IRIS: Inverse Rendering of Indoor Scenes from Low Dynamic Range Images

Supplementary Material

Abstract

This supplementary document shows additional details of our method and more results. We refer readers to our webpage, which shows more results that allow for easy comparisons with the baseline methods on all scenes we use.

A. Relighting and Object Insertion Results

Our method estimates accurate surface material and spatially varying HDR illumination from LDR images, enabling various applications such as relighting and object insertion. We provide the qualitative results of real-world scenes in Figure 1, Figure 2, Figure 3, Figure 6, Figure 7, where we sample novel camera trajectories and render the scene at different time steps. The results demonstrate effective modeling of specular reflections on smooth surfaces (like 'mirror' and 'whiteboard') upon introducing new light sources. Moreover, our method accurately simulates inter-reflections between the scene and the inserted objects, significantly elevating the realism of object insertion. To summarize, we show that IRIS can render real-world scenes under various illumination from different viewpoints. For more interactive visualizations and comparisons, please check our supplementary webpage https://iris-ldr.github.io.

B. Additional Evaluation Results

In addition to physically-based inverse rendering techniques like FIPT, methods based on neural radiance fields (NeRF) [7] strive for scene disentanglement by representing indoor scenes' incident radiance fields with a 5D network [13] without constraints. Recent NeRF-based approaches like I^2-SDF [15], NeILF++ [14], and NeFII [11], much like FIPT, rely on pre-calculated irradiance, and focus on surface rendering to reconstruct scene materials and/or lighting. However, these methods typically account for only single-bounce light transport, leading to compromised quality in both material and lighting reconstruction. The complete metrics of inverse rendering are shown in Table 1 and the complete metrics of novel-view synthesis and relighting are listed in Table 2. Our method achieves comparable novel-view synthesis results and outperforms other baselines for relighting. The results underscore the effectiveness of our method in accurately decomposing intrinsic elements from LDR images. As for computational efficiency, the whole training takes 57 mins on a single RTX 4090, compared to 298 mins for NeILF [13] and 50 mins for FIPT [12].

Table 1. **BRDF-emission comparison on synthetic data.** FIPT-LDR* is provided with the GT emitter mask as additional input. The best metrics among LDR methods are highlighted in bold.

	Method	$\mathbf{k}_d$	$\mathbf{a}'$	σ	$\mathbf{L}_e$	
		PSNR ↑			IoU ↑	L2 ↓
Kitchen	Li et al [6]	15.75	12.64	10.15	0.43	1.410
	NeILF [13]	16.63	13.73	14.77	—	—
	FIPT-LDR*	15.77	8.97	5.94	**0.58**	0.450
	Ours	**23.22**	**17.52**	**20.35**	**0.58**	**0.203**
	FIPT-HDR [12]	34.34	27.05	24.55	0.88	0.010
Bedroom	Li et al [6]	18.90	15.10	11.38	0.34	2.784
	NeILF [13]	16.85	13.99	16.03	—	—
	FIPT-LDR*	18.38	9.60	5.82	**0.77**	0.245
	Ours	**26.44**	**20.95**	**26.47**	**0.77**	**0.043**
	FIPT-HDR [12]	28.98	25.86	23.53	0.92	0.004
Livingroom	Li et al [6]	16.78	14.71	11.42	0.17	3.610
	NeILF [13]	16.06	13.86	15.95	—	—
	FIPT-LDR*	11.59	8.93	4.08	**0.77**	0.240
	Ours	**18.09**	**15.45**	**25.28**	**0.77**	**0.103**
	FIPT-HDR [12]	28.42	27.47	30.44	0.95	0.005
Bathroom	Li et al [6]	15.50	13.60	12.24	0.45	1.351
	NeILF [13]	17.85	14.49	**21.09**	—	—
	FIPT-LDR*	16.21	11.46	4.12	**0.62**	0.187
	Ours	**21.56**	**17.74**	13.43	**0.62**	**0.135**
	FIPT-HDR [12]	28.06	23.54	26.97	0.68	0.080

C. Qualitative Results of Synthetic Scenes

To verify the effectiveness of inverse rendering, we compare IRIS with several baselines on synthetic scenes provided by FIPT [12], which provide ground-truth geometry, material properties, and lighting. Figure 4 shows the qualitative results of inverse rendering, including image reconstruction, material reflectance $\mathbf{a}'$, roughness σ, and emission maps. While NeILF [13] achieves accurate rendering, it bakes significant shading effects into its diffuse albedo map. Li et al. [6] generate a noisy BRDF from a single image input. FIPT* tends to underestimate illumination intensity, overestimating the reflectance $\mathbf{a}'$ as compensation. In contrast, our method successfully recovers high-quality HDR emission from LDR input, resulting in precise intrinsic decomposition.

D. Additional Ablation Study

Our method explicitly models the HDR–LDR conversion and estimates the CRF from input images, and thus achieves better inverse rendering quality. To further validate the design choices, we conduct an ablation study on the CRF modeling strategy and evaluate inverse rendering from input images with varying exposure levels, which is collected with the strategy described in Appendix F. We visualize the CRFs

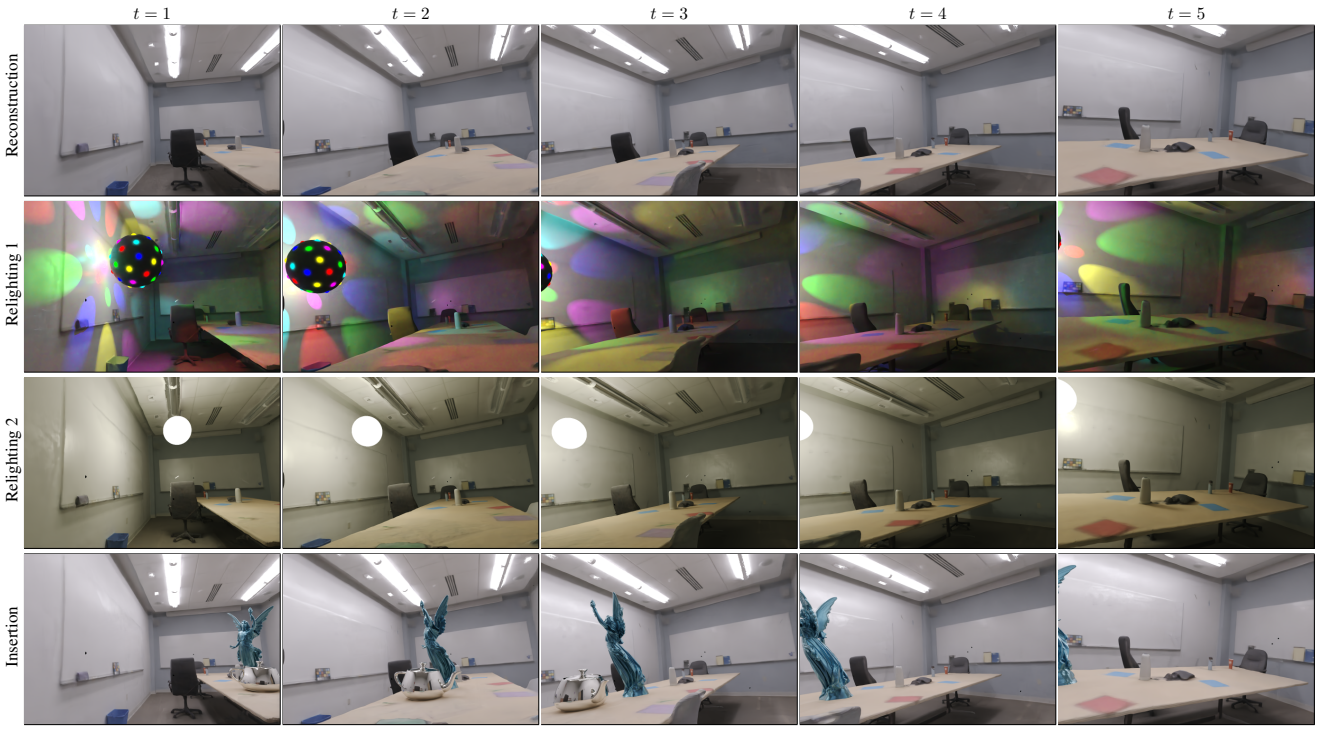

Figure 1. **Relighting and object insertion in 'conference room'.** The inserted new light sources are reflected on the whiteboard surface, demonstrating the accuracy of the material estimation of IRIS.

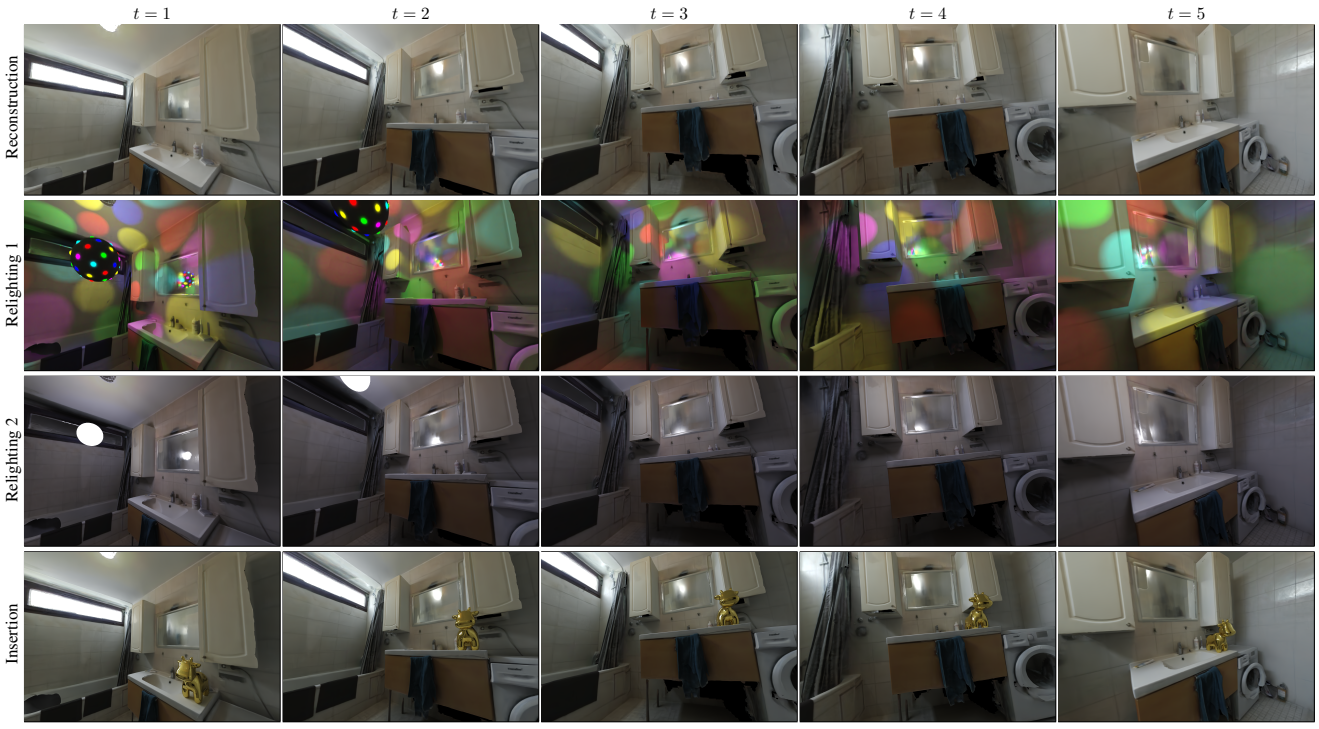

Figure 2. **Relighting and object insertion in 'bathroom'.** The mirror is estimated as a low-roughness surface, and it reflects the new light sources and enhances the realism of relighting significantly. The inserted object also exhibits reflection of HDR lighting recovered by IRIS.

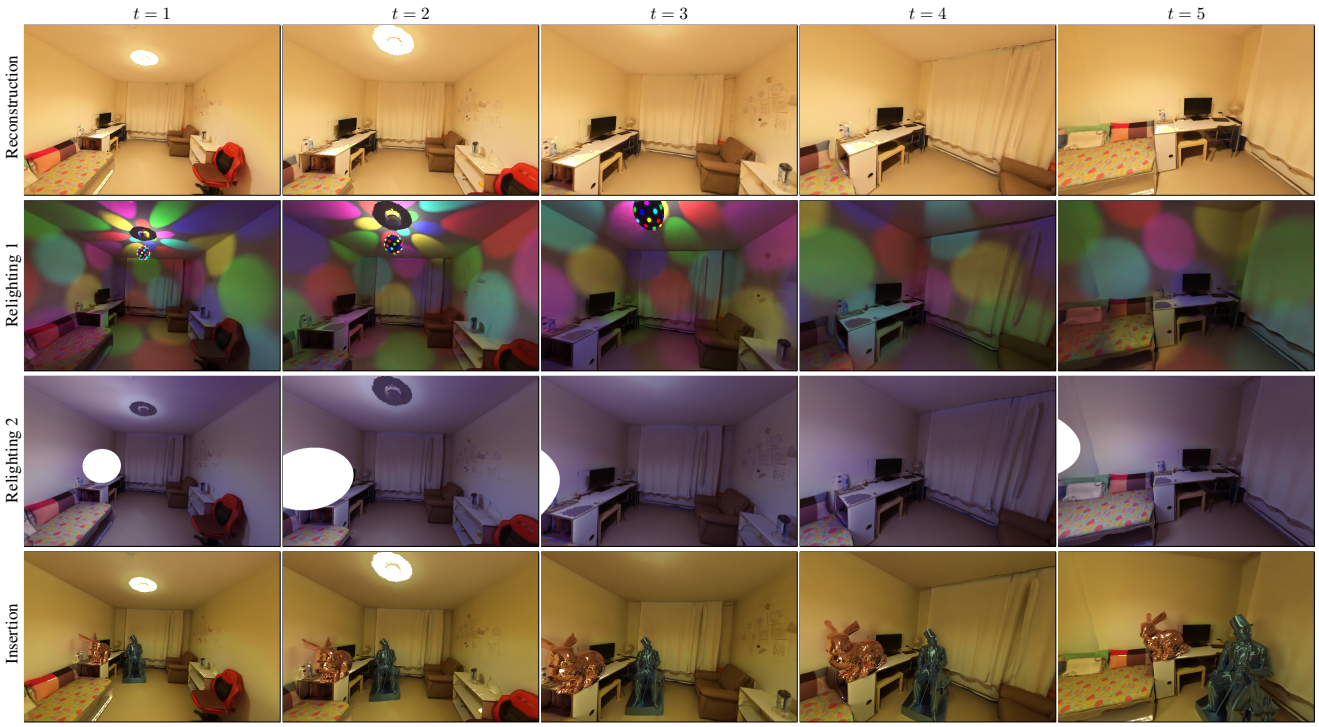

Figure 3. **Relighting and object insertion in 'bedroom'.** The Disco ball rotates and casts colorful lights in different directions, creating realistic relighting results in the real-world scene.

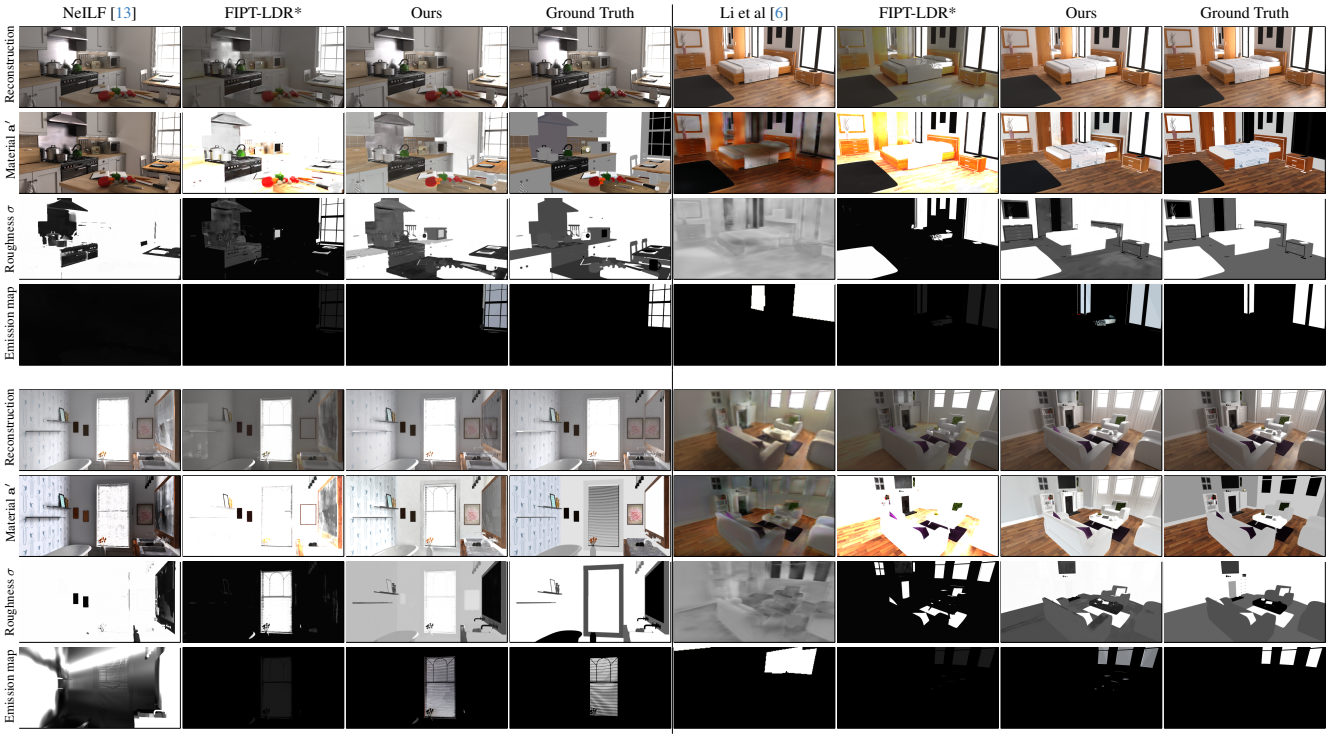

Figure 4. **Intrinsic decomposition of synthetic scenes [12].** From top to bottom, we show reconstruction, material reflectance $\mathbf{a}'$, roughness σ, and emission maps. For the emission map, we show normalized HDR emission, such that it is not saturated and differences become visible. With LDR images as input, IRIS successfully recovers the HDR lighting and accurate surface material.

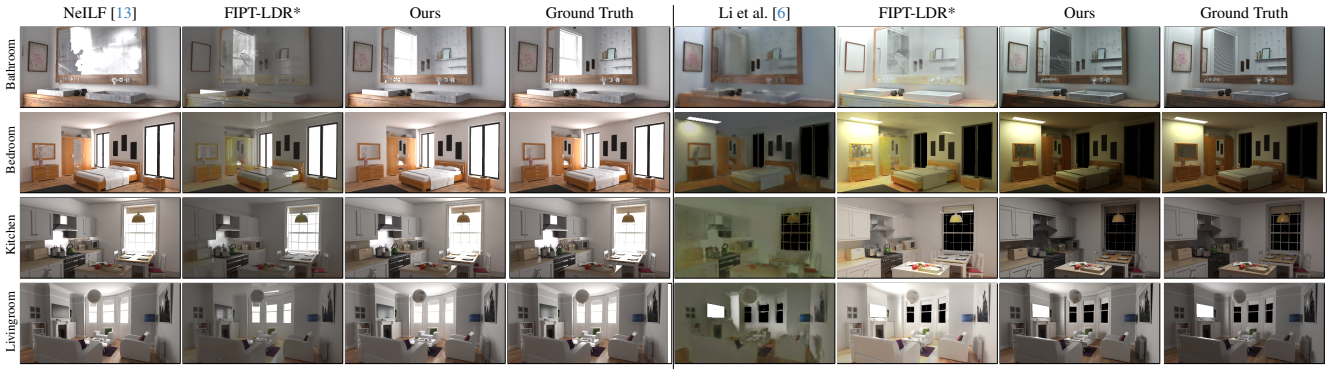

Figure 5. **Novel-view synthesis and relighting results on synthetic scenes [12].** The novel view synthesis results are shown in the left four columns, and the relighting of the same novel view are shown in the right four columns.

Table 2. **Complete quantitative results of novel view synthesis and relighting on synthetic scenes**

	Method	Kitchen PSNR ↑	Kitchen SSIM ↑	Kitchen LPIPS ↓	Bedroom PSNR ↑	Bedroom SSIM ↑	Bedroom LPIPS ↓	Livingroom PSNR ↑	Livingroom SSIM ↑	Livingroom LPIPS ↓	Bathroom PSNR ↑	Bathroom SSIM ↑	Bathroom LPIPS ↓
NVS	NeILF [13]	29.309	0.910	**0.187**	**29.651**	**0.944**	0.095	**34.653**	0.959	0.099	26.509	0.783	0.339
	I^2-SDF [15]	24.993	0.898	0.234	25.845	0.916	0.150	27.955	**0.962**	**0.091**	24.967	0.698	0.483
	FIPT-LDR*	16.372	0.776	0.381	14.536	0.784	0.389	16.146	0.805	0.361	13.665	0.609	0.616
	Ours	**29.730**	**0.916**	0.192	28.765	0.940	**0.094**	31.368	0.954	0.104	**28.008**	**0.802**	**0.335**
	FIPT-HDR [12]	29.059	0.924	0.180	27.670	0.940	0.095	28.524	0.951	0.109	29.788	0.792	0.358
Relight	Li22 [6]	21.755	0.815	0.381	23.662	0.851	0.342	**21.631**	0.841	0.395	22.887	0.747	0.475
	FIPT-LDR*	11.932	0.715	0.283	13.132	0.701	0.334	9.198	0.710	0.345	12.240	0.694	0.473
	Ours	**23.818**	**0.873**	**0.143**	**25.483**	**0.892**	**0.166**	18.478	**0.906**	**0.127**	**23.664**	**0.856**	**0.254**
	FIPT-HDR [12]	27.597	0.886	0.115	28.411	0.878	0.155	32.543	0.964	0.078	27.497	0.881	0.208

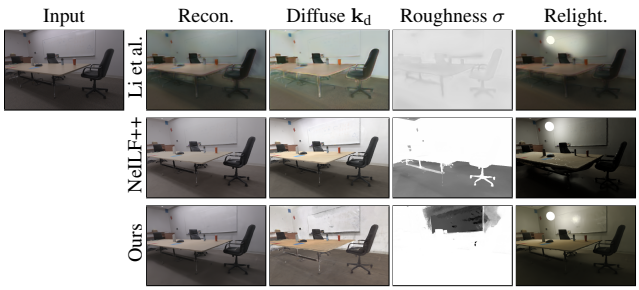

Figure 6. **Comparison with Li et al [6] and NeILF++ [14].**

Figure 7. **Relighting and object insertion in kitchen.** From top to bottom, we visualize the reconstruction (1st row), relighting (2nd), and object insertion (3rd).

Figure 8. **Visualizing albedo a during the training.** We show that leveraging data-driven IRISformer [16] estimation (left) provides us good albedo initialization (center), and final result is refined with physically-based rendering model.

Table 3. **Ablation of CRF modeling.**

Method	PSNR ↑ $\mathbf{k}_d$	PSNR ↑ $\mathbf{a}'$	PSNR ↑ σ	L_2 ↓ CRF
Constant exposure	24.24	19.11	**27.42**	4.074
Mean CRF $\bar{g}$	23.61	19.55	15.25	4.240
Gamma $1/2.2$	23.65	20.05	15.72	3.683
Full model	**26.82**	**23.43**	26.63	**1.363**

estimation with different modeling techniques in Figure 9, corresponding to Table 3. The results show that from input images captured with varying exposure, our method can

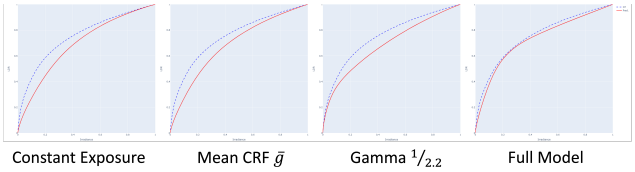

Figure 9. **CRF comparison visualization.** The blue dash lines are the ground-truth CRF, and the red lines are the estimated CRF after the optimization. We compare with three variants of CRF modeling settings. We show that the full model with varying exposure and learnable CRF model can approximate the ground truth quite well.

recover ground-truth CRF, demonstrating the effectiveness and importance of CRF modeling. We parametrize the CRF as a continuous and monotonically increasing function across the domain $(0, 1)$, sample 1024 points between 0 and 1, and calculate the L2 distance between the function values and the ground truth. We compare with three CRF alternatives: (1) constant exposure input, (2) Mean CRF $\bar{g}$ (the mean CRF from 201 empirical CRF functions measured in the real world [2]), and (3) Gamma $1/2.2$ ($g(x) = x^{1/2.2}$, as used in FIPT [12]). Our method outperforms the single exposure approach, suggesting the benefits of using varying exposure values to enhance dynamic range. It also achieves better results than constant CRF functions, justifying joint CRF optimization's merits.

E. Factorized Light Transport

We follow the rendering equation [3] to model physically-based light transport for realistic rendering:

$$\mathbf{L}_\mathrm{o}(\mathbf{x}, \boldsymbol{\omega}_\mathrm{o}) = \mathbf{L}_\mathrm{e}(\mathbf{x}, \boldsymbol{\omega}_\mathrm{o}) + \int_{\Omega^+} \mathbf{L}_\mathrm{i}(\mathbf{x}, \boldsymbol{\omega}_\mathrm{i}) f(\mathbf{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) d\boldsymbol{\omega}_\mathrm{i}, \quad (1)$$

where $\mathbf{L}_\mathrm{o}$ is the radiance observed along a ray $(\mathbf{x}, \boldsymbol{\omega}_\mathrm{o})$ for a 3D position $\mathbf{x}$ and a direction $\boldsymbol{\omega}_\mathrm{o}$, $\mathbf{L}_\mathrm{e}$ is the emission term, $\mathbf{L}_\mathrm{r} = \int_{\Omega^+} \mathbf{L}_\mathrm{i}(\mathbf{x}, \omega_\mathrm{i}) f(\mathbf{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) d\boldsymbol{\omega}_\mathrm{i}$ is the reflectance term, and $f(\mathbf{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o})$ is the BRDF. While $\mathbf{L}_\mathrm{i}$ encapsulates recursive incident radiance computation, we represent spatially-varying materials using the Cook–Torrance BRDF [1]:

$$f(\mathbf{x}, \boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}) = \frac{\mathbf{k}_\mathrm{d}(\mathbf{x})}{\pi} (\mathbf{n} \cdot \boldsymbol{\omega}_\mathrm{i})_+ + \frac{F \cdot D \cdot G}{4(\mathbf{n} \cdot \boldsymbol{\omega}_\mathrm{o})}, \quad (2)$$

where $D(\mathbf{h}, \mathbf{n}, \sigma(\mathbf{x}))$ describes the distribution of microfacet orientations, $G(\boldsymbol{\omega}_\mathrm{i}, \boldsymbol{\omega}_\mathrm{o}, \mathbf{n}, \sigma(\mathbf{x}))$ encodes the masking and shadowing effects between microfacets, and $F(\boldsymbol{\omega}_\mathrm{i}, \mathbf{h}, \mathbf{k}_\mathrm{s}(\mathbf{x}))$ is the Fresnel reflection term. The recursive integral in Equation (1) is computationally expensive and usually approximated with Monte–Carlo path tracing [3, 5] with multiple bounces. The rendering equation can be accelerated by factorizing the BRDF term from the integral [4, 9, 10]:

$$\mathbf{L}_\mathrm{r}(\mathbf{x}, \boldsymbol{\omega}_\mathrm{o}) = \mathbf{k}_\mathrm{d} \mathbf{L}_\mathrm{d}(\mathbf{x}) + \mathbf{k}_\mathrm{s} \mathbf{L}_\mathrm{s}^0(\mathbf{x}, \boldsymbol{\omega}_\mathrm{o}, \sigma) + \mathbf{L}_\mathrm{s}^1(\mathbf{x}, \boldsymbol{\omega}_\mathrm{o}, \sigma), \quad (3)$$

Table 4. **Notation table.**

Symbol	Description
$(\cdot)_+$	dot product clamped to positive value
$\boldsymbol{\omega}_\mathrm{i}$	incident light direction
$\boldsymbol{\omega}_\mathrm{o}$	outgoing light direction
$\mathbf{h}$	half vector $(\boldsymbol{\omega}_\mathrm{i} + \boldsymbol{\omega}_\mathrm{o})/\|\boldsymbol{\omega}_i + \boldsymbol{\omega}_o\|_2$
$\mathbf{n}$	surface normal
$\mathbf{x}$	3D position
$\mathbf{a}(\mathbf{x})$	surface albedo (base color)
$m(\mathbf{x})$	surface metallicness
$\sigma(\mathbf{x})$	surface roughness
$\mathbf{k}_\mathrm{d}(\mathbf{x})$	diffuse reflectivity $\mathbf{a}(\mathbf{x})(1 - m(\mathbf{x}))$
$\mathbf{k}_\mathrm{s}(\mathbf{x})$	specular reflectivity $\mathbf{a}(\mathbf{x})m(\mathbf{x}) + 0.04(1 - m(\mathbf{x}))$
$D(\cdot)$	GGX normal distribution
$F(\cdot)$	Schlick's approximation of Fresnel coefficients
$G(\cdot)$	Geometry (shadow-masking) term

where we decompose the reflectance term into a diffuse shading term $\mathbf{L}_\mathrm{d}(\mathbf{x}) = \int_{\Omega^+} \mathbf{L}_\mathrm{i}(\mathbf{x}, \boldsymbol{\omega}_\mathrm{i}) \frac{(\mathbf{n} \cdot \boldsymbol{\omega}_\mathrm{i})_+}{\pi} d\boldsymbol{\omega}_\mathrm{i}$, as well as two specular terms

$$\mathbf{L}_\mathrm{s}^0(\mathbf{x}, \boldsymbol{\omega}_\mathrm{o}, \sigma) = \int_{\Omega^+} \mathbf{L}_\mathrm{i}(\mathbf{x}, \boldsymbol{\omega}_\mathrm{i}) \frac{F_0 DG}{4(\mathbf{n} \cdot \boldsymbol{\omega}_\mathrm{o})} d\boldsymbol{\omega}_\mathrm{i}, \quad (4)$$

$$\mathbf{L}_\mathrm{s}^1(\mathbf{x}, \boldsymbol{\omega}_\mathrm{o}, \sigma) = \int_{\Omega^+} \mathbf{L}_\mathrm{i}(\mathbf{x}, \boldsymbol{\omega}_\mathrm{i}) \frac{F_1 DG}{4(\mathbf{n} \cdot \boldsymbol{\omega}_\mathrm{o})} d\boldsymbol{\omega}_\mathrm{i}, \quad (5)$$

where $F_0 = 1 - F_1$ and $F_1 = (1 - \mathbf{h} \cdot \boldsymbol{\omega}_\mathrm{i})^5$. $\mathbf{k}_\mathrm{d}(\mathbf{x})$ is diffuse reflectance and $\mathbf{k}_\mathrm{s}(\mathbf{x})$ is specular reflectance calculated from the BRDF. $\mathbf{L}_\mathrm{s}^*$ is further approximated by linearly interpolating the shading maps pre-computed at various roughness σ levels: $\mathbf{L}_\mathrm{s}^*(\cdot, \sigma) = \mathtt{LERP}(\{\mathbf{L}_\mathrm{s}^*(\cdot, \sigma_i)\}_{i=1}^6, \sigma)$, where $\{\sigma_i\}_{i=1}^6$ is uniformly sampled between $(0, 1)$. With the factorization formulation, the shading maps $\mathbf{L}_\mathrm{d}$, $\{\mathbf{L}_\mathrm{s}^0(\cdot, \sigma_i), \mathbf{L}_\mathrm{s}^1(\cdot, \sigma_i)\}_{i=1}^6$ can be pre-computed and allow for more efficient and stable optimization of material properties and HDR lighting.

F. Implementation Details

To clarify the equations in the paper, we describe the mathematical expressions and associated physical meanings in Table 4.

Varying exposure data generation. In real-world photography pipelines, exposure levels are adjusted by manipulating camera settings, such as shutter speed, aperture size, and ISO, to capture bright and dark regions. While the FIPT dataset [12] assumes single exposure and utilizes a simplistic camera response function (CRF) model defined as $\mathtt{CRF}(x) = x^{1/2.2}$, our approach simulates a capturing process that is both more realistic and challenging. To create LDR images of synthetic scenes for CRF metric calculation, we split the HDR images of the same scene into five exposure levels $\{\Delta t_i\}_{i=1}^5$, s.t. $\Delta t_i < \Delta t_{i+1}$, where the brightest HDR image corresponds to Δt_0, and conversely, the darkest

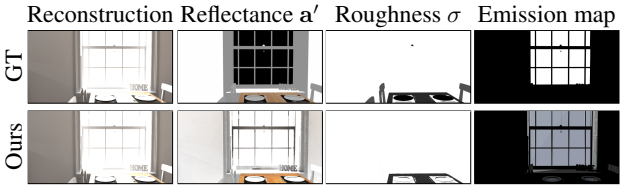

Figure 10. **Failure cases.**

to Δt_5, effectively mimicking an auto-exposure mechanism. Subsequently, we apply each exposure level to the HDR image and convert it into LDR format with the CRF derived from real-world sensors [2].

Direct illumination $\mathbf{L}_{\mathrm{e}}(\mathbf{x})$. We first identify the mesh faces $\{\mathbf{f}_i\}$ of emitters with the emitter mask $M_{\mathrm{e}}(\mathbf{x})$ defined on the mesh faces. We associate a learnable 3-dimensional parameter for each face: $\mathbf{e}(\mathbf{f}) \in \mathbb{R}^3$, representing the emitted light radiance. These parameters are then optimized during the HDR emission restoration phase.

BRDF. The surface material is represented as a neural field: $(\mathbf{a}, m, \sigma) = \mathbf{F}(\mathbf{x})$, the model architecture of which is based on Instant-NGP [8].

Shading Baking. Intuitively, the ray tracing continues if it encounters a non-emissive, specular surface (identified by a roughness threshold of 0.6), and stops otherwise. The radiance at the endpoint adheres to Eq. 11 in the main paper. The view-independent term $\mathbf{L}_{\mathrm{SLF}}(\mathbf{x})$ effectively approximates the global illumination on diffuse surfaces, which also expedites the rendering process. This is because rays typically reach diffuse surfaces within a few bounces, eliminating the need for further path tracing:

$$\begin{aligned} \mathbf{L}_{\mathrm{i}}(\mathbf{x}, \boldsymbol{\omega}) &= \mathbf{L}_{\mathrm{end}}(\mathbf{x}_n) \prod_{i=1}^{n-1} f(\mathbf{x}_{i+1} \rightarrow \mathbf{x}_i) \\ \text{s.t. } \sigma(\mathbf{x}_i) &\leq 0.6, M_{\mathrm{e}}(\mathbf{x}_i) = 0, \forall i < n, \end{aligned} \tag{6}$$

where $\{\mathbf{x}_i\}_{i=1}^{n}$ are the intersected points along the paths.

G. Limitations

Our emitter mask estimation may be inaccurate, especially when images are largely saturated. An incorrect mask cannot be recovered from as masks are not further optimized. Our CRF model is global, and it cannot capture complex non-local tone-mapping or while-balance changes. Addressing these issues would allow for a truly practical method for inverse rendering, which is left for future work.

References

[1] Robert L Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (ToG)*, 1982. 5

[2] Michael D. Grossberg and Shree K. Nayar. Modeling the space of camera response functions. *IEEE TPAMI*, 26(10): 1272–1282, 2004. 5, 6

[3] James T Kajiya. The rendering equation. In *SIGGRAPH*, 1986. 5

[4] Jaroslav Krivánek and Pascal Gautron. *Practical global illumination with irradiance caching*. Springer, 2022. 5

[5] Eric Lafortune. *Mathematical models and Monte Carlo algorithms for physically based rendering*. PhD thesis, Katholieke Universiteit Leuven, 1996. 5

[6] Zhengqin Li, Jia Shi, Sai Bi, Rui Zhu, Kalyan Sunkavalli, Miloš Hašan, Zexiang Xu, Ravi Ramamoorthi, and Manmohan Chandraker. Physically-based editing of indoor scene lighting from a single image. In *ECCV*, 2022. 1, 3, 4

[7] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1

[8] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM TOG*, 2022. 6

[9] Dario Seyb, Peter-Pike Sloan, Ari Silvennoinen, Michał Iwanicki, and Wojciech Jarosz. The design and evolution of the UberBake light baking system. *ACM Transactions on Graphics (TOG)*, 2020. 5

[10] Zian Wang, Jonah Philion, Sanja Fidler, and Jan Kautz. Learning indoor inverse rendering with 3D spatially-varying lighting. In *ICCV*, 2021. 5

[11] Haoqian Wu, Zhipeng Hu, Lincheng Li, Yongqiang Zhang, Changjie Fan, and Xin Yu. NeFII: Inverse rendering for reflectance decomposition with near-field indirect illumination. In *CVPR*, 2023. 1

[12] Liwen Wu, Rui Zhu, Mustafa B Yaldiz, Yinhao Zhu, Hong Cai, Janarbek Matai, Fatih Porikli, Tzu-Mao Li, Manmohan Chandraker, and Ravi Ramamoorthi. Factorized inverse path tracing for efficient and accurate material-lighting estimation. In *ICCV*, 2023. 1, 3, 4, 5

[13] 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. 1, 3, 4

[14] Jingyang Zhang, Yao Yao, Shiwei Li, Jingbo Liu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. NeILF++: Inter-reflectable light fields for geometry and material estimation. In *ICCV*, 2023. 1, 4

[15] Jingsen Zhu, Yuchi Huo, Qi Ye, Fujun Luan, Jifan Li, Dianbing Xi, Lisha Wang, Rui Tang, Wei Hua, Hujun Bao, and Rui Wang. I^2-SDF: Intrinsic indoor scene reconstruction and editing via raytracing in neural SDFs. In *CVPR*, 2023. 1, 4

[16] Rui Zhu, Zhengqin Li, Janarbek Matai, Fatih Porikli, and Manmohan Chandraker. IRISformer: Dense vision transformers for single-image inverse rendering in indoor scenes. In *CVPR*, 2022. 4