PS-NeRF: Neural Inverse Rendering for Multi-view Photometric Stereo

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

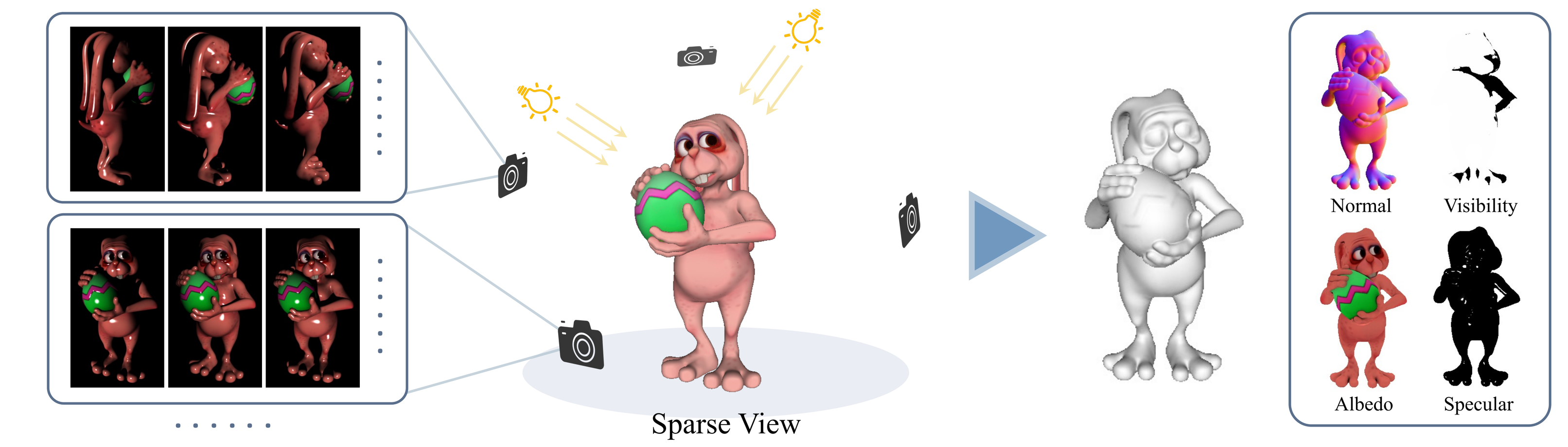
[1]The University of Hong Kong [2]FNii and SSE, CUHK-Shenzhen
[3]Nanyang Technological University [4]MIT-IBM Watson AI Lab

ECCV TEL AVIV 2022

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

Problem and Contribution

Idea:
Given multi-view and multi-light images of an object taken from M sparse views, PS-NeRF simultaneously reconstructs its shape, materials, and lights.



Contributions:

- A neural inverse rendering method for multi-view photometric stereo which jointly optimizes shape, BRDFs, and lights based on a shadow-aware differentiable rendering layer.
- We propose to regularize the implicit surface with normals estimated from multi-light images, which significantly improves surface reconstruction, especially for sparse input views (e.g., 5 views).
- Our method achieves state-of-the-art results.

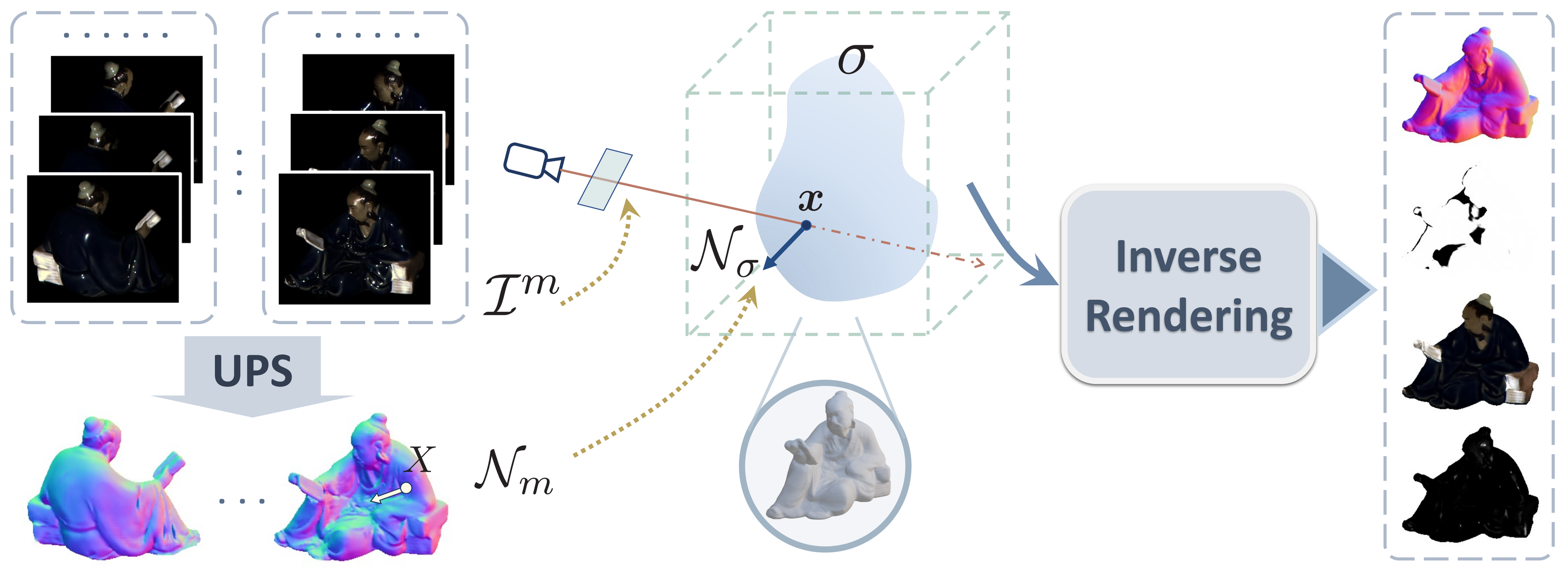
Comparison with Existing Neural Rendering Methods:

Method	Input	Shape	BRDF	Lighting	Shadow
NeRV [1]	MVI (Fixed lighting)	Density	Microfacet model	Known Envmap	Yes
PhySG [2]	MVI (Fixed lighting)	SDF	Microfacet (SGs)	Unknown Envmap (SGs)	No
NeRFactor [3]	MVI (Fixed lighting)	Density	Learned BRDF	Unknown Envmap	Yes
NeRD [4]	MVI (Varying lighting)	Density	Microfacet (SGs)	Unknown Envmap (SGs)	No
NRF [5]	MVI (Co-located light)	Density	Microfacet model	Known Co-located light	Yes
KB22 [6]	MVI (Multi-light)	Density	No	No	No
Ours	MVI (Multi-light)	Density	Mixture of SGs	Unknown Multi-light	Yes

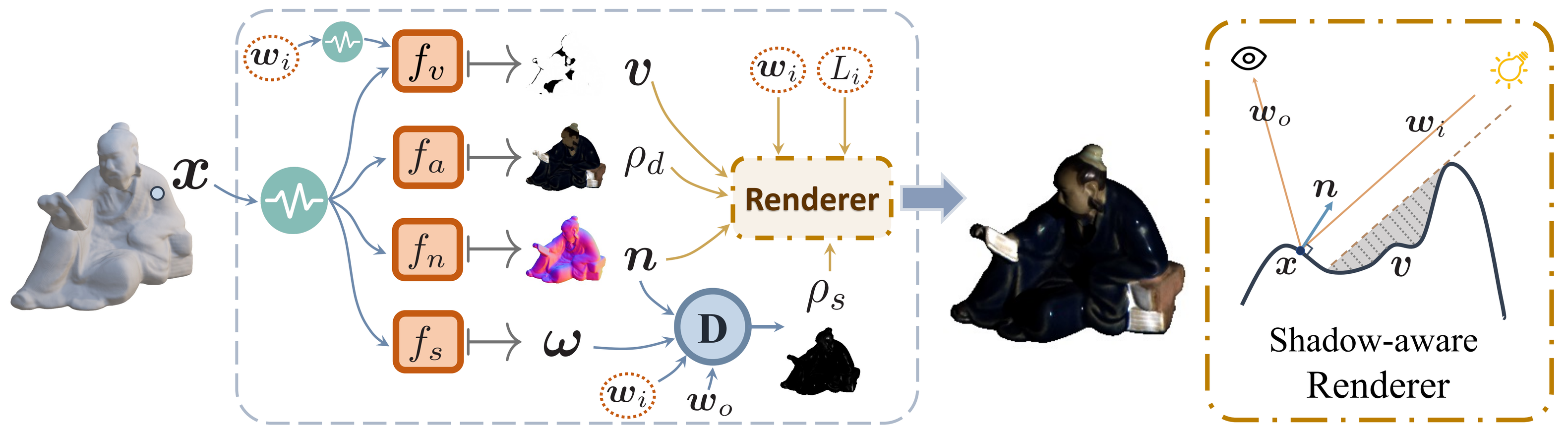
- Our method is the only one that explicitly models surface reflectances and lights under a MVPS setup.
- Our results demonstrate that incorporating multi-light information appropriately can produce a far more accurate shape reconstruction.

Method

Overview:
Inspired by the recent success of neural radiance field [9] for 3D scene representation, we represent the object shape with a density field. Our method consists of two stages to make full use of multi-view multi-light images.



Stage I – Initial Shape Modeling:
We estimate a guidance normal map $\mathcal{N}_m$ for each view, which is used to supervise the normals derived from the density field. This direct normal supervision is expected to provide a strong regularization on the density field, leading to an accurate surface.



Stage II – Joint Optimization with Inverse Rendering:
Based on the learned density field as the shape prior, we jointly optimize the surface normals, materials, and lights using a shadow-aware rendering layer.

- We model normals, BRDFs, and light visibility with MLPs. The weights of the MLPs and lights are jointly optimized to fit the input images.
- The rendering equation with directional light and visibility is:

$$\hat{I}(\boldsymbol{w}_o, \boldsymbol{w}_i; \boldsymbol{x}) = f_v(\boldsymbol{w}_i; \boldsymbol{x}) L_i(\boldsymbol{w}_i) f_r(\boldsymbol{w}_o, \boldsymbol{w}_i; \boldsymbol{x})(\boldsymbol{w}_i \cdot \boldsymbol{n}).$$

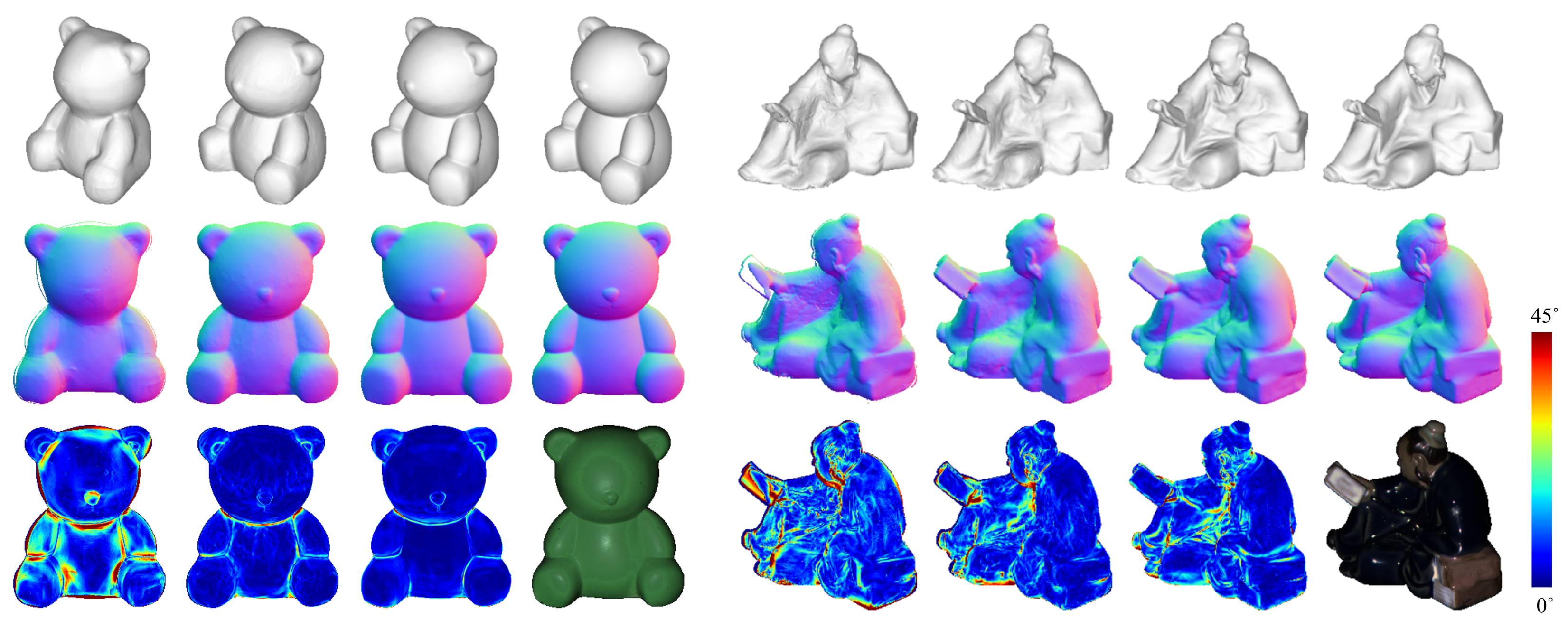
Experiments & Results

Comparison with MVPS Methods:

Quantitative results on DiLiGenT-MV benchmark

Method	Chamfer Dist↓ BEAR	BUDDHA	COW	POT2	READING	Average	Normal MAE↓ BEAR	BUDDHA	COW	POT2	READING	Average
PJ16 [7]	19.58	11.77	**9.25**	24.82	22.62	17.61	12.78	14.68	13.21	15.53	12.92	13.83
LZ20 [8]	8.91	13.29	14.01	7.40	24.78	13.68	4.39	11.45	**4.14**	6.70	8.73	7.08
Ours	**8.65**	**8.61**	10.21	**6.11**	**12.35**	**9.19**	**3.54**	**10.87**	4.42	**5.93**	**8.42**	**6.64**

Qualitative results of shape and normal on DiLiGenT-MV benchmark



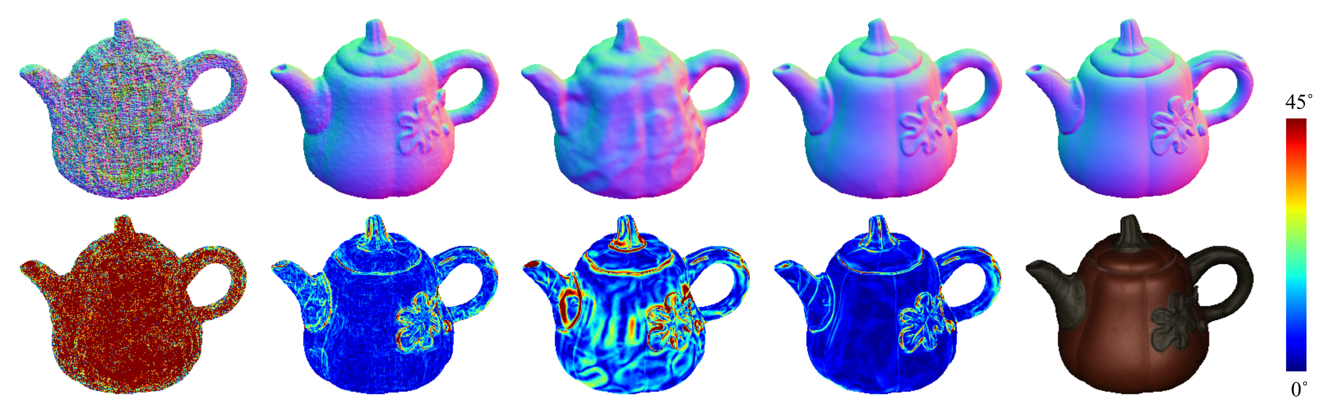
PJ16 [7] LZ20 [8] Ours GT
BEAR
PJ16 [7] LZ20 [8] Ours GT
READING

Effectiveness of Normal Regularization:

Quantitative analysis on normal supervision

Method	Normal MAE↓ BEAR	BUDDHA	COW	POT2	READING	BUNNY	Chamfer Dist↓ BEAR	BUDDHA	COW	POT2	READING	BUNNY
NeRF	73.45	59.62	55.10	69.41	55.55	49.24	66.68	29.28	70.07	42.28	48.26	19.67
NeRF^{+N}	7.03	13.50	8.26	7.93	14.01	12.04	16.02	9.65	12.04	7.25	13.31	5.44
UNISURF	6.51	17.13	8.26	13.04	19.68	10.04	9.24	9.83	13.25	10.21	62.89	6.89
UNISURF^{+N}	**4.26**	**11.29**	**5.05**	**6.37**	**9.58**	**7.76**	**7.24**	**8.93**	**11.33**	**5.76**	**12.83**	**5.32**

Qualitative comparison on w/ & w/o normal supervision



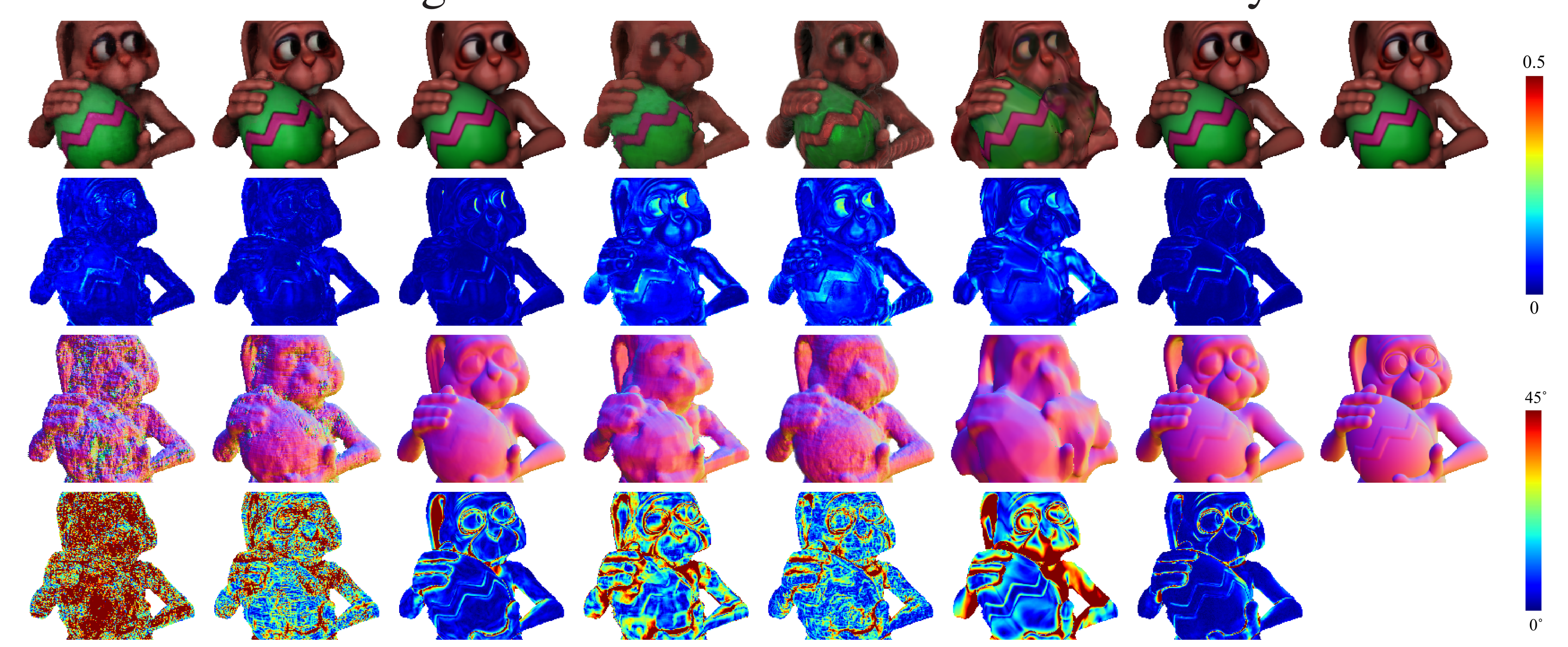
NeRF NeRF^{+N} UNISURF UNISURF^{+N} GT

Comparison with Neural Rendering Based Methods:

Comparison of shape reconstruction results on real and synthetic datasets

Method	DiLiGenT-MV BEAR MAE↓	BEAR CD↓	BUDDHA MAE↓	BUDDHA CD↓	COW MAE↓	COW CD↓	POT2 MAE↓	POT2 CD↓	READING MAE↓	READING CD↓	Synthetic BUNNY MAE↓ PS	BUNNY MAE↓ Env	BUNNY CD↓ PS	BUNNY CD↓ Env	ARMADILLO MAE↓ PS	ARMADILLO MAE↓ Env	ARMADILLO CD↓ PS	ARMADILLO CD↓ Env
NeRF [9]	73.90	66.68	59.89	29.28	55.14	70.07	69.71	42.28	55.75	48.26	49.42	48.22	19.67	20.09	44.27	41.54	20.95	24.34
KB22 [6]	53.19	66.18	39.72	17.92	85.11	82.43	87.30	63.82	70.13	86.79	31.64	37.00	9.61	18.32	35.88	51.08	14.50	7.58
UNISURF [10]	6.48	9.24	17.11	9.83	8.25	13.25	13.05	10.21	19.72	62.89	10.00	11.46	6.89	8.74	8.12	10.12	3.76	3.96
NeRFactor [3]	12.68	26.21	25.71	26.97	17.87	50.65	15.46	29.00	21.24	47.36	21.97	21.50	17.29	18.29	36.07	19.27	7.86	18.44
NeRD [4]	19.49	13.90	30.41	18.54	33.18	38.62	28.16	9.00	30.83	30.05	17.51	19.12	11.08	13.44	19.36	18.43	7.02	9.34
PhySG [2]	11.22	19.07	26.31	21.66	11.53	22.23	13.74	32.29	25.74	46.90	23.53	25.66	22.23	21.87	14.46	19.53	8.27	12.08
Ours	**3.21**	**7.24**	**10.10**	**8.93**	**4.08**	**11.33**	**5.67**	**5.76**	**8.83**	**12.83**	**5.14**	–	**5.32**	–	**5.18**	–	**3.61**	–

Novel view rendering and normal estimation results on the synthetic dataset



NeRF [9] KB22 [6] UNISURF [10] NeRFactor [3] NeRD [4] PhySG [2] Ours GT

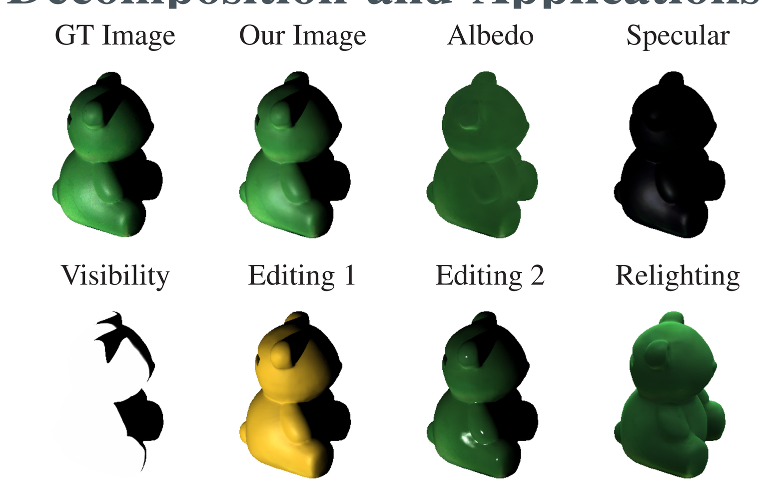
Normal Improvement:

Method	BEAR	BUDDHA	COW	POT2	READING	BUNNY
SDPS-Net	7.52	11.47	9.57	7.98	15.94	10.65
Stage I	4.26	11.29	5.05	6.39	9.58	7.79
Ours	**3.25**	**10.20**	**4.12**	**5.73**	**8.87**	**5.24**

Light Improvement:

Method	BEAR	BUDDHA	COW	POT2	READING	BUNNY
SDPS-Net	4.90	7.17	8.55	4.73	9.09	8.96
Ours	**2.27**	**2.75**	**2.59**	**2.89**	**4.26**	**1.53**

Decomposition and Applications:



References:

[1] NeRV [Srinivasan *et al.*, CVPR21]
[2] PhySG [Zhang *et al.*, CVPR21]
[3] NeRFactor [Zhang *et al.*, TOG21]
[4] NeRD [Boss *et al.*, ICCV21]
[5] NRF [Bi *et al.*, arXiv20]
[6] KB22 [Kaya *et al.*, WACV22]
[7] PJ16 [Park *et al.*, TPAMI16]
[8] LZ20 [Li *et al.*, TIP20]
[9] NeRF [Mildenhall *et al.*, ECCV20]
[10] UNISURF [Oechsle *et al.*, ICCV21]