SIR: Multi-view Inverse Rendering with Decomposable Shadow Under Indoor Intense Lighting

Xiaokang Wei[1], Zhuoman Liu[1], Ping Li[1], Yan Luximon[*,1]*

[1]The Hong Kong Polytechnic University

{xiaokang.wei, zhuo-man.liu}@connect.polyu.hk, {p.li, yan.luximon}@polyu.edu.hk

Abstract—**3D inverse rendering in indoor scenes with strong light sources presents a significant challenge, primarily due to the substantial ambiguity in material recovery caused by the complex interaction between lighting and shadows. To address this, we propose a novel approach that integrates an implicit-explicit shadow predictor with a three-stage material estimation process. Our method enhances shadow realism by accurately predicting light interactions, while our material estimation process improves SVBRDF quality under challenging lighting conditions. Extensive experiments demonstrate the effectiveness of our method in both quantitative and qualitative metrics, enabling realistic object insertion and material replacement with proper shadow rendering under strong indoor light sources.**

Index Terms—**inverse rendering, multi-view HDR images, lighting estimation, implicit-explicit shadow predictor, relighting**

I. INTRODUCTION

Inverse rendering, which deduces scene properties such as geometry, lighting, and materials from images, has broad applications like photo-realistic editing and augmented reality. However, indoor scenes with complex lighting and materials present significant challenges.

The first challenge is scene geometry and lighting representation. **i)** For geometry, methods range from point clouds and meshes to newer techniques like neural radiance fields (NeRFs), with varying efficiency and computational demands. **ii)** For lighting, the choice between low dynamic range (LDR) and high dynamic range (HDR) impacts rendering accuracy, with HDR providing more detailed radiance measurements.

The second challenge is material estimation, particularly shadow and albedo decomposition under intense indoor lighting. Shadows are prominent due to the strong light sources and complex object interactions. This requires accurate shadow decomposition for realistic scene editing. Existing methods have not effectively solved this problem, especially when shadows do not shift with changes in light source position.

To address these, we propose SIR (Shadow Inverse Rendering), which efficiently decomposes shadows in indoor scenes without explicit supervision. SIR uses an SDF-based neural radiance field for scene representation and multi-view HDR images to capture lighting. For unsupervised shadow decomposition, we introduce a three-stage material estimation process: **1)** albedo initialization with hard shadows, **2)** albedo

refinement with differentiable soft shadows, and **3)** roughness refinement, along with instance-level BRDF regularization. This improves SVBRDF quality and enables scene editing.

In summary, the main contributions are:

- We propose SIR, a new inverse rendering framework with an explicit-implicit shadow predictor to decompose shadow and albedo in multi-view indoor scenes.
- We present a three-stage material estimation strategy, incorporating differentiable shadow and BRDF regularization.
- Our method exhibits superior inverse rendering accuracy on synthetic and real-world indoor datasets, demonstrated through successful scene editing.
- We propose a multi-view, intense-light indoor scene inverse rendering dataset with shadow ground-truths.

II. RELATED WORKS

A. Neural Scene Representations

Recent advances in neural scene representations have improved geometry and radiance recovery for inverse rendering. NeRF [1] generates new viewpoints using a single MLP but struggles with geometric accuracy due to volume rendering ambiguities [2]. Methods like NeuS [3] and VolSDF [4] address this with SDFs, while 3D Gaussian splatting [5] has greatly accelerated the process, as shown in GS-IR [6]. However, challenges remain in capturing complex indoor topologies and fine details. To tackle these, we adopt an SDF-based neural radiance field optimized for topology, lighting, and materials. To overcome these challenges, we adopt an SDF-based neural radiance field optimized for geometry, lighting, and materials.

B. Lighting Estimation

Accurate lighting estimation is crucial for inverse rendering, especially in indoor scenes. Most existing methods focus on single images, primarily aiming to integrate virtual objects into real scenes rather than significantly altering the scene's illumination [7]. Traditional approaches, such as single environment maps and spherical lobes [8], often fail to capture spatial variations and high-frequency lighting details. Recent advancements have sought to improve 3D lighting representations. For instance, per-pixel spatially-varying spherical Gaussians (SVSG)[9] effectively capture high-frequency effects and outperform spherical harmonics (SH) in depicting indoor lighting details. Neural-PIL[10], a pre-integrated lighting

---

*Corresponding author
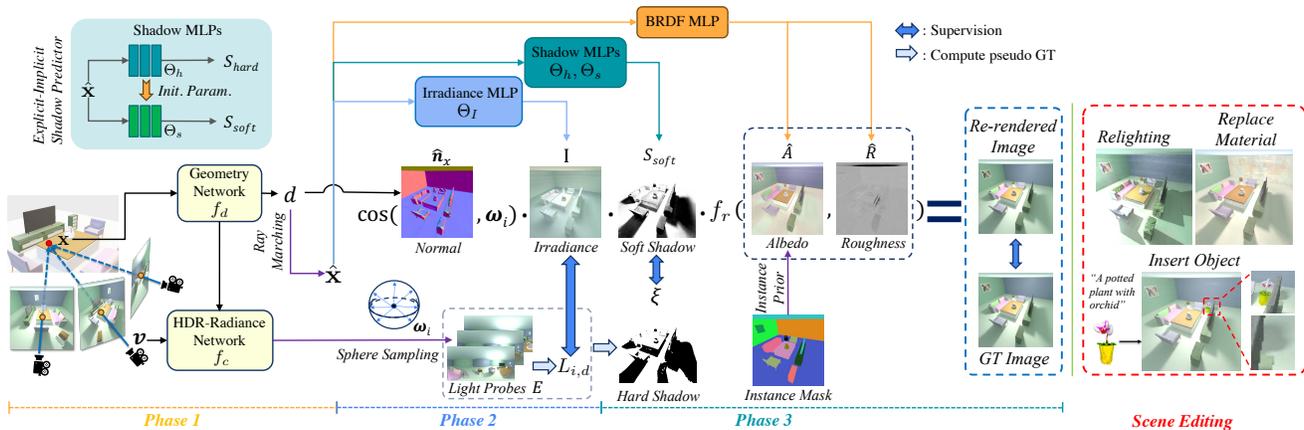
Fig. 1. The pipeline has three phases: **1)** A ray with direction $\boldsymbol{v}$ and point $\mathbf{x}$ is sampled from posed HDR images. The geometry network f_d learns the signed distance d to find surface point $\hat{\mathbf{x}}$, and the HDR-radiance network f_c learns radiance $\hat{C}$. **2)** Diffuse lighting $L_{i,d}$ is integrated from the neural light probes field E to learn irradiance I. **3)** Hard shadow S_{hard} is learned using Θ_h with pseudo ground truth ξ. The implicit shadow predictor Θ_s is optimized with Θ_h as prior knowledge. Instance-level BRDF regularizers are applied, optimizing the rendering equation to update $\hat{A}$, $\hat{R}$, and S_{soft}.

network based on image-based lighting (IBL), further enhances global illumination representation compared to SGs and SH. However, challenges like spatial instability and limited HDR information persist [11, 12].

To address these issues, we adopt a neural HDR-radiance field to represent IBL at any spatial point. This approach enables a more accurate and detailed depiction of indoor lighting, with a particular focus on physically accurate HDR lighting prediction.

C. Material Estimation

Material estimation in inverse rendering can be divided into object-level and scene-level tasks. Object-level estimation [6, 13, 14, 15] focuses on individual objects under controlled conditions, dealing with fewer variables and simpler lighting. Scene-level estimation [16, 17], however, is far more complex due to diverse lighting, multiple materials, and the presence of shadows in entire scenes.

Scene-level estimation faces additional challenges based on the input method. Single-view approaches [9], while simpler and less data-intensive, often suffer from ill-posed problems, leading to ambiguous or inaccurate results in complex scenes. In contrast, multi-view methods [13, 14, 15, 16] leverage images from multiple perspectives, providing richer information and significantly reducing ambiguity.

To address the challenges of scene-level material estimation, our work employs multi-view images, enabling accurate extraction of material properties in complex indoor scenes.

III. METHODS

A. Preliminary

The traditional rendering equation [18] calculates the outgoing radiance L_o at surface point $\hat{\mathbf{x}}$ by integrating over the hemisphere Ω^+ of incident light directions $\boldsymbol{\omega}_i$. It is given by:

$$L_o(\hat{\mathbf{x}}, \boldsymbol{\omega}_o) = \int_{\Omega^+} L_i(\hat{\mathbf{x}}, \boldsymbol{\omega}_i) f_r(\hat{\mathbf{x}}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) V(\hat{\mathbf{x}}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \hat{\boldsymbol{n}}_x) \mathrm{d}\boldsymbol{\omega}_i \quad (1)$$

where L_i is incoming radiance, f_r is the BRDF describing light reflection, and V is the visibility term indicating occlusion.

B. Geometry and Light Probes Field

For inverse rendering, it is important to estimate geometry and lighting accurately, especially in indoor scenes. We use VolSDF [4] for geometry representation and replace LDR input with HDR images to capture more lighting details. The geometry network predicts the signed distance, while the HDR-radiance network predicts radiance based on position and view direction. The scene is rendered using volume rendering [1].

In addition, we introduce a neural light probes field to model spatially-varying lighting, inspired by Unity's light probes. This field is queried from a pre-trained HDR-radiance field to estimate the lighting at any location in the scene. By placing a virtual camera at a specific position, we obtain the corresponding light probe E. The diffuse incoming lighting $L_{i,d}$ at a point $\mathbf{x}$ is computed by integrating the incident radiance over the hemisphere:

$$L_{i,d}(\mathbf{x}) = \int_{\Omega^+} E(\mathbf{x}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \boldsymbol{n}_x) \mathrm{d}\boldsymbol{\omega}_i \quad (2)$$

Here, $E(\mathbf{x}, \boldsymbol{\omega}_i)$ is the radiance at point $\mathbf{x}$ in direction $\boldsymbol{\omega}_i$, obtained by querying the HDR-radiance field. This method allows us to efficiently capture spatially varying lighting and model diffuse incoming lighting for any given point.

To further enhance the lighting representation, we introduce an *irradiance field* $\Theta_I(\mathbf{x})$, which learns irradiance values I at any spatial position $\mathbf{x}$, supervised by pre-computed incoming lighting. This field models continuous and spatially-varying global illumination in indoor scenes, significantly improving the lighting model's accuracy in complex environments.

C. Explicit Shadow Predictor

Shadows represent the occlusion between geometry and lighting. In indoor scenes, where primary light sources are strong, shadows play a key role. Traditional graphics methods use ray marching to compute hard shadows, but for inverse rendering, where light positions are unknown, this becomes challenging. Gradient optimization tends to bake shadows into the albedo, leading to ambiguities during material estimation [19].

In HDR environments, the light intensity near light sources can differ significantly from other areas. To detect hard shadows, we set a threshold μ to distinguish between light and non-light areas. If the light intensity at a point is below μ, it is classified as being in shadow ($\xi = 0$). This intensity is derived from the pre-trained HDR-radiance field. Formally:

$$\xi = \begin{cases} 1, & \text{if } \Gamma_{max}(\mathbf{x}) \geq \mu, \\ 0, & \text{if } \Gamma_{max}(\mathbf{x}) < \mu, \end{cases} \quad (3)$$

where $\Gamma_{max}(\mathbf{x})$ is the maximum incoming radiance intensity at a point $\mathbf{x}$ over all directions.

This approach detects hard shadows, even with multiple light sources or indirect lighting. To accelerate computation, ξ is used as pseudo ground truth, and a *hard shadow field* Θ_h learns shadow distributions at any 3D location. However, limited sample rays may cause noise in edge areas. To improve accuracy, a differentiable shadow module (Sec. III-D) refines rendering using the hard shadow field as prior knowledge.

D. Material Estimation

Physically based rendering (PBR) provides an accurate way to model light-material interactions [20, 21]. We use the microfacet BRDF model from Unreal Engine [21] to approximate surface reflectance and introduce a BRDF MLP f_r to model albedo $\hat{A}$ and roughness $\hat{R}$. However, directly optimizing the BRDF MLP can lead to non-convergence, especially for roughness, due to insufficient viewpoints and self-occlusion [14]. Additionally, shadows can be inadvertently baked into the albedo during optimization [15].

We propose a *three-stage* material estimation strategy combining Monte Carlo rendering and light probe fields [15] to recover albedo $\hat{A}$, roughness $\hat{R}$, and soft shadows S_{soft}. The rendering equation is:

$$L_o(\hat{\mathbf{x}}, \boldsymbol{\omega}_o) = \int_{\Omega^+} E(\hat{\mathbf{x}}, \boldsymbol{\omega}_i)(\frac{\hat{A}}{\pi} + f_s(\hat{\mathbf{x}}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o))(\boldsymbol{\omega}_i \cdot \hat{\boldsymbol{n}}_x)V(\hat{\mathbf{x}}, \boldsymbol{\omega}_i)d\boldsymbol{\omega}_i \quad (4)$$

$$= L_{o,d}(\hat{\mathbf{x}}) + L_{o,s}(\hat{\mathbf{x}}, \boldsymbol{\omega}_o) \quad (5)$$

Here, $\hat{\boldsymbol{n}}_x$ is the normal at the surface point $\hat{\mathbf{x}}$, $\boldsymbol{\omega}_i$ is the incident light direction, and $\boldsymbol{\omega}_o$ is the view direction. According to the Lambertian model [18], albedo $\hat{A}$ contributes to the diffuse component, while roughness $\hat{R}$ affects the specular component f_s in the Microfacet model. The predicted rendering result $\hat{L}_o$ is split into diffuse $\hat{L}_{o,d}$ and specular $\hat{L}_{o,s}$, with the specular component computed using Monte Carlo integration.

We optimize the parameters of the rendering equation in the following stages:

Stage 1: Albedo Initialization with Explicit Shadow S_{hard} Current neural inverse rendering methods [13, 14, 15] use spherical uniform sampling for ambient occlusion, which fails to capture cast shadows accurately (see in Appendix). Inspired by shadow mapping and indoor scene lighting, we propose extracting and normalizing the visibility term separately from the rendering integral, using shadows cast by the main light source for visibility approximation.

This approach contrasts with random sampling of transmittance, which may misrepresent light attenuation. As ambient

occlusion from indirect light is already accounted for in irradiance, it doesn't need to be considered again.

Using the coarse albedo $\hat{A}$ predicted by f_r, we add a hard shadow term ξ to the diffuse radiance, preserving inherent surface colors without shadow artifacts.

$$L_{o,d}(\hat{\mathbf{x}}) = \frac{\hat{A}}{\pi} \int_{\Omega^+} E(\hat{\mathbf{x}}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \hat{\boldsymbol{n}}_x)V(\hat{\mathbf{x}}, \boldsymbol{\omega}_i)\,d\boldsymbol{\omega}_i \quad (6)$$

$$\approx \frac{\hat{A}}{\pi} \xi \int_{\Omega^+} E(\hat{\mathbf{x}}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \hat{\boldsymbol{n}}_x)\,d\boldsymbol{\omega}_i \quad (7)$$

We pre-compute the hard shadow and irradiance terms with MLPs Θ_h and Θ_I to minimize unnecessary ray samples and improve computational efficiency. Thus, the predicted diffuse radiance is $\hat{L}_{o,d} = \frac{\hat{A}}{\pi}S_{hard}I$.

Stage 2: Albedo refinement with implicit shadow S_{soft}. As discussed in Sec. III-C, hard shadows can introduce noise in the rendering results. To mitigate this, we propose an implicit shadow predictor S_{soft} using an MLP to model the *soft shadow field* $\Theta_s(\mathbf{x})$. This soft shadow representation extends beyond the binary nature of hard shadows. The initial parameters of Θ_s are inherited from the hard shadow field Θ_h. We replace the hard shadow with the differentiable soft shadow in the diffuse radiance equation, which results in $\hat{L}_{o,d} = \frac{\hat{A}}{\pi}S_{soft}I$. The soft shadow is updated by optimizing the diffuse radiance. Since albedo and shadows interact, we introduce an instance-level regularizer $\mathcal{L}_{albedo}$ to ensure uniform luminance of the albedo across instances. This regularizer is formulated as:

$$\mathcal{L}_{albedo} = \Big| \sum_{i=1}^{K} \left(\hat{A} - \Phi_{inv}\left(\frac{\sum_p \Phi(\hat{A}) \odot M_i}{\sum_p M_i} \right) \right) \Big| \quad (8)$$

where $\Phi(\cdot)$ and $\Phi_{inv}(\cdot)$ convert between RGB and HSV, M_i is the instance segmentation mask, K is the number of instance classes, p is the minibatch of 3D points, and $\odot$ denotes element-wise multiplication.

Stage 3: Roughness refinement. In the first two stages, roughness $\hat{R}$ is predicted by the BRDF MLP f_r. Since roughness has a similar instance-level assumption, we add a roughness smooth regularizer $\mathcal{L}_{rough}$ to make roughness similar at the instance level.

$$\mathcal{L}_{rough} = \Big| \sum_{i=1}^{K} (\hat{R} - \frac{\sum_p \hat{R} \odot M_i}{\sum_p M_i}) \Big| \quad (9)$$

IV. EXPERIMENTS

Our SIR pipeline is validated through experiments conducted on both synthetic and real-world indoor datasets.

A. Indoor Datasets

We developed two indoor datasets for evaluating our method:
1) Synthetic Dataset: An extended DM-NeRF dataset [23] with six indoor scenes rendered using Blender Cycles. It includes 300 training and 200 testing images (400×400 pixels), with modified object locations, colors, and properties (*e.g.*, albedo, roughness, shadows, and instance masks).
2) Real-World Dataset: The dataset includes two indoor scenes with complex materials and lighting, captured with a professional camera. Each scene has 120 HDR images (three exposures, $\frac{1}{15000}$s to $\frac{1}{8}$s) resized to 512×512 pixels, with 85% for training and 15% testing, challenging inverse rendering.
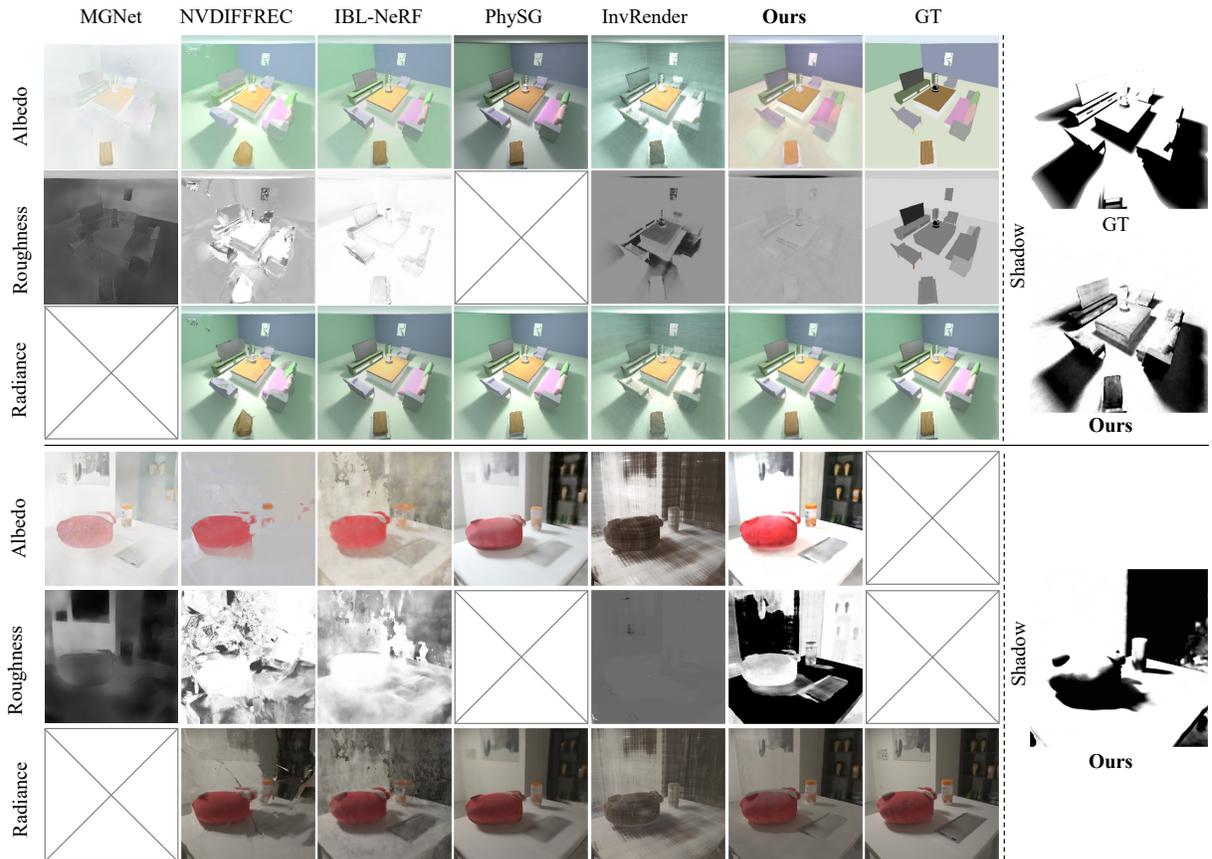
Fig. 2. Qualitative results of all methods on two datasets (*Syn: Restroom, Real: Office*).

TABLE I
QUANTITATIVE RESULTS FOR ALL METHODS, AVERAGED ACROSS 6 SYNTHETIC SCENES.

	Albedo			Roughness	Shadow	View Synthesis		
	PSNR↑	SSIM ↑	LPIPS ↓	MSE ↓	MSE ↓	PSNR↑	SSIM ↑	LPIPS ↓
MGNet [22]	12.9030	0.6676	0.3787	0.0994	-	-	-	-
NVDIFFREC [15]	16.6377	0.7906	0.3736	0.0531	-	23.8048	0.8606	0.1863
IBL-NeRF [16]	16.7773	0.8468	0.2161	0.0564	-	24.5258	**0.9263**	**0.0844**
PhySG [13]	10.5322	0.7076	0.3838	-	-	26.1667	0.9202	0.1014
InvRender [14]	12.5401	0.6740	0.4935	**0.0412**	-	23.8582	0.7521	0.3923
SIR (Ours)	**20.2767**	**0.8600**	**0.2154**	0.0445	**0.0541**	**28.5456**	0.9258	0.0964

TABLE II
QUANTITATIVE RESULTS ON *view synthesis* FOR ALL METHODS,
AVERAGED OVER 2 REAL-WORLD SCENES.

	PSNR↑	SSIM ↑	LPIPS ↓
NVDIFFREC [15]	22.7146	0.7798	0.4456
IBL-NeRF [16]	**24.2781**	0.7491	0.5225
PhySG [13]	22.1492	0.7796	0.4581
InvRender [14]	20.1327	0.6694	0.6384
SIR (Ours)	22.8176	**0.8345**	**0.2931**

B. Implementation

All neural fields in our network are implemented as multi-layer perceptrons (MLPs) with ReLU activations. The geometry network f_d and HDR-radiance network f_c follow the VolSDF [4] setup: f_d is an 8-layer MLP with 256 hidden units, and f_c is a 4-layer MLP with 256 hidden units. The irradiance, shadow, and BRDF MLPs have 4 layers and 256 hidden units each. We apply positional encoding with 10 and 4 frequency components for 3D locations and directions, respectively.

The SDF-based neural radiance field is implemented in PyTorch and optimized using Adam with a 5×10^{-4} learning rate for 250K iterations. For irradiance and shadow estimation, we use the same learning rate for 10K iterations, sampling 512 rays at each surface point. Material estimation phases use Adam with a 10^{-3} learning rate for 25K iterations, sampling 128 rays for the BRDF term. Training on an NVIDIA GeForce RTX 3090 GPU takes about 40 hours: 16 hours for phase one, 8 for phase two, and 16 for phase three.

C. Inverse Rendering

We compare our method with five well-known inverse rendering approaches: 1) MGNet [24], 2) NVDIFFREC [15], 3) IBL-NeRF [16], 4) PhySG [13], and 5) InvRender [14]. NVDIFFREC uses a DMTet-based approach, IBL-NeRF is
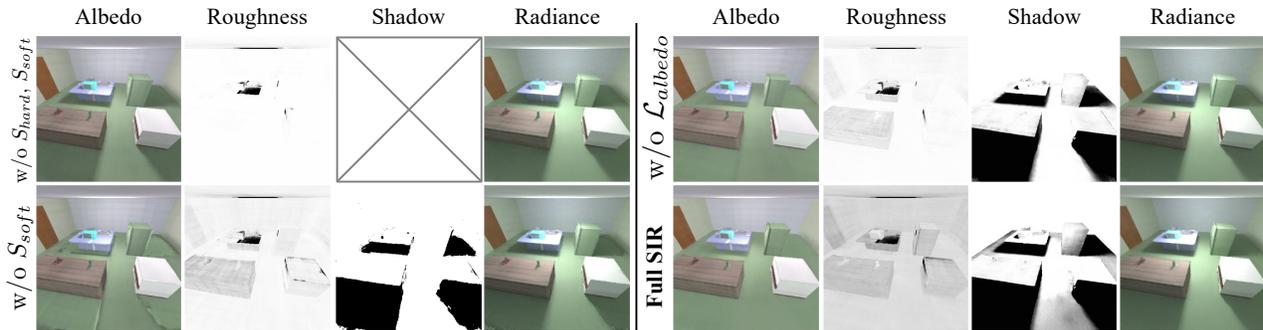
Fig. 3. Qualitative results of ablation study (*Kitchen* in synthetic dataset).

TABLE III
QUANTITATIVE RESULTS OF ABLATION STUDY ON SIR, AVERAGED ACROSS 6 SYNTHETIC SCENES.

| | | Albedo | | Roughness | Shadow | | View Synthesis | |
	PSNR↑	SSIM ↑	LPIPS ↓	MSE ↓	MSE ↓	PSNR↑	SSIM ↑	LPIPS ↓
(1) w/o S_{hard}, S_{soft}	16.6411	0.8084	0.2919	0.0712	-	28.2254	0.9221	0.0974
(2) w/o S_{soft}	19.3836	0.8052	0.3061	0.0627	0.0650	27.8201	0.8946	0.1897
(3) w/o $\mathcal{L}_{albedo}$	19.2597	0.8454	0.2406	0.0676	0.0863	28.3610	**0.9258**	**0.0924**
Full SIR	**20.2767**	**0.8600**	**0.2154**	**0.0445**	**0.0541**	**28.5456**	0.9258	0.0964

NeRF-based, and PhySG and InvRender employ SDF-based neural radiance fields. For fair comparison, we adapt InvRender [14] and PhySG [13] to align with our VolSDF-based geometry representation [4]. See the Appendix for details.

For the synthetic dataset, we report PSNR, SSIM [25], LPIPS [26], and MSE for roughness and shadow predictions. Since our method uniquely recovers shadows, we report only MSE for shadow predictions, using binary ground truth shadows from Blender for metric computation.

Our method outperforms baselines in material decomposition, especially shadow extraction, leading to superior view synthesis results for indoor scenes (Table I, Fig. 2). In real-world datasets, it handles complex lighting conditions effectively, providing reliable material estimations (Table II, Fig. 2). Although our PSNR is slightly lower than IBL-NeRF, we show advantages in SSIM and LPIPS, indicating better perceptual quality with less distortion.

D. Ablation Study

The key innovation of our method is the capacity to effectively isolate shadows while ensuring that these shadows are not incorporated into the albedo under an indoor scene. Therefore, this section is dedicated to evaluating the impact of shadow terms, differentiable soft shadow, and albedo regularizer:
(1) Removing shadow terms: We remove both S_{hard} and S_{soft} in phases 2, 3 to demonstrate the necessity of introducing shadow terms in inverse rendering.
(2) Removing soft shadow: Without the differentiable shadow S_{soft} in phase 3, the pipeline re-renders images using hard shadows S_{hard}.
(3) Removing albedo regularizer: This ablation study evaluates the impact of instance-level albedo consistency.

Table III (1) clearly shows that removing the shadow term in the rendering equation leads to a significant decline in novel view synthesis performance, adversely affecting albedo

estimation as well. The qualitative results in Fig. 3 further highlight that the absence of shadow terms results in incorrect albedo outputs with baked shadows, negatively impacting synthesis results. However, introducing a binary (hard) shadow still leads to suboptimal synthesis, as shown in Table III (2), particularly around shadow edges (see Fig. 3). Therefore, learning a differentiable soft shadow after an initial hard shadow is necessary to preserve shadow details in the synthesis images. Additionally, as demonstrated in Table III (3), the integration of an instance-level albedo regularizer is effective in maintaining albedo consistency for each instance, contributing to more accurate synthesis results. Moreover, the MSE values on roughness in Table III clearly show that correctly decomposing shadows and albedo prevents the tendency for roughness to diverge, leading to accurate roughness recovery.

E. Scene Editing

Once SIR has decomposed the intrinsic properties of indoor scenes, it enables the realistic rendering of novel views by manipulating these properties. To showcase its effectiveness, we developed two editing applications: 1) object insertion and 2) material replacement. Detailed explanations of these applications can be found in the Appendix.

V. CONCLUSION

We propose SIR, a multi-view inverse rendering method on indoor scenes, effectively addressing the challenges of material and lighting decomposition by explicitly isolating shadows. Leveraging posed HDR images and an SDF-based radiance field, SIR significantly enhances realism in material estimation and scene editing, surpassing previous methods with accurate shadow estimation. The innovative incorporation of shadows with a three-stage material estimation process substantially improves the quality of SVBRDFs. Extensive evaluations on diverse indoor datasets demonstrate the superiority of SIR in both quantitative and qualitative aspects over existing methods.

REFERENCES

[1] 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," *ECCV*, 2020.

[2] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun, "Nerf++: Analyzing and improving neural radiance fields," *arXiv preprint arXiv:2010.07492*, 2020.

[3] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang, "Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," *NeurIPS*, 2021.

[4] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman, "Volume rendering of neural implicit surfaces," *NeurIPS*, 2021.

[5] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis, "3d gaussian splatting for real-time radiance field rendering," *TOG*, 2023.

[6] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia, "Gs-ir: 3d gaussian splatting for inverse rendering," in *CVPR*, 2024, pp. 21644–21653.

[7] Kevin Karsch, Varsha Hedau, David Forsyth, and Derek Hoiem, "Rendering synthetic objects into legacy photographs," *TOG*, 2011.

[8] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde, "Fast spatially-varying indoor lighting estimation," in *CVPR*, 2019.

[9] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker, "Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image," in *CVPR*, 2020.

[10] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch, "Neural-pil: Neural pre-integrated lighting for reflectance decomposition," *NeruIPS*, 2021.

[11] Shuran Song and Thomas Funkhouser, "Neural illumination: Lighting prediction for indoor environments," in *CVPR*, 2019.

[12] Pratul P Srinivasan, Ben Mildenhall, Matthew Tancik, Jonathan T Barron, Richard Tucker, and Noah Snavely, "Lighthouse: Predicting lighting volumes for spatially-coherent illumination," in *CVPR*, 2020.

[13] 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 *CVPR*, 2021.

[14] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou, "Modeling indirect illumination for inverse rendering," in *CVPR*, 2022.

[15] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler, "Extracting triangular 3d models, materials, and lighting from images," in *CVPR*, 2022.

[16] Changwoon Choi, Juhyeon Kim, and Young Min Kim, "Ibl-nerf: Image-based lighting formulation of neural radiance fields," *Computer Graphics Forum*, 2023.

[17] Zhen Li, Lingli Wang, Mofang Cheng, Cihui Pan, and Jiaqi Yang, "Multi-view inverse rendering for large-scale real-world indoor scenes," in *CVPR*, 2023.

[18] James T Kajiya, "The rendering equation," in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, 1986.

[19] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg, "Shape, light, and material decomposition from images using monte carlo rendering and denoising," *NeurIPS*, 2022.

[20] Matt Pharr, Wenzel Jakob, and Greg Humphreys, *Physically based rendering: From theory to implementation*, MIT Press, 2023.

[21] Brian Karis and Epic Games, "Real shading in unreal engine 4," *Proc. Physically Based Shading Theory Practice*, 2013.

[22] Jingsen Zhu, Fujun Luan, Yuchi Huo, Zihao Lin, Zhihua Zhong, Dianbing Xi, Rui Wang, Hujun Bao, Jiaxiang Zheng, and Rui Tang, "Learning-based inverse rendering of complex indoor scenes with differentiable monte carlo raytracing," in *SIGGRAPH Asia*, 2022.

[23] Bing Wang, Lu Chen, and Bo Yang, "Dm-nerf: 3d scene geometry decomposition and manipulation from 2d images," *ICLR*, 2023.

[24] Jingsen Zhu, Fujun Luan, Yuchi Huo, Zihao Lin, Zhihua Zhong, Dianbing Xi, Rui Wang, Hujun Bao, Jiaxiang Zheng, and Rui Tang, "Learning-based inverse rendering of complex indoor scenes with differentiable monte carlo raytracing," in *SIGGRAPH Asia*, 2022.

[25] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," in *ICLR*, 2018.

[26] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018.

[27] 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*, 2007.

[28] Robert L Cook and Kenneth E. Torrance, "A reflectance model for computer graphics," *TOG*, 1982.

[29] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger, "Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction," *NeurIPS*, 2022.

[30] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan, "Ref-nerf: Structured view-dependent appearance for neural radiance fields. in 2022 ieee," in *CVPR*, 2022.

[31] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall, "Dreamfusion: Text-to-3d using 2d diffusion," in *ICLR*, 2023.