Learning Object-Centric Neural Scattering Functions for Free-viewpoint Relighting and Scene Composition: Supplementary Material

Hong-Xing Yu*[1], Michelle Guo*[1], Alireza Fathi[2], Yen-Yu Chang[1], Eric Ryan Chan[1], Ruohan Gao[1], Thomas Funkhouser[2], Jiajun Wu[1]

[1] *Stanford University* [2] *Google Research* * *Contributed equally.*

Reviewed on OpenReview: *https://openreview.net/forum?id=NrfSRtTpN5*

Our supplementary material consists of:

A. Supplementary video.

B. Implementation details.

C. Additional experiment results.

A Supplementary Video

In the supplementary video, we first give an overview of our method, Object-centric Neural Scattering Functions (OSFs), and explain how it learns cumulative radiance transfer. Then, we show demos of free-viewpoint relighting on both opaque and translucent objects. Finally, we show how we compose multiple OSFs into a new scene configuration.

B Implementation Details

We use a multilayer perception (MLP) with rectified linear activations. The predicted density σ is view-invariant, while the cumulative radiance transfer function ρ is dependent on the incoming and outgoing light directions. We use an eight-layer MLP with 256 channels to predict σ, and a four-layer MLP with 128 channels to predict ρ. Following NeRF (Mildenhall et al., 2020), we similarly apply positional encoding to our inputs and employ a hierarchical sampling procedure to recover higher quality appearance and geometry of learned objects. For positional encoding, we use $W = 10$ to encode the position $\mathbf{x}$ and $W = 4$ to encode the incoming and outgoing directions $(\boldsymbol{\omega}_{\text{light}}, \boldsymbol{\omega}_{\text{out}})$, where W is the highest frequency level. To avoid ρ from saturating in training, we adopt a scaled sigmoid (Brock et al., 2016) defined as $S'(\rho) = \delta(S(\rho) - 0.5) + 0.5$ with $\delta = 1.2$. We use a batch size of 2048 rays.

For synthetic datasets, we sample $N_c = 64$ coarse samples and $N_f = 128$ fine samples per ray. For real world datasets, we sample $N_c = 64$ coarse samples and $N_f = 64$ fine samples per ray. We use the Adam optimizer (Kingma & Ba, 2014) with a learning rate of 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$.

C Additional Experiments

C.1 Ablation study on translucency

We study how well OSFs can model different levels of translucency. To evaluate this, we use the Stanford Bunny model with different translucency intensities including $\{0, 0.1, 0.3, 0.6, 1\}$, where 0 means opaque and 1 means highly translucent with strong subsurface scattering. For each of these 5 objects, we generate images with identical camera poses and light directions. We show results in Figure S1 and Table S1. From the figure,

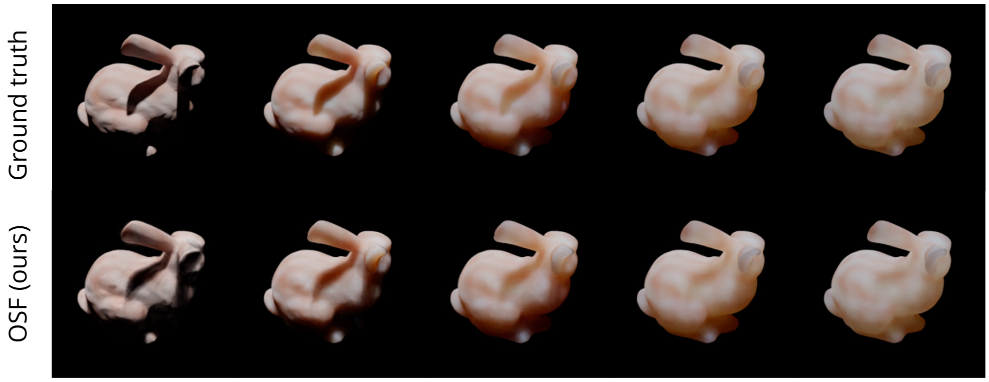

Figure S1: Qualitative examples of ablation study on translucency.

SSS	PSNR↑	SSIM↑	LPIPS↓
0	30.01	0.89	0.067
0.1	34.55	0.94	0.030
0.3	37.46	0.97	0.011
0.6	39.63	0.98	0.006
1	40.91	0.99	0.004

Table S1: Ablation study on modeling different levels of translucency. The leftmost column shows the intensity of subsurface scattering.

	PSNR↑	SSIM↑	LPIPS↓
One light	34.26	0.92	0.034
Two lights	32.40	0.93	0.027

Table S2: Free viewpoint relighting with two distant light sources.

we see that OSFs can model all objects well, especially translucent objects that have smoother shadows. A major reason is that the appearances of translucent objects vary smoothly w.r.t. changing view angles and lighting directions, and thus a learned neural implicit model is suitable to represent them. Opaque concave objects can have harsh self-shadows which are very high-frequency signals that are difficult for neural models to represent and interpolate (Rahaman et al., 2019).

C.2 Free viewpoint relighting with two light sources

Since OSFs learn radiance transfer functions, they support relighting with multiple distant lights due to the linearity of radiance. To demonstrate this, we evaluate OSFs on free-viewpoint relighting with two light sources, while all models are trained with only one light source. We use synthetic opaque objects. For each object, we generate a new test set with the same camera poses and light directions as the original one-light test set, but we add an additional light source. We show results in Figure S2 and in Table S2. From the figure, we observe that the visual results are faithful for both settings. From the table, we validate this via numerical metrics. We note that the widely-used environment maps are essentially collections of distant lights.

References

Andrew Brock, Theodore Lim, James M Ritchie, and Nick Weston. Generative and discriminative voxel modeling with convolutional neural networks. *arXiv preprint arXiv:1608.04236*, 2016.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

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 *European Conference on Computer Vision*, 2020.

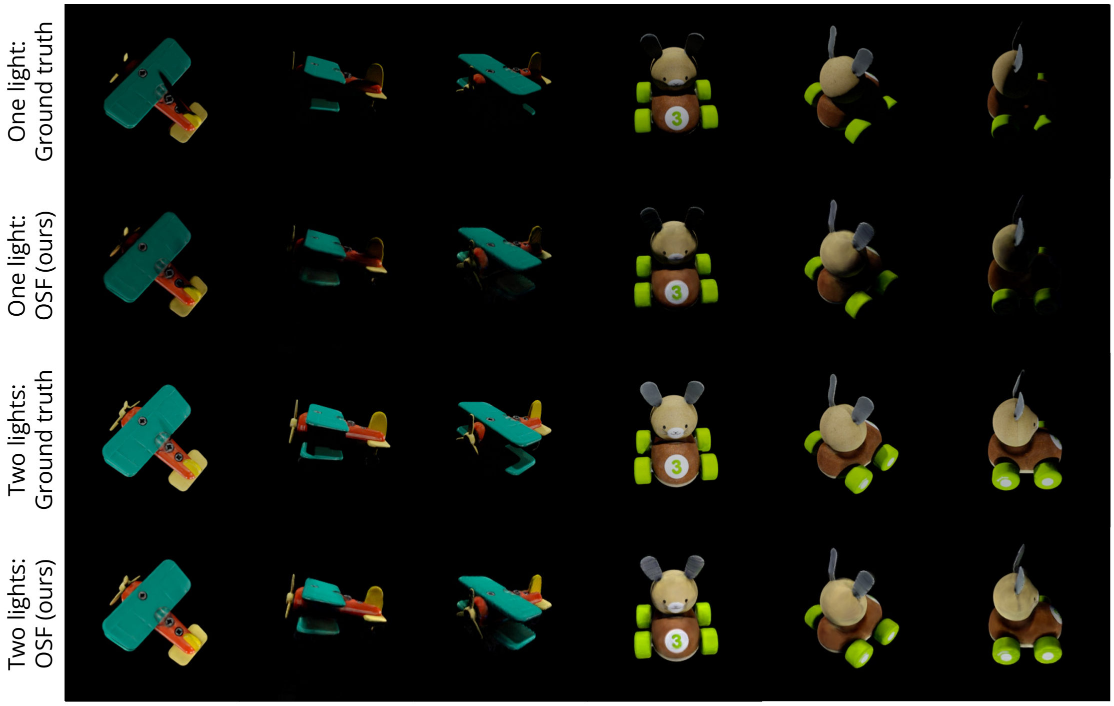

Figure S2: Qualitative examples of relighting with two distant light sources.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pp. 5301–5310. PMLR, 2019.