

GUS-IR: Gaussian Splatting with Unified Shading for Inverse Rendering

Zhihao Liang, *Student Member, IEEE*, Hongdong Li, Kui Jia, Kailing Guo and Qi Zhang,

Abstract—Recovering the intrinsic physical attributes of a scene from images, generally termed as the inverse rendering problem, has been a central and challenging task in computer vision and computer graphics. In this paper, we present GUS-IR, a novel framework designed to address the inverse rendering problem for complicated scenes featuring rough and glossy surfaces. This paper starts by analyzing and comparing two prominent shading techniques popularly used for inverse rendering, *forward shading* and *deferred shading*, effectiveness in handling complex materials. More importantly, we propose a unified shading solution that combines the advantages of both techniques for better decomposition. In addition, we analyze the normal modeling in 3D Gaussian Splatting (3DGS) and utilize the shortest axis as normal for each particle in GUS-IR, along with a depth-related regularization, resulting in improved geometric representation and better shape reconstruction. Furthermore, we enhance the probe-based baking scheme proposed by GS-IR to achieve more accurate ambient occlusion modeling to better handle indirect illumination. Extensive experiments have demonstrated the superior performance of GUS-IR in achieving precise intrinsic decomposition and geometric representation, supporting many downstream tasks (such as relighting, retouching) in computer vision, graphics, and extended reality.

Index Terms—3D Gaussian Splatting, Inverse Rendering, Unified Shading.

1 INTRODUCTION

INVERSE rendering is a long-standing and challenging task in computer vision and computer graphics, aiming to recover intrinsic physical attributes (*e.g.* scene geometry, surface material, and environment lighting) of a 3D scene from multiple observations. The main challenges arise from the optimization in uncontrolled environments with unknown illumination, which leads to ill-posedness in decomposing multiple strongly coupled physical attributes. Traditional methods [1], [2], [3], [4], [5], [6], [7] struggled to simultaneously and accurately estimate the geometry, material, and illumination of a complex scene.

Recently, neural rendering techniques such as Neural Radiance Fields (NeRF) [8] have shown remarkable promise in scene geometric reconstruction tasks [9], [10], [11], [12], [13]. However, most existing methods have overlooked the interaction between intrinsic physical attributes and illumination. Their implicit representations, which adopt an overly simplified appearance model as a function of view direction only, limit their applicability to many downstream tasks like relighting. Although recent methods [14], [15], [16], [17] introduce differentiable physical-based rendering to decompose physical attributes (*e.g.* geometry, surface

materials, and environmental lighting), they still face challenges related to low rendering speed and explicit editing, especially when it can not be rendered at interactive rates.

3D Gaussian Splatting (3DGS) [18] has recently emerged as a promising technique to model 3D scenes and significantly boost the rendering speed to a real-time level. It combines explicit and more compact scene representation with differentiable rasterization [19], [20], resulting in a fast and remarkable performance for novel view synthesis. It is natural and essential to introduce physical-based rendering to 3DGS to achieve efficient inverse rendering tasks [21], [22], [23]. Despite their impressive relighting performance, these approaches still face challenges in geometric estimation and shading schemes. Firstly, during the 3DGS optimization, the adaptive control of the Gaussian density may lead to loose geometry, making it difficult to estimate accurate scene's normal. One key step in using 3DGS for inverse rendering is accurately representing the geometry to describe the transportation of illumination. Although GS-IR [22] introduces a depth-related regularization, it ignores the geometric attributes of the Gaussian ellipsoid by attaching a learnable vector as the normal for each particle, struggling to compactly associate the normal optimization with the geometry. Consequently, it is necessary to model normal and better reconstruct geometry via the geometric attributes of the ellipsoid.

Besides, another critical aspect of employing 3DGS for inverse rendering involves the shading scheme. Depending on where physical-based rendering is incorporated within the 3DGS-based framework, the shading scheme can be divided into forward shading [23] (*i.e.* shading each particle *before* 'splatting') and deferred shading [22] (*i.e.* shading each pixel *after* 'splatting'). These different shading schemes have a substantial impact on material decomposition and overall rendering quality. Specifically, forward shading emphasizes

- Zhihao Liang and Kailing Guo are with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China (*e-mail:* eezhihaoliang@mail.scut.edu.cn; guokl@scut.edu.cn).
- Hongdong Li is with the Australian National University, Canberra 2600, Australia (*e-mail:* hongdong.li@anu.edu.au).
- Kui Jia is with the School of Data Science, The Chinese University of Hong Kong, Shenzhen 518172, China (*e-mail:* kuijia@cuhk.edu.cn).
- Qi Zhang (corresponding author) is with VIVO, Xi'an 710055, China (*e-mail:* nwpuqzhang@gmail.com).

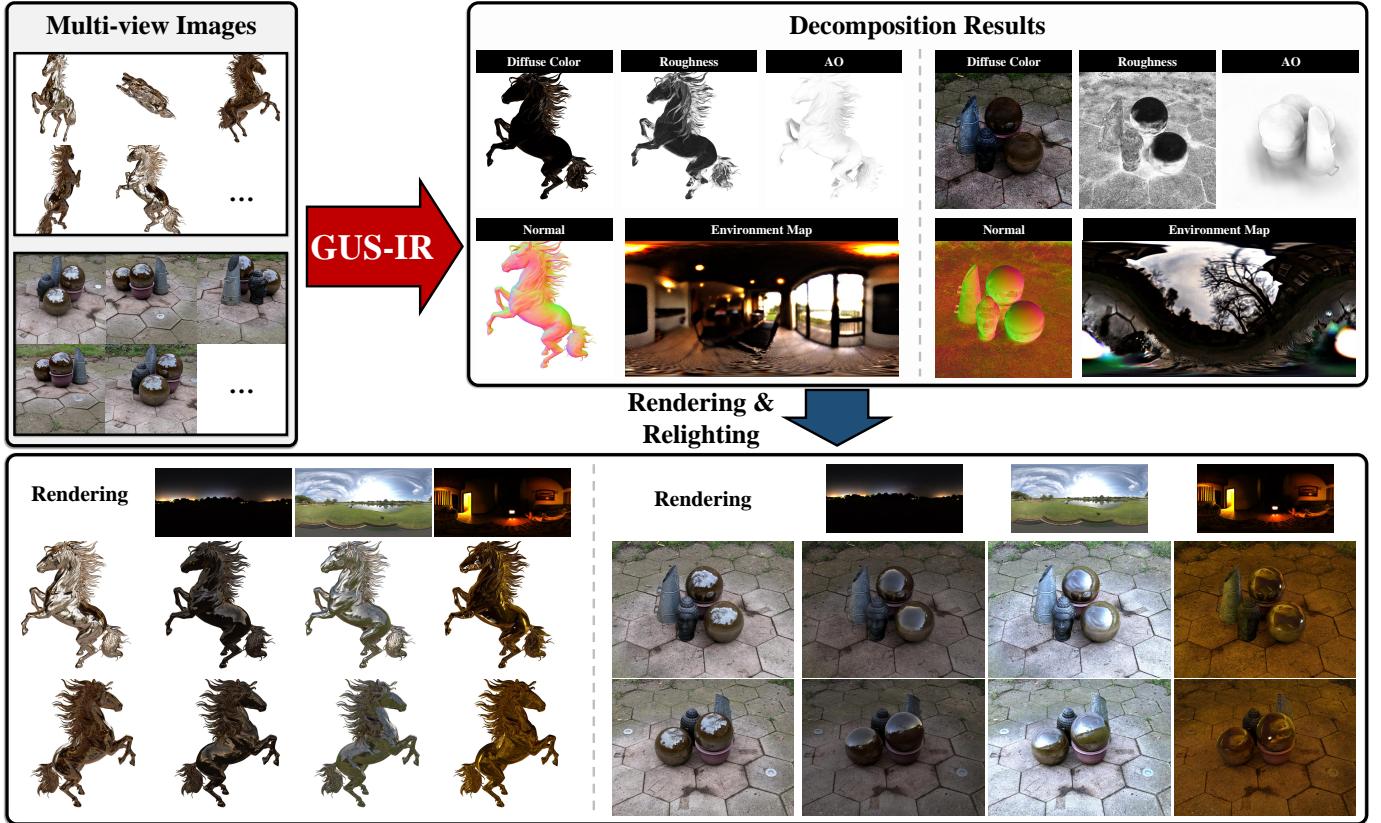


Fig. 1. Given multi-view captured images of a complex scene featuring rough and glossy surfaces, we propose **GUS-IR** (Gaussian Splatting with Unified Shading for Inverse Rendering), which utilizes 3D Gaussians to recover high-quality physical properties (e.g., normal, material, illumination) under unknown illumination. This enables us to perform advanced applications (e.g. relighting), resulting in outstanding inverse rendering results. Better viewed on screen with zoom-in, we successfully recovered the glossy surfaces of the marble balls placed in the center of the garden.

the representation of diffuse colors, making it well-suited for complex scenes featuring rough surfaces, while deferred shading is more adept at capturing glossy objects with high specular regions. More details can be found in Fig. 3. Thus, it is critical to define a suitable shading scheme to make the inverse rendering feasible for different cases, particularly for complex scenes with rough surfaces and specular regions.

In this paper, we are motivated to present a powerful 3DGS-based inverse rendering framework, GUS-IR (Gaussian Splatting with Unified Shading for Inverse Rendering), to achieve high-quality intrinsic decomposition under unknown illumination. To represent the normal in 3DGS, we treat each particle as a surfel [24] and utilize its shortest axis as the orientation (*i.e.* normal) for each particle. Then we find that forward and deferred shading schemes are beneficial for reconstructing rough and glossy surface materials, respectively. In GUS-IR, we thus unify both shading schemes to facilitate the intrinsic physical attributes decomposition of complex scenes containing various surface materials. By leveraging the benefits of unification, GUS-IR addresses the challenges posed by glossy surfaces in complex scenes. Additionally, we improve the baking scheme proposed in GS-IR, which leverages dense binary cubemaps to cache the occlusion during baking, achieving better ambient occlusion modeling. The contributions of this work can be summarized as follows:

- We present *GUS-IR* that unifies several shading

mechanisms to achieve intrinsic decomposition of glossy objects and complex scenes;

- We revisit forward and deferred shading schemes, and propose a unified shading scheme to handle complex scenes and better capture highlight details on glossy surfaces;
- We propose the shortest-axis normal modeling to represent reliable geometry, along with the depth-related regularization for better reconstruction.
- We improve the probe-based baking solution proposed in GS-IR to better model ambient occlusion and handle indirect illumination.

We demonstrate the superiority of our method to baseline methods qualitatively and quantitatively on various challenging scenes, including the TensoIR Synthesis [16], Shiny Blender [25], Glossy Blender [26], Mip-NeRF 360 [27], and Ref-NeRF Real [25] datasets.

This work extends our preliminary work as described in [22]. In this journal version, we first give a formal formulation of the rendering equation for Gaussians and analyze the advantages of two shading schemes (forward shading and deferred shading). We then propose a unified shading framework that systematically consolidates two shading mechanisms, which outperforms other shading methods on both diffuse and specular regions. Finally, we replace the SH coefficients used in the baking solution of GS-IR with dense binary cubemaps to better handle ambient occlusion. To val-

idate this framework, we re-conducted all our experiments including new tests on glossy objects, achieving superior performance on view synthesis, relighting, and geometry reconstruction.

2 RELATED WORKS

2.1 Neural Representation and Rendering

Recently, a surge of neural rendering techniques, exemplified by Neural Radiance Field (NeRF) [8], have succeeded in modeling light fields, supporting remarkable results in novel-view synthesis [28], [29], [30], [31], 3D and 4D reconstruction [11], [12], [13], [32], generation [33], [34], [35], and other advanced tasks [36], [37], [38]. Instead of the triangular or quadrilateral meshes widely used in the traditional graphics pipeline, these methods adopt the implicit neural representation (INR) to model the geometry and appearance. Specifically, for 3D reconstruction, NeuS [9] and VolSDF [10] use Multi-Layer Perceptron (MLP) to parameterize a signed distance field (SDF) to represent the geometry of the target object or scene, which can be friendly optimized as well as extracting an explicit mesh through iso-surface extraction algorithms [39], [40]. However, MLP incurs a substantial payload in rendering, which is further exacerbated by the ray-casting technique adopted by NeRF. To address these inefficient, several methods proposed to use explicit structures to manage features and accelerate sampling, such as volumes [31], [41], [42], hash grids [43], tri-planes [30], [44], [45], points [46] even mesh [47]. While these methods alleviate the payload, they struggle with high-fidelity real-time rendering, getting rid of inefficient ray-casting is essential to solve this dilemma. To this end, NVDiffRec [48], [49], [50] attempts to achieve efficient reconstruction using a mesh with a rasterization pipeline in the traditional form. However, it struggles with complex topologies and scene reconstruction. To address these challenges, 3DGS [18] incorporates explicit 3D Gaussian primitives with specially designed tile-based rasterization to achieve high-fidelity real-time rendering performance. 3DGS manages a set of particles, each of which maintains a 3D opacity field, and shades the pixel by parallelly ‘splatting’ particles onto the screen space. Several contemporary works further fundamentally improve 3DGS [24], [51], [52], [53], [54] or apply it to advanced tasks [55], [56]. For inverse rendering, GS-IR [22], GaussianShader [23], and RelightableGS [21] achieve promising results based on 3DGS. For modeling the normal in 3DGS, GS-IR and RelightableGS attach a learnable orientation for each particle as the normal, and GaussianShader associates a predicted residual on the shortest axis as the normal. Different from them, we follow 2DGS [24], treating each particle as a flat surfel and using the shortest axis as the normal in GUS-IR. Our experimental results validate these normal representations and show that our scheme is superior to other schemes in terms of both decomposition and relighting results.

2.2 Inverse Rendering

Inverse rendering aims to decompose the physical attributes (*e.g.* geometry, materials, and illumination conditions) from observations. Importantly, in inverse rendering tasks, appearance tends to be treated as the physical interaction [57],

[58], [59] of the surface and illumination, rather than a fitted spatially view-dependent function. However, the inherent ambiguity between observation and the underlying properties makes this problem extremely ill-posed. To alleviate the challenge, some approaches propose constraints to obtain well-defined illumination conditions, such as fixing illumination conditions, rotating the target object, or co-locating the light source with the moving camera [60], [61], [62], [63], [64], [65]. Inspired by the success of NeRF, recent approaches [14], [15], [16], [29], [50], [66], [67], [68], [69], [70] leverage neural representation and rendering to alleviate the ill-posedness in the optimization and aim to simulate physical interactions from multiple observations under unknown conditions. As pioneers, NRF [29] and NeRD [67] follow NeRF and use an MLP to parameterize the physical field (*e.g.* volumetric density, normal, BRDF) of the target scene. Instead of a view-dependent MLP, they conduct the physical-based rendering equation for shading. NeRV [15], NeRFactor [14], InvRender [69] further model the visibility or shadows in the target scene. In addition, PhySG [66] uses Spherical Gaussians (SGs) to represent illumination and BRDFs, TensoIR [16] adopts the tensor factorization [44] to compute visibility and indirect illumination efficiently. However, these methods are limited to the object-level and almost require several hours even days for inverse rendering. Based on 3DGS, GaussianShader [23] handle the decomposition of reflective objects, GS-IR [22] and RelightableGS [21] respectively leverage precomputation and ray-tracing techniques to handle the visibility. Notably, the above methods can be easily transferred to complex scenes for inverse rendering. In this work, we propose a 3DGS-based framework to decompose the geometry, materials, and illumination conditions for rough and glossy objects, even complex scenes.

3 3DGS FOR INVERSE RENDERING

In this section, we review GS-IR [22] and give the technical backgrounds and math symbols that are necessary for the presentation of our proposed method.

3.1 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [18] explicitly represents a 3D scene as a set of particles, each particle stores the parameters that express a 3D Gaussian signal g with a view-dependent color c modeled by spherical harmonics (SH), a mean (*i.e.* position) $\mu \in \mathbb{R}^3$ and an anisotropic covariance matrix $\Sigma \in \mathbb{R}^{3 \times 3}$, where the covariance matrix $\Sigma = \mathbf{R} \mathbf{S} \mathbf{S}^\top \mathbf{R}^\top$ is obtained by a rotation matrix \mathbf{R} and scaling \mathbf{S} . \mathbf{S} and \mathbf{R} refer to a diagonal matrix $\text{diag}(s_x, s_y, s_z)$ and a rotation matrix constructed from a unit quaternion q . During rasterization, 3DGS conducts perspective projection and splats the 3D Gaussian signal g into a 2D Gaussian signal \tilde{g} with mean vector $\tilde{\mu}$ and covariance $\tilde{\Sigma}$ on the screen space according to the extrinsic matrix \mathbf{T} and intrinsic matrix \mathbf{K} . After obtaining all 2D Gaussian signals in screen space, 3DGS

shades the pixel via α -blending N 2D Gaussian signals covering the pixel \mathbf{u} , and we finally obtain the color $\hat{\mathbf{c}}(\mathbf{u})$:

$$\begin{aligned} \alpha_k &= \tilde{g}(\mathbf{u} | \tilde{\mu}_k, \tilde{\Sigma}_k) o_k, \quad T_k = \prod_{j=1}^{k-1} (1 - \alpha_j), \\ \hat{\mathbf{c}}(\mathbf{u}) &= \sum_{k=1}^N w_k \mathbf{c}_k, \quad w_k = T_k \alpha_k \end{aligned} \quad (1)$$

where o_k and \mathbf{c}_k are the opacity and color associated with the k -th point.

3.2 Gaussian Splatting for Inverse Rendering

For inverse rendering, some methods [21], [22], [23] represent the view-dependent appearance as the physical interaction between illumination and materials, rather than spherical harmonics used in 3DGS. To model the appearance of a given surface point \mathbf{x} with normal \mathbf{n} from view direction \mathbf{v} , they follow the classic rendering equation [58] to formulate the outgoing radiance L_o via integrating the response of the incident radiance over the upper hemisphere Ω :

$$\begin{aligned} L_o(\mathbf{x}, \mathbf{v}) &= \int_{\Omega} L_i(\mathbf{x}, \mathbf{l}) f_r(\mathbf{x}, \mathbf{l}, \mathbf{v})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l}, \\ f_r(\mathbf{x}, \mathbf{l}, \mathbf{v}) &= f_r^d(\mathbf{x}) + f_r^s(\mathbf{x}, \mathbf{l}, \mathbf{v}), \end{aligned} \quad (2)$$

where f_r describes the Bidirectional Reflectance Distribution Function (BRDF), mostly formulated by the Cook-Torrance microfacet model [57], [59]. f_r can be divided into diffuse term f_r^d and specular term f_r^s depending on whether related to incidence \mathbf{l} and view direction \mathbf{v} . In different workflows, the reflection in f_r is modeled by different physical attributes, such as Dielectric F0 coefficients with metallic [22], and specular color [23]. In GUS-IR, for each particle, we define learnable diffuse color α and specular color s to represent its base color and reflection color, respectively. After dividing the f_r into diffuse and specular terms, the rendering equation Eq. (2) can be rewritten as the combination of diffuse component L_o^d and specular component L_o^s as:

$$\begin{aligned} L_o(\mathbf{x}, \mathbf{v}) &= L_o^d(\mathbf{x}) + L_o^s(\mathbf{x}, \mathbf{v}) \\ L_o^d(\mathbf{x}) &= f_r^d(\mathbf{x}) \int_{\Omega} L_i(\mathbf{x}, \mathbf{l})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l}, \\ L_o^s(\mathbf{x}, \mathbf{v}) &= \int_{\Omega} L_i(\mathbf{x}, \mathbf{l}) f_r^s(\mathbf{x}, \mathbf{l}, \mathbf{v})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l}. \end{aligned} \quad (3)$$

Specifically, previous methods almost adopt the image-based lighting (IBL) model and assume that direct illumination comes from a distance, thus incident radiance $L_i(\mathbf{x}, \mathbf{l})$ degenerates as $L_i(\mathbf{l})$ that is only related to incidence \mathbf{l} , and the direct light sources are represented by an HDRI. They further use split-sum approximation [71] to solve the intractable integral in Eq. (3) as:

$$\begin{aligned} L_o^d(\mathbf{x}) &= f_r^d(\mathbf{x}) \underbrace{\int_{\Omega} L_i(\mathbf{l})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l}}_{I^d}, \\ L_o^s(\mathbf{x}, \mathbf{v}) &= \int_{\Omega} L_i(\mathbf{l}) f_r^s(\mathbf{x}, \mathbf{l}, \mathbf{v})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l} \\ &\approx \underbrace{\int_{\Omega} f_r^s(\mathbf{x}, \mathbf{l}, \mathbf{v})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l}}_R \underbrace{\int_{\Omega} p(\mathbf{l}, \mathbf{v}) L_i(\mathbf{l})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l}}_{I^s}, \end{aligned} \quad (4)$$

where $p(\mathbf{l}, \mathbf{v})$ is related to the roughness of surface point \mathbf{x} and denotes the importance (*i.e.* probability density function, PDF) of incidence \mathbf{l} over the upper hemisphere Ω . Note that I^d and I^s denote the incident radiance integral over the upper hemisphere and the reflected lobe. I^d , I^s , and R can be precomputed in advance from the HDRI and stored in look-up tables.

In addition, GS-IR treats the irradiance I^d in the diffuse component as the combination of direct irradiance I^{d-dir} and indirect irradiance $I^{d-indir}$:

$$\begin{aligned} I^d(\mathbf{x}) &= \int_{\Omega} L_i(\mathbf{x}, \mathbf{l})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l} \\ &= \int_{\Omega_V} L_i^{dir}(\mathbf{x}, \mathbf{l})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l} + \int_{\Omega_O} L_i^{indir}(\mathbf{x}, \mathbf{l})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l} \\ &\approx (1 - O(\mathbf{x})) I^{d-dir}(\mathbf{x}) + O(\mathbf{x}) I^{d-indir}(\mathbf{x}), \end{aligned} \quad (5)$$

where Ω_V and Ω_O refer to the visible and occluded domain on the upper hemisphere Ω , respectively. In GS-IR, the ambient occlusion $O(\mathbf{x})$ is pre-calculated and cached in occlusion volumes \mathcal{V}^O during the baking stage, which can be efficiently recovered in the decomposition stage. In the baking stage, GS-IR freezes the geometry and regularly places occlusion volumes \mathcal{V}^O in the bounded 3D space. Each volume $\mathcal{V}^O \subset \mathcal{V}^O$ first captures a depth cubemap $\{D_i^{\mathcal{V}^O}\}_{i=1}^6$ and then converts it into a binary occlusion cubemap $\{O_i^{\mathcal{V}^O}\}_{i=1}^6$ through a manually set distance threshold, finally cache the low-frequent occlusion in the form of spherical harmonics (SH) coefficients f_i^O . In summary, the ambient occlusion of point \mathbf{x} can be expressed as:

$$O(\mathbf{x}) = \sum_{l=0}^{deg} \sum_{m=-l}^l f_{\mathbf{x}(lm)}^O Y_{lm}(\mathbf{n}), \quad (6)$$

where \mathbf{n} refers to the normal of point \mathbf{x} , deg denotes the degree of SH, and $\{Y_{lm}(\cdot)\}$ is a set of real basis of SH. Notably, $f_{\mathbf{x}}^O$ are the SH coefficients at point \mathbf{x} , which is obtained by conducting normal-aware masked-trilinear interpolation in the decomposition stage.

4 METHOD

In this section, we present our novel framework GUS-IR for the inverse rendering of glossy objects and complex scenes, as shown in Fig. 2. Given a set of images with unknown illumination, our GUS-IR recovers Gaussians with Unified Shading for Inverse Rendering, achieving impressive intrinsic decomposition and relighting on both glossy objects and complex scenes. Specifically, we first discuss different ways of normal modeling in Sec. 4.1. Next, we revisit the rendering equation and propose the unification of forward and deferred shading schemes in Sec. 4.2. Then, we introduce an improved baking strategy that models ambient occlusion for handling indirect illumination more accurately compared to GS-IR in Sec. 4.3. Finally, we summarize the supervision of the GUS-IR in Sec. 4.4.

4.1 Normal Modeling

The first key problem faced by introducing 3DGS for inverse rendering is reconstructing geometry, and normal estimation is crucial. Since each particle in 3DGS models

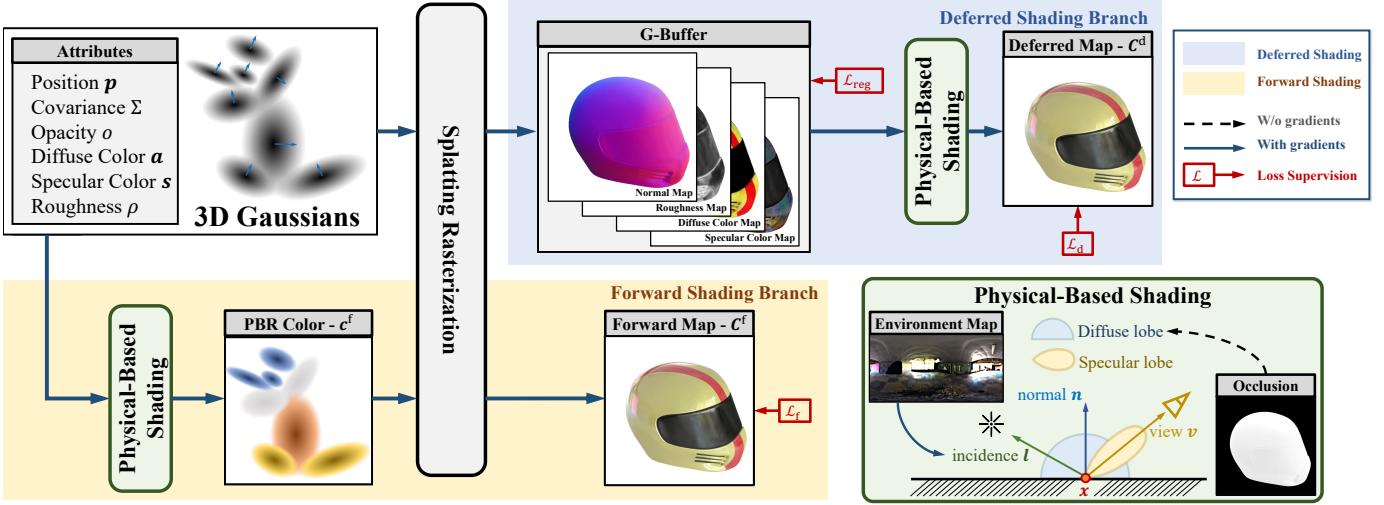


Fig. 2. Overview of GUS-IR. During optimization, GUS-IR simultaneously conducts forward and deferred shading schemes and supervises the rendering results produced by both schemes. We use the shortest axis towards the view as the particle's normal for forward shading and render a normal map for deferred shading.

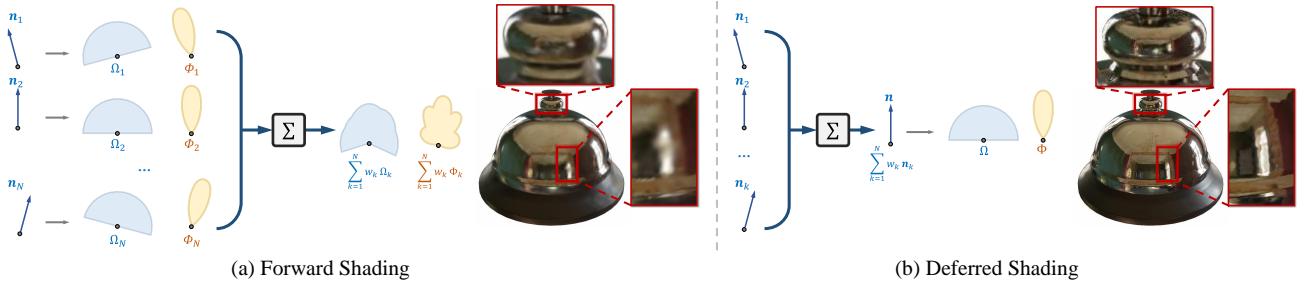


Fig. 3. Difference between forward and deferred shading schemes. Σ denotes the accumulation in the volume rendering. n_k, Ω_k, Φ_k are the normal, upper hemisphere, and BRDF lobe of the k -th particle, respectively. The table bell examples show that **deferred shading** can better capture and reconstruct glossy details than **forward shading**.

an outward attenuating opacity field instead of a well-defined geometry, 3DGS fails to natively represent normal. To overcome this limitation, previous methods [21], [22] attach each particle with a learnable normal. In contrast to them, GaussianShader [23] proposes to add a learnable residual on the shortest axis of the particle as the normal.

In GUS-IR, we directly treat the shortest axis of the particle as normal to achieve a tight geometric relation. To supervise the learning of normal in GUS-IR, we first adopt the linear interpolation scheme proposed in GS-IR to produce a reliable depth map and derive a pseudo-normal map for supervision. Then, unlike GS-IR, we directly run differentiable physical-based rendering in the first stage to better handle glossy surfaces. Please note that we do not stop the gradients of normal and the depth map in these supervisions like GS-IR. Experiments in Sec. 5.3 validate these normal representations and demonstrate the effectiveness of our adopted scheme for geometry reconstruction.

4.2 Shading Unification for Inverse Rendering

In inverse rendering, the view-dependent appearance is modeled as the physical interaction between illumination and materials as mentioned at Sec. 3. It is straightforward to use the physical-based rendering function (Eq. (2)) to

shade each particle, then perform volume rendering to accumulate the color of the pixel, which is called *forward shading*. According to Eq. (2) and Eq. (1), given N particles splatted on the pixel u , the result of forward shading $\hat{c}_{\text{for}}(u)$ can be formulated as:

$$\hat{c}_{\text{for}}(u) = \sum_{k=1}^N w_k \int_{\Omega_k} L_i(\mu_k, l) f_r(\mu_k, l, v)(l \cdot n_k) dl, \quad (7)$$

where Ω_k denotes the upper hemisphere of the k -th particle with position μ_k and normal n_k . In contrast, *deferred shading* first leverages volume rendering to obtain the geometric and physical attributes of pixel u , then obtain the pixel color $\hat{c}_{\text{def}}(u)$ through a one-times shading:

$$\begin{aligned} n &= \sum_{i=k}^N w_i n_i, \quad x = \sum_{i=k}^N w_i \mu_i, \\ \hat{c}_{\text{def}}(u) &= \int_{\Omega} L_i(x, l) \left(\sum_{i=k}^N w_i f_r(\mu_i, l, v) \right) (l \cdot n) dl, \end{aligned} \quad (8)$$

where x and n are the expected position and normal of the pixel u , the upper hemisphere Ω is centered at x around n .

To better deduce the difference in shading results, we only consider direct illumination and adopt the IBL model to simplify the lighting condition. Then we assume that

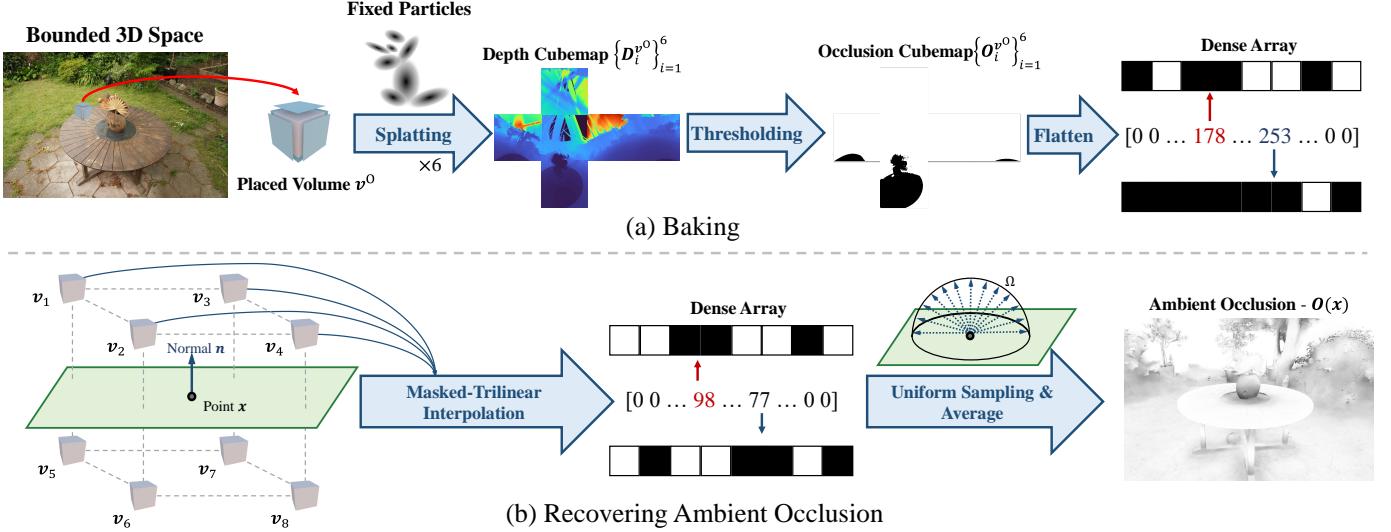


Fig. 4. Overview of caching occlusion in the baking stage and recovering occlusion in the decomposition from volumes. We employ the dense structure to bake occlusion volumes for modeling ambient occlusion. For each volume v^0 , we run six passes to render the depth cubemap $\{D_i^{v^0}\}_{i=1}^6$ and convert the cubemap into a binary occlusion cubemap $\{O_i^{v^0}\}_{i=1}^6$ through a manually set distance threshold. Then we flatten the binary occlusion cubemap and store it in a dense byte array. Given a surface point x with normal n , we conduct normal-aware masked-trilinear interpolation to get the occlusion cubemap, averaging the uniform sampling occlusion as the ambient occlusion $O(x)$.

$\sum_{k=1}^N w_k = 1$ in volume rendering and all particles have the same material (*i.e.* spatially constant). Given these conditions, the indirect radiance $L_i(\mathbf{x}, \mathbf{l})$ degenerates as $L_i(\mathbf{l})$ that is only related to incidence \mathbf{l} , and the BRDFs of all particles can be express as $f_r(\mathbf{l}, \mathbf{v})$. In this way, the shading results in Eq. (7) and Eq. (8) can be rewritten as:

$$\begin{aligned} \hat{c}_{\text{for}}(\mathbf{u}) &= \sum_{k=1}^N w_k \int_{\Omega_k} L_i(\mathbf{l}) f_r(\mathbf{l}, \mathbf{v})(\mathbf{l} \cdot \mathbf{n}_k) d\mathbf{l}, \\ \hat{c}_{\text{def}}(\mathbf{u}) &= \int_{\Omega} L_i(\mathbf{l}) \left(\sum_{i=k}^N w_k f_r(\mathbf{l}, \mathbf{v}) \right) (\mathbf{l} \cdot \mathbf{n}) d\mathbf{l} \quad (9) \\ &= \int_{\Omega} L_i(\mathbf{l}) f_r(\mathbf{l}, \mathbf{v})(\mathbf{l} \cdot \mathbf{n}) d\mathbf{l}. \end{aligned}$$

Obviously, the result of deferred shading converges to Eq. (2), while forward shading still maintains the form of volume rendering in Eq. (1). We further decompose them into diffuse and specular terms: $f_r(\mathbf{l}, \mathbf{v}) = f_r^d + f_r^s(\mathbf{l}, \mathbf{v})$. Note that the f_r^d is constant under our assumption. The forward shading result $\hat{c}_{\text{for}}(\mathbf{u})$ can be rewritten as:

$$\begin{aligned} \hat{c}_{\text{for}}(\mathbf{u}) &= \sum_{k=1}^N w_k \int_{\Omega_k} L_i(\mathbf{l}) (f_r^d + f_r^s(\mathbf{l}, \mathbf{v})) (\mathbf{l} \cdot \mathbf{n}_k) d\mathbf{l} \\ &= \sum_{k=1}^N w_k f_r^d \int_{\Omega} L_i(\mathbf{l}) (\mathbf{l} \cdot \mathbf{n}_k) d\mathbf{l} + \\ &\quad \sum_{k=1}^N w_k \int_{\Omega} L_i(\mathbf{l}) f_r^s(\mathbf{l}, \mathbf{v}) (\mathbf{l} \cdot \mathbf{n}_k) d\mathbf{l} \quad (10) \\ &\approx f_r^d \sum_{k=1}^N w_k I_k^d + \sum_{k=1}^N w_k R_k I_k^s, \\ &= \hat{c}_{\text{for}}^d(\mathbf{u}) + \hat{c}_{\text{for}}^s(\mathbf{u}), \end{aligned}$$

where I_k^d , I_k^s and R_k respectively denote the relevant pre-computed results in the split-sum approximation of the k -th

particle. The deferred shading result in Eq. (9) is convergence to surface shading, thus we directly adopt the split-sum approximation (Eq. (4)) to decompose the results as:

$$\begin{aligned} \hat{c}_{\text{def}}(\mathbf{u}) &= \int_{\Omega} L_i(\mathbf{l}) (f_r^d + f_r^s(\mathbf{l}, \mathbf{v})) (\mathbf{l} \cdot \mathbf{n}) d\mathbf{l}, \\ &= f_r^d \int_{\Omega} L_i(\mathbf{l}) (\mathbf{l} \cdot \mathbf{n}) d\mathbf{l} + \int_{\Omega} L_i(\mathbf{l}) f_r^s(\mathbf{l}, \mathbf{v}) (\mathbf{l} \cdot \mathbf{n}) d\mathbf{l} \\ &\approx f_r^d I^d + R I^s \\ &= \hat{c}_{\text{def}}^d(\mathbf{u}) + \hat{c}_{\text{def}}^s(\mathbf{u}). \end{aligned} \quad (11)$$

In a nutshell, the decomposed results of forward and deferred shading schemes are:

$$\begin{aligned} \hat{c}_{\text{for}}^d(\mathbf{u}) &= f_r^d \sum_{k=1}^N w_k I_k^d, & \hat{c}_{\text{for}}^s(\mathbf{u}) &= \sum_{k=1}^N w_k R_k I_k^s, \quad (12) \\ \hat{c}_{\text{def}}^d(\mathbf{u}) &= f_r^d I^d, & \hat{c}_{\text{def}}^s(\mathbf{u}) &= R I^s. \end{aligned}$$

Fig. 3 demonstrates more intuitive comparisons. Even though we assume all particles have the same material, different particles covering the pixel \mathbf{u} have various normals $\{\mathbf{n}_k\}_{k=1}^N$, resulting in various upper hemispheres $\{\Omega_k\}_{k=1}^N$ and BRDF lobes $\{\Phi_k\}_{k=1}^N$. Thus the integration domain of incidence in forward shading contributing to the results indeed exceeds a hemisphere and a lobe in terms of diffuse and specular terms, which means that the forward shading scheme tends to produce fuzzy results and achieves better results than the deferred shading scheme in scenes featuring rough surfaces. However, the highlights on the glossy surface are almost contributed by the incidence within extremely narrow lobes according to the importance sampling principle [73]. Therefore, under the premise that all materials are constant, the forward shading scheme further requires ensuring all particles' upper hemispheres (*i.e.* normals) are constant to achieve identical rendering results, which is impractical. The forward shading scheme is weak at handling glossy surfaces, which is the strength of the

TABLE 1

Quantitative Comparisons on object-level datasets (*i.e.* TensoIR, Shiny Blender, and Glossy Blender datasets). The methods in the second block support relighting applications, and the methods marked with \dagger achieve real-time rendering. Our method achieves the best results on TensoIR and Glossy Blender datasets.

Method	TensoIR [16]				Shiny Blender [25]				Glossy Blender [26]		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	MAE \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	MAE \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Ref-NeRF [25]	37.75	0.982	0.026	7.585	35.96	0.967	0.059	13.94	27.86	0.878	0.114
3DGSt [18]	40.82	0.990	0.014	-	30.37	0.947	0.083	-	24.96	0.872	0.104
2DGSt [24]	39.23	0.987	0.019	4.165	29.47	0.946	0.084	13.647	26.00	0.915	0.087
NDR \dagger [49]	30.19	0.962	0.052	6.079	26.36	0.886	0.261	17.047	22.11	0.856	0.165
NeRO [26]	29.08	0.949	0.067	4.969	29.84	0.962	0.072	2.360	28.76	0.943	0.062
ENVIDR [72]	33.37	0.961	0.049	8.027	35.85	0.983	0.036	4.608	29.23	0.943	0.065
TensoIR [16]	35.09	0.976	0.041	4.100	29.56	0.927	0.130	5.408	24.85	0.872	0.149
RelightGS [21]	37.57	0.983	0.020	6.078	27.31	0.920	0.120	6.115	24.21	0.893	0.106
GS-IR \dagger [22]	35.33	0.974	0.039	4.947	26.88	0.887	0.141	9.464	23.58	0.813	0.168
GShader \dagger [23]	37.54	0.984	0.022	6.525	31.94	0.957	0.068	6.115	26.04	0.911	0.104
Ours \dagger	37.61	0.985	0.020	4.485	34.28	0.973	0.056	2.193	29.44	0.953	0.057

deferred shading scheme. Our experiments in Sec. 5.3 verify this analysis. In addition, the deferred shading scheme supports modifications on the G-Buffer to change the rendering results, which is friendly to the editing task.

Based on it, we are motivated to propose shading unification in GUS-IR for inverse rendering. As shown in Fig. 2, GUS-IR consists of forward and deferred shading branches. During training, we perform both branches simultaneously and employ the same supervision. Benefiting from this combination of two branches, GUS-IR achieves remarkable results in most complex scenes, whether the scenes mainly exist on rough or glossy surfaces. In the inference, we only maintain the deferred shading branch.

4.3 Occlusion Volumes

Modeling occlusion is essential to handle indirect illumination. Ambient occlusion (AO) represents the accessibility of a surface point, and is an efficient and promising way to tackle the low-frequent occlusion. To model the AO in a scene represented by particles in 3DGS, as mentioned at Sec. 3, GS-IR uses spherical harmonics (SH) bases to convolve the binary occlusion cubemap $\{\mathcal{D}_i^o\}_{i=1}^6$ and obtain SH coefficients f^o to cache low-frequency occlusion in regularly placed volumes. However, the SH convolution cannot perfectly approximate the result of the hemispherical integration of the occlusion cubemap and suffers from the ‘leaking’ problem. To this end, in GUS-IR, we decided to directly store the binary occlusion cubemap in each volume and obtain AO through runtime sampling. As shown in Fig. 4, we follow GS-IR, regularly place the volumes in the bounded 3D scene, and get the occlusion cubemap $\{\mathcal{D}_i^{v^o}\}_{i=1}^6$ in the same way. In contrast to GS-IR which compresses the occlusion information in a set of SH coefficients, we flatten the occlusion cubemap into a dense byte array, where each byte can represent 8 pixels’ occlusion.

In the decomposition stage, we follow GS-IR and conduct normal-aware masked-trilinear interpolation to get the occlusion cubemap for each particle. For a particle with position \mathbf{x} and normal \mathbf{n} , we uniformly sample directions of its upper hemisphere centered at \mathbf{n} and query the occlusion on this array based on the directions. Then we average the values as the particle’s ambient occlusion $O(\mathbf{x})$. The

ambient occlusion $O(\mathbf{x})$ is used in Eq. (5). For deferred shading, we conduct volumetric accumulation to obtain the pixel’s ambient occlusion. The whole recovery still achieves real-time performance.

4.4 Losses

In GUS-IR, we simultaneously supervise the forward and deferred shading results in Eq. (10) and Eq. (11):

$$\begin{aligned}\mathcal{L}_f &= (1 - \lambda)\mathcal{L}_1(\hat{\mathcal{C}}_{\text{for}}, \mathbf{C}) + \lambda\mathcal{L}_{\text{D-SSIM}}(\hat{\mathcal{C}}_{\text{for}}, \mathbf{C}), \\ \mathcal{L}_d &= (1 - \lambda)\mathcal{L}_1(\hat{\mathcal{C}}_{\text{def}}, \mathbf{C}) + \lambda\mathcal{L}_{\text{D-SSIM}}(\hat{\mathcal{C}}_{\text{def}}, \mathbf{C}),\end{aligned}\quad (13)$$

where $\mathbf{C}, \hat{\mathcal{C}}_{\text{for}}, \hat{\mathcal{C}}_{\text{def}}$ denote the reference image, the images rendered by performing forward and deferred shading schemes, respectively.

Following GS-IR [22] and GaussianShader [23], we also and leverage the gradient derived from the rendered depth map to supervise the learned normal and encourage the volumetric-accumulated alpha to approach 0 or 1:

$$\begin{aligned}\mathcal{L}_n &= \frac{1}{\|\mathbf{C}\|} \sum_{\mathbf{u} \in \mathbf{C}} \left[1 - \hat{\mathbf{n}}(\mathbf{u})^\top \cdot \nabla \hat{d}(\mathbf{u}) \right], \\ \mathcal{L}_\alpha &= \frac{1}{\|\mathbf{C}\|} \sum_{\mathbf{u} \in \mathbf{C}} [\log(\alpha(\mathbf{u})) + \log(1 - \alpha(\mathbf{u}))], \\ \mathcal{L}_{\text{reg}} &= \lambda_n \mathcal{L}_n + \lambda_\alpha \mathcal{L}_\alpha,\end{aligned}\quad (14)$$

where $\hat{\mathbf{n}}(\mathbf{u})$ denotes the volumetric accumulated normal of pixel \mathbf{u} , $\nabla \hat{d}(\mathbf{u})$ denotes the gradient to the depth of pixel \mathbf{u} , $\lambda_n = 0.1$ and $\lambda_\alpha = 0.001$. The summary loss \mathcal{L} is:

$$\mathcal{L} = \mathcal{L}_f + \mathcal{L}_d + \mathcal{L}_{\text{reg}}. \quad (15)$$

5 EXPERIMENTS

5.1 Implementation

Dataset & Metrics We conduct experiments on three object-level datasets (*i.e.* TensorIR Synthetic [16], Shiny Blender [25] datasets, and Glossy Blender [26] datasets) and two real scene-level datasets (*i.e.* Mip-NeRF 360 [27] and Ref-NeRF Real [25] datasets), which contain 18 objects and 12 publicly available scenes, covering both rough and glossy surfaces. Specifically, TensoIR and Shiny Blender datasets provide

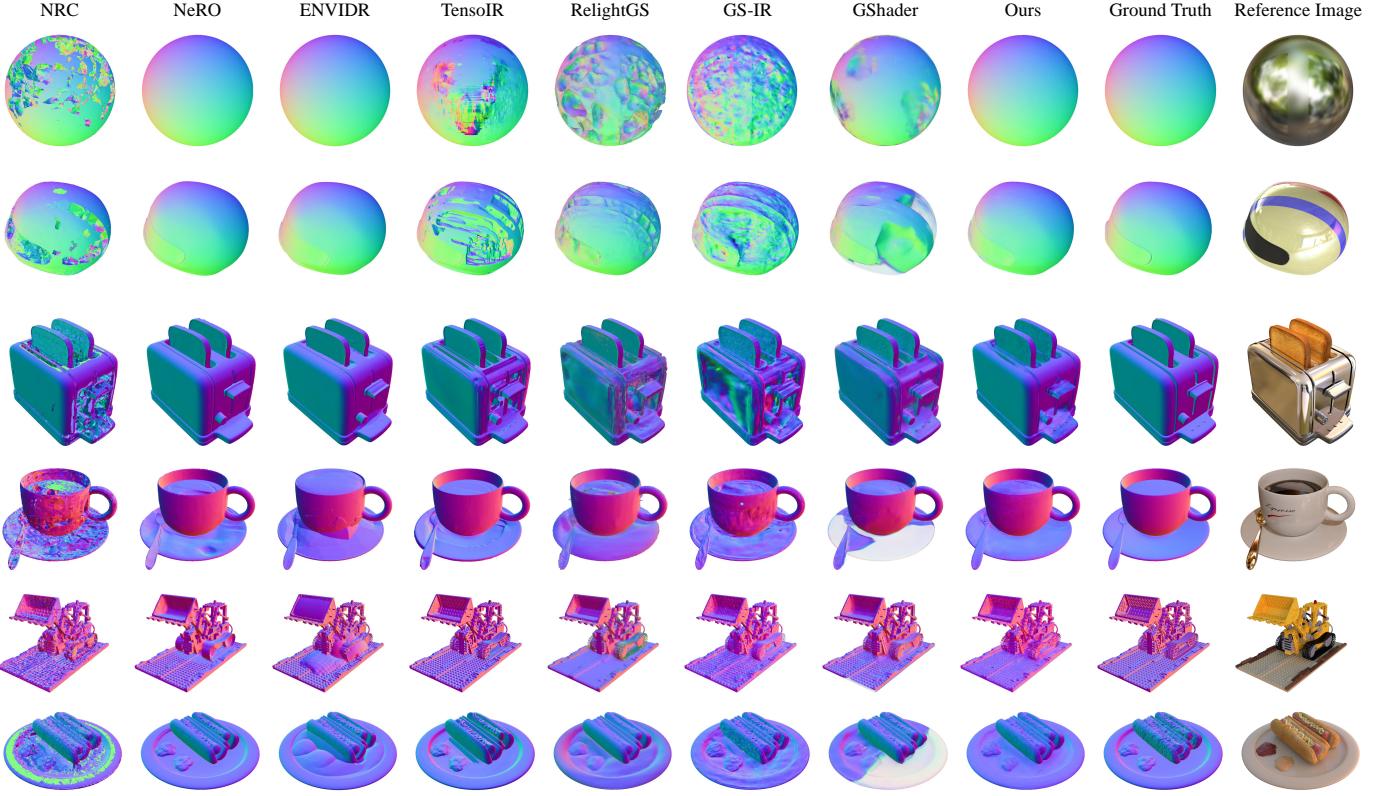


Fig. 5. Normal estimation comparisons on TensoIR [16] and Shiny Blender [25] datasets. We exemplify the normal estimation results of our GUS-IR and other cutting-edge methods.

TABLE 2

Quantitative Comparisons of Relighting on TensoIR and Glossy Blender datasets. The methods marked with \dagger achieve real-time rendering.

Method	TensoIR [16]			Glossy Blender [26]		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
NeRO [26]	26.20	0.891	0.096	28.77	0.950	0.057
ENVIDR [72]	24.81	0.869	0.095	19.21	0.789	0.115
TensoIR [16]	28.58	0.944	0.081	15.19	0.786	0.213
RelightGS [21]	24.41	0.890	0.100	20.22	0.856	0.112
NDR \dagger [49]	19.88	0.879	0.104	24.54	0.859	0.135
GS-IR \dagger [22]	24.37	0.885	0.096	17.50	0.818	0.136
GShader \dagger [22]	22.42	0.872	0.103	21.61	0.874	0.105
Ours \dagger	27.53	0.919	0.085	25.17	0.922	0.069

normal references for evaluating the quality of geometry reconstruction. TensoIR and Glossy Blender datasets further support the relighting evaluation. Generally, we evaluate the novel-view synthesis results on all datasets in terms of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [75]. To evaluate the relighting results, we use the same metrics as novel-view synthesis. In addition, we leverage mean angular error (MAE) to evaluate the normal estimation quality.

Implementation Details In the first stage, we follow the training schedule of 3DGS [18] to optimize the GUS-IR for 30K iterations using the Adam optimizer [76] and use the loss defined in Eq. (15) for supervision. For caching the oc-

clusion in the baking stage, we regularly place 128^3 volumes in the space $[-1.5, 1.5]^3$ for objects and 256^3 volumes in the space $[-8, 8]^3$ for scenes to cache the occlusion and model the indirect irradiance. After the baking stage, we only use \mathcal{L}_f and \mathcal{L}_d defined in Eq. (13) to optimize GUS-IR for 5K iterations.

5.2 Comparisons

We conduct a comprehensive comparison against cutting-edge novel-view synthesis and inverse rendering methods on public datasets. All methods take multi-view images captured under unknown lighting conditions as input.

Object-level In Tab. 1, we show the novel-view synthesis and normal estimation performance of all methods on the object-level datasets (*i.e.* TensoIR [16], Shiny Blender [25], and Glossy Blender [26]). Our methods almost surpass other relightable methods. The qualitative comparisons of normal estimation and novel-view synthesis are respectively shown in Fig. 5 and Fig. 6. It is worth noting that these datasets contain complex rough and glossy surfaces. Both quantitative and qualitative results demonstrate that our method can effectively decompose any surface. Notably, GUS-IR is one of the few methods that achieve real-time rendering. In addition, our method also achieves remarkable relighting performance as shown in Tab. 2. Fig. 8 shows the relighting results of GUS-IR on objects.

Scene-level In Tab. 3, we show the novel-view synthesis comparisons on the scene-level datasets, including Mip-NeRF 360 [27] and Ref-NeRF Real [25] datasets. The results

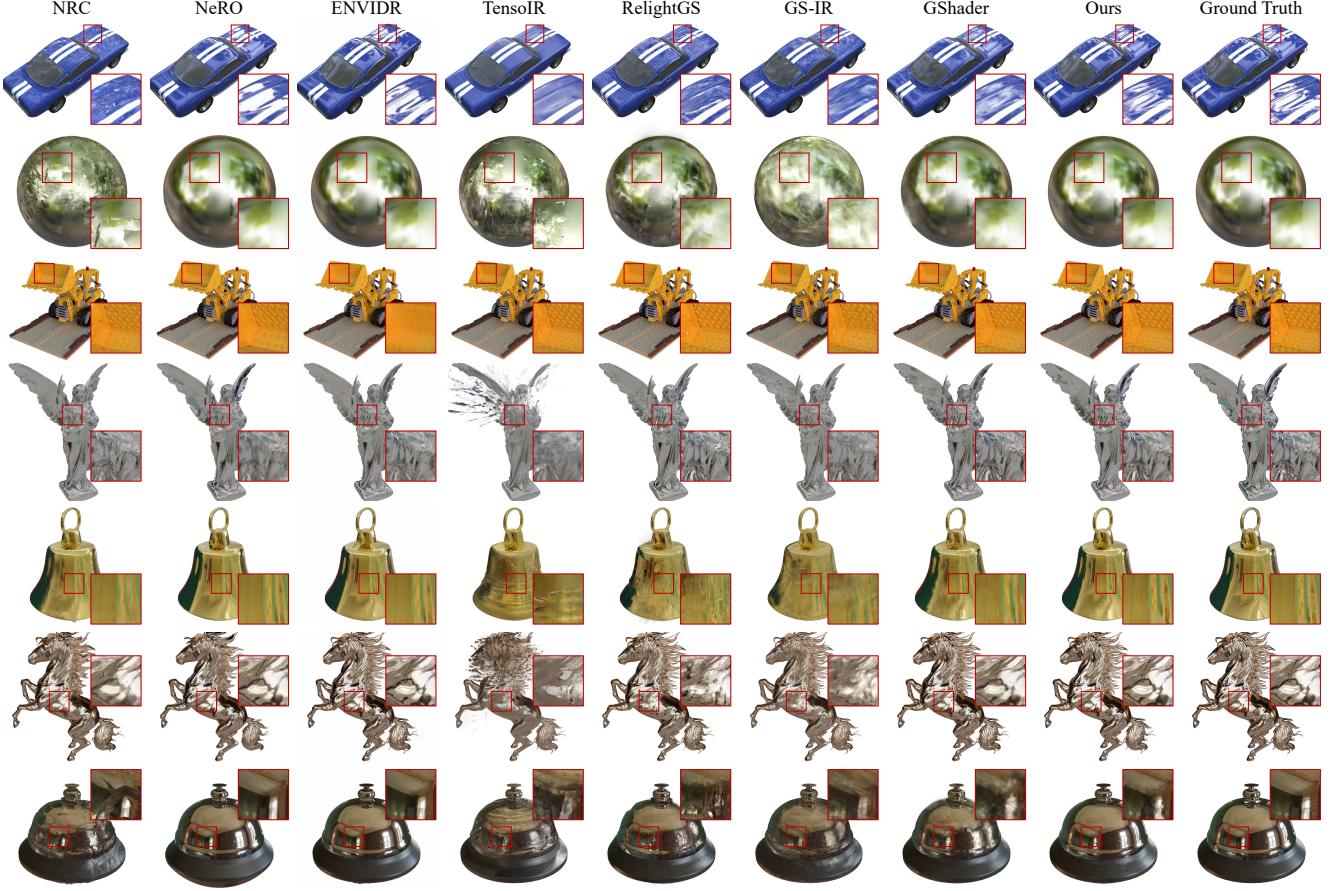


Fig. 6. Novel-view synthesis comparisons on object-level datasets (*i.e.* TensoIR, Shiny Blender, and Glossy Blender datasets). We exemplify the rendering results of our GUS-IR and other cutting-edge methods.

TABLE 3

Quantitative Comparisons on scene-level datasets (*i.e.* Mip-NeRF 360 and Ref-NeRF Real datasets). The methods in the second block support relighting applications, and the methods marked with \dagger achieve real-time rendering. Please note that our GUS-IR achieves comparable performance on the Ref-NeRF Real dataset against other methods that only focus on novel-view synthesis.

Method	Mip-NeRF 360 Outdoor [27]			Mip-NeRF 360 Indoor [27]			Ref-NeRF Real [25]		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Mip-NeRF 360 [27]	24.43	0.694	0.278	31.49	0.918	0.179	24.27	0.649	0.276
UniSDF [74]	24.77	0.723	0.241	31.28	0.901	0.181	23.70	0.635	0.266
3DGS \dagger [18]	24.64	0.731	0.234	30.41	0.917	0.190	23.67	0.632	0.288
2DGS \dagger [24]	24.33	0.709	0.284	30.39	0.922	0.183	23.65	0.634	0.285
GS-IR \dagger [22]	23.45	0.671	0.284	27.80	0.870	0.248	23.32	0.625	0.283
GaussianShader \dagger [16]	22.80	0.665	0.297	26.61	0.878	0.243	22.96	0.624	0.294
Ours \dagger	23.76	0.696	0.276	28.98	0.902	0.222	23.67	0.646	0.282

show that GUS-IR achieves the best performance among relightable methods. Notably, GUS-IR achieves comparable performance to other methods that only focus on novel-view synthesis. Fig. 10 shows the relighting results of GUS-IR on scenes.

In summary, GUS-IR achieves high-quality real-time inverse rendering by incorporating unified shading into the 3DGS framework, which effectively decomposes the physical properties of rough and glossy surfaces in complex scenes. We further show the decomposition on the glossy surfaces of GUS-IR in Fig. 7. The relighting results in Fig. 8 and Fig. 10 verify that GUS-IR can reasonably decompose

physical properties and produce harmonious relighting effects.

5.3 Ablation Studies

To evaluate the efficacy of our proposed schemes, we design elaborate experiments on the same five datasets as in the above comparison for ablation. We analyze the impact of different shading schemes, normal representations, and modeling indirect illumination on the final performance.

Analysis on Different Shading Schemes Employing a correct shading scheme is critical to inverse rendering, as it

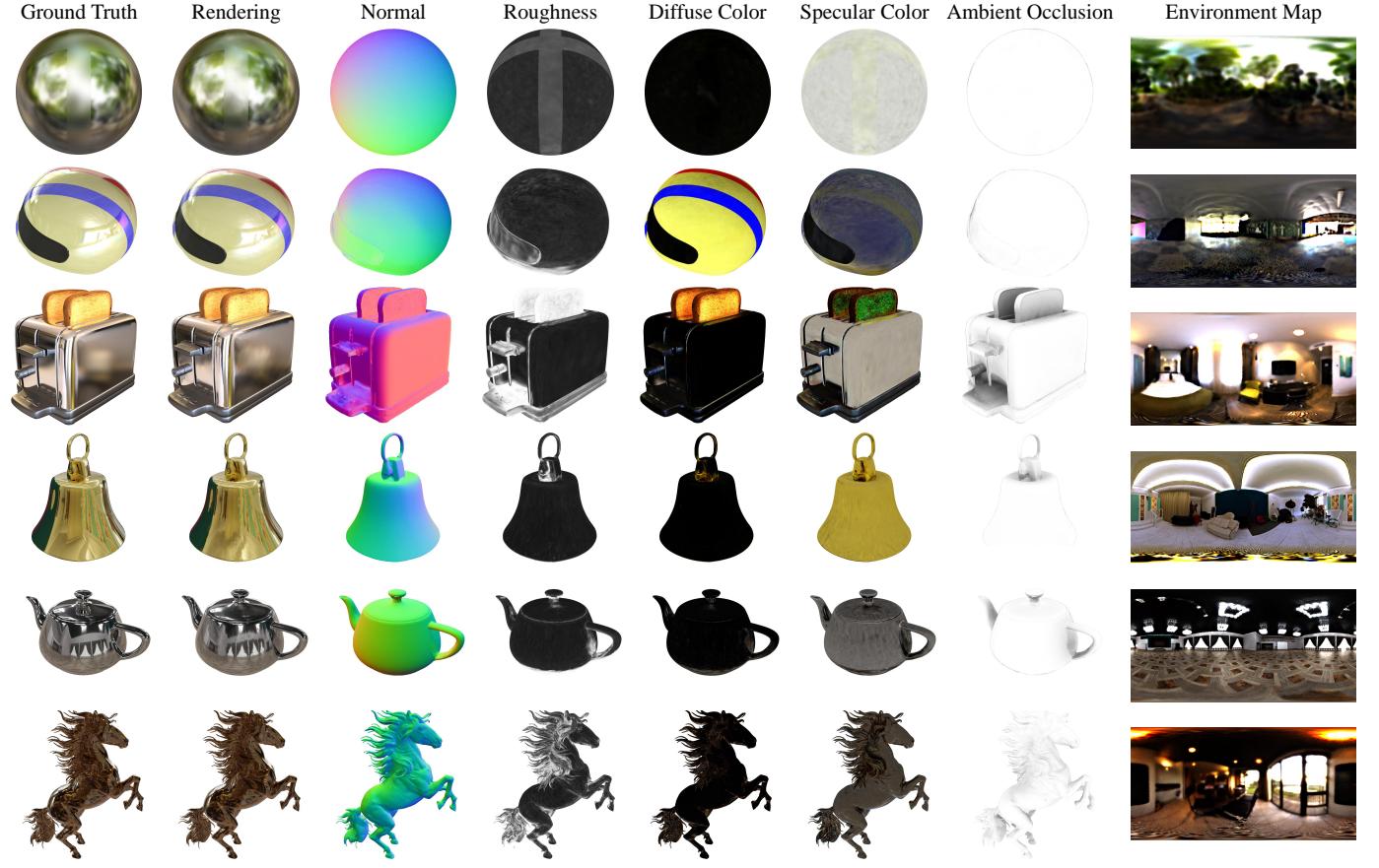


Fig. 7. Decomposition results of GUS-IR on Shiny Blender [25] and Glossy Blender [26] datasets, including objects with glossy surfaces.

TABLE 4

Ablation on object-level datasets (*i.e.* TensoIR [16], Shiny Blender [25], and Glossy Blender [26] datasets). *Only Forward* and *Only Deferred* respectively refer to only employing forward or deferred shading schemes. Other alternatives use the unified shading scheme. *Attach Normal* indicates attaching a learnable normal for each particle. *Residual Normal* denotes associating a predicted residual on the shortest axis as the normal for each particle. *W/o Indirect* denotes ignoring ambient occlusion and indirect illumination.

Method	TensoIR [16]				Shiny Blender [25]				Glossy Blender [26]		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	MAE \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	MAE \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Only Forward	37.74	0.985	0.020	4.680	33.15	0.965	0.067	2.436	28.64	0.928	0.084
Only Deferred	36.35	0.980	0.029	4.769	33.58	0.970	0.059	2.370	29.12	0.950	0.059
Attach Normal	37.67	0.984	0.022	5.433	33.43	0.968	0.060	2.457	28.89	0.946	0.063
Residual Normal	37.62	0.985	0.021	4.761	34.17	0.971	0.058	2.221	29.43	0.952	0.063
W/o Indirect	37.37	0.983	0.024	4.485	34.20	0.970	0.058	2.193	29.28	0.951	0.058
Ours	37.61	0.985	0.020	4.485	34.28	0.973	0.056	2.193	29.44	0.953	0.057

TABLE 5

Analysis of the impact of different conditions to relighting on object-level datasets (*i.e.* TensoIR [16] and Glossy Blender [25] datasets).

Method	TensoIR [16]			Glossy Blender [26]		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Only Forward	27.15	0.917	0.086	23.16	0.873	0.105
Only Defer	26.83	0.915	0.087	23.99	0.903	0.073
Attach Normal	26.12	0.913	0.089	22.81	0.881	0.092
Residual Normal	26.32	0.914	0.089	24.72	0.911	0.073
Ours	27.53	0.919	0.085	25.17	0.922	0.069

relates to the decomposition quality. As stated in Sec. 4.2, our employed unified shading scheme not only restores

the diffuse material of rough surfaces but also captures the highlight details, handling glossy surfaces. The results in Tab. 4 and Tab. 6 show that the forward shading slightly outperforms other schemes on the TensoIR and Mip-NeRF 360 indoor datasets that are full of rough surfaces. However, it is significantly inferior to other schemes on the Shiny Blender and Glossy Blender datasets, mainly containing objects with glossy surfaces. The relighting results on Tab. 5 further verify the efficacy of our scheme.

Analysis on Normal Representations Modeling normals is the primary problem faced when using 3DGS for 3D reconstruction and inverse rendering. The results in Tab. 4, Tab. 6, and Tab. 5 show that our employed scheme (using the shortest axis as the normal for each particle) is almost



Fig. 8. Relighting results on object-level datasets. We perform relighting experiments on objects using the geometry, and material recovered from our GUS-IR method. We test our method under different lighting conditions. The results show that our method handles both rough and glossy surfaces well.

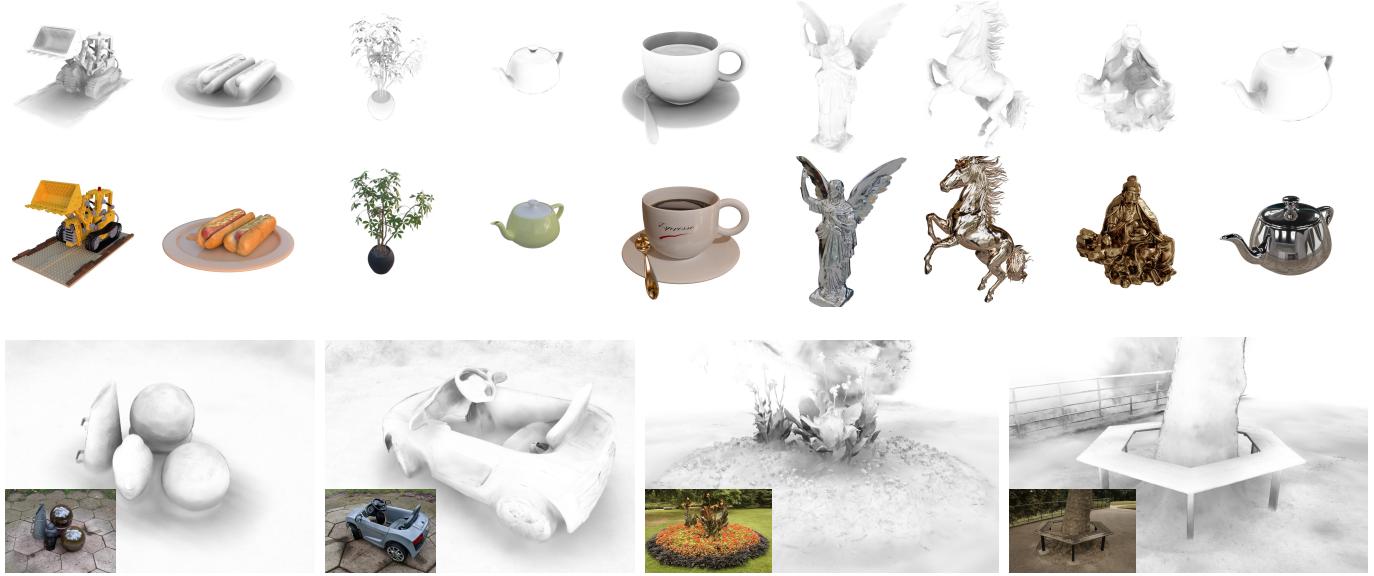


Fig. 9. Visualization of the recovered ambient occlusion. The visualization highlights the intricate shadowing and occlusion details captured by our GUS-IR method, emphasizing the performance of our approach in modeling indirect illumination.

superior to *Attach Normal* and *Residual Normal* in novel-view synthesis, especially in relighting tasks. It is worth noting that the normal estimation (*i.e.* MAE) of our employed scheme significantly outperforms *Attach Normal* and *Residual Normal*, and the *Residual Normal* surpasses *Attach Normal Normal*.

Analysis on Indirect Illumination To verify the efficacy of

modeling indirect illumination, we compare our results with those of the first stage, skipping baking and subsequent stages, referred to *W/o Indirect*. Results in Tab. 4, Tab. 6 verify that modeling diffuse indirect illumination using our proposed scheme as illustrated at Sec. 4.3 improves the novel-view synthesis results. Fig. 9 showcases the ambient occlusion produced by our proposed scheme, which verifies that our scheme can capture the complex shadowing and

TABLE 6

Ablation on scene-level datasets (*i.e.* Mip-NeRF 360 [27] and Ref-NeRF Real [25] datasets). *Only Forward* and *Only Deferred* respectively refer to only employing forward or deferred shading schemes, Other alternatives use the unified shading scheme. *Attach Normal* indicates attaching a learnable normal for each particle. *Residual Normal* denotes associating a predicted residual on the shortest axis as the normal for each particle. *W/o Indirect* denotes ignoring ambient occlusion and indirect illumination.

Method	Mip-NeRF 360 Outdoor [27]			Mip-NeRF 360 Indoor [27]			Ref-NeRF Real [25]		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Only Forward	23.72	0.695	0.281	29.08	0.904	0.223	22.89	0.619	0.305
Only Defer	23.53	0.679	0.285	28.84	0.891	0.233	23.49	0.641	0.278
Attach Normal	23.89	0.701	0.269	28.93	0.902	0.225	23.02	0.635	0.282
Residual Normal	23.79	0.699	0.273	28.90	0.901	0.223	23.66	0.648	0.280
W/o Indirect	23.66	0.693	0.280	28.52	0.900	0.225	23.56	0.642	0.285
Ours	23.76	0.696	0.276	28.98	0.902	0.222	23.67	0.646	0.282

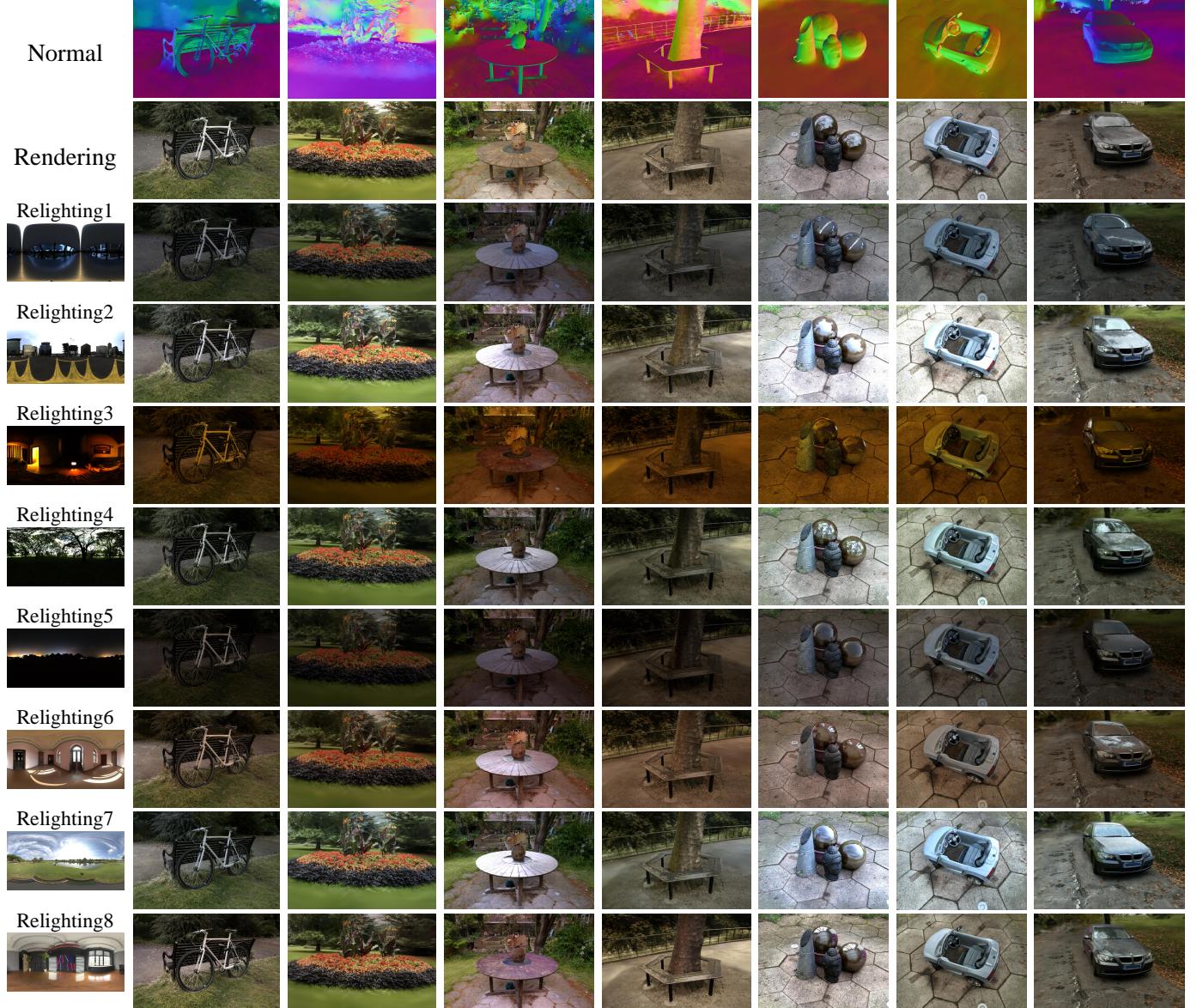


Fig. 10. Relighting results on scene-level datasets. We perform relighting experiments on scenes using the geometry, and material recovered from our GUS-IR method. The results in the second and third columns on the right show that GUS-IR resolves the glossy marble surfaces and the car metal shell, reflecting clear environment illumination.

occlusion.



(a) Relighting of GUS-IR (b) Relighting Reference

Fig. 11. GUS-IR fails to estimate smooth and accurate normals on the waist body of the bell, leading to distorted relighting results.

6 CONCLUSION

In this paper, we introduce a new GUS-IR framework for recovering Gaussians with Unified Shading for Inverse Rendering from a set of images with natural illumination. Our approach achieves impressive intrinsic decomposition and relighting on both glossy objects and complex scenes. We conduct a thorough comparison and analysis of two prominent shading schemes, namely forward shading and deferred shading, commonly used for 3DGS-based inverse rendering tasks. Based on this analysis, we propose a unified shading solution that combines the advantages of both shading schemes, leading to better intrinsic decomposition suitable for complex scenes with rough surfaces and glossy objects with high specular regions. Additionally, we incorporate the shortest-axis normal modeling with depth-related regularization to represent reliable geometry, resulting in improved shape reconstruction. Furthermore, we enhance the probe-based baking solution to better model ambient occlusion and handle indirect illumination. Through extensive experiments on various challenging datasets, including complex scenes and glossy objects, we demonstrate the superiority of our method over state-of-the-art methods qualitatively and quantitatively. Our results highlight the efficacy of GUS-IR in achieving high-quality intrinsic decomposition, particularly for glossy surfaces.

Limitation In GUS-IR, we take the shortest axis as normal for each particle. Although our scheme significantly outperforms the other two schemes (*Attach Normal* and *Residual Normal*), it still struggles to accurately reconstruct the normal of the glossy surfaces compared to the method based on implicit nerual field to construct SDF (e.g. NeRO [26]). This leads to distorted relighting results on smooth and glossy surfaces as shown in Fig. 11, damaging the relighting quality. This is the reason that our method is notably inferior to NeRO on the Glossy Blender dataset in Tab. 2. We believe it is valuable to improve the geometry representation of 3DGS and address this limitation in future work.

REFERENCES

- [1] S. Shen, "Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes," *IEEE transactions on image processing*, vol. 22, no. 5, pp. 1901–1914, 2013.
- [2] E. Zheng, E. Dunn, V. Jovic, and J.-M. Frahm, "Patchmatch based joint view selection and depthmap estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1510–1517.
- [3] S. Galliani, K. Lasinger, and K. Schindler, "Massively parallel multiview stereopsis by surface normal diffusion," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 873–881.
- [4] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4104–4113.
- [5] J. L. Schönberger, E. Zheng, J.-M. Frahm, and M. Pollefeys, "Pixel-wise view selection for unstructured multi-view stereo," in *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 501–518.
- [6] A. Romanoni and M. Matteucci, "Tapa-mvs: Textureless-aware patchmatch multi-view stereo," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 10 413–10 422.
- [7] Q. Xu and W. Tao, "Planar prior assisted patchmatch multi-view stereo," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 12 516–12 523.
- [8] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *European Conference on Computer Vision*, 2020, pp. 405–421.
- [9] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," *arXiv preprint arXiv:2106.10689*, 2021.
- [10] L. Yariv, J. Gu, Y. Kasten, and Y. Lipman, "Volume rendering of neural implicit surfaces," *Advances in Neural Information Processing Systems*, vol. 34, pp. 4805–4815, 2021.
- [11] J. Wang, P. Wang, X. Long, C. Theobalt, T. Komura, L. Liu, and W. Wang, "Neuris: Neural reconstruction of indoor scenes using normal priors," in *European Conference on Computer Vision*. Springer, 2022, pp. 139–155.
- [12] Z. Yu, S. Peng, M. Niemeyer, T. Sattler, and A. Geiger, "Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction," *Advances in neural information processing systems*, vol. 35, pp. 25 018–25 032, 2022.
- [13] Z. Liang, Z. Huang, C. Ding, and K. Jia, "Helixsurf: A robust and efficient neural implicit surface learning of indoor scenes with iterative intertwined regularization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13 165–13 174.
- [14] X. Zhang, P. P. Srinivasan, B. Deng, P. Debevec, W. T. Freeman, and J. T. Barron, "Nerfactor: Neural factorization of shape and reflectance under an unknown illumination," *ACM Transactions on Graphics (ToG)*, vol. 40, no. 6, pp. 1–18, 2021.
- [15] P. P. Srinivasan, B. Deng, X. Zhang, M. Tancik, B. Mildenhall, and J. T. Barron, "Nerv: Neural reflectance and visibility fields for relighting and view synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7495–7504.
- [16] H. Jin, I. Liu, P. Xu, X. Zhang, S. Han, S. Bi, X. Zhou, Z. Xu, and H. Su, "Tensoro: Tensorial inverse rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 165–174.
- [17] Y. Zhuang, Q. Zhang, X. Wang, H. Zhu, Y. Feng, X. Li, Y. Shan, and X. Cao, "A pre-convolved representation for plug-and-play neural illumination fields," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 7, 2024, pp. 7828–7836.
- [18] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics (ToG)*, vol. 42, no. 4, pp. 1–14, 2023.
- [19] M. Zwicker, H. Pfister, J. Van Baar, and M. Gross, "Ewa splatting," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 3, pp. 223–238, 2002.
- [20] ———, "Surface splatting," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 371–378.
- [21] J. Gao, C. Gu, Y. Lin, H. Zhu, X. Cao, L. Zhang, and Y. Yao, "Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing," *arXiv preprint arXiv:2311.16043*, 2023.
- [22] Z. Liang, Q. Zhang, Y. Feng, Y. Shan, and K. Jia, "Gs-ir: 3d gaussian splatting for inverse rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 21 644–21 653.
- [23] Y. Jiang, J. Tu, Y. Liu, X. Gao, X. Long, W. Wang, and Y. Ma, "Gaussianshader: 3d gaussian splatting with shading functions

- for reflective surfaces," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5322–5332.
- [24] B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao, "2d gaussian splatting for geometrically accurate radiance fields," in *ACM SIGGRAPH 2024 Conference Papers*, 2024, pp. 1–11.
- [25] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, "Ref-nerf: Structured view-dependent appearance for neural radiance fields," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2022, pp. 5481–5490.
- [26] Y. Liu, P. Wang, C. Lin, X. Long, J. Wang, L. Liu, T. Komura, and W. Wang, "Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images," *ACM Transactions on Graphics (TOG)*, vol. 42, no. 4, pp. 1–22, 2023.
- [27] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5470–5479.
- [28] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 5855–5864.
- [29] S. Bi, Z. Xu, P. Srinivasan, B. Mildenhall, K. Sunkavalli, M. Hašan, Y. Hold-Geoffroy, D. Kriegman, and R. Ramamoorthi, "Neural reflectance fields for appearance acquisition," *arXiv preprint arXiv:2008.03824*, 2020.
- [30] S. Fridovich-Keil, G. Meanti, F. R. Warburg, B. Recht, and A. Kanazawa, "K-planes: Explicit radiance fields in space, time, and appearance," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 12479–12488.
- [31] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, and A. Kanazawa, "Plenoxels: Radiance fields without neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5501–5510.
- [32] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, "Nerfies: Deformable neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5865–5874.
- [33] E. R. Chan, M. Monteiro, P. Kellnhofer, J. Wu, and G. Wetzstein, "pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 5799–5809.
- [34] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "Dreamfusion: Text-to-3d using 2d diffusion," *arXiv preprint arXiv:2209.14988*, 2022.
- [35] Z. Wang, C. Lu, Y. Wang, F. Bao, C. Li, H. Su, and J. Zhu, "Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [36] K. Zhang, N. Kolkin, S. Bi, F. Luan, Z. Xu, E. Shechtman, and N. Snavely, "Arf: Artistic radiance fields," in *European Conference on Computer Vision*. Springer, 2022, pp. 717–733.
- [37] X. Huang, Q. Zhang, Y. Feng, H. Li, X. Wang, and Q. Wang, "Hdrr-nerf: High dynamic range neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18398–18408.
- [38] L. Ma, X. Li, J. Liao, Q. Zhang, X. Wang, J. Wang, and P. V. Sander, "Deblur-nerf: Neural radiance fields from blurry images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12861–12870.
- [39] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *ACM SIGGRAPH Computer Graphics*, vol. 21, no. 4, pp. 163–169, 1987.
- [40] T. Ju, F. Losasso, S. Schaefer, and J. Warren, "Dual contouring of hermite data," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002, pp. 339–346.
- [41] P. Hedman, P. P. Srinivasan, B. Mildenhall, J. T. Barron, and P. Debevec, "Baking neural radiance fields for real-time view synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5875–5884.
- [42] C. Sun, M. Sun, and H.-T. Chen, "Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5459–5469.
- [43] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics (ToG)*, vol. 41, no. 4, pp. 1–15, 2022.
- [44] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, "Tensorf: Tensorial radiance fields," in *European Conference on Computer Vision*. Springer, 2022, pp. 333–350.
- [45] A. Cao and J. Johnson, "Hexplane: A fast representation for dynamic scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 130–141.
- [46] Q. Xu, Z. Xu, J. Philip, S. Bi, Z. Shu, K. Sunkavalli, and U. Neumann, "Point-nerf: Point-based neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5438–5448.
- [47] Z. Huang, Z. Liang, H. Zhang, Y. Lin, and K. Jia, "Sur2f: A hybrid representation for high-quality and efficient surface reconstruction from multi-view images," *arXiv preprint arXiv:2401.03704*, 2024.
- [48] T. Shen, J. Gao, K. Yin, M.-Y. Liu, and S. Fidler, "Deep marching tetrahedra: a hybrid representation for high-resolution 3d shape synthesis," *Advances in Neural Information Processing Systems*, vol. 34, pp. 6087–6101, 2021.
- [49] J. Munkberg, J. Hasselgren, T. Shen, J. Gao, W. Chen, A. Evans, T. Müller, and S. Fidler, "Extracting triangular 3d models, materials, and lighting from images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8280–8290.
- [50] J. Hasselgren, N. Hofmann, and J. Munkberg, "Shape, light, and material decomposition from images using monte carlo rendering and denoising," *Advances in Neural Information Processing Systems*, vol. 35, pp. 22856–22869, 2022.
- [51] Z. Yu, A. Chen, B. Huang, T. Sattler, and A. Geiger, "Mip-splatting: Alias-free 3d gaussian splatting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 19447–19456.
- [52] Z. Liang, Q. Zhang, W. Hu, Y. Feng, L. Zhu, and K. Jia, "Analytic-splatting: Anti-aliased 3d gaussian splatting via analytic integration," *arXiv preprint arXiv:2403.11056*, 2024.
- [53] Z. Yu, T. Sattler, and A. Geiger, "Gaussian opacity fields: Efficient and compact surface reconstruction in unbounded scenes," *arXiv preprint arXiv:2404.10772*, 2024.
- [54] B. Zhang, C. Fang, R. Shrestha, Y. Liang, X. Long, and P. Tan, "Rade-gs: Rasterizing depth in gaussian splatting," *arXiv preprint arXiv:2406.01467*, 2024.
- [55] G. Wu, T. Yi, J. Fang, L. Xie, X. Zhang, W. Wei, W. Liu, Q. Tian, and X. Wang, "4d gaussian splatting for real-time dynamic scene rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 20310–20320.
- [56] Q. Gao, Q. Xu, Z. Cao, B. Mildenhall, W. Ma, L. Chen, D. Tang, and U. Neumann, "Gaussianflow: Splatting gaussian dynamics for 4d content creation," *arXiv preprint arXiv:2403.12365*, 2024.
- [57] R. L. Cook and K. E. Torrance, "A reflectance model for computer graphics," *ACM Transactions on Graphics (ToG)*, vol. 1, no. 1, pp. 7–24, 1982.
- [58] J. T. Kajiya, "The rendering equation," in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, 1986, pp. 143–150.
- [59] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance, "Microfacet models for refraction through rough surfaces," in *Proceedings of the 18th Eurographics conference on Rendering Techniques*, 2007, pp. 195–206.
- [60] Y. Dong, G. Chen, P. Peers, J. Zhang, and X. Tong, "Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, pp. 1–12, 2014.
- [61] R. Xia, Y. Dong, P. Peers, and X. Tong, "Recovering shape and spatially-varying surface reflectance under unknown illumination," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, pp. 1–12, 2016.
- [62] S. Bi, Z. Xu, K. Sunkavalli, M. Hašan, Y. Hold-Geoffroy, D. Kriegman, and R. Ramamoorthi, "Deep reflectance volumes: Relightable reconstructions from multi-view photometric images," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 2020, pp. 294–311.
- [63] S. Bi, Z. Xu, K. Sunkavalli, D. Kriegman, and R. Ramamoorthi, "Deep 3d capture: Geometry and reflectance from sparse multi-view images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5960–5969.

- [64] F. Luan, S. Zhao, K. Bala, and Z. Dong, "Unified shape and svbrdf recovery using differentiable monte carlo rendering," in *Computer Graphics Forum*, vol. 40, no. 4. Wiley Online Library, 2021, pp. 101–113.
- [65] G. Nam, J. H. Lee, D. Gutierrez, and M. H. Kim, "Practical svbrdf acquisition of 3d objects with unstructured flash photography," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–12, 2018.
- [66] K. Zhang, F. Luan, Q. Wang, K. Bala, and N. 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*, 2021, pp. 5453–5462.
- [67] M. Boss, R. Braun, V. Jampani, J. T. Barron, C. Liu, and H. Lensch, "Nerd: Neural reflectance decomposition from image collections," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 684–12 694.
- [68] M. Boss, V. Jampani, R. Braun, C. Liu, J. Barron, and H. Lensch, "Neural-pil: Neural pre-integrated lighting for reflectance decomposition," *Advances in Neural Information Processing Systems*, vol. 34, pp. 10 691–10 704, 2021.
- [69] Y. Zhang, J. Sun, X. He, H. Fu, R. Jia, and X. Zhou, "Modeling indirect illumination for inverse rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18 643–18 652.
- [70] K. Zhang, F. Luan, Z. Li, and N. Snavely, "Iron: Inverse rendering by optimizing neural sdbs and materials from photometric images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5565–5574.
- [71] B. Karis and E. Games, "Real shading in unreal engine 4," *Proc. Physically Based Shading Theory Practice*, vol. 4, no. 3, p. 1, 2013.
- [72] R. Liang, H. Chen, C. Li, F. Chen, S. Panneer, and N. Vijaykumar, "Enviridr: Implicit differentiable renderer with neural environment lighting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 79–89.
- [73] E. Veach, *Robust Monte Carlo Methods for Light Transport Simulation*. Stanford University, 1998.
- [74] F. Wang, M.-J. Rakotosaona, M. Niemeyer, R. Szeliski, M. Pollefeys, and F. Tombari, "Unisdf: Unifying neural representations for high-fidelity 3d reconstruction of complex scenes with reflections," *arXiv preprint arXiv:2312.13285*, 2023.
- [75] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.
- [76] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.