RobIR: Robust Inverse Rendering for High-Illumination Scenes – *Supplemental Material* –

Ziyi Yang[1] **Yanzhen Chen**[1] **Xinyu Gao**[1] **Yazhen Yuan**[2]
Yu Wu[2] **Xiaowei Zhou**[1] **Xiaogang Jin**[1†]

[1]Zhejiang University [2]Tencent

A Overview

This supplementary document provides some implementation details and further results that accompany the paper.

- Section B introduces the differences between the dataset used by our method and those used by previous methods.
- Section C introduces more details of the SG approximation for the rendering equation.
- Section D provides additional results, including more visualizations and results on more datasets.

B High-illumination Dataset

Currently, neural field-based inverse rendering methods, such as InvRender [11], NeRFactor [10], and TensoIR [5], generally use scenes with almost no high-intensity ambient light (See Fig. 3). The advantage of these scenes is that the object's BRDF estimation is not affected by self-occlusion shadows, making albedo and color quite similar. As a result, even if each part of the BRDF estimation is somewhat messy, plausible results can still be obtained. However, when the scene has intense illumination and shadows, these methods will fail to correctly model the object's BRDF. Therefore, to more accurately evaluate the robustness of inverse rendering, we choose a more challenging high-illumination dataset.

C SG Approximation for the Rendering Equation

Following the methodology from [9], we employ the inner product of SGs to approximate the computation of the rendering equation. The position $\mathbf{x}$ is dropped in the following equation due to the distant illumination assumption. Specifically, the term $\boldsymbol{\omega}_i \cdot \mathbf{n}$ is approximated by a SG as follows:

$$\boldsymbol{\omega}_i \cdot \mathbf{n} \approx G\left(\boldsymbol{\omega}_i; 0.0315, \mathbf{n}, 32.7080\right) - 31.7003. \tag{1}$$

As for the specular component f_s, we employ the simplified Disney BRDF model as previous methods [2, 6, 1]:

$$\begin{aligned} f_s\left(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i\right) &= \mathcal{M}\left(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i\right) \mathcal{D}(\mathbf{h}), \\ \mathbf{h} &= \frac{\boldsymbol{\omega}_o + \boldsymbol{\omega}_i}{\left\|\boldsymbol{\omega}_o + \boldsymbol{\omega}_i\right\|_2}, \end{aligned} \tag{2}$$

where $\mathcal{M}$ represents the Fresnel with shadowing effects, and $\mathcal{D}$ is the normalized distribution function.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

To simplify the computation, we assume an isotropic specular BRDF, and adapt $\mathcal{D}$ and $\mathcal{M}$ as follows:

$$
\begin{aligned}
\mathcal{M}\left(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i\right) &= \frac{\mathcal{F}\left(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i\right) \mathcal{G}\left(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i\right)}{4\left(\mathbf{n} \cdot \boldsymbol{\omega}_o\right)\left(\mathbf{n} \cdot \boldsymbol{\omega}_i\right)} \\
\mathcal{F}\left(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i\right) &= \boldsymbol{s} + (1 - \boldsymbol{s}) \cdot 2^{-(5.55473 \boldsymbol{\omega}_o \cdot \mathbf{h} + 6.8316)(\boldsymbol{\omega}_o \cdot \mathbf{h})}, \\
\mathcal{G}\left(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i\right) &= \frac{\boldsymbol{\omega}_o \cdot \mathbf{n}}{\boldsymbol{\omega}_o \cdot \mathbf{n}(1-k)+k} \cdot \frac{\boldsymbol{\omega}_i \cdot \mathbf{n}}{\boldsymbol{\omega}_i \cdot \mathbf{n}(1-k)+k}, \\
k &= \frac{(r+1)^2}{8}, \\
\mathcal{D}(\mathbf{h}) &= G\left(\mathbf{h}; \mathbf{n}, \frac{2}{r^4}, \frac{1}{\pi r^4}\right),
\end{aligned}
$$

where $s \in [0, 1]^3$ is the specular factor, and r denotes the roughness. Finally, we can compute the rendering equation through the fast inner product of SGs [8].

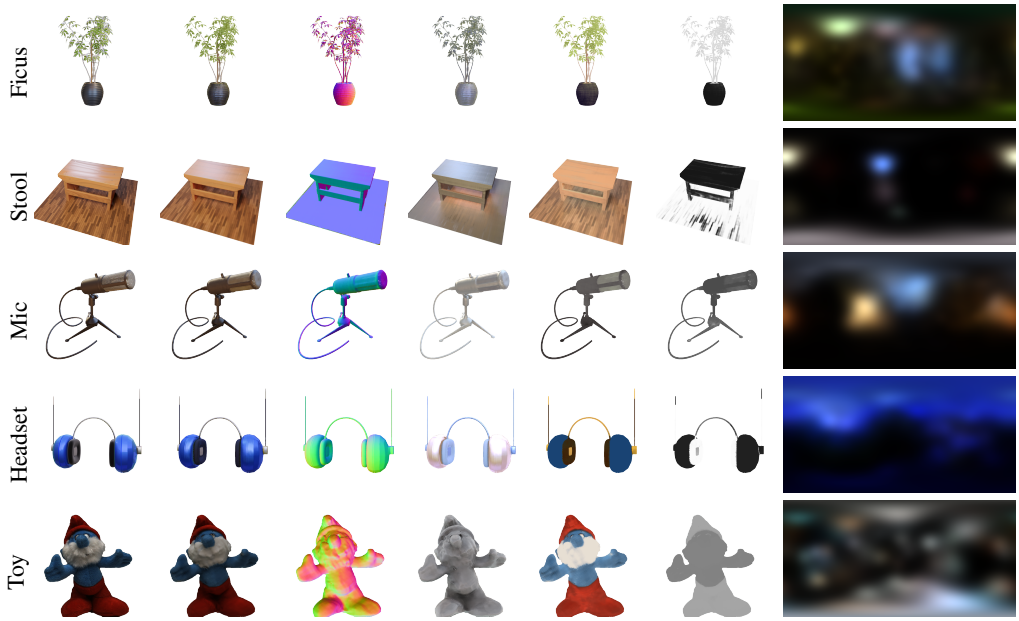

Figure 1: **Other results of our method.** In each scene, we present the input ground-truth image (a), our rendering result (b), normal (c), light (d), albedo (e), and roughness (f) obtained through our method. These experiments illustrate the generalizability of our method across diverse datasets and demonstrate its ability to produce high-quality results.

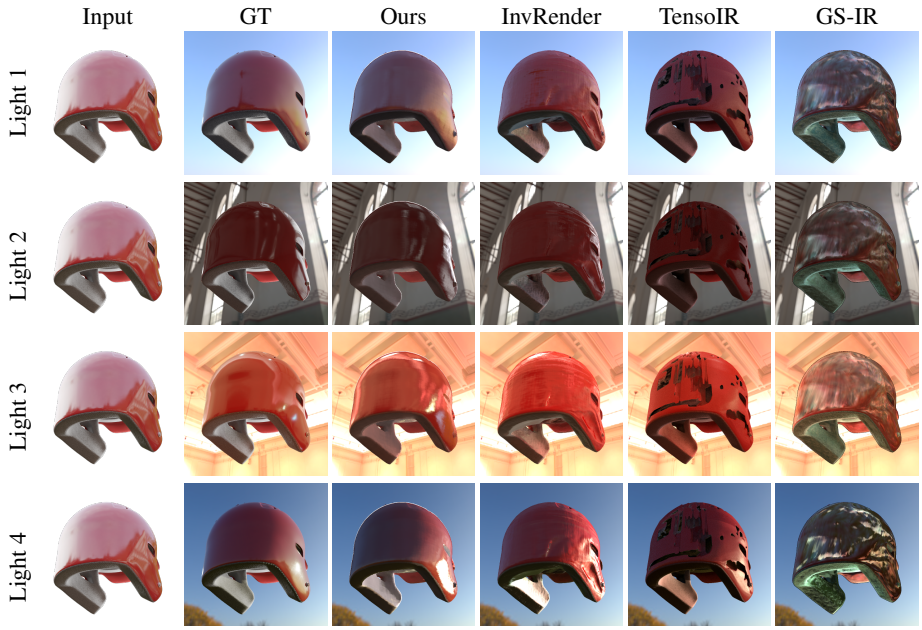

Figure 2: **Helmet Relighting.** Our method achieves high-quality relighting results in scenarios with specular highlights and slight specular reflections.

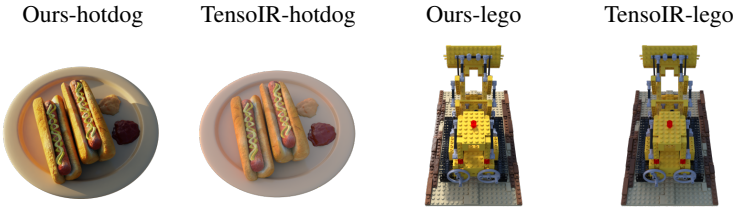

Figure 3: **Dataset Comparison.** We choose a more challenging high-illumination dataset, which exposed the inability of previous neural field-based inverse rendering methods to decouple shadows from the object's PBR materials.

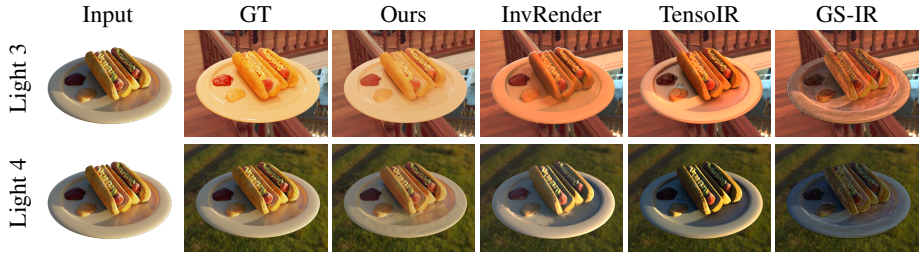

Figure 4: **Hotdog Relighting.** Our method achieves high-quality relighting results in scenarios with severe shadows.

Method	Hotdog			Lego			Helmet		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NVDiffrec	20.60	0.8872	0.1777	18.52	0.8299	0.1616	12.06	0.7866	0.2274
InvRender	15.76	0.8575	0.2029	20.75	0.8606	0.1656	19.50	0.8761	0.1697
TensoIR	16.01	0.8496	0.2047	20.74	0.8493	0.1541	16.95	0.8341	0.1759
Relightable-GS	15.34	0.8453	0.2111	20.07	0.8030	0.1580	14.97	0.7946	0.1963
GS-IR	9.72	0.6382	0.3139	13.03	0.6860	0.2386	13.72	0.7774	0.2538
Ours-no aces	22.24	0.8582	0.1779	22.00	0.8675	0.1497	19.98	0.9173	0.1202
Ours-no rve	19.04	0.8487	0.1489	20.17	0.8659	0.1437	14.30	0.8769	0.1493
Ours-Log	19.01	0.8570	0.1560	21.43	0.8596	0.1504	21.87	0.9078	0.1012
Ours	24.25	0.9185	0.0970	24.63	0.9175	0.1071	24.14	0.9427	0.1122

Method	Truck			Stool			Average		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NVDiffrec	19.59	0.9010	0.1437	13.69	0.7213	0.2722	16.89	0.8252	0.1965
InvRender	21.68	0.9023	0.1440	17.90	0.8822	0.1440	19.12	0.8757	0.1652
TensoIR	24.06	0.9375	0.1071	24.81	0.8692	0.1269	20.51	0.8679	0.1537
Relightable-GS	19.17	0.8720	0.1528	18.62	0.8567	0.1296	17.63	0.8343	0.1696
GS-IR	19.03	0.8379	0.1729	18.92	0.8695	0.1061	14.88	0.7618	0.2171
Ours-no aces	20.81	0.9177	0.1111	21.15	0.8647	0.1518	21.24	0.8851	0.1421
Ours-no rve	20.59	0.9200	0.1108	18.46	0.8814	0.1511	18.51	0.8786	0.1408
Ours-Log	24.04	0.9418	0.0997	19.29	0.8755	0.1395	21.13	0.8883	0.1294
Ours	27.46	0.9592	0.0647	24.98	0.9136	0.1051	25.09	0.9303	0.0972

Table 1: **Quantitative albedo comparison on synthetic dataset**. We compare our method to several previous approaches: NVDiffrec [4], InvRender [11], TensoIR [5], Relightable-GS [3] and GS-IR [7]. We report PSNR, SSIM, LPIPS(VGG) and color each cell as best, second best and third best.

D Additional Results

More qualitative results. Our method can effectively remove shadows baked into albedo and roughness, thanks to our accurate modeling of each decomposition component. Therefore, our method can certainly handle scenes with less intense lighting. Fig. 1 shows the results of our method on real-world datasets and some synthetic datasets, including scenes with shadows and specular, as well as diffuse objects. Our method can robustly perform inverse rendering in any situation without baking shadows and illumination into PBR materials.

Per-scene albedo results. We present the complete metrics of our method compared to other methods in Tab. 1. The estimated albedo in our method surpasses existing SOTA methods in every synthetic scene.

More relighting. We show more relighting results in Fig. 2 and Fig. 4. The two scenes demonstrate that our method can accurately estimate the BRDF of the object under scenes with specular highlights and severe shadows.

Additional comparison with NVDiffrecMC. We show more albedo, roughness, and environment map comparison with NVDiffrecMC [4] in Figs. 5-7.

Visualization and evaluation on tone mapping. We first visualize the vanilla ACES and sRGB tone mapping in Fig. 8, indicating that ACES curve has much wider input range. Then in Fig. 9, we show the scene-specific tone mapping curve with different γ, enabling the ACES curve to fit other settings with different tone mapping methods. Finally, we evaluate the optimized ACES curve (with $\gamma = 0.42$) in *chessboard* scene with GT tone mapping (sRGB) in Fig. 10. The results show that our scene-specific ACES tone mapping can stretch to sRGB curve, demostrating the effectiveness of our method.

References

[1] Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *European Conference on Computer Vision*, pages 294–311. Springer, 2020.

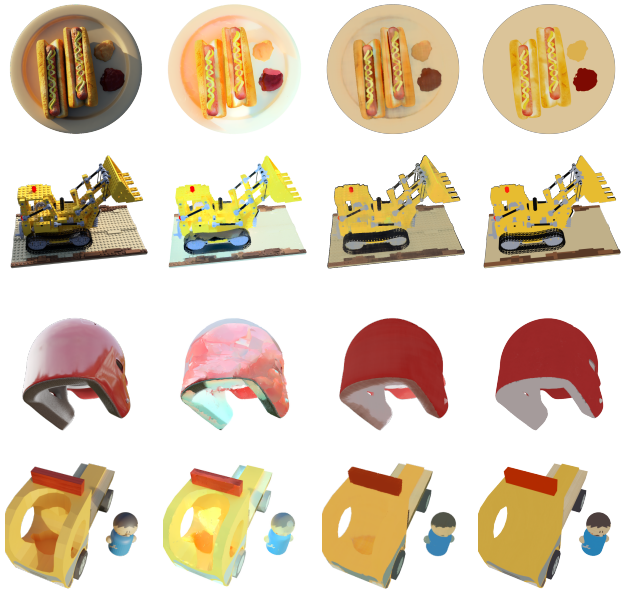

Figure 5: **Albedo comparison with NvDiffRecMC on synthetic scenes.** NvDiffRecMC cannot achieve the decouple of shadow, indirect illumination, and the PBR materials of the objects.

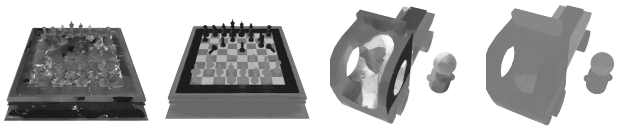

Figure 6: **Roughness comparison with NvDiffRecMC.** NvDiffRecMC cannot obtain high-quality roughness in high illumination scenes.

Figure 7: **Environment map comparison with NvDiffRecMC.**

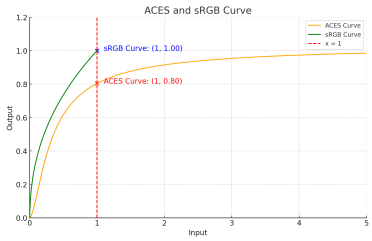

Figure 8: **Comparison on ACES and sRGB curve.**

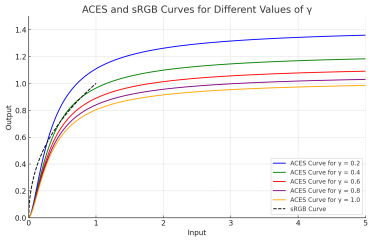

Figure 9: **Visualization of ACES tone mapping with different γ.**

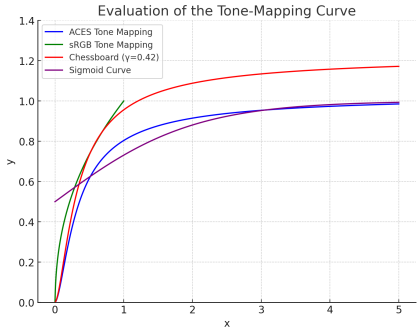

Figure 10: **Evaluation tone mapping in chessboard.** The ACES tone mapping with $\gamma = 0.42$ matches well with the sRGB curve.

[2] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *Acm Siggraph*, volume 2012, pages 1–7. vol. 2012, 2012.

[3] Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing. *arXiv:2311.16043*, 2023.

[4] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. Shape, Light, and Material Decomposition from Images using Monte Carlo Rendering and Denoising. *arXiv:2206.03380*, 2022.

[5] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensoir: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[6] Zhengqin Li, Zexiang Xu, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Learning to reconstruct shape and spatially-varying reflectance from a single image. *ACM Transactions on Graphics (TOG)*, 37(6):1–11, 2018.

[7] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering. *arXiv preprint arXiv:2311.16473*, 2023.

[8] Julian Meder and Beat D. Brüderlin. Hemispherical gaussians for accurate light integration. In *International Conference on Computer Vision and Graphics*, 2018.

[9] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5453–5462, 2021.

[10] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (TOG)*, 40(6):1–18, 2021.

[11] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18643–18652, 2022.