

Differentiable Inverse Rendering with Interpretable Basis BRDFs

-Supplemental Document-

Hoon-Gyu Chung Seokjun Choi Seung-Hwan Baek
POSTECH

In this supplemental document, we provide additional results and details in support of our findings in the main manuscript.

Contents

1. Additional Details for the Proposed Method	2
1.1. Simplified Disney BRDF Model	2
1.2. 2D Gaussian Splatting	2
1.3. Loss Functions	2
1.4. Point Initialization	3
1.5. Basis BRDF Initialization	3
1.6. Adaptive Density Control	3
1.7. Scheduling	3
1.8. Runtime	3
2. Dataset	3
2.1. Synthetic Photometric Dataset	3
2.2. Real-world Photometric Dataset	4
3. Additional Ablation Study	4
3.1. Specular-weighted Photometric Loss	4
3.2. Softmax Function with Temperature	4
3.3. Sparsity Regularizer	5
3.4. Mask Loss	5
3.5. Number of Initialized Basis BRDFs	5
3.6. Threshold for Basis BRDF Merge	5
3.7. Thresholds for Basis BRDF Removal	6
4. Additional Discussions	7
4.1. Shadow Computation by Occlusion	7
4.2. Global Illumination	7
4.3. Reconstruction under Environment Maps	7
4.4. Similarity of Close Gaussians	7
4.5. Opacity Distribution of Gaussians	7
5. Additional Results	7
5.1. Novel-view Relighting Results	7
5.2. Point Relighting Results	7
5.3. Reflectance Editing Results	7
5.4. Additional Results with Synthetic Photometric Dataset	9
5.5. Additional Results with Real-world Photometric Dataset	13

1. Additional Details for the Proposed Method

1.1. Simplified Disney BRDF Model

We use the simplified Disney BRDF model [8] having parameters: \mathbf{b}_i , roughness σ_i , and metallic m_i . The BRDF model has three major terms: distribution function D , Fresnel term F , and geometry attenuation term G to represent realistic specularly as follows:

$$f_i(\mathbf{i}, \mathbf{o}) = \frac{1 - m_i}{\pi} \mathbf{b}_i + \frac{D(\mathbf{h}; \sigma_i) F(\mathbf{o}, \mathbf{h}; \mathbf{b}_i, m_i) G(\mathbf{i}, \mathbf{o}, \mathbf{n}; \sigma_i)}{4(\mathbf{n} \cdot \mathbf{i})(\mathbf{n} \cdot \mathbf{o})}. \quad (1)$$

We compute normal distribution function D using a Spherical Gaussian function:

$$D(\mathbf{h}; \sigma_i) = S(\mathbf{h}; \mathbf{n}, \frac{2}{\sigma_i^2}, \frac{1}{\pi \sigma_i^2}) = \frac{1}{\pi \sigma_i^2} \exp\left(\frac{2}{\sigma_i^2}(\mathbf{h} \cdot \mathbf{n} - 1)\right). \quad (2)$$

We use Schlick's approximation for the Fresnel term:

$$F(\mathbf{o}, \mathbf{h}; \mathbf{b}_i, m_i) = F_0 + (1 - F_0)(1 - \mathbf{o} \cdot \mathbf{h})^5 \quad (3)$$

where $F_0 = 0.04(1 - m_i) + m_i \mathbf{b}_i$. The geometry attenuation term is computed by multiplication of two GGX functions [7]:

$$G(\mathbf{i}, \mathbf{o}, \mathbf{n}; \sigma_i) = G_{GGX}(\mathbf{i} \cdot \mathbf{n}) G_{GGX}(\mathbf{o} \cdot \mathbf{n}), \quad (4)$$

where the GGX function is defined as follows:

$$G_{GGX}(z) = \frac{z}{(1 - \alpha)z + \alpha}, \quad \alpha = \frac{(1 + \sigma_i)^2}{8} \quad (5)$$

1.2. 2D Gaussian Splatting

We use 2D Gaussians [4] with learnable geometric parameters: $\{\mathbf{p}, \mathbf{t}, \mathbf{s}\}$. Each 2D Gaussian is defined on the local tangent frame in world coordinate, which is parameterized as:

$$P(a, b) = \mathbf{p} + s_a \mathbf{t}_a a + s_b \mathbf{t}_b b = \begin{bmatrix} \mathbf{t}\mathbf{s} & \mathbf{p} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} (a, b, 1)^\top, \quad (6)$$

where \mathbf{p} is the location of Gaussian center, $\mathbf{t} = [\mathbf{t}_a, \mathbf{t}_b, \mathbf{t}_c] \in \mathbb{R}^{3 \times 3}$ denotes the rotation matrix, and $\mathbf{s} = \text{diag}(s_a, s_b, 0) \in \mathbb{R}^{3 \times 3}$ denotes the scaling matrix of 2D Gaussian. \mathbf{t}_c denotes the surface normal of 2D Gaussian. Then, 2D Gaussian-filtered distance value is computed as:

$$\mathcal{G}(\mathbf{a}) = \exp\left(-\frac{a^2 + b^2}{2}\right), \quad (7)$$

where $\mathbf{a} = (a, b)$ denotes a point that is intersected with the ray $\mathbf{r}(u)$ coming from corresponding pixel u and Gaussian. A point \mathbf{a} is defined on the local tangent frame in ab space. Then, we render an image with following equation:

$$I(u) = \sum_{i=1}^M L_i \alpha_i \mathcal{G}_i(\mathbf{r}(u)) \prod_{j=1}^{i-1} (1 - \alpha_j \mathcal{G}_j(\mathbf{r}(u))), \quad (8)$$

where $u \in U$ is a pixel, M is the number of Gaussians projected onto pixel u . $\{L_i\}_{i=1}^M$ and $\{\alpha_i\}_{i=1}^M$ are the radiance and opacity values of the depth-sorted i -th Gaussian, respectively.

1.3. Loss Functions

We optimize Gaussian parameters G and basis BRDF parameters R by minimizing the following loss function:

$$\mathcal{L} = \mathcal{L}_{\text{render}} + \lambda_{\text{geom}} \mathcal{L}_{\text{geom}} + \lambda_{\text{mask}} \mathcal{L}_{\text{mask}} + \lambda_{\text{sparse}} \mathcal{L}_{\text{sparse}}. \quad (9)$$

- $\mathcal{L}_{\text{render}}$ is the specular-weighted rendering loss consisting of \mathcal{L}_1 loss and SSIM loss between the reconstructed image I and the observed image I' as follows:

$$\mathcal{L}_{\text{render}} = \frac{1}{|U|} \sum_{u \in U} H(u) ((1 - \lambda_s) \mathcal{L}_1(u) + \lambda_s \mathcal{L}_{\text{SSIM}}(u)), \quad (10)$$

where λ_s is a balancing weight and $H(u)$ is a specular weight map.

- To promote stable optimization, we incorporate geometric regularization term $\mathcal{L}_{\text{geom}}$ following the approach in 2DGS [4], specifically focusing on depth distortion and normal consistency. Depth distortion loss aims to compact the Gaussian splats by minimizing the depth distance between ray-splat intersections. Normal consistency loss aligns the 2D splats with the object’s surface to approximate a smooth surface. The geometric regularization is formulated as:

$$\mathcal{L}_{\text{geom}} = \sum_{i,j} t_i t_j |z_j - z_i| + \sum_{i=1} t_i (1 - \mathbf{n}_i^T \tilde{\mathbf{n}}_i) \quad (11)$$

where i indexes the splats intersected along the ray, $t_i = \alpha_i \mathcal{G}_i(\mathbf{r}(u)) \prod_{j=1}^{i-1} (1 - \alpha_j \mathcal{G}_j(\mathbf{r}(u)))$, z_i is the depth of the splat, \mathbf{n}_i is the normal of the rendered normal image, and $\tilde{\mathbf{n}}_i$ is the normal computed from the rendered depth image.

- $\mathcal{L}_{\text{mask}}$ is the cross entropy loss between rendered mask M and ground-truth mask M' .
- To obtain spatially separated interpretable basis BRDFs, we apply sparsity regularizer $\mathcal{L}_{\text{sparse}}$ with Gaussian weights and weight images.

1.4. Point Initialization

For the synthetic dataset, we utilize masked-based point sampling for the initial point cloud. We sample 3D points uniformly and project these points on the image plane to estimate points that fall within masked region [10]. We employ COLMAP output as a point cloud initialization for real-world photometric images. This point cloud initialization enables stable and accurate reconstruction of 3D geometry.

1.5. Basis BRDF Initialization

For basis BRDF initialization, we perform k-means clustering to obtain the initial base color for each basis. We utilize all input pixel values as input and compute the centers of clusters with the number of initialized basis BRDFs. We set \mathbf{b}_i as center of cluster, $\sigma_i = 0.5$ and $m_i = 0.0$. This basis BRDF initialization enables efficient and accurate reconstruction of spatially-varying BRDFs with interpretable basis BRDFs. We obtain similar reconstruction results for two initial metallic parameters (0.2, 0.0): relighting PSNR is (31, 32) dB, and normal MAE is (9.7, 9.8) degrees for the synthetic dataset.

1.6. Adaptive Density Control

We follow the adaptive density control policy of 2D Gaussian splatting [4]. We repeat point upsampling and pruning based on the gradient of Gaussian parameters to achieve high-quality reconstruction. To optimize stable geometry and interpretable basis BRDFs efficiently, we increase the densification interval to 500.

1.7. Scheduling

We optimize Gaussian and basis BRDF parameters simultaneously without sparsity regularizers before 5000 iterations. We apply Gaussian weight sparsity regularization after 5000 iterations and weight image sparsity regularization after 9000 iterations. Weight image regularization constraints the sparsity of basis BRDF weights stronger than Gaussian weight regularization. Basis merging and removal occur at 500-step intervals, beginning at 6000 steps.

1.8. Runtime

Reconstruction takes 27 min. without sparsity regularizers (Sec.3.1 in the main paper) and 33 min. with sparsity regularizers (Sec.3.2 in the main paper).

2. Dataset

2.1. Synthetic Photometric Dataset

We test our method on synthetic photometric dataset, following the configuration of mobile flash photography [6]. We rendered 4 complex scenes of multiple objects with Blender using mesh and image texture data from IRON [9] and Objaverse [3]. We rendered 300 view images with co-located point lights and used 200/100 views for training/testing, respectively. We also render ground-truth normal images for normal evaluation.

2.2. Real-world Photometric Dataset

We test our method on real-world photometric dataset, which is captured by a mobile phone with flashlight. We utilize Proshot application to capture raw images in dng format. We fix exposure, focal length, ISO and white balance to capture images with consistent settings. We perform checkerboard calibration to obtain camera intrinsic parameters and rectify distorted images. We also capture color checker images to estimate white balance and flash light intensity simultaneously. Chrome ball calibration [5] is performed to estimate the accurate position of flash light, assuming a point light source. We treat flash light position as a fixed parameter. Flash light intensity is modeled as a fixed scalar value with the distance-aware light fall-off term.

3. Additional Ablation Study

Method	Synthetic photometric dataset			
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	MAE \downarrow
Proposed	31.78	0.9754	0.0234	9.81
w/o spec	30.82	0.9748	0.0299	10.29
w/o softmax	30.19	0.9720	0.0381	10.17
w/o $\mathcal{L}_{\text{sparse}}$	34.13	0.983	0.0180	9.85
w/o $\mathcal{L}_{\text{mask}}$	32.51	0.9790	0.0186	11.05
basis 9	31.51	0.9727	0.0257	10.21
basis 15	32.13	0.9744	0.0242	9.82

Table 1. **Quantitative comparison of ablation studies for a synthetic photometric dataset.** We achieve the highest reconstruction quality without sparsity loss, however it produces non-interpretable basis BRDFs and weight maps. Our proposed method enables reconstruction of both high-fidelity and interpretable basis BRDFs simultaneously.

3.1. Specular-weighted Photometric Loss

We evaluate the importance of specular-weighted photometric loss of our method. Table 1 shows that a specular-weight map improves both image and normal reconstruction quality. Especially, the specular region is weighted by potentially-specular weight map, resulting in an accurate reconstruction of specularity.

3.2. Softmax Function with Temperature

We evaluate the impact of softmax function with low temperature. Table 1 and Figure 1 show that high temperature $T = 1$, results in non-interpretable basis BRDFs and non-intuitive weight images. We employ low temperature softmax function to optimize spatially separated weight maps with interpretable basis BRDFs, enabling intuitive scene editing.



Figure 1. **Impact of softmax function with low temperature.** Low temperature softmax function results in interpretable basis BRDFs and images. We employ $T = 0.0125$ in our experiment.

3.3. Sparsity Regularizer

Table 1 shows that sparsity regularizer downgrades the reconstruction quality while it produces interpretable basis BRDFs. Without sparsity regularizer, optimized basis BRDFs are non-interpretable, and weight images are not spatially separated.

3.4. Mask Loss

Table 1 shows that we achieve high-quality reconstruction results without mask loss. However, we obtain non-interpretable basis BRDFs and weight maps as shown in Figure 2. Mask loss regularizes Gaussians to align with the object surface that edge does not include the color of the background. We apply mask loss to reconstruct more interpretable basis BRDFs and accurate surface normal.

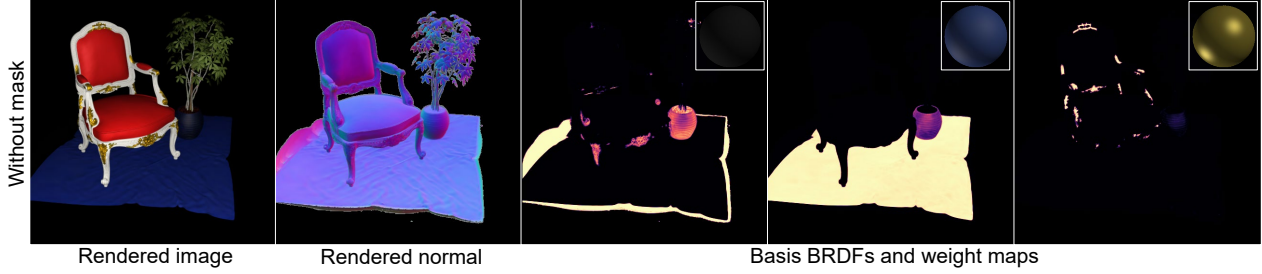


Figure 2. **Impact of the mask input.** Without mask loss, weight maps are non-intuitive, and the edge is unclear.

3.5. Number of Initialized Basis BRDFs

We evaluate the impact of the number of initialized basis BRDFs. We test our method on complex scenes with multi-object which need sufficient basis BRDFs. Table 1 and Figure 3 show that using a small number of basis BRDFs results in low reconstruction quality and non-interpretable basis BRDFs.

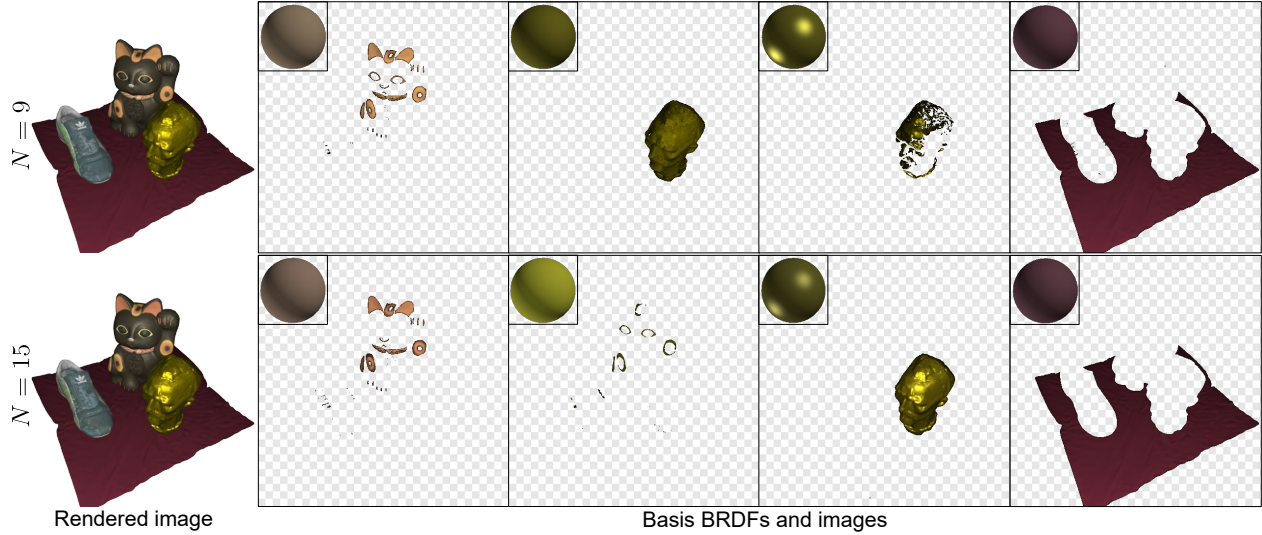


Figure 3. **Impact of the number of initialized basis BRDFs.** With a small number of initialized basis BRDFs, basis BRDFs are non-interpretable, and weight maps are not spatially separated. We utilize $N = 12$ in our experiment.

3.6. Threshold for Basis BRDF Merge

Figure 4 shows the result of optimized basis BRDFs and images depending on the merge threshold τ_{merge} . If we set a large value for the threshold τ_{merge} , basis BRDFs are not merged when they are non-intuitive with duplicates. If we set a small value for the threshold τ_{merge} , basis BRDFs are merged when they are not similar. We find appropriate merge threshold τ_{merge} for each complex scene to estimate optimal basis BRDFs and weight images.

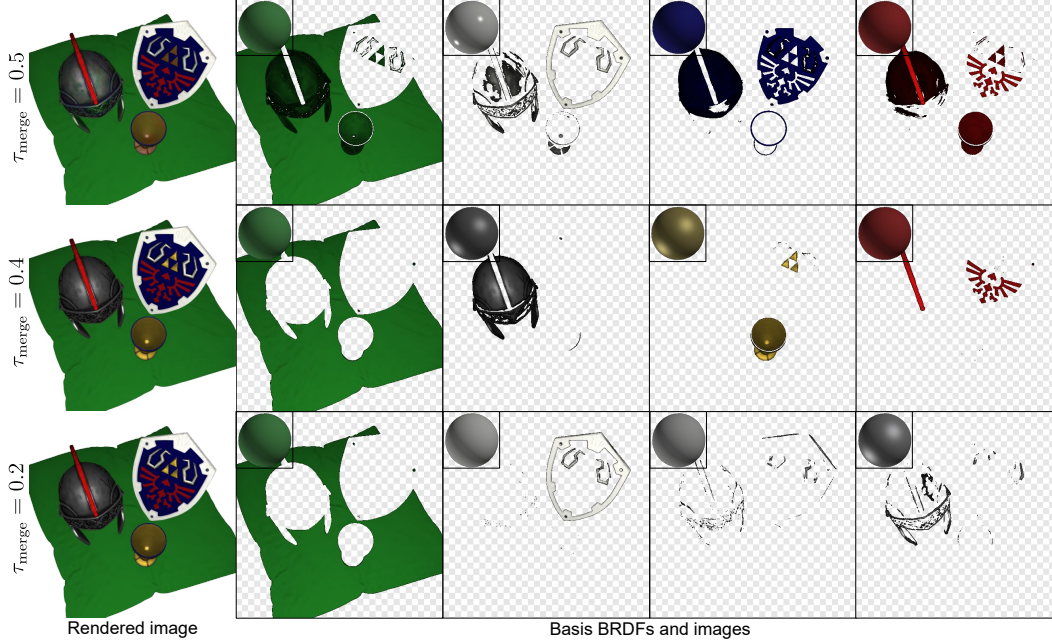


Figure 4. **Importance of merge threshold.** Merge threshold τ_{merge} directly affects the interpretability and scalability of basis BRDFs. If τ_{merge} is too large, distinguishable basis BRDFs are merged excessively. If τ_{merge} is too small, similar basis BRDFs are not merged. We utilize $\tau_{\text{merge}} = 0.4$ for this scene.

3.7. Thresholds for Basis BRDF Removal

Figure 5 shows the result of optimized basis BRDFs and weight images depending on the removal thresholds $\tau_{\text{removal-weight}}$ and $\tau_{\text{removal-number}}$. $\tau_{\text{removal-weight}}$ determines whether weight image pixels are contributing to the reconstruction. $\tau_{\text{removal-number}}$ determines whether basis BRDFs are significant depending on the portion between valid weight pixels and total pixels. If $\tau_{\text{removal-weight}}$ is too small, many pixel are considered valid with non-interpretable basis BRDFs. If $\tau_{\text{removal-number}}$ is too big, necessary basis BRDFs are removed.

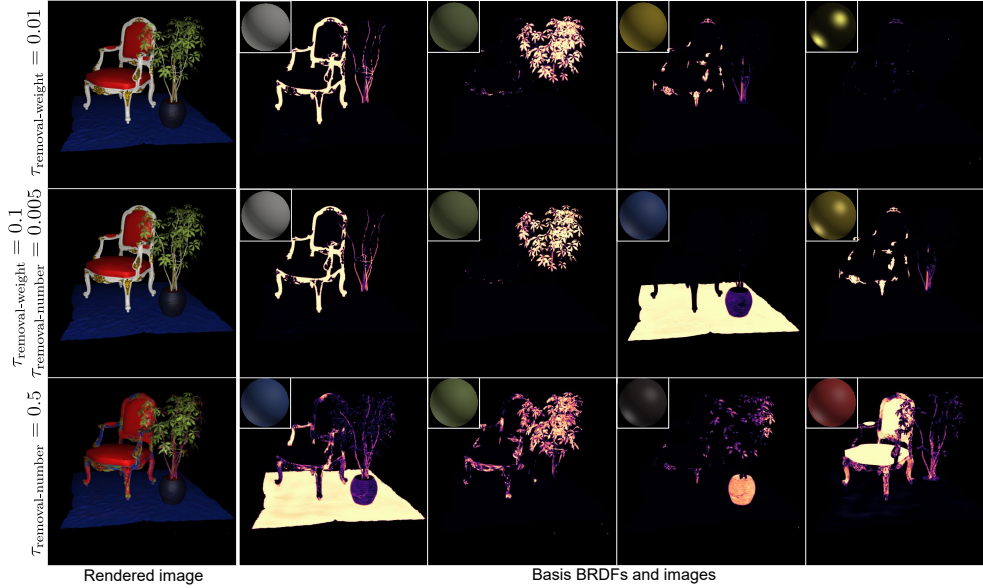


Figure 5. **Importance of removal thresholds.** Removal thresholds $\tau_{\text{removal-weight}}$ and $\tau_{\text{removal-number}}$ determine which unnecessary basis BRDF to be removed. With small $\tau_{\text{removal-weight}}$, unnecessary basis BRDFs are not removed, resulting in non-interpretable basis BRDFs. If $\tau_{\text{removal-number}}$ is too big, necessary basis BRDFs are removed, and reconstruction fails. We utilize $\tau_{\text{removal-weight}} = 0.1$ and $\tau_{\text{removal-number}} = 0.005$ for this scene.

4. Additional Discussions

4.1. Shadow Computation by Occlusion

Our method utilizes captured images with multi-view flash photography to collect sufficient light-angular samples and neglect shadow by occlusion, which does not exist when light and camera are co-located. We achieve high-fidelity reconstruction of geometry and spatially-varying BRDFs even on real-world photometric dataset captured by a mobile phone with the flash light. However, our method can be further improved with shadow computation by rendering a shadow image [1]. By multiplying a shadow image and a rendered image, we can render a photo-realistic image with shadow handling.

4.2. Global Illumination

We only consider direct illumination from a point light source, neglecting global illumination. Global illumination may be handled by employing residual MLPs or ray tracing. This would enable to handle indirect illumination and inter-reflection between specular objects, recovering accurate geometry and spatially varying BRDFs.

4.3. Reconstruction under Environment Maps

Our method supports differentiable forward rendering with arbitrary environmental lighting as shown in Fig. 1(d) in the main paper. Thus, environment-light inverse rendering is possible in principle. However, challenges exist: ambiguity of unknown illumination and BRDF, and the complexity of environment map optimization, where learned priors may help.

4.4. Similarity of Close Gaussians

We observed that spatially neighboring Gaussians have similar normals and weights thanks to our normal consistency loss: The $k = 3$ -nearest-neighbor Gaussians have the low standard deviations of 24.06° for normals and 0.21 for weights. For qualitative results, we refer to Fig.5 and Fig.6 in the main paper.

4.5. Opacity Distribution of Gaussians

For the scene of Fig. 8, 70% of Gaussians have opacity values in the range $[0.9, 1.0]$ and are aligned with the surface within a average distance of 0.016. The objects are located inside of the sphere with radius 2.

5. Additional Results

5.1. Novel-view Relighting Results

We compare our method with IRON [9], DPIR [2], GS³ [1] in quantitative manner. Table 2 demonstrates that we achieve the highest accuracy with surface normal estimation and the second highest accuracy with novel view relighting. We focus on reconstructing interpretable BRDFs and geometry by leveraging spatial coherence for sufficient light-view angular sampling, whereas GS³ prioritizes relighting by employing residual networks with inaccurate geometry. Our method achieves both accurate geometry and interpretable basis BRDF reconstruction that enable high-fidelity novel view relighting and intuitive scene editing.

Method	Synthetic photometric dataset			
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	MAE \downarrow
IRON	24.09	0.8714	0.0924	15.71
DPIR	26.87	0.954	0.0594	17.04
GS ³	38.93	0.9944	0.0038	15.28
Ours	31.78	0.9754	0.0234	9.81

Table 2. **Comparison between our method and other baselines.** We achieve high-fidelity novel-view relighting results with accurate geometry reconstruction.

5.2. Point Relighting Results

We achieve realistic relighting results with a moving point light source as shown in Figure 6.

5.3. Reflectance Editing Results

We achieve intuitive reflectance editing with interpretable basis BRDFs in Figure 7.

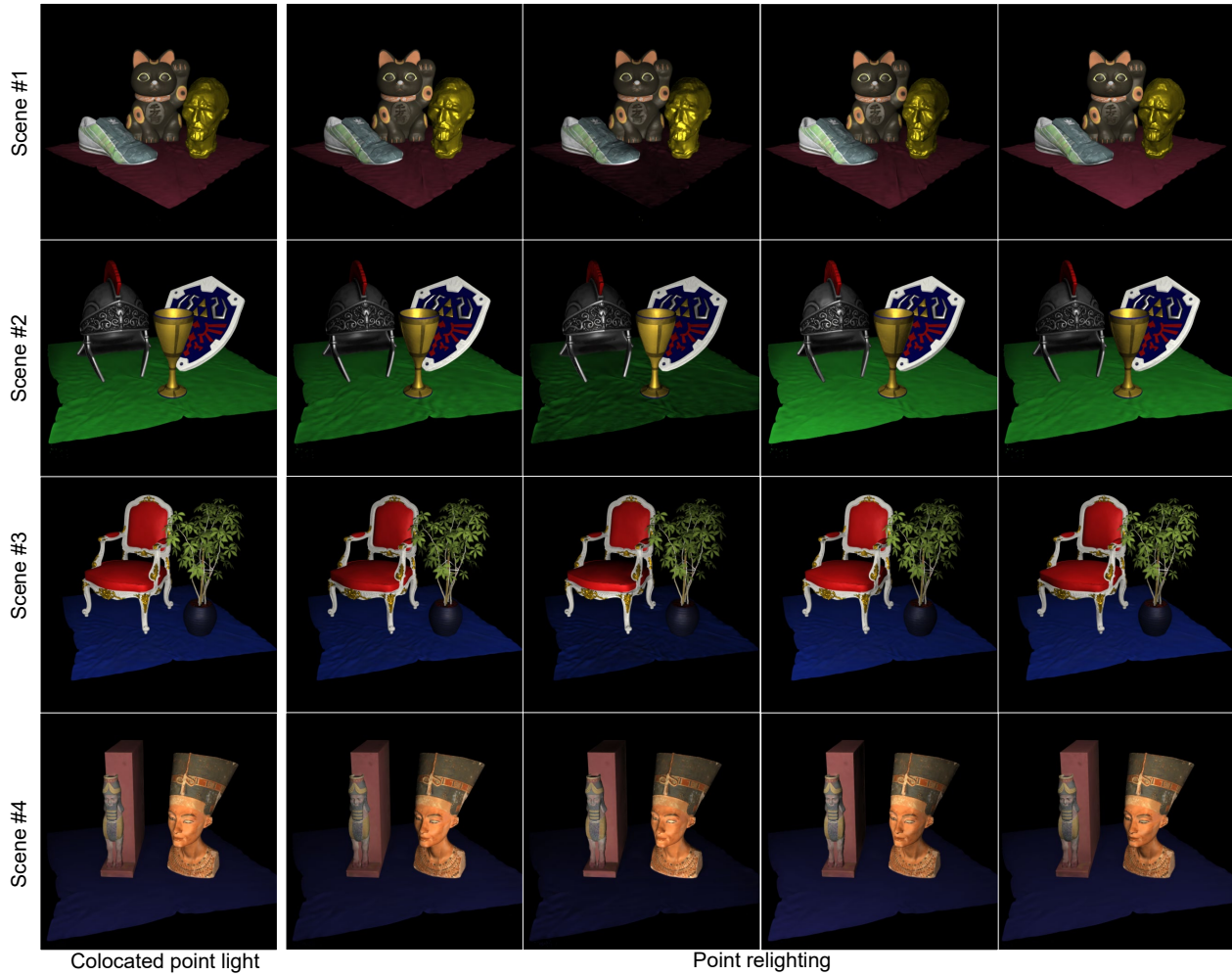


Figure 6. **Relighting with point light sources.** We render 4 complex scenes with a moving point light source.

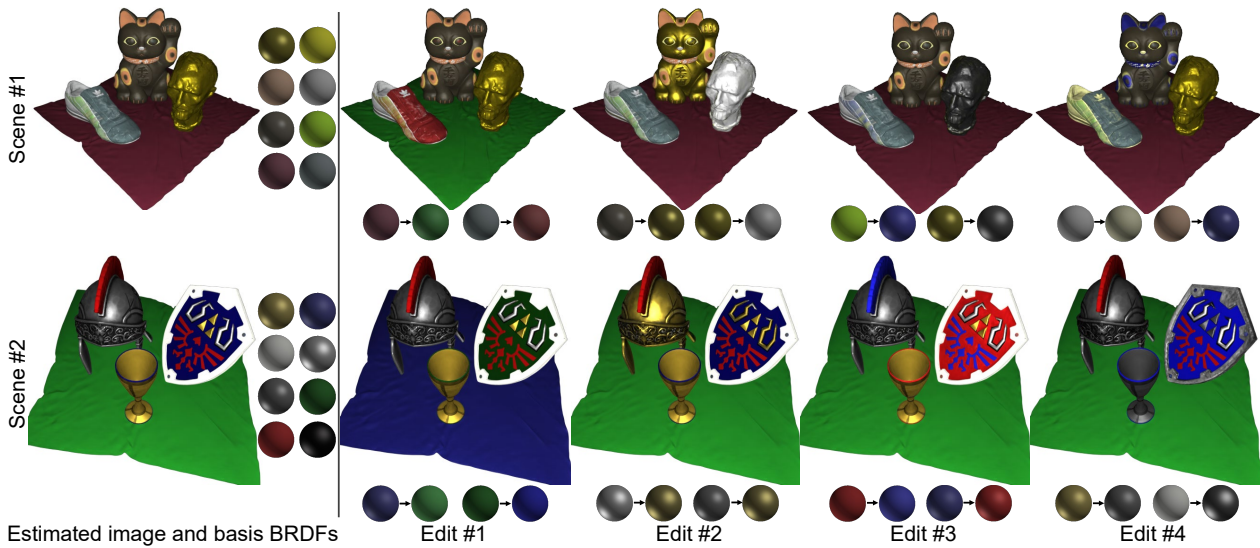


Figure 7. **Intuitive reflectance editing.** We efficiently edit reflectance by modifying parameters of basis BRDFs.

5.4. Additional Results with Synthetic Photometric Dataset

Figure 8, Figure 9, Figure 10 and Figure 11 show visualizations of 4 complex scenes in a synthetic photometric dataset. Our method reconstructs accurate geometry and interpretable basis BRDFs. We provide the visualization of a rendered image, estimated normal, depth map, basis BRDFs, basis BRDF images, and basis BRDF weight images. It demonstrates that our method successfully recovers complex geometry and spatially-varying BRDFs with interpretable basis BRDFs. Weight images are spatially separated which enables intuitive scene editing.

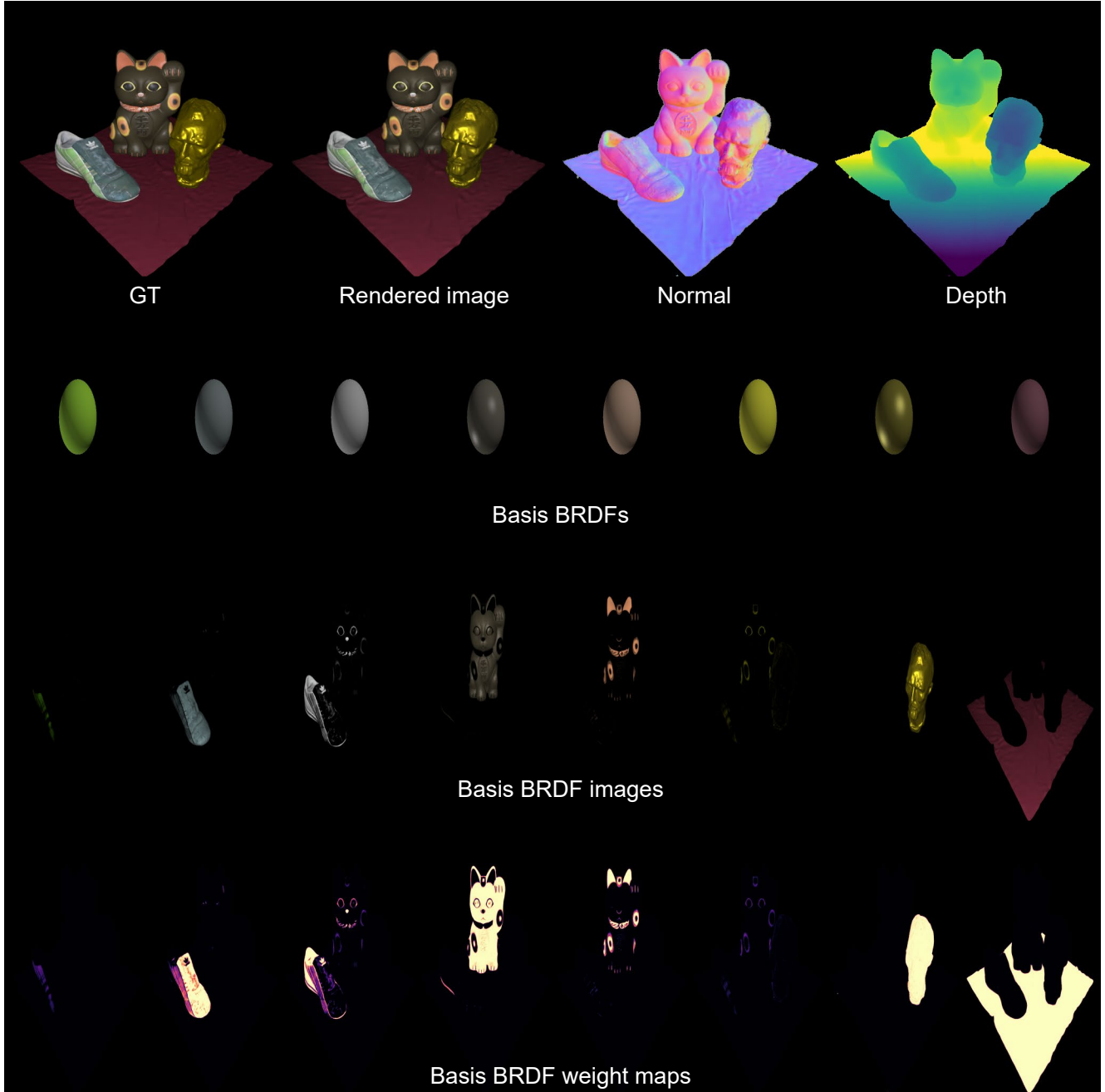


Figure 8. Reconstruction results of scene #1 on the synthetic photometric dataset.

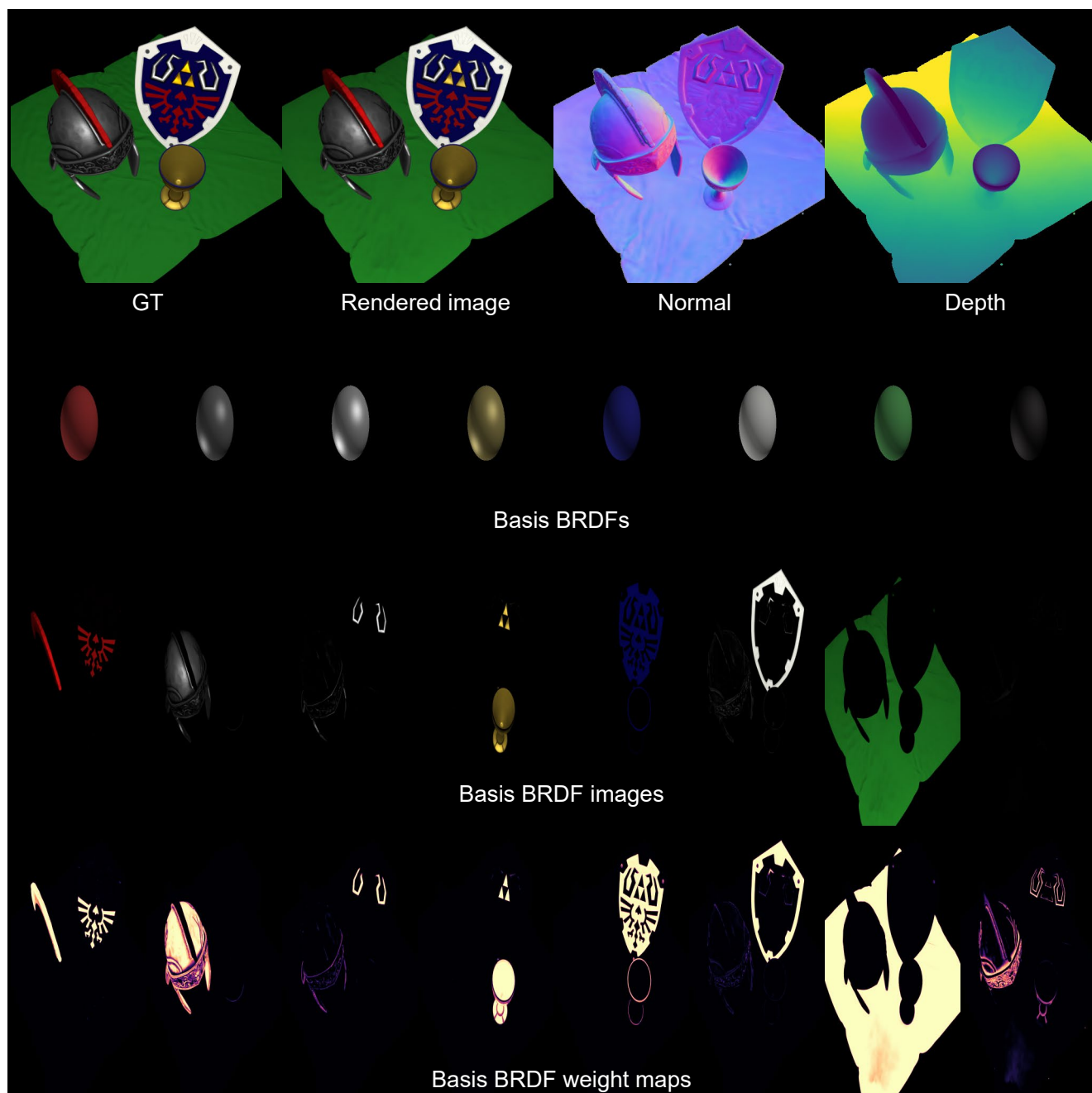


Figure 9. **Reconstruction results of scene #2 on the synthetic photometric dataset.**

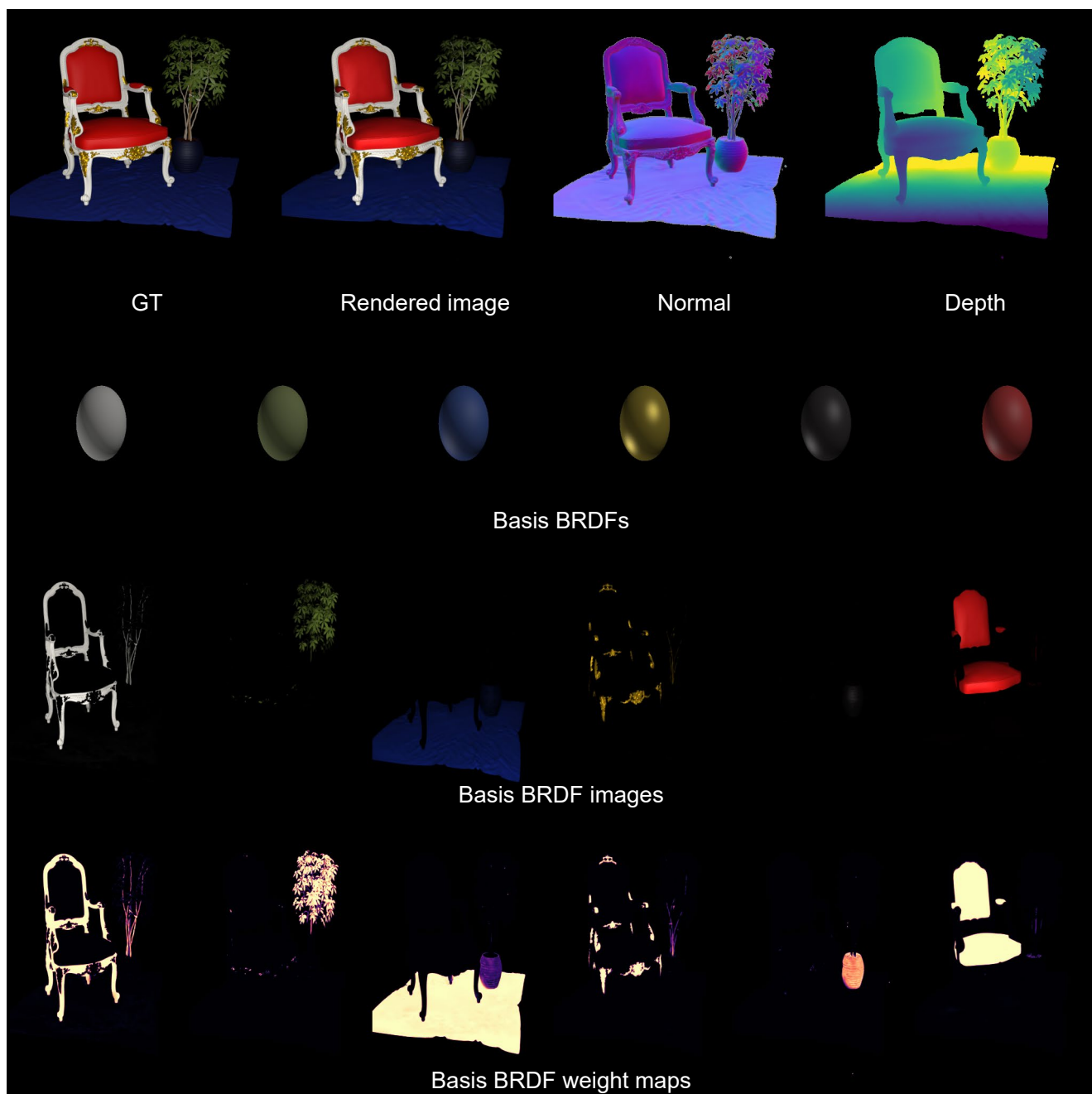


Figure 10. **Reconstruction results of scene #3 on the synthetic photometric dataset.**

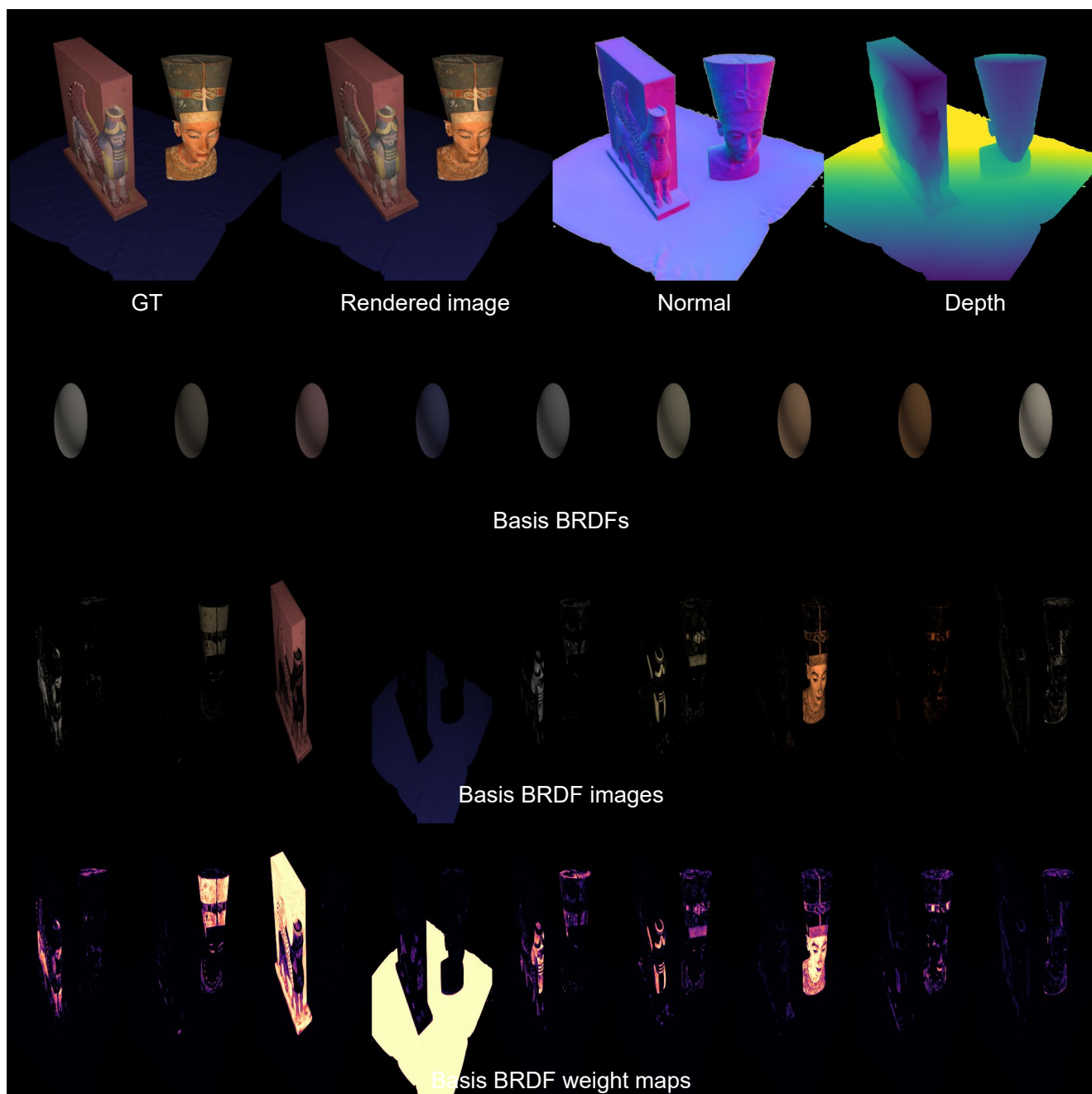


Figure 11. **Reconstruction results of scene #4 on the synthetic photometric dataset.**

5.5. Additional Results with Real-world Photometric Dataset

Figure 12 and Figure 13 show a visualization of a complex scene in a real-world photometric dataset. Our method achieves accurate reconstruction of geometry and interpretable basis BRDFs from multi-view flash images.

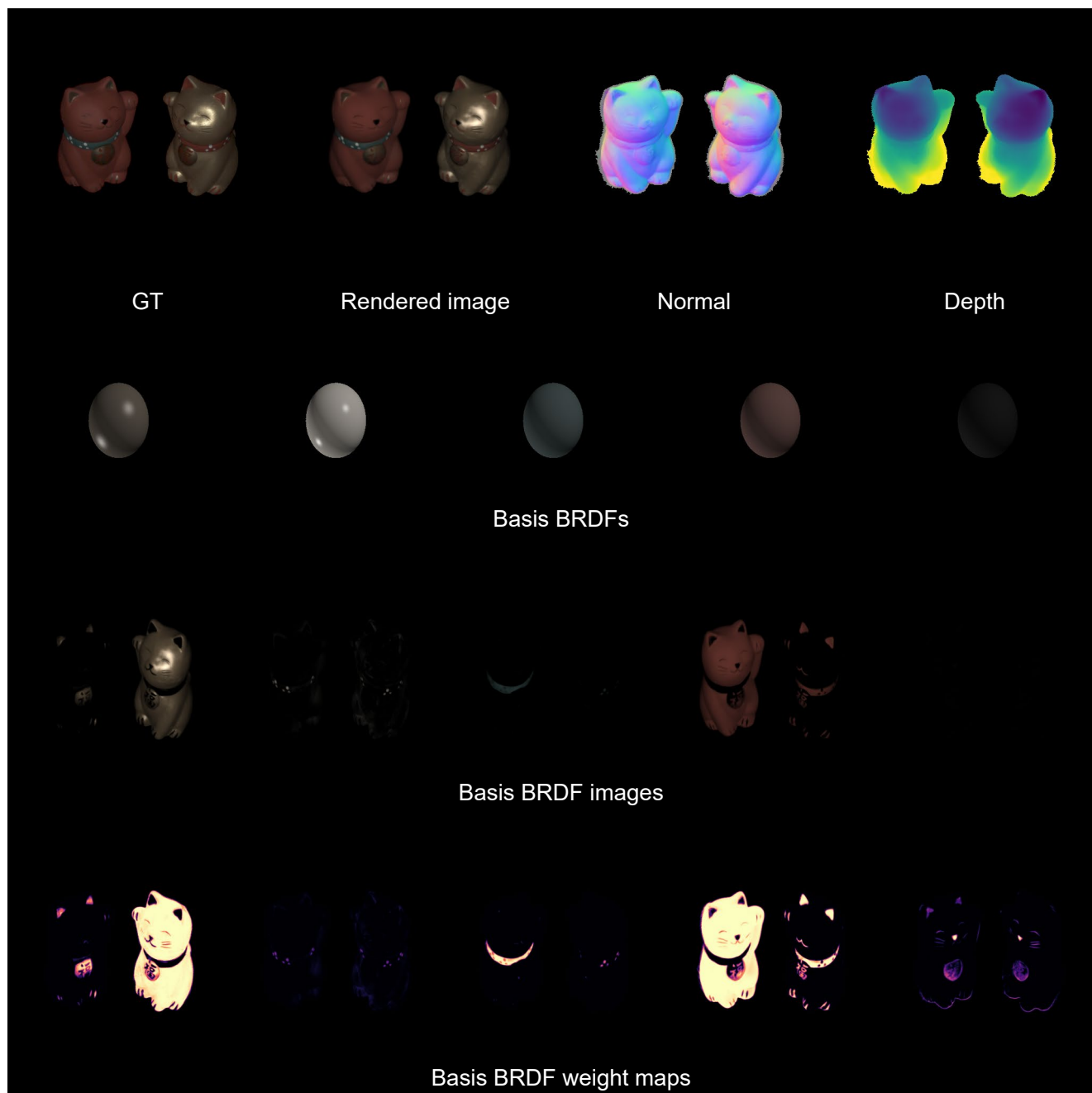


Figure 12. Reconstruction results of scene #1 on the real-world photometric dataset.

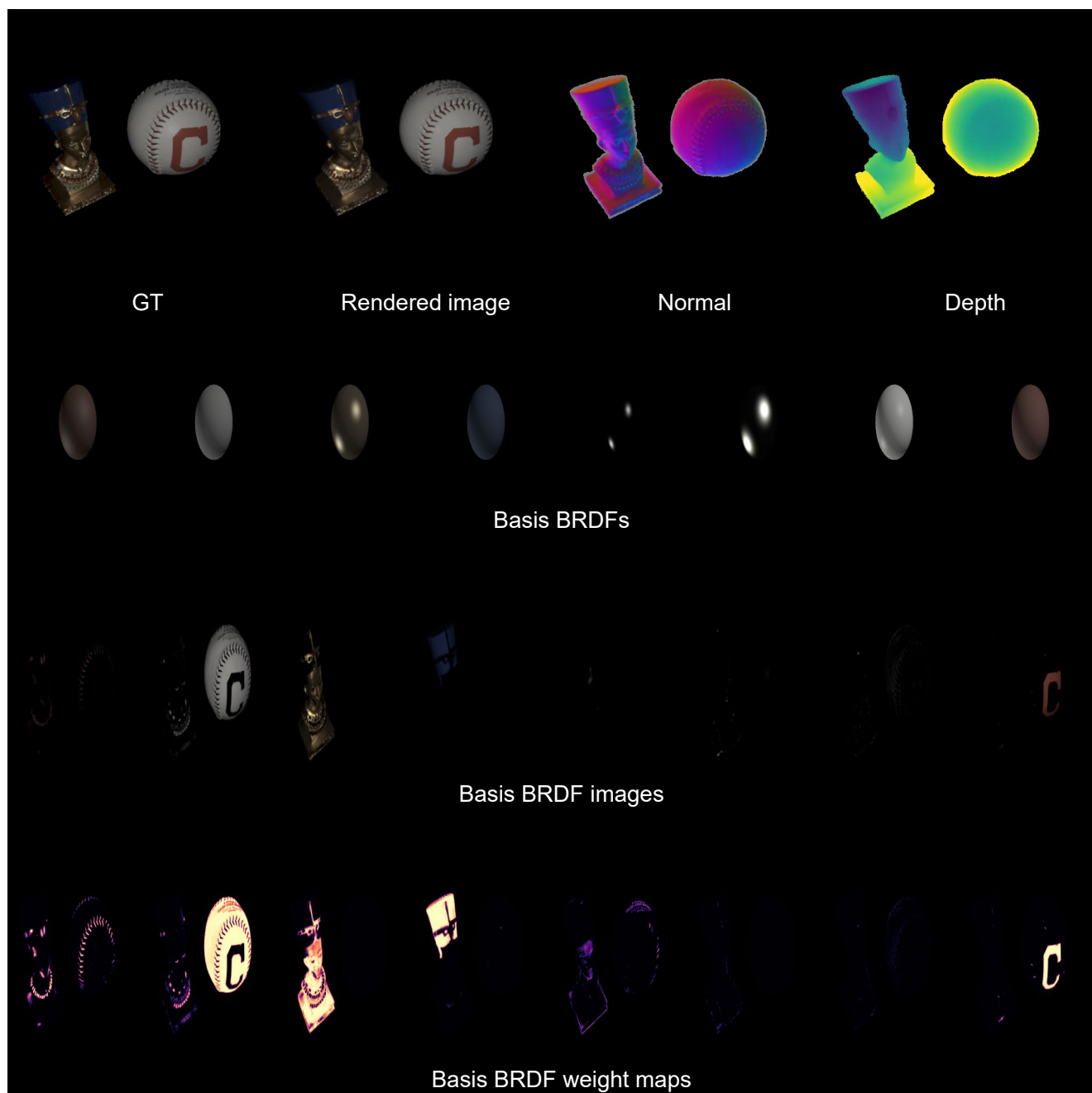


Figure 13. **Reconstruction results of scene #2 on the real-world photometric dataset.**

References

- [1] Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. Gs³: Efficient relighting with triple gaussian splatting. In *SIGGRAPH Asia 2024 Conference Papers*, 2024. 7
- [2] Hoon-Gyu Chung, Seokjun Choi, and Seung-Hwan Baek. Differentiable point-based inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4399–4409, 2024. 7
- [3] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13142–13153, 2023. 3
- [4] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 2, 3
- [5] Hendrik PA Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics (TOG)*, 22(2):234–257, 2003. 4
- [6] Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H Kim. Practical svbrdf acquisition of 3d objects with unstructured flash photography. *ACM Trans. Graph.*, 37(6):1–12, 2018. 3
- [7] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. Microfacet models for refraction through rough surfaces. *Rendering techniques*, 2007:18th, 2007. 2
- [8] 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 *European Conference on Computer Vision*, pages 700–716. Springer, 2022. 2
- [9] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5565–5574, 2022. 3, 7
- [10] Qiang Zhang, Seung-Hwan Baek, Szymon Rusinkiewicz, and Felix Heide. Differentiable point-based radiance fields for efficient view synthesis. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–12, 2022. 3