ReNeRF: Relightable Neural Radiance Fields with Nearfield Lighting (Supplementary Material)

Yingyan Xu[1,2] Gaspard Zoss[2] Prashanth Chandran[2]
Markus Gross[1,2] Derek Bradley[2] Paulo Gotardo[2*]
[1]ETH Zürich [2]DisneyResearch|Studios
{yingyan.xu,grossm}@inf.ethz.ch, gotardop@gmail.com
{gaspard.zoss,prashanth.chandran,derek.bradley}@disneyresearch.com

1. ReNeRF Architecture

While each input 3D point $\mathbf{x}_i$ is position-encoded with 10 frequency bands, the corresponding input lighting direction ω_i is position-encoded with a 5th-order spherical harmonics (SH) basis. Both inputs are fed into the ReNeRF MLP, whose architecture is illustrated in Fig. 1. The initial (top) eight layers and the density output branch follow the original NeRF design [6]. The intermediate NeRF features $\mathbf{f}_i$ are concatenated with the SH-encoded input light direction and fed into the OLAT MLP. The coefficient matrix $\mathbf{C}$ of the spherical codebook is implicitly modeled within the first layer of the OLAT MLP. This MLP comprises a sequence of 4 Linear-ReLU layers whose output dimensions are indicated in the figure. Finally the output layer produce the resulting diffuse and specular layers of color separately. The output specular color is *omnidirectional* and encoded using 3rd-order spherical harmonics to enforce smoothness of specular reflection *across viewing directions* [7]. Note that, without polarization filters on the lights and cameras, the constant (0th-order) SH coefficient remains ambiguous and cannot be separated from the diffuse color; for this reason, we set this term to 0 and output only the remaining 8 SH coefficients for our 3rd-order SH specular model.

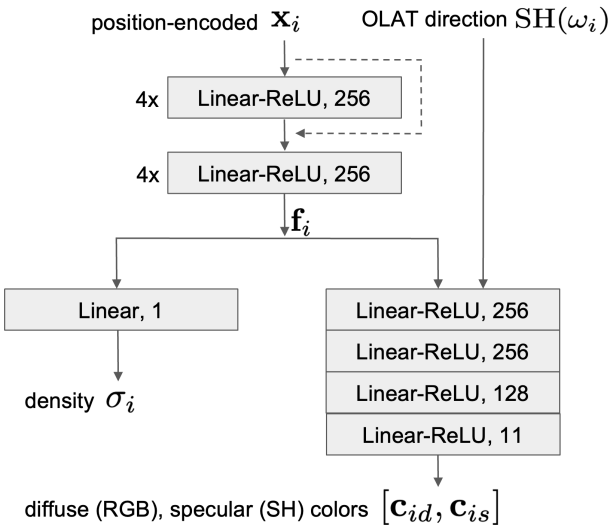

Figure 1. The architecture of the ReNeRF MLP comprises 9 standard NeRF layers (top) to model the density σ_i and 4 MLP layers (bottom-right) to model the OLAT MLP, which outputs diffuse and omnidirectional specular colors separately.

2. Distant *vs* Nearfield Codebook Supervision

As mentioned in the main text, the baseline version of our method, dubbed ReNeRF$_{(D)}$, makes the usual assumption of distant lighting. This means that, during training, every sampled point $\mathbf{x}_i$ is assigned a much smaller number of OLAT basis directions ω_i, which are spatially constant. In other words, this baseline method can only inject supervision on 32 locations over its learned OLAT spherical codebook. In contrast, by explicitly modeling nearfield lighting, the proposed ReNeRF model assigns a distinct set of N_p OLAT directions for each sampled point $\mathbf{x}_i$, thus injecting supervision over a larger area of the spherical codebook during training. The density of sampled OLAT basis directions $\omega_i(\mathbf{x}_i, \mathbf{p})$, over the frontal hemisphere in latlong format, is visualized as heat maps in Fig. 2, for two values of N_p (number of point sources $\mathbf{p}$ per LED-bar area light). As a result, ReNeRF can better interpolate and extrapolate novel lighting directions at test time, including lighting from point sources that are closer to the captured objects. Note that, in Fig. 2, the OLAT directions sampled by ReNeRF$_{(D)}$ (red crosses) are not always located at the center of the samples taken by the proposed nearfield model, ReNeRF. This is due to another problem of directional (distant lighting) models such as ReNeRF$_{(D)}$: they use lighting directions relative to the center of the mirror sphere used to calibrate the environment map. In contrast, the proposed ReNeRF derives the

*Now at Google

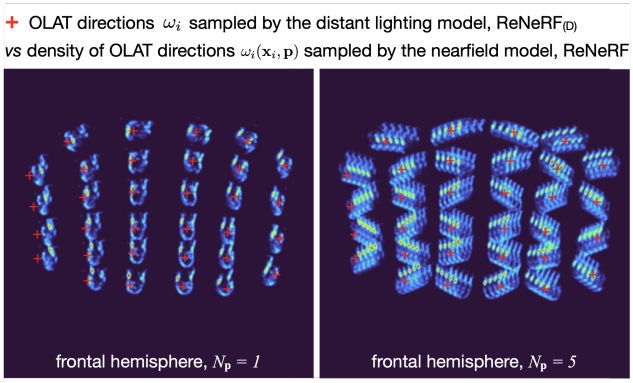

Figure 2. Sampled OLAT directions ω_i for a batch of point samples $\mathbf{x}_i$. Our baseline model ReNeRF$_{(D)}$, assuming distant lighting, always samples the same set of 32 directions (red crosses). In contrast, with our nearfield ReNeRF model, each point $\mathbf{x}_i$ has a number N_p of distinct OLAT basis directions, thus injecting supervision over a broader area of the learned spherical codebook (red colors indicate greater density than blue colors).

lighting direction (OLAT basis) using the calibrated 3D position of the light source and the actual 3D scene point being evaluated by the network. This fact becomes even more important when the OLAT sources are very near the captured scene.

3. Results

ReNeRF *vs* mesh-based IBRL. When modeling challenging geometry like that of our furry dog, traditional mesh-based approaches provide inaccurate, shrink-wrapped surfaces, Fig. 4 (bottom). In contrast, NeRF-based methods are known to provide much better depth estimates in such cases, Fig. 4 (top). This fact illustrates a key advantage of extending IBRL from 2D to 3D using a NeRF, rather than using the 2D texture space of a 3D mesh [8, 3].

Rendering with near point lights. When rendering with a point light that is moving very close towards the captured scene, the inverse-square-root-distance factor causes the light intensity to increase rapidly as distance decreases, thus saturating the rendered images. To better visualize the nearfield lighting effect, we normalize the maximum image intensity over the sequence of renderings (keep it constant) in Figs. 4 and 7 (main text) and in the supplementary video. Note however, that this effect is optional.

Rendering with LatLong environment maps. The ReNeRF models trained with our in-studio lighting setup only learn a transport function for incoming, *frontal* light. Thus, when relighting with a latlong environment map, we first need to mask out the back hemisphere (left half) of the latlong, Fig. 6, whose light transport the ReNeRF model cannot extrapolate well. We show in Fig. 7 some failure cases where the main light sources are from the back side. For rendering efficiency, we also downsample the latlong into 144 directions uniformly spaced on the frontal hemisphere (right half), and additionally drop dim lighting directions (lowest 10% in intensity) that contribute little to the rendered image appearance. As most of the environment maps that we use have only one dominant light source (Fig. 6), using more directions does not give noticeable improvement in image quality. During rendering, for each sampled point $\mathbf{x}_i$, we evaluate the NeRF MLP only once and the OLAT MLP N times, where N is the number of directions of incoming (distant) lighting. We compute a weighted sum of the output colors, with weights given by the intensity of the incoming light directions. We show two additional environment relighting results on real datasets in Fig. 5 as continuation of Fig. 5 in the main text. For more environment relighting results, please refer to the supplementary video.

Quantitative relighting evaluation. To compare relightable NeRF models, we trained ReNeRF and the baseline methods on a reduced dataset including only half of the 32 area-OLAT sources, holding out for validation the images captured under even-numbered LED-bars shown in Fig. 8. For each of the five datasets in Fig. 5 (main text), we trained ReNeRF, ReNeRF$_{(D)}$, the distant lighting model of Li *et al*. [4], NeRF-SHL (without the specialized eye component), NeRFactor [9] and a simplified NeRF-W [5] with 64D and 128D lighting code. ReNeRF, ReNeRF$_{(D)}$, NeRF-SHL, and NeRF-W models are trained for 120K iterations. We run 10K optimization iterations at test time for NeRF-W models. For NeRFactor, instead of assuming unknown illumination as in the original paper, we use the calibrated light positions. We use an MLP to predict spatially-varying roughness for a microfacet BRDF in stead of using a data-driven BRDF, which statistically gives similar results as indicated in the original NeRFactor paper. We compute up-to-scale error (PSNR) values in linear RGB space for Table 1 (main text). Example validation images and the corresponding renderings and rendering errors obtained by these methods are shown in Fig. 11 for real datasets (studio area-OLATs lighting) and in Fig. 3 (synthetic environment lighting). On average, ReNeRF does a better job at generalizing to the lighting of the validation OLATs. NeRF-SHL often cannot reproduce sharp shadows and tends to render with more uniform lighting. In addition, NeRF-SHL factors out diffuse albedo in the output, making it difficult to model subsurface scattering, e.g., it shows larger errors in the wax candles in Fig. 11. Also, both NeRF-SHL and ReNeRF$_{(D)}$ model distant lighting and cannot render nearfield lighting effects as ReNeRF does. NeRFactor produces more visual artifacts due to the imperfect pretrained NeRF surfaces. While NeRF-W models achieve good PSNR scores. They

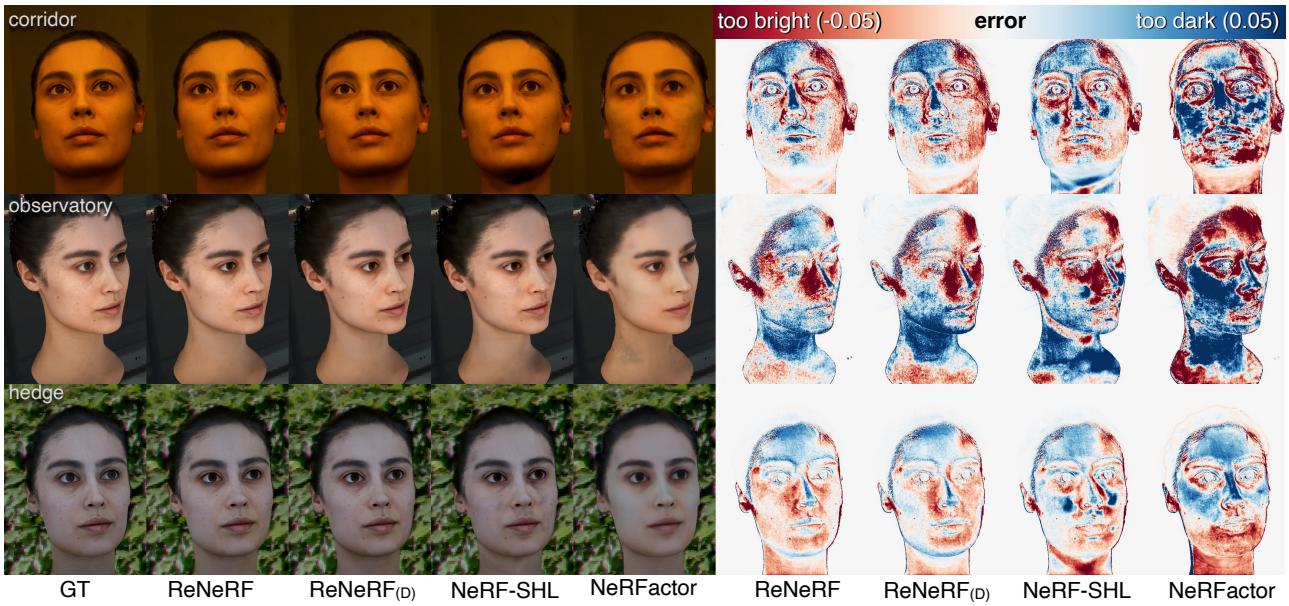

Figure 3. Example validation images under environment lighting on a synthetic dataset [1], renderings by four relightable NeRF methods, and their respective color-coded (signed) rendering error, with $error = real - rendered$ being negative (red) for pixels rendered too bright.

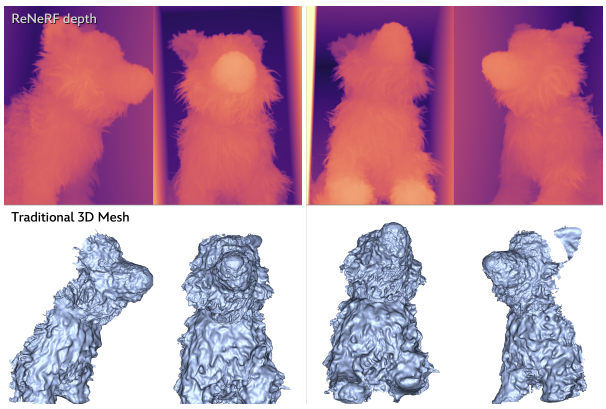

Figure 4. For scenes with thin 3D structures like the fur of our toy dog, a NeRF-based IBRL approach can leverage much better geometry estimates than a mesh-based IBRL approach; in such case, mesh-based methods [2] generate highly inaccurate surfaces.

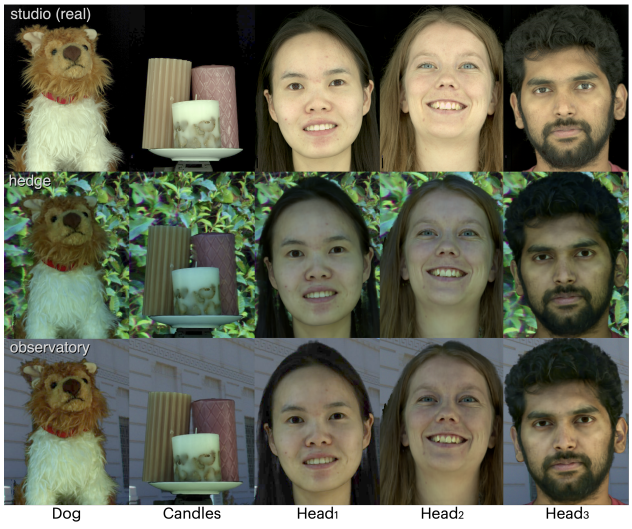

Figure 5. ReNeRF models rendered within two new, distant lighting environments encoded by LatLong maps (shown on the background). The real full-ON image (top) is also shown for reference.

do not provide an explicit and intuitive mechanism to control lighting. Moreover, as shown in Fig. 9, NeRF-W fails to *extrapolate* while ReNeRF produces reasonable extrapolation. Also, the specular reflections on the eyes of NeRF-W tend to fade away and reappear at shifted locations during *interpolation*.

Training with Sparser Lights. In contrast with models requiring dense light stage setups [3, 8], a ReNeRF can be trained using a small number of area lights. It shows reasonably good performance even when trained with only half of our 32 OLATs. As shown in Table 1, the model trained on 16 OLATs, $\text{ReNeRF}_{(16)}$, shows just slightly lower PSNRs on the validation OLAT images for the Toy Dog and Head datasets, compared with the best model, $\text{ReNeRF}_{(32)}$. The Candles scene is harder to model due to highly reflective surfaces. Example validation images and renders are shown in Fig. 12. $\text{ReNeRF}_{(16)}$ still closely matches the ground truth, being capable of modeling some challenging lighting arriving at grazing angles, while the main degradation is in

Figure 6. LatLong environment maps showing the LED bars in our in-studio capture setup, in full-ON condition (top), and the distant lighting used for relighting the ReNeRF models in Fig. 6 of the main manuscript and in the supplementary video.

specular highlights and shadows. Also, when animating a smoothly moving point light, it struggles to produce sharp specular reflection in the eyes, leading to ghosting, as shown in Fig. 10.

Figure 7. Failure relighting results using full LatLong environment maps where the main light sources are from the back side.

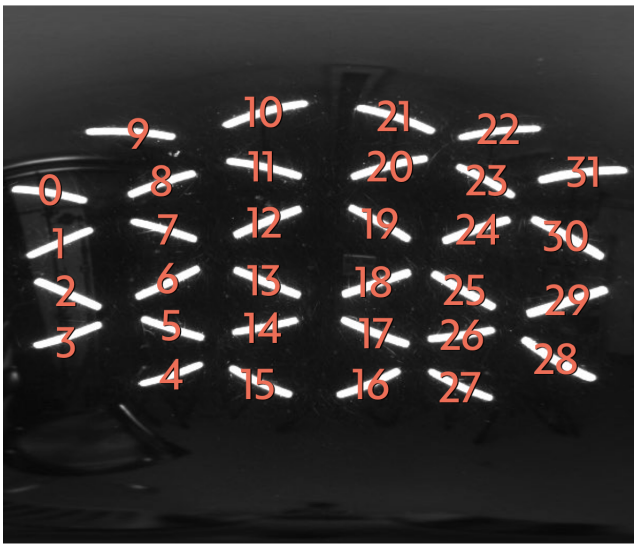

Figure 8. Frontal hemisphere of full-ON incoming light in LatLong format. For quantitative evaluation and comparison, we hold out validation images from even-numbered LED bars and train ReNeRF and the baseline methods using only images taken under odd-numbered area OLATs.

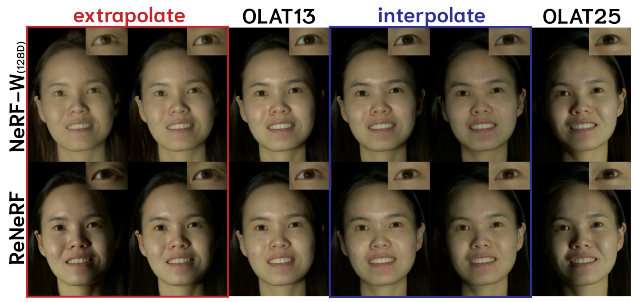

Figure 9. Interpolation/Extrapolation: ReNeRF vs. NeRF-W.

Table 1. PSNRs on validation OLATs.

Method	Dog	Candles	Head_1	Head_2	Head_3
$\text{ReNeRF}_{(32)}$	33.59	34.74	38.04	35.85	40.58
$\text{ReNeRF}_{(16)}$	31.17	29.49	34.04	33.04	38.14

References

[1] Animatable digital double of louise by eisko©. http://www.eisko.com. 3

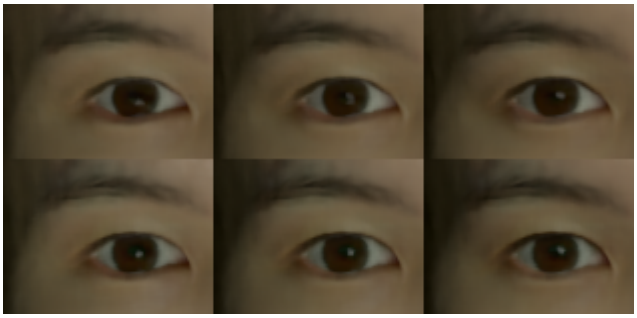

Figure 10. Reflections on the eyes, caused by a moving point light. When trained with half of the OLATs, the resulting $\text{ReNeRF}_{(16)}$ model produces noticeable ghosting (top row) in comparison with the sharper reflections rendered by $\text{ReNeRF}_{(32)}$ (bottom row).

[2] Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. High-quality single-shot capture of facial geometry. *ACM TOG*, 29(3), 2010. 3

[3] Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. Deep relightable appearance models for animatable faces. *ACM TOG*, 40(4), 2021. 2, 3

[4] Gengyan Li, Abhimitra Meka, Franziska Mueller, Marcel C. Buehler, Otmar Hilliges, and Thabo Beeler. EyeNeRF: A hybrid representation for photorealistic synthesis, animation and relighting of human eyes. *ACM TOG (Proc. Siggraph)*, 41(4), 2022. 2

[5] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *CVPR*, 2021. 2

[6] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1

[7] Daoye Wang, Prashanth Chandran, Gaspard Zoss, Paulo Bradley, and Paulo Gotardo. MoRF: Morphable radiance fields for multiview neural head modeling. In *Proc. SIGGRAPH*, 2022. 1

[8] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, Jonathan T. Barron, Ravi Ramamoorthi, and William T. Freeman. Neural light transport for relighting and view synthesis. *ACM TOG*, 40(1), 2021. 2, 3

[9] Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. NeRFactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM TOG (Proc. Siggraph Asia)*, 40(6), 2021. 2

Figure 11. Example validation images left out of training, renderings by four relightable NeRF methods (trained on 16 OLATs), and their respective color-coded (signed) rendering error, with $error = real - rendered$ being negative (red) for pixels rendered too bright.

Figure 12. Example validation images comparing the rendering quality when training on 32 *versus* 16 area OLATs.