L-Tracing: Fast Light Visibility Estimation on Neural Surfaces by Sphere Tracing

Ziyu Chen[1†], Chenjing Ding[2,3], Jianfei Guo[3], Dongliang Wang[2,3], Yikang Li[2,3], Xuan Xiao[4], Wei Wu[2,3], and Li Song[1‡]

[1] Department of Electronic Engineering, Shanghai Jiao Tong University
[2] SenseTime Research, {dingchenjing,wangdongliang,wuwei}@senseauto.com
[3] Shanghai AI Laboratory, {guojianfei,liyikang}@pjlab.org.cn
[4] Department of Mechanical Engineering, Tsinghua University
x-xiao20@mails.tsinghua.edu.cn

1 Additional Proof

In this section, we prove the linear convergence of L-Tracing. As shown in Fig. 3 of the main body, the signed distance of the k_{th} tracing point $\mathbf{O}_k$ is f_k, the signed distance between $\mathbf{O}_k$ and the surface point is $\mathbf{Q}$ is ϵ_k, according to the defination of the signed distance, we know $|f_k| \leq |\epsilon_k|$, besides, f_k and ϵ_k have the same sign. If the ray is encountered with a convex surface, we have $|f_k| \geq |\epsilon_k sin\theta|$, if the surface is concave, we have $|f_k| \geq |\epsilon_k sin\alpha|$, Since our tracing point always starts from outside of the surface, we have $\epsilon_0 > 0$. Thus we formulate our problem as follows:

For a series ϵ_k, where $k \in \mathbb{N}$, the equation $\epsilon_{k+1} = \epsilon_k - f_k$ is satisfied, besides, $|\epsilon_k| \geq |f_k| \geq |\epsilon_k sin\phi|$, ϕ equals to θ for the convex surface and α for the concave surface, $\phi \in (0, 90°]$, if $\epsilon_0 > 0$, f_k and ϵ_k have the same sign. Prove: the series ϵ_k converges to 0 linearly.

Firstly, we prove series ϵ_k is convergent by applying mathematical induction, when $k = 0$, we have:

$$\epsilon_0 > 0 \tag{1}$$

If $\epsilon_k > 0$, we have $\epsilon_k \geq f_k > 0$, thus:

$$\epsilon_{k+1} = \epsilon_k - f_k \geq 0 \tag{2}$$

According to mathematical induction, we derive that: for $\forall k \in \mathbb{N}$, $\epsilon_k \geq 0$ is satisfied. Since f_k have the same sign with ϵ_k, for $\forall k \in \mathbb{N}$, $f_k \geq 0$ is satisfied, thus:

$$\epsilon_{k+1} - \epsilon_k = -f_k \leq 0 \tag{3}$$

From Eqn. 3, we know ϵ_k is a monotonic decreasing series, according to monotone convergence theorem, series ϵ_k is convergent.

[†] The work is done when Ziyu Chen <ziyu.sjtu@gmail.com> is an intern at SenseTime.
[‡] Corresponding author: Li Song. <song_li@sjtu.edu.cn>

Secondly, we prove series ϵ_k converges to 0. Since $f_k \geq 0$ and $\epsilon_k sin\phi \geq 0$, we get $f_k \geq \epsilon_k sin\phi$, as:

$$0 \leq \epsilon_{k+1} = \epsilon_k - f_k \leq \epsilon_k(1 - sin\phi) \tag{4}$$

from Eqn. 4 we get:

$$0 \leq \epsilon_n \leq (1 - sin\phi)^n \epsilon_0 \tag{5}$$

for $\forall q > 0$, we find:

$$N = \lceil \frac{log(\frac{q}{\epsilon_0})}{log(1 - sin\phi)} \rceil, \quad N \in \mathbb{N} \tag{6}$$

when $n > N$, we get:

$$|\epsilon_n - 0| \leq (1 - sin\phi)^n \epsilon_0 < q \tag{7}$$

Thus the series ϵ_k converges to 0. Thirdly, we prove that the convergence is in a linear speed, the limitation is described as:

$$\lim_{k \to \infty} \frac{|\epsilon_{k+1} - 0|}{|\epsilon_k - 0|} = \frac{|\epsilon_k - f_k|}{|\epsilon_k|} \tag{8}$$

since $0 \leq f_k \leq \epsilon_k$, we get:

$$\lim_{k \to \infty} \frac{|\epsilon_{k+1} - 0|}{|\epsilon_k - 0|} = 1 - \frac{f_k}{\epsilon_k} \tag{9}$$

as $k \to \infty$, we have: $\epsilon_k \to 0$ and $f_k \to 0$, the local surface is regarded as a plane if k reaches the limitation. Since ϕ is the angle between the surface and the ray, we get:

$$\lim_{k \to \infty} \frac{\epsilon_{k+1}}{\epsilon_k} = 1 - sin\phi < 1 \tag{10}$$

Now we complete the proof, the series ϵ_k converges to 0 linearly.

2 Ablation Study

In Fig. 2, we visualize the light visibility estimated by L-Tracing in different iterations(20, 40, 80). Similar to Section 5.1 of the main body, we visualize the "mean" light visibility of the observed object's surface to show the average intensity of the light reaching the surface, we also list the light visibility under the single point light with no ambient illumination(OLAT). Results show that the light visibility images estimated under different tracing iterations are close to each other, except for slight differences that we visualize in Fig. 1. Since these differences induce a small drop in the overall quality of estimated light visibility and recovered surface material properties, we set the L-Tracing iterations to 20 when comparing with related approaches.

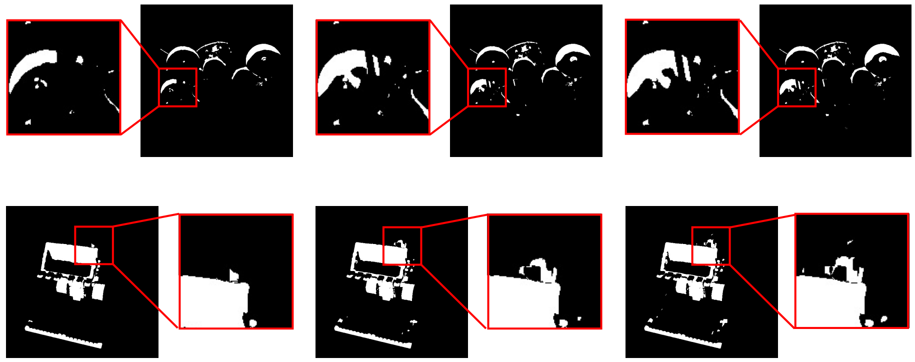

Fig. 1. Details of Light Visibility. We select one region on the light visibility images of "Drums OLAT2" and "Lego OLAT2". Although the light visibility estimated in 20 iterations losses some subtle details compared to those from 40 and 80 iterations, it has very little impact on the performance of our framework.

3 Implementation details

Our framework is implemented in PyTorch [5], there are two training stages in our framework: stage one is the training of neural implicit surfaces, we optimize the MLPs that model the signed distance function and spatial color. In stage two, we do reflectance and illumination decomposition on the trained neural surface with L-Tracing as the light visibility estimation method, we optimize the Albedo MLP and Material Latent MLP. The BRDF MLP is pre-trained on MERL dataset [2] and fixed during stage two.

3.1 Network Architecture

For stage one, we refer to [6] for network architecture designing, the neural surface MLP f that models the SDF consists of 8 layers with the layer size of 256, we use a skip connection that connects the input and the fourth layer's output. The positional encoding frequency of the input coordinates is 8. For stage two, all introduced MLPs contain 4 layers with layer size of 128, a skip connection connects the input to the third layer. Similar to [7], the final output is activated by the Sigmoid function in Albedo MLP, and the Softplus(β=1) function in Material Latent MLP. The encoding frequency for the spatial location is 10 in Albedo MLP and Material Latent MLP. According to [4, 7],we constrain the albedo to range $[0.03, 0.8]$, making it useful for learning reflectance. Our environment illumination is in the resolution of 16×32, each light source includes 3 parameters learning the RGB channels, thus the illumination is modeled by trainable parameters with the tensor size of 512×3.

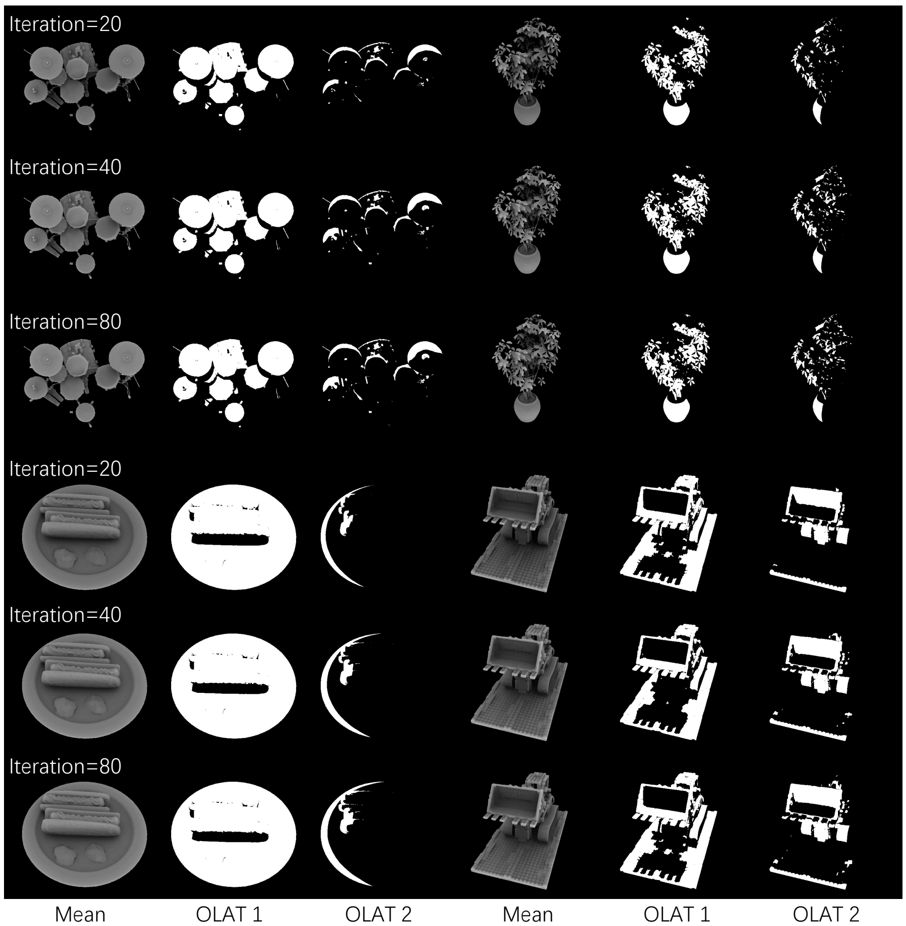

Fig. 2. Visual Results of Ablation Study. We visualize the light visibility estimated by L-Tracing in different tracing iterations. The light visibility images estimated in 20 iterations are close the those in 40 and 80 iterations, indicating that it rarely needs 80 iterations to compute accurate light visibility.

3.2 Training Details

In both stages, the resolution of the training images is 512 × 512. For shape reconstruction, we sample a batch of rays with the batch size of 1024 in each iteration, the model is optimized for 150k iterations, we linearly warm up the learning rate from 0 to 5×10^{-4} in the first 5k iterations, then use cosine decay scheduler controlling it to 5×10^{-5} in the rest iterations. For reflectance factorization, the training iterations is 5k and the batch size of the sampled rays is 2048. The learning rate is 5×10^{-3} during all iterations. L-Tracing is applied for light visibility estimation in stage two, we set the tracing iterations to 20. We test our framework on one Nvidia GeForce RTX 3090, it takes 8 hours for the training of shape reconstruction and takes only 20-30 minutes for the training of reflectance factorization.

4 Additional Results

In Tab. 1, we report the metrics on novel view synthesis quality for the four scenes in NeRFactor's synthetic dataset. We also report the albedo estimation quality in Tab. 2 and the HDR relighting image quality in Tab. 3 for the same scenes. Each scene of the four scenes includes 8 validation views, among them each view includes one ground truth albedo and 8 ground truth HDR relighting images. The additional visualization of estimated light visibility of "Hotdog" and "Lego" is shown in Fig. 4. We visualize the recovered albedo of "Hotdog" and "Lego" in Fig. 3, and visualize the relighting images of "Drums" and "Ficus" in Fig. 5. As for real-world datasets, we did experiments on DTU MVS datasets [1] and real scene pinecone(captured by authors of [3]) , the results are shown in Fig. 6.

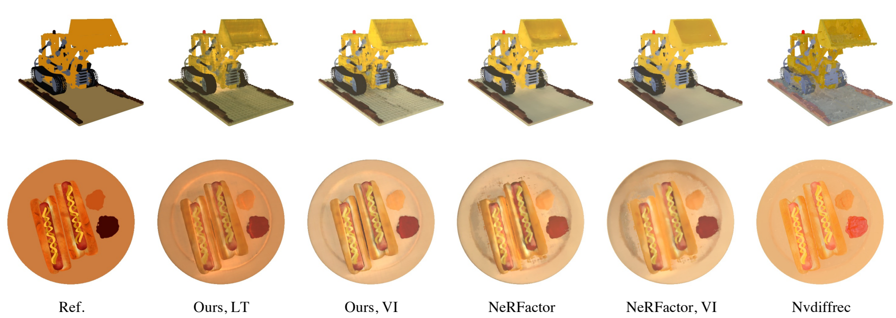

Fig. 3. Albedo Recovery. We visualize the recovered albedo of "Lego" and "Hotdog". The albedo from ours and NeRFactor are rescaled to eliminate the scale ambiguity. The results of Nvdiffrec are provided by Nvdiffrec's [4] authors.

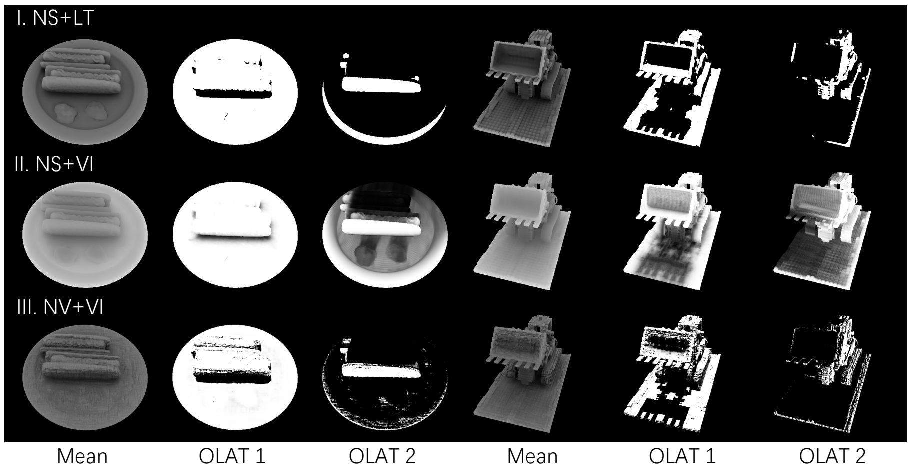

Fig. 4. Light Visibility Visualization. The light visibility is estimated on the NeRFactor synthetic dataset. The color denotes the lighting intensity on the surface. "Mean" denotes the mean light visibility across all light sources, while "OLAT1" and "OLAT2" denotes the light visibility under two different single light source points.

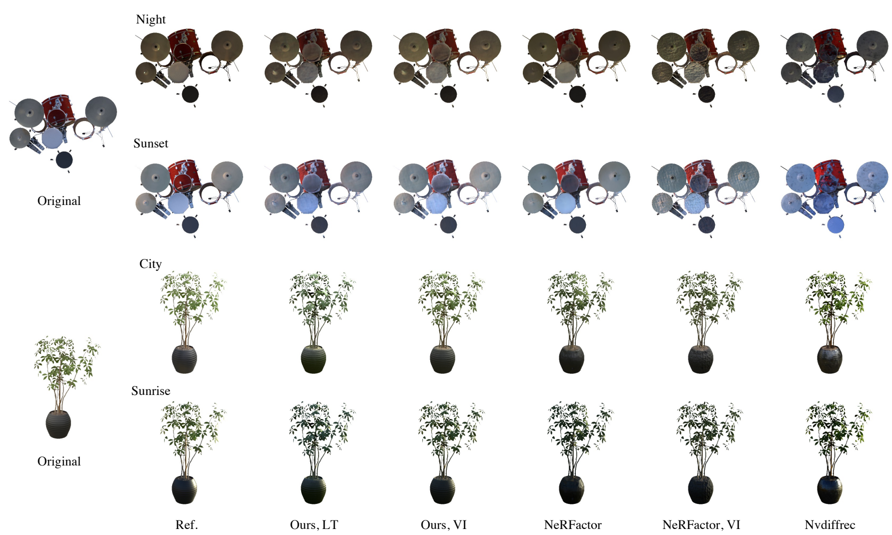

Fig. 5. Novel View Relighting. We visualize the relighting results of "Drums" and "Ficus". The images are produced by rendering the learned surface and reflectance with novel HDR light probes. The results of Nvdiffrec are provided by Nvdiffrec's [4] authors.

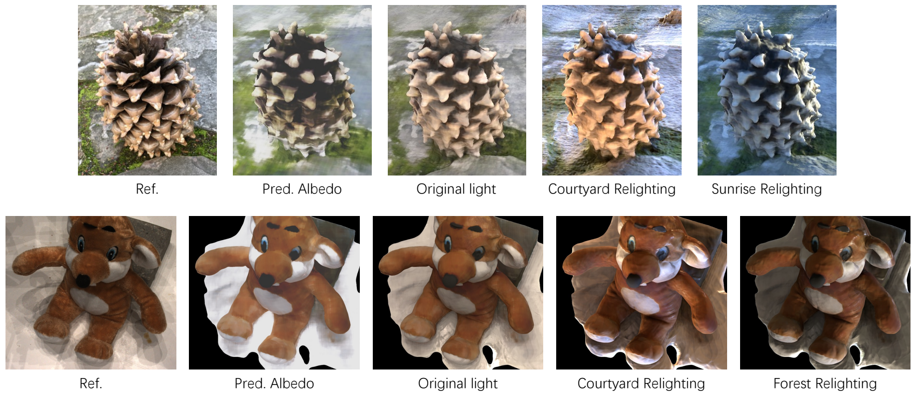

Fig. 6. Real-World Results. "Original light" means rendering the observed object with learned surface reflectance and learned environment illumination. We choose pinecone from NeRF [3] and scan105 from DTU MVS datasets [1].

Table 1. Evaluation on Novel View Synthesis. The metrics for each scene are the arithmetic mean over 8 validation views. The results of NeRF are taken from Table 4 of Nvdiffrec.

PSNR↑

Method	Drums	Ficus	Hotdog	Lego	Avg
NeRF [3]	27.670	28.050	36.710	31.890	31.080
NeuS [6]	24.806	29.211	36.106	30.652	30.193
Nvdiffrec [4]	27.550	26.783	36.023	32.213	30.640
NeRFactor [7]	27.134	35.090	34.451	30.799	31.869
NeRFactor, VI [7]	25.136	34.614	31.594	29.808	30.288
Ours, VI	25.861	36.989	35.750	30.316	32.229
Ours, LT	25.628	34.714	34.836	29.596	31.194

SSIM↑

Method	Drums	Ficus	Hotdog	Lego	Avg
NeRF [3]	0.951	0.957	0.971	0.944	0.956
NeuS [6]	0.932	0.931	0.976	0.938	0.944
Nvdiffrec [4]	0.958	0.970	0.980	0.955	0.965
NeRFactor [7]	0.950	0.979	0.945	0.902	0.944
NeRFactor, VI [7]	0.920	0.977	0.883	0.882	0.916
Ours, VI	0.941	0.988	0.957	0.903	0.947
Ours, LT	0.938	0.986	0.955	0.893	0.943

LPIPS↓

Method	Drums	Ficus	Hotdog	Lego	Avg
NeRF [3]	0.069	0.055	0.058	0.075	0.064
NeuS [6]	0.057	0.041	0.085	0.141	0.081
Nvdiffrec [4]	0.054	0.033	0.040	0.049	0.044
NeRFactor [7]	0.065	0.026	0.097	0.116	0.076
NeRFactor, VI [7]	0.081	0.027	0.162	0.117	0.097
Ours, VI	0.066	0.013	0.063	0.095	0.059
Ours, LT	0.066	0.018	0.070	0.098	0.063

Table 2. Evaluation on Albedo. Referring to [4], before measuring the errors, we rescaled the albedos from all tested methods to eliminate the scale ambiguity.

PSNR↑

Method	Drums	Ficus	Hotdog	Lego	Avg
Nvdiffrec [4]	20.418	35.452	27.559	21.393	26.205
NeRFactor [7]	23.211	36.043	27.265	24.790	27.829
NeRFactor, VI [7]	23.029	36.101	26.773	24.844	27.686
Ours, VI	22.704	36.161	27.069	24.443	27.594
Ours, LT	22.616	35.440	26.950	24.179	27.296

SSIM↑

Method	Drums	Ficus	Hotdog	Lego	Avg
Nvdiffrec [4]	0.909	0.986	0.944	0.877	0.929
NeRFactor [7]	0.935	0.990	0.928	0.920	0.943
NeRFactor, VI [7]	0.932	0.990	0.919	0.921	0.941
Ours, VI	0.928	0.989	0.926	0.897	0.935
Ours, LT	0.927	0.989	0.924	0.890	0.933

LPIPS↓

Method	Drums	Ficus	Hotdog	Lego	Avg
Nvdiffrec [4]	0.084	0.015	0.076	0.138	0.078
NeRFactor [7]	0.063	0.011	0.097	0.089	0.065
NeRFactor, VI [7]	0.066	0.011	0.107	0.096	0.070
Ours, VI	0.070	0.011	0.093	0.129	0.076
Ours, LT	0.075	0.011	0.099	0.134	0.080

Table 3. Evaluation on Relighting. Each listed scene includes 8 validation views, each view includes 8 HDR probe relighting images. The metrics for each scene are the arithmetic mean over 64 relighting images.

PSNR↑

Method	Drums	Ficus	Hotdog	Lego	Avg
Nvdiffrec [4]	23.112	28.404	29.029	21.461	25.502
NeRFactor [7]	23.648	28.409	25.314	26.694	26.016
NeRFactor, VI [7]	22.558	28.804	23.109	25.180	24.913
Ours, VI	22.130	31.144	25.463	24.776	25.878
Ours, LT	22.986	28.737	25.665	24.957	25.586

SSIM↑

Method	Drums	Ficus	Hotdog	Lego	Avg
Nvdiffrec [4]	0.923	0.978	0.930	0.849	0.919
NeRFactor [7]	0.921	0.953	0.912	0.873	0.915
NeRFactor, VI [7]	0.884	0.953	0.801	0.841	0.870
Ours, VI	0.911	0.981	0.921	0.856	0.917
Ours, LT	0.916	0.977	0.922	0.866	0.920

LPIPS↓

Method	Drums	Ficus	Hotdog	Lego	Avg
Nvdiffrec [4]	0.070	0.023	0.086	0.111	0.073
NeRFactor [7]	0.077	0.036	0.121	0.124	0.090
NeRFactor, VI [7]	0.095	0.051	0.203	0.137	0.122
Ours, VI	0.080	0.018	0.093	0.111	0.075
Ours, LT	0.077	0.026	0.100	0.116	0.080

References

1. Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., Aanæs, H.: Large scale multi-view stereopsis evaluation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 406–413 (2014) 5, 7
2. Matusik, W.: A data-driven reflectance model. Ph.D. thesis, Massachusetts Institute of Technology (2003) 3
3. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: European conference on computer vision. pp. 405–421. Springer (2020) 5, 7
4. Munkberg, J., Hasselgren, J., Shen, T., Gao, J., Chen, W., Evans, A., Müller, T., Fidler, S.: Extracting triangular 3d models, materials, and lighting from images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8280–8290 (2022) 3, 5, 6, 7, 8
5. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019), `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf` 3
6. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. Advances in Neural Information Processing Systems **34**, 27171–27183 (2021) 3, 7
7. Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T., Barron, J.T.: Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics (TOG) **40**(6), 1–18 (2021) 3, 7, 8