillars), Clay Banks (wood wall), Kate Darmody (lamp), Huy Phan (blue couch), Lisa Anna (blue wall), Mircea Solomiea (crosswalk), mos design (japan), Raimond Klavins (castle), Prateek Katyal (books)

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2025. Scaling In-the-Wild Training for Diffusion-based Illumination Harmonization and Editing by Imposing Consistent Light Transport. In *Proc. ICLR*.