Supplementary Materials for NeILF++

Jingyang Zhang[1] Yao Yao[2*] Shiwei Li[1] Jingbo Liu[1] Tian Fang[1]
David McKinnon[1] Yanghai Tsin[1] Long Quan[3*]

[1]Apple [2]Nanjing University [3]The Hong Kong University of Science and Technology
[1]{jingyang_zhang, shiwei, jingbo, fangtian, dmckinnon, ytsin}@apple.com
[2]yaoyao@nju.edu.cn [3]quan@cse.ust.hk

1. Method

1.1. BRDF Parameterization

In Sec. 3.1 we introduce the D, F, and G term of the specualr BRDF. The normal dsitribution term D is approximated by Spherical Gaussian whose sharpness is controlled by the roughness r:

$$D(\mathbf{h}; r) = S(\mathbf{h}, \frac{1}{\pi r^4}, \mathbf{n}, \frac{2}{r^4}) = \frac{1}{\pi r^4} \exp(\frac{2}{r^4}(\mathbf{h} \cdot \mathbf{n} - 1)). \tag{1}$$

The Fresnel term F is given by:

$$\begin{aligned} F(\boldsymbol{\omega}_o, \mathbf{h}; \mathbf{b}, m) &= F_0 + (1 - F_0)(1 - (\boldsymbol{\omega}_o \cdot \mathbf{h})^5), \\ \text{where } F_0 &= 0.04(1 - m) + \mathbf{b}m. \end{aligned} \tag{2}$$

The geoemtry term G is modeled by two GGX function [4]:

$$\begin{aligned} &G(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o, \mathbf{n}; r) = G_{\text{GGX}}(\boldsymbol{\omega}_i \cdot \mathbf{n}) G_{\text{GGX}}(\boldsymbol{\omega}_o \cdot \mathbf{n}), \\ &\text{where } G_{\text{GGX}}(z) = \frac{2z}{(2 - r^2)z + r^2}. \end{aligned} \tag{3}$$

1.2. Loss Terms

In this section we introduce the details of the loss terms besides L_{vol}, L_{phys} and L_{ref}.

Eikonal loss. This loss is adopted by most of the SDF-based systems [7] It expects the expectation of gradient magnitude is 1:

$$L_{\text{Eik}} = |\|\nabla \mathbb{G}(\mathbf{x})\| - 1|. \tag{4}$$

Hessian loss. This loss discourages the direction of the gradient to change rapidly by minimizing the norm of the Hessian matrix [8]:

$$L_{\text{Hess}} = \|\mathbf{H}\mathbb{G}(\mathbf{x})\|_1, \tag{5}$$

where $\|\cdot\|_1$ is the element-wize matrix 1-norm and $\mathbf{H}f$ is the Hessian matrix of the a function f.

Minimal surface loss. This loss minimize the elastic energy of the surface to produce compact interpolation or extrapolation of the surface for unobserved region [8]. To calculate the area of surface given by an implicit function, we model the surface $\mathcal{S}$ as a differentiation of the object interior $\mathcal{V}$:

$$\begin{aligned} \mathcal{V} &= \{\mathbf{x} \in \Omega | \mathbb{G}(\mathbf{x}) < 0\}, \\ \mathcal{S} &= \partial\mathcal{V} = \{\mathbf{x} \in \Omega | \mathbb{G}(\mathbf{x}) = 0\}, \end{aligned} \tag{6}$$

where Ω is the whole 3D space. And the surface area can be derived as:

$$\begin{aligned} \text{volume}(\mathcal{V}) &= \int_\Omega H(\mathbb{G}(\mathbf{x})) d\mathbf{x}, \\ \text{area}(\mathcal{S}) &= \int_\Omega \|\nabla H(\mathbb{G}(\mathbf{x}))\| \\ &= \int_\Omega \delta(\mathbb{G}(\mathbf{x})) \|\nabla \mathbb{G}(\mathbf{x})\|, \end{aligned} \tag{7}$$

where H is the Heaviside function and δ is the Dirac function. In practice, the gradient magnitude is 1 so can be omitted, and we use a regularized form of the Dirac function parameterized by a sharpness term ϵ. The loss is given by:

$$L_{\text{surf}} = \delta_\epsilon(\mathbb{G}(\mathbf{x})), \text{ where } \delta_\epsilon(z) = \frac{\epsilon \pi^{-1}}{\epsilon^2 + z^2}. \tag{8}$$

Point cloud supervision. We can optionally introduce an oriented point cloud to facilitate the convergence in early stage [8]. The distance values and the gradient directions are fit to the point cloud at the locations of the data points. Let $\mathbf{x}_D$ be the points in a point cloud D with normal $\mathbf{n}_D$, the point cloud loss is given by:

$$L_{\text{pcd}} = |\mathbb{G}(\mathbf{x}_D)| + (1 - \frac{\mathbf{n}_D \cdot \nabla \mathbb{G}(\mathbf{x}_D)}{\|\nabla \mathbb{G}(\mathbf{x}_D)\|}). \tag{9}$$

Bilateral smoothness. If the color $\mathbf{I}_\mathbf{p}$ of a pixel $\mathbf{p}$ does not change rapidly, we expect that the roughness and the

*Research done while at Apple.

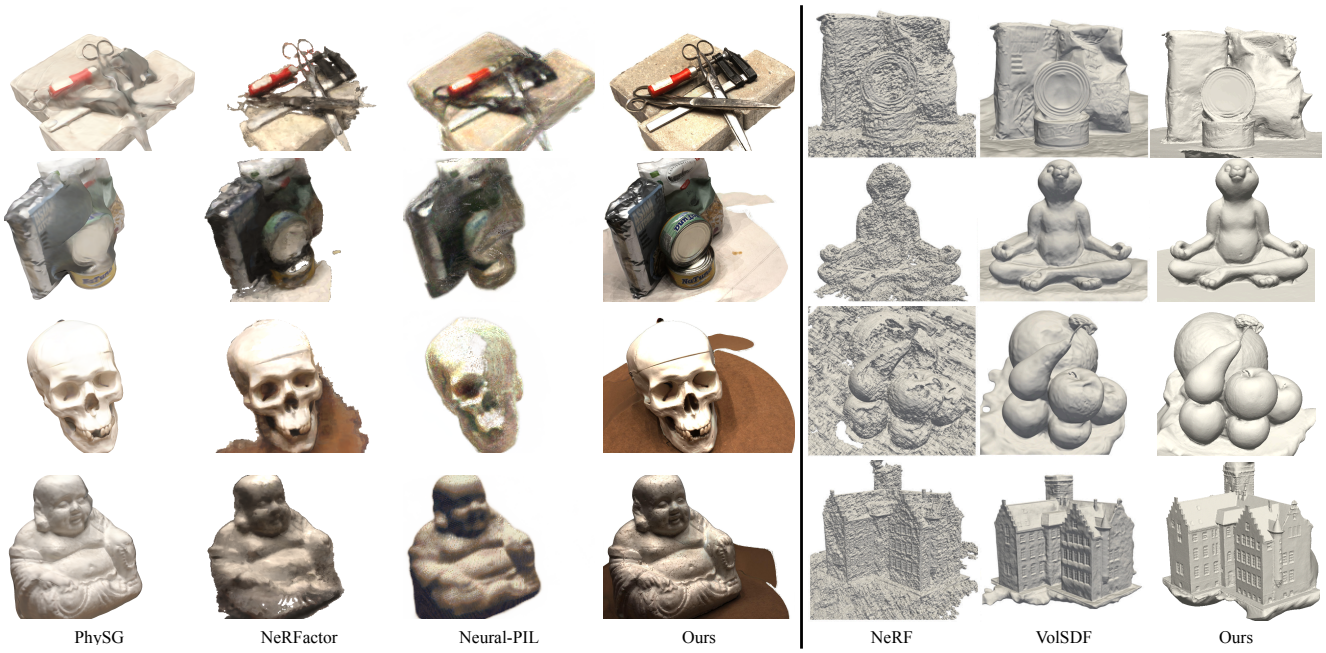

Figure 1. Qualitative comparisons on the DTU [2] dataset with previous methods. The geometry results of previous methods are taken from [6].

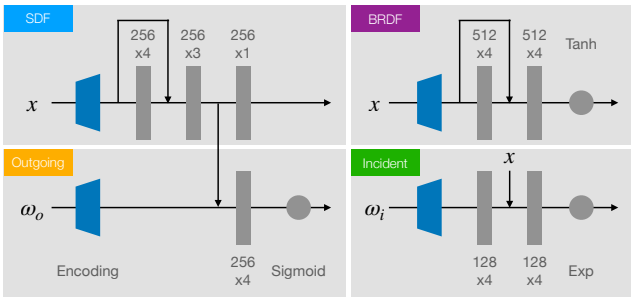

Figure 2. The network architecture for the 4 fields in the *MLP* setting.

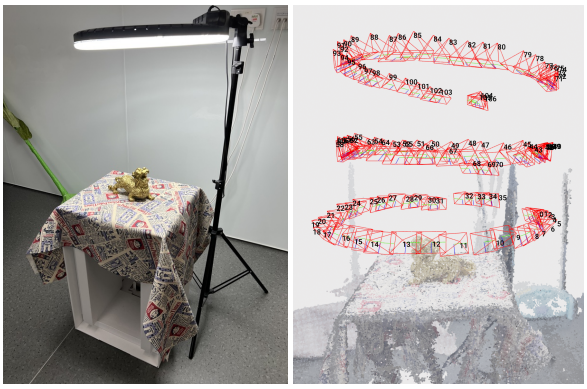

Figure 3. Capturing setup for the NeILF-HDR dataset. The left image shows the lighting setup, and the right image illustrates the camera trajectories.

metallic the corresponding 3D surface point $\mathbf{x_p}$ are smooth [5]. The bilateral smoothness loss is given by:

$$L_{\text{smth}} = (\|\nabla r(\mathbf{x_p})\| + \|\nabla m(\mathbf{x_p})\|) \exp(-\|\nabla \mathbf{I_p}\|). \quad (10)$$

Lambertian assumption. We encourage the surface to be Lambertian as a tie break rule during material optimization [5]. The loss is given by:

$$L_{\text{Lam}} = |r - 1| + |m|. \quad (11)$$

In practice, the weight for this loss is set to be very small to prevent interfering the material estimation.

2. Network Architecture

The detailed network architectures for the 4 fields in the *MLP* setting are illustrated in Fig. 2. Note that although the two light fields have the same input, in practice we still separate them because they have different frequency response with respect to each input. The outgoing light field is expected to be more sensitive to the position input. while the incident light field acts in the opposite way.

The *NGP* setting which uses hash grid [3] as the encoder has less number of layers in general. The SDF, the outgoing radiance field and the BRDF field each has 2, 4, 2 linear layers with 64 hidden units, and the skip connections in the SDF and the BRDF field are removed.

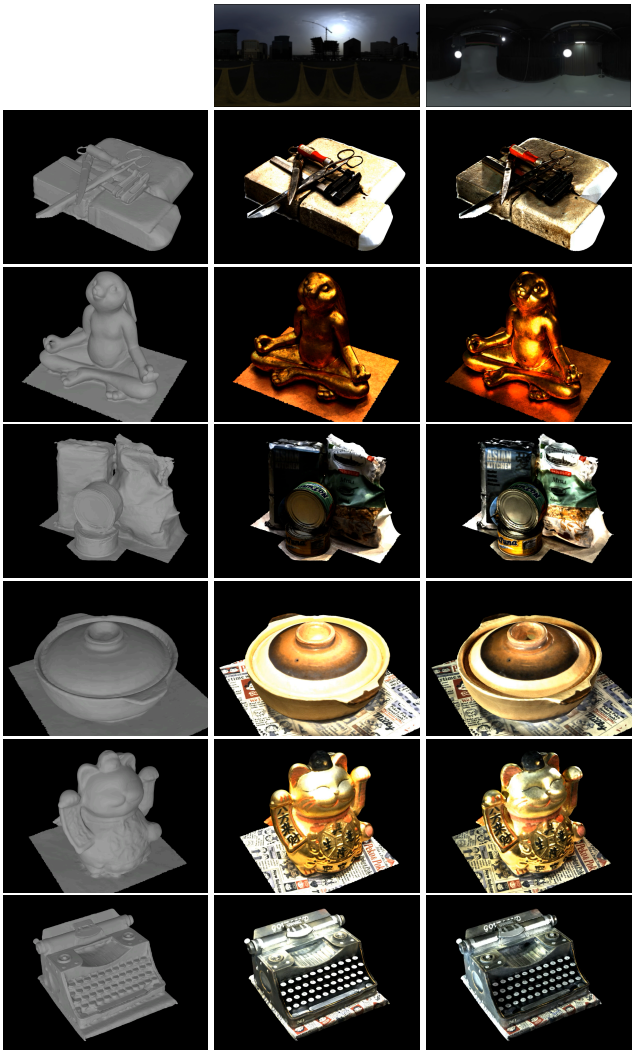

Figure 4. Geometry and relighting results of real world objects.

3. NeILF-HDR Dataset

In this paragraph we elaborate more details of this dataset. The capturing setup and the camera trajectories are shown in Fig. 3. Besides Fig. 5 and 6 of the main paper, more examples of captured objects can be found in Fig. 6. (Semi-) GT camera parameters will be provided, which were recovered by a state-of-the-art SfM system. The focal length in EXIF was used to initialize the intrinsics before bundle adjustment. Note that our scene is densely captured and rich-textured tablecloths are used to ensure high-quality feature matching and SfM reconstruction. The GT geometry is not provided at the current stage.

	PhySG	SG-ENV	Pix-ENV	Ne-ENV	NeILF	NeILF++
Base Color	14.83	**18.68**	11.93	10.61	15.94	17.95
Roughness	7.23	11.25	14.37	14.63	16.49	**21.90**
Metallic	8.57	15.66	14.59	14.94	16.97	**19.08**

Table 1. Quantitative comparisons on BRDF on NeILF-synthetic.

Geometry						Render			
L_{Hess}+L_{surf}	✗	✗	✓	✓	w/ joint	L_{smth}+L_{Lam}	✗	✓	w/ joint
L_{pcd}	✗	✓	✗	✓					
Chamfer ↓	2.161	0.824	2.576	0.760	0.739	PSNR ↑	27.88	27.68	28.61
Normal ↓	34.09	21.92	32.85	19.64	19.42				

Table 2. Ablation study on loss functions on DTU.

4. Additional Results

4.1. Comparison with Previous Methods

We qualitatively compare the geometry and novel view rendering of our method with previous methods, and the results are shown in Fig. 1.

We also quantitatively compare the accuracy of material estimation with previous baselines in Tab. 1. The results of the baselines are taken from the original NeILF paper who has already compared their method with re-implemented PhySG (SG-ENV) and NeRFactor (Pix-ENV). More details could be found in the NeILF paper [5].

4.2. Ablation study on loss functions

The ablation studies are performed on geometry and material initialization stages of the NGP setting. We report the mean scores over the 15 DTU scenes in Tab. 2: 1) The L_{Hess}+L_{Surf}, L_{pcd} all improve the geometry accuracy. Especially, the point cloud prior is important for the convergence under NGP setting, which is also mentioned in the main paper. 2) L_{smooth}+L_{Lam} does not necessarily improve the rendering quality, which is consistent with results and discussions in the NeILF paper.

4.3. Ablation study on joint optimization

We first clarify the difference between *before* and *w/o joint*: In *w/o joint*, we 1) run *SDF Init.*; 2) run *Joint* with only geometry losses; 3) run *Mat. Init.*; 4) fix geometry and run *Joint* with only material losses (please refer to Tab. 1 for the training stages). While in *before joint* we only run 1) and 3). Comparisons between *before* and *w/ joint* are shown in the last two columns in each group in Tab. 2, where all the metrics get improved after finetuning.

And we additionally study the joint optimization on NeILF-HDR using the novel view rendering PSNR. The *w. joint* setting still performs better than *w/o joint* (22.76 v.s. 22.53), which is consistent with the result in DTU.

4.4. Geometry and Material Estimation

Additional qualitative results of geometry and material estimation results on the DTU [2] and the NeILF-HDR dataset are shown in Fig. 5 and 6. Note that the metallic for the DTU scenes is often overestimated because of LDR

input. On the NeILF-HDR dataset, the estimated metallic is more reasonable.

4.5. Relighting

We export the material parameters from the BRDF field to a UV map and relight the objects by new environment maps using Blender [1]. The results are shown in Fig. 4. For animated results, please refer to the supplementary video.

References

[1] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. 4

[2] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *CVPR*, 2014. 2, 3, 5

[3] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv:2201.05989*, 2022. 2

[4] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*, pages 195–206, 2007. 1

[5] Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Neilf: Neural incident light field for physically-based material estimation. In *ECCV*, 2022. 2, 3

[6] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *arXiv preprint arXiv:2106.12052*, 2021. 2

[7] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *NeurIPS*, 2020. 1

[8] Jingyang Zhang, Yao Yao, Shiwei Li, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Critical regularizations for neural surface reconstruction in the wild. In *CVPR*, 2022. 1

Figure 5. Qualitative results on the DTU [2] dataset. The metallic is often overestimated because of LDR input.

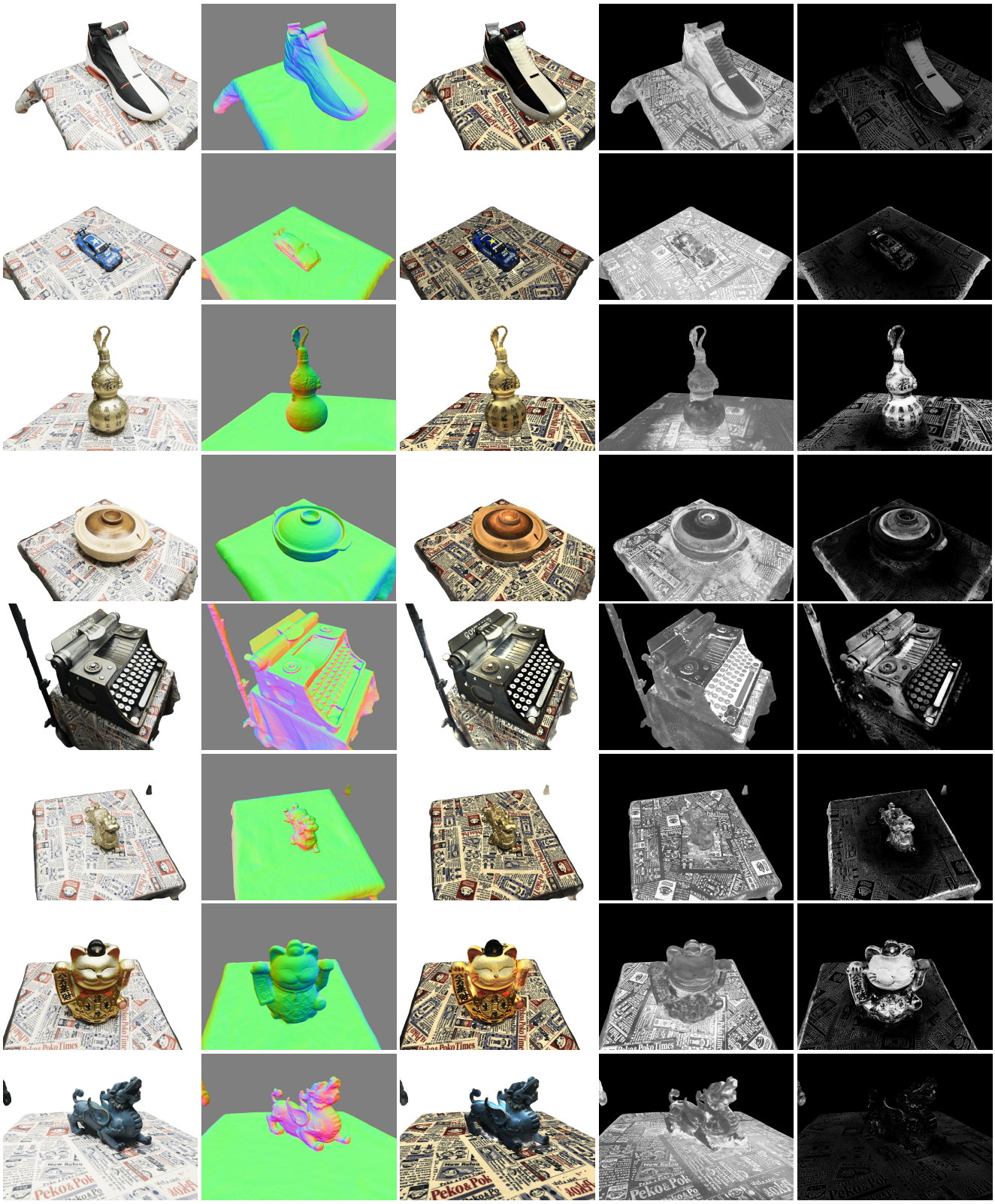

Figure 6. Qualitative results on the NeILF-HDR dataset.