S^3-NeRF: Neural Reflectance Field from Shading and Shadow under a Single Viewpoint

Wenqi Yang[1] Guanying Chen[2] Chaofeng Chen[3] Zhenfang Chen[4] Kwan-Yee K. Wong[1]

[1]The University of Hong Kong [2]FNii and SSE, CUHK-Shenzhen
[3]Nanyang Technological University [4]MIT-IBM Watson AI Lab

NEURAL INFORMATION PROCESSING SYSTEMS

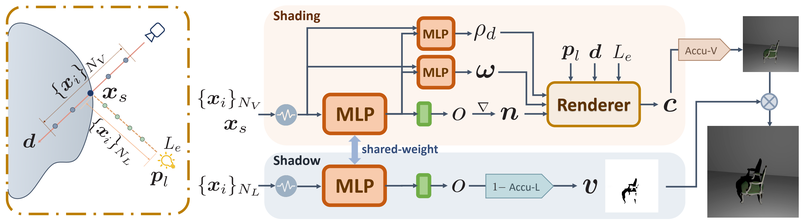
Code & Data & Model:
https://ywq.github.io/s3nerf/

Idea

- "Dual problem" of Multi-View Scene reconstruction – utilize **single-view** images captured under **different point lights** to reconstruct a scene.
- Existing single-view methods – only recover a **2.5D** scene representation (*i.e.*, a normal / depth map for the visible surface).
- Ours – Learn a **3D** neural reflectance field.
- MVS – Rely on multi-view photo-consistency to infer scene geometry.
- Ours – Exploit two information-rich monocular cues – **shading** & **shadow**.

Method

Overview:
Given N images captured from a single viewpoint under different near point lights, S^3-NeRF targets at recovering the geometry and materials of the scene.



We first apply root-finding to locate the surface intersection point $\boldsymbol{x}_s$.
- N_V points on the camera ray are sampled around the surface to generate accumulated shading values.
- N_L points are sampled on the surface-to-light segment to calculate the light visibility ($O(N_L)$ MLP queries per ray, calculated in an online manner).

Rendering Model:

We consider non-Lambertian surfaces with spatially-varying BRDFs. The rendering equation for a surface point $\boldsymbol{x}$ viewed from a direction $\boldsymbol{d}$ under a near point light $(\boldsymbol{p}_l, L_e)$ can be written as

$$f_c(\boldsymbol{d}, \boldsymbol{p}_l, L_e; \boldsymbol{x}) = \underbrace{L_{int}(\boldsymbol{p}_l, L_e; \boldsymbol{x})}_{\text{Light Intensity}} \underbrace{f_m(\boldsymbol{d}, \boldsymbol{w}_i(\boldsymbol{p}_l; \boldsymbol{x}); \boldsymbol{x})}_{\text{BRDF Value}} \underbrace{\max\left(\boldsymbol{w}_i(\boldsymbol{p}_l; \boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}), 0\right)}_{\text{Shading}}.$$

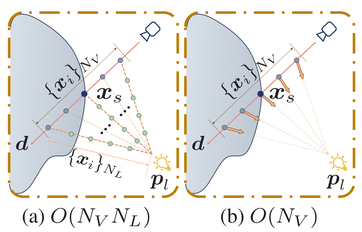
We propose to adopt a joint volume and surface rendering strategy:

$$\boldsymbol{C}_v(\boldsymbol{r}) = f_v(\boldsymbol{p}_l; \boldsymbol{x}_s) \sum_{i=1}^{N_V} o(\boldsymbol{x}_i) \prod_{j<i} \left(1 - o\left(\boldsymbol{x}_j\right)\right) f_c(\boldsymbol{x}_i, \boldsymbol{d}, \boldsymbol{p}_l, L_e),$$

$$\boldsymbol{C}_s(\boldsymbol{r}) = f_v(\boldsymbol{p}_l; \boldsymbol{x}_s) f_c(\boldsymbol{d}, \boldsymbol{p}_l, L_e; \boldsymbol{x}_s).$$

Alternative Shadow Modeling:

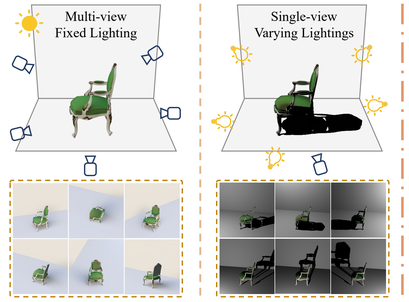
(a) Calculate light visibilities for all N_V points sampled along the ray (computationally expensive).
(b) Adopt an MLP to directly regress light visibility of a point to reduce the queries for each ray.



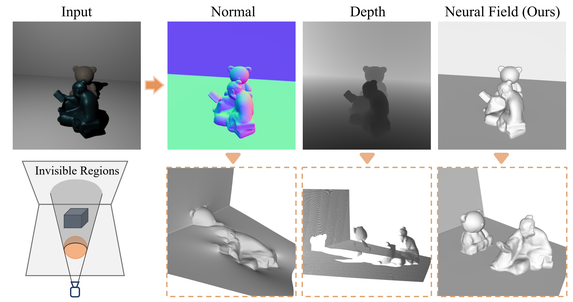
(a) $O(N_V N_L)$ (b) $O(N_V)$

Overview

Contributions:
- Novel problem: 3D neural reflectance field optimization from single-view images captured under different point lights.
- Exploit monocular shading and shadow cues to jointly recover the geometry and BRDFs of a scene.
- Adopt an efficient online shadow computation to fully exploit the information-rich shading and shadow cues.
- Experiments on multiple challenging datasets show that S^3-NeRF can faithfully reconstruct a complete scene geometry from single-view images and is robust to depth discontinuity.



(a) Different capturing setups



(b) Comparison of different scene representations

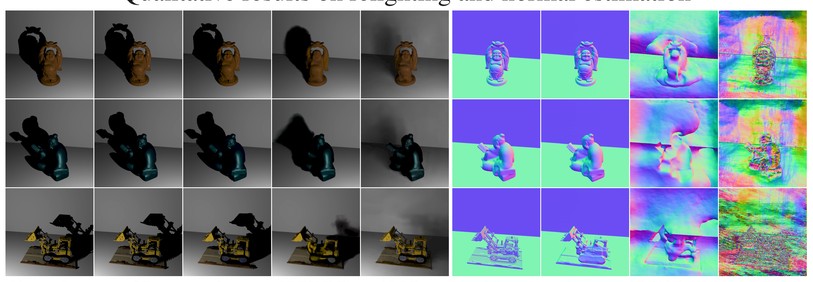
Experiments & Results

Comparison with Neural Radiance Field Methods:

Quantitative results on relighting and normal estimation

Method	BUDDHA PSNR↑	BUDDHA MAE↓	READING PSNR↑	READING MAE↓	BUNNY PSNR↑	BUNNY MAE↓	CHAIR PSNR↑	CHAIR MAE↓	LEGO PSNR↑	LEGO MAE↓	HOTDOG PSNR↑	HOTDOG MAE↓
NeRF* [1]	38.57	70.12	39.50	72.60	37.41	68.35	35.25	88.46	**35.56**	91.09	**39.80**	72.07
UNISURF* [2]	41.51	54.86	40.54	60.59	38.48	54.27	34.98	47.79	34.55	45.81	38.64	51.00
Ours	**43.42**	**2.44**	**43.13**	**2.03**	**40.43**	**1.72**	**36.33**	**1.83**	35.54	**6.49**	38.01	**2.50**

Qualitative results on relighting and normal estimation

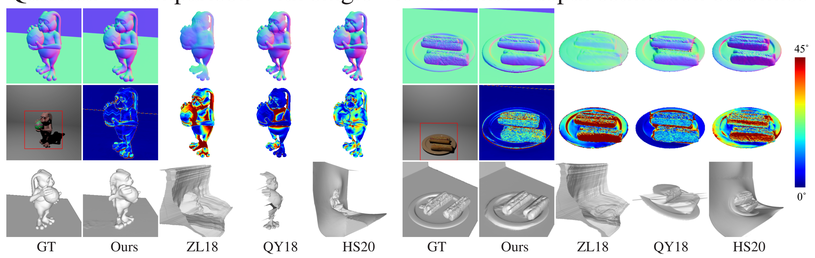


Comparison with Single-view Shape Estimation Methods:

Quantitative comparison (only object regions)

Method	BUDDHA MAE↓	BUDDHA Depth L1↓	READING MAE↓	READING Depth L1↓	BUNNY MAE↓	BUNNY Depth L1↓	CHAIR MAE↓	CHAIR Depth L1↓	LEGO MAE↓	LEGO Depth L1↓	HOTDOG MAE↓	HOTDOG Depth L1↓
ZL18 [3]	37.51	19.84	37.29	25.97	31.40	17.68	39.53	41.19	46.82	34.56	39.74	18.02
QY18 [4]	**12.25**	3.81	40.84	26.13	14.21	4.10	29.68	15.95	33.08	17.87	16.81	8.98
HS20 [5]	18.39	6.47	27.11	18.94	16.92	10.96	29.56	13.99	33.54	13.27	27.25	13.22
Ours	14.24	**1.50**	**7.00**	**2.09**	**9.40**	**1.63**	**17.43**	**4.74**	**31.13**	**7.31**	**14.65**	**1.68**

Qualitative comparison with single-view normal / depth estimation baselines

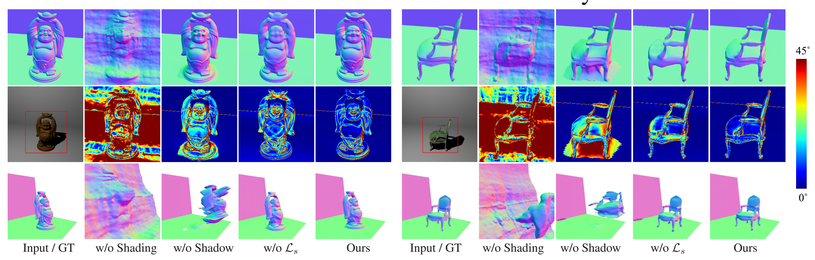


Ablation Study:

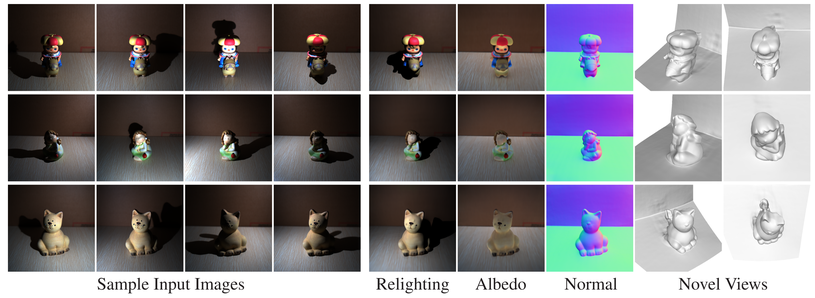
Quantitative results for the ablation study

Method	Train View CHAIR MAE↓	Train View CHAIR PSNR↑	Train View BUNNY MAE↓	Train View BUNNY PSNR↑	Train View BUDDHA MAE↓	Train View BUDDHA PSNR↑	Novel Views CHAIR MAE↓	Novel Views CHAIR PSNR↑	Novel Views BUNNY MAE↓	Novel Views BUNNY PSNR↑	Novel Views BUDDHA MAE↓	Novel Views BUDDHA PSNR↑
w/o shading	32.49	33.71	40.18	38.72	35.68	41.43	–	–	–	–	–	–
w/o shadow	3.39	30.81	2.26	33.45	3.33	34.30	12.24	22.31	11.93	24.43	16.60	23.27
w/o $\mathcal{L}_s$	2.48	35.85	2.75	39.73	3.77	43.04	**5.10**	**28.58**	6.27	29.11	8.50	28.61
Ours	**1.83**	**36.33**	**1.72**	**40.43**	**2.44**	**43.42**	5.45	26.82	**6.11**	**29.55**	**6.89**	**31.53**

Visual results for the ablation study



Results on Real Scenes:



Capturing Setup:

References:

[1] NeRF [Mildenhall *et al.*, ECCV20]
[2] UNISURF [Oechsle *et al.*, ICCV21]
[3] ZL18 [Li *et al.*, TOG18]
[4] QY18 [Quéau *et al.*, JMIV18]
[5] HS20 [Santo *et al.*, ECCV20]