GAINS: Gaussian-based Inverse Rendering from Sparse Multi-View Captures

Patrick Noras[1,2], Jun Myeong Choi[3], Didier Stricker[1,2], Pieter Peers[4], and Roni Sengupta[3]

[1] University of Kaiserslautern-Landau, Kaiserslautern, Germany
[2] German Research Center for Artificial Intelligence, Kaiserslautern, Germany
[3] University of North Carolina at Chapel Hill, NC, USA
[4] College of William & Mary, VA, USA

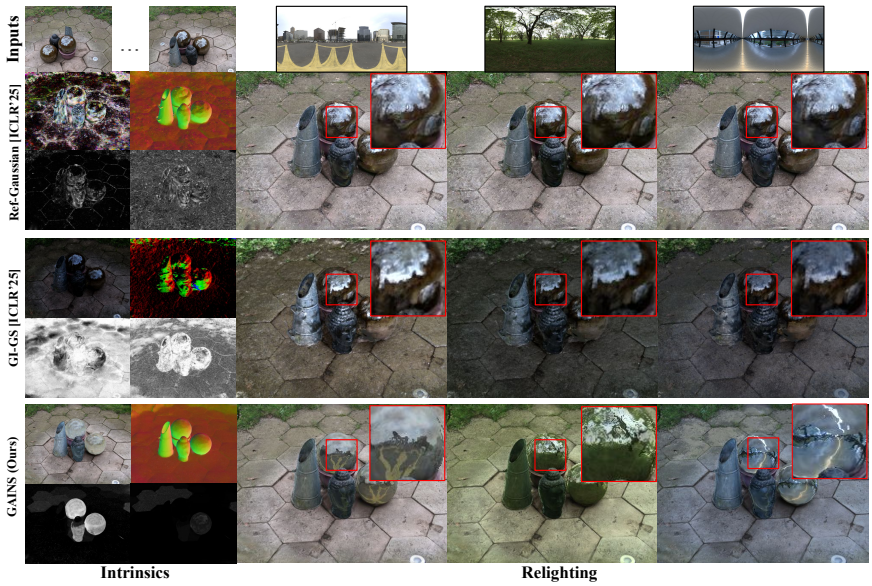

Fig. 1: We introduce **GAINS**, GAussian-based INverse rendering from Sparse multi-view captures, which synergizes foundation models for monocular depth/normal, segmentation, intrinsic image decomposition (IID), and diffusion as priors, to better disambiguate reflectance from lighting, leading to better intrinsics, novel view synthesis and relighting compared to existing state-of-the-art approaches such as Ref-Gaussian [39] and GI-GS [3]. Prior methods often overfit diffuse (*e.g.*, missing yellow reflection from the ground in the first relighting example for GI-GS) and/or specular reflections (*e.g.*, for both Ref-Gaussian and GI-GS the reflected details remain unchanged under different lighting conditions). In contrast, **GAINS** improves estimation of material properties, leading to better relighting in novel views.

Abstract. Recent advances in Gaussian Splatting-based inverse rendering extend Gaussian primitives with shading parameters and physically grounded light transport, enabling high-quality material recovery from dense multi-view captures. However, the accuracy of these methods degrades under sparse-view settings, where limited observations lead to

severe ambiguity between geometry, reflectance, and lighting. We introduce GAINS (Gaussian-based Inverse rendering from Sparse multi-view captures), a two-stage inverse rendering framework that leverages foundation models as priors to stabilize geometry and material estimation. The core technical contribution of this paper is an inverse rendering framework that unifies foundation model priors with physically-based representations in an optimization scheme. GAINS first refines geometry using monocular depth, normal, and diffusion priors, and then employs segmentation, intrinsic image decomposition (IID), and diffusion priors to regularize material recovery. Extensive experiments on synthetic and real-world datasets show that GAINS significantly improves material parameter accuracy, relighting quality, and novel-view synthesis compared to state-of-the-art Gaussian-based inverse rendering methods. While GAINS outperforms and remains competitive across a wide range of objects captured with 4 to 32 cameras, the improvement is particularly pronounced under sparse-view settings, where ambiguity is high and learning-based priors become especially beneficial. Project page: https://patrickbail.github.io/gains/

1 Introduction

Inverse rendering (IR) aims to recover the intrinsic 3D properties of a scene (*i.e.*, geometry, material properties, and lighting) from multi-view images. IR serves as a key step toward physically grounded 3D scene understanding with many downstream applications such as novel-view synthesis, relighting, material and shape editing, etc. Over the past decades, inverse rendering has seen remarkable progress, evolving from early methods [26, 27, 41] that use simple intrinsic representations such as surface normals, albedo, and spherical harmonics lighting, to modern approaches that employ rich and realistic intrinsic representations such as 3D Gaussian primitives, physically-based BRDFs, and realistic illumination [6, 7, 15, 17, 19, 30, 34, 44]. Simultaneously, the scope of inverse rendering has broadened from simple, controlled objects like faces to complex real-world scenes containing intricate geometry and diverse materials.

Despite these advances, scaling inverse rendering from simple objects to complex scenes with intricate details typically requires dense multi-view captures. Dense observations provide strong geometric and photometric constraints that help disambiguate material properties, lighting, and shape, all of which are entangled in the image formation process. However, when viewpoints are sparse, these constraints weaken, leading to overfitting and degraded performance. In such settings, existing methods often fail to recover accurate intrinsic properties, producing inconsistent material and lighting estimates. Moreover, traditional smoothness priors are inadequate for resolving the reflectance-lighting ambiguity, resulting in poor novel-view synthesis (NVS) and relighting.

To address these challenges, we propose GAINS (GAussian-based INverse rendering from Sparse multi-view captures), a novel inverse rendering framework designed to operate robustly under sparse-view settings. We focus on a

Gaussian splatting-based framework since state-of-the-art approaches [3, 28, 39] found Gaussian splatting to be more effective for inverse rendering both in terms of quality and efficiency. GAINS follows the standard two-stage inverse rendering pipeline, where we first estimate geometry (Stage I), followed by an estimation of material properties and lighting (Stage II). The key technical contribution of this work lies in developing an inverse rendering framework that integrates physically-based modeling with multiple foundation models serving as priors within an optimization pipeline.

Geometry estimation forms the foundation of physically-based rendering, and thus, high-quality geometry is crucial for accurate material and lighting recovery. In Stage I, we leverage learning-based priors from *monocular depth and normal prediction* networks as well as a *latent diffusion model*, drawing inspiration from recent progress in sparse-view Gaussian-based reconstruction methods [37,38], as detailed in Sec. 4. Each of these priors contributes towards improving geometric estimation accuracy, which further improves novel-view synthesis, BRDF estimation, and relighting. In Stage II, we introduce three complementary learning-based priors for improved material estimation: (1) A *segmentation guidance* that enforces multi-view consistency and reduces noise in material maps. However, segmentation guidance lacks intrinsic knowledge of surface reflectance, limiting its generalization to novel views and lighting conditions; (2) An *Intrinsic Image Decomposition (IID) prior*, implemented using a state-of-the-art intrinsic image decomposition network, that provides a strong initialization for albedo, at the cost of poor cross-view consistency and weak generalization; and (3) a *latent diffusion prior* that offers strong generalization capabilities. However, the latent diffusion prior struggles with material consistency and multi-view consistency. Only by combining all three complementary priors (Sec. 5), GAINS recovers robust material properties and lighting while achieving stable novel-view synthesis and relighting, even under sparse input conditions, as shown in Fig. 1.

We conduct extensive experiments on two synthetic benchmarks (Shiny Blender [32], extended with ground-truth albedo and relighting, and Synthetic4Relight [48], which additionally includes ground-truth roughness), as well as one real-world dataset [32]. GAINS consistently outperforms state-of-the-art Gaussian-based inverse rendering methods across geometry and material estimation tasks by a significant margin. For example, GAINS outperforms state-of-the-art Ref-Gaussian [39] in relighting on Synthetic4Relight [48] by 28.16 vs. 24.29 in PSNR and on Shiny Blender [32] by 22.29 vs. 17.26 in PSNR. While the improvements are especially pronounced under sparse-view conditions, our method also achieves competitive results with dense inputs. We further perform detailed ablation studies to analyze the contributions of each foundation model prior and demonstrate how their combination yields the most robust and physically consistent inverse rendering results.

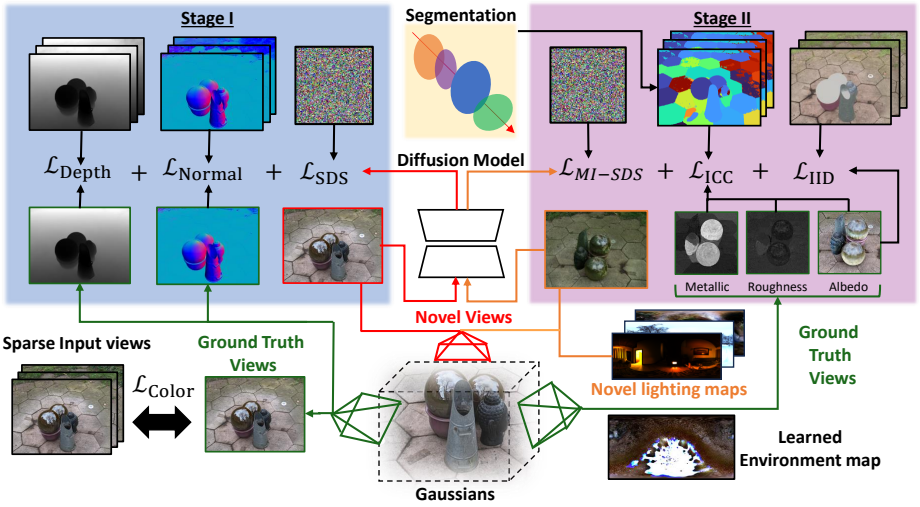

Fig. 2: GAINS follows a two-stage inverse rendering pipeline. In Stage I, we enhance geometry using learning-based priors from monocular depth, normal, and diffusion predictors. In Stage II, we introduce three complementary priors: segmentation, intrinsic image decomposition (IID), and diffusion, to improve material estimation, novel-view synthesis, and relighting. Each prior provides distinct benefits: segmentation boosts cross-view consistency of specular parameters but degrades albedo; IID improves albedo accuracy but remains view-inconsistent; diffusion enhances generalization to novel view and relighting but lacks material consistency. **GAINS** integrates these priors to achieve stable, high-quality material properties, leading to better relighting under novel views.

2 Related Work

Inverse Rendering (IR) decomposes images into geometry, material properties, and illumination. Classical methods estimate surface shape and spatially varying material properties from controlled captures using analytic BRDF models and global illumination optimization [4, 25, 36]. With the advent of learning-based scene representations such as NeRF [20], neural IR frameworks [11, 29, 35, 42, 47] integrate simplified BRDF and lighting models to recover intrinsic scene parameters from multi-view inputs. Notably, NeRFactor [47] estimates shape, materials, and lighting under unknown illumination using smoothness and BRDF priors, while GaNI [35] and related work jointly optimize geometry and material parameters with near-field lighting and neural radiance caching.

Inverse Rendering with Gaussian Splatting. Recently, 3D Gaussian Splatting (3DGS) [14] emerged as an efficient explicit radiance representation, which subsequently has been extended to IR [3, 5, 10, 16, 28, 39]. GaussianShader [10] augments Gaussian primitives with BRDF parameters, while Ref-Gaussian [39] and GI-GS [3] integrate physically-based rendering (PBR) with deferred shading and ray tracing. While Gaussian splatting based IR frameworks provide higher efficiency and fidelity than neural implicit representations, they assume dense multi-view captures and/or object-centric settings, making them less practical

for real-world sparse-view scenarios.

Inverse Rendering with Sparse Views. Earlier approaches for sparse-view or single-view IR rely on learned priors from synthetic data [15, 26, 27] without explicit geometric and light-transport modeling leading to poor generaliation to complex real world scenes. Recent works like RelitLRM [45] leverage Large Reconstruction Models (LRMs) and generative priors for object relighting, but they can not estimate intrinsic components (*e.g.*, material properties), nor do they generalize to complex real-world scenes.

Recent approaches for shape reconstruction from sparse views using NeRF- or 3DGS-based representations increasingly leverage learning-based priors, such as monocular depth, normal, or diffusion-based guidance [9,33,37,38,49], demonstrating that combining learned priors with explicit geometric modeling leads to significant performance gains. However, a sparse-input inverse rendering framework that jointly integrates such learning-based priors with explicit shape representations and physically based rendering remains largely unexplored.

3 Overview

Our goal is to reconstruct robust and accurate geometry and material properties from a sparse set of captured RGB images $\{I_i\}_{i=1}^{N}$ (with N as low as 4-16 views) under known camera intrinsics and extrinsics $C_i = \{K_i, R_i, t_i\}$. Our pipeline builds on 2D Gaussian Splatting (2DGS) [8], and adapts it for robust Inverse Rendering (IR) from sparse input views. We opt for 2DGS over 3D Gaussian Splatting [14] due to its more accurate geometry reconstruction which is essential for accurate material property estimation.

Scene representation. As in 2DGS, we parameterize the geometry as: $\mathcal{G} = \{\mu_j, S_j, R_j, \sigma_j\}_{j=1}^{M}$, where μ_j describes the 3D position of each primitive, along with opacity σ_j, scaling vector S_j and rotation matrix R_j; the surface normal n_j is also determined by R_j.

Material and lighting representation. Following previous Gaussian splatting-based IR methods [3,28,39], we use a physically-based rendering (PBR) deferred shading approach. We employ a principled BRDF model [1] and store for each Gaussian primitive the corresponding BRDF parameters: $\mathcal{M} = \{\boldsymbol{a_j}, r_j, m_j\}_{j=1}^{M}$, where $\boldsymbol{a_j} \in [0,1]^3$ is the albedo, $r_j \in [0,1]$ the roughness and $m_j \in [0,1]$ the metallicity. For efficiency, we employ the split-sum approximation [12] to model the diffuse and specular surface reflectance:

$$L_{\text{diffuse}} = (1-m)\cdot a \int_{\Omega} L(\omega_i)\frac{\omega_i \cdot n}{\pi} d\omega_i \tag{1}$$

$$L_{\text{specular}} \approx \underbrace{\int_{\Omega} \frac{DFG}{4(\omega_o \cdot n)} \omega_i d\omega_i}_{\text{precomp. BRDF LUT}} \cdot \underbrace{\int_{\Omega} D\ L(\omega_i) d\omega_i}_{\text{prefiltered Env. Map}}, \tag{2}$$

where $L(\omega_i)$ is the direct incident lighting stored as a 128×128 environment cubemap, n the surface normal. D the GGX microfacet distribution, F Fresnel

term, and G geometric shadowing term.

Optimization. The goal of IR is to optimize shape $\mathcal{G}$, material reflectance $\mathcal{M}$, and lighting L of the scene to minimize a re-rendering loss with respect to the input images. Similarly to prior work, we solve the optimization in two stages: (1) optimize for shape $\mathcal{G}$, followed by (2) a joint estimation of material parameters $\mathcal{M}$ and lighting L. To improve robustness of the two-stage pipeline under sparse input views, we augment each stage with appropriate additional loss terms. We improve geometry estimation (Sec. 4) by leveraging additional learning-based priors from a latent diffusion model and a monocular depth and normal estimation model following established best practices from recent sparse-view Gaussian-based geometry reconstruction methods [9, 37, 38]. We stabilize material and lighting estimation using three complementary learning-based priors: segmentation (Sec. 5.1), IID (Intrinsic Image Decomposition; Sec. 5.2), and diffusion priors (Sec. 5.3).

4 GAINS Stage I: Shape Estimation

We start with the standard 2DGS shape estimation loss, originally formulated for dense input views: $\mathcal{L} = \mathcal{L}_{\text{col}} + \lambda_{DD}\mathcal{L}_{DD} + \lambda_{NC}\mathcal{L}_{NC}$, where, $\mathcal{L}_{\text{col}} = (1 - \lambda)\mathcal{L}_1 + \lambda\mathcal{L}_{\text{D-SSIM}}$ is a re-rendering loss; $\mathcal{L}_{DD}$ is an optional depth distortion loss, reducing depth ambiguity by bringing intersected Gaussians closer along each ray, and $\mathcal{L}_{NC} = 1 - {N_i}^T\tilde{N}_i$ is the normal consistency loss with $\tilde{N}_i$ being surface normals obtained from depth.

When applied to sparse input views, the above loss tends to overfit to the captured viewpoints, resulting in significant novel viewpoint artifacts. To better guide the IR optimization towards a robust and generalizable solution, we introduce additional learning based priors: (1) depth and normal guidance, and (2) diffusion guidance.

Depth and Normal Guidance While learning-based depth priors have proven effective for geometry reconstruction [9, 37, 38, 49], their role in IR remains unclear, as hallucinations or systematic biases in monocular depth estimation may adversely affect material and lighting estimation. We include a depth loss by enforcing similarity to a monocularly estimated depth $\hat{D}_i$ [13]: $\mathcal{L}_{DC} = 1 - PCC(D_i, \hat{D}_i)$, where $PCC(\cdot)$ is the Pearson Correlation Coefficient. Additionally, we also add a local depth ranking loss $\mathcal{L}_{DR}$ [9] to prevent geometric collapse from hard constraints and alleviate long-range ambiguity. Moreover, given the importance of accurate surface normals in material estimation, we impose normal smoothness via a total variation loss and an L1 loss on rendered normals N_i and surface reconstructed normals $\hat{N}_i$. The reference normals $\tilde{N}_i$ are obtained from a monocular normal estimator [13]. Together we obtain the normal guided loss $\mathcal{L}_N = \mathcal{L}_1(\tilde{N}_i, \hat{N}_i) + 0.1 \cdot \mathcal{L}_1(\tilde{N}_i, N_i) + TV(\tilde{N}_i, \hat{N}_i)$. As shown in the ablation study (Tab. 3), incorporating normal guidance improves high-frequency geometric fidelity and leads to more stable material estimation and rendering quality.

Diffusion Guidance Inspired by recent successes in leveraging diffusion models for zero-shot 3D reconstruction [18, 40], we guide the shape reconstruction

through Score Distillation Sampling (SDS). Unlike depth and normal guidance, which provide pixel-aligned geometric constraints, the diffusion prior regularizes the reconstruction in unseen viewpoints C_j by encouraging rendered images to lie on the manifold of the objects. Concretely, We sample 100 novel viewpoints for which we render and perform a forward diffusion, yielding noisy latents: $Z_j = \alpha_t \mathcal{R}_{geo}(C_j, \mathcal{G}) + \sigma_t \epsilon$, where timestep $t \sim \mathcal{U}(0.02, 0.98)$ and noise $\epsilon \sim \mathcal{N}(\mathbf{0}, I_0)$. The SDS loss is then calculated as: $\mathcal{L}_{SDS} = \mathbb{E}_{t,\epsilon}\Big[w(t)||(\epsilon_\phi(Z_j; t) - \epsilon)||_2^2\Big]$. However, the SDS loss is not effective during early stages of the optimization when the shape reconstruction is far from the target shape. We therefore only include the SDS loss after iteration 10000 (out of 16000). With a guidance scale of 100, the diffusion prior primarily suppresses high-frequency artifacts; because the geometry is mostly converged, hallucinations are limited.

Outlier Removal If ground-truth alpha masks are provided, we additionally employ an Outlier Removal (OR) strategy. Here, a Binary-Cross Entropy (BCE) loss on reconstructed and ground truth alpha mask is employed, along with a KNN-based floater removal mechanism at every 1000 iterations. If no alpha masks are provided, we assume the scene to reconstruct the background as well.

Summary The total loss function for shape estimation in Stage I is:

$$\mathcal{L}_{col} + \lambda_{DC}\mathcal{L}_{DC} + \lambda_{DR}\mathcal{L}_{DR} + \lambda_{NC}\mathcal{L}_{NC} + \lambda_N\mathcal{L}_N + \lambda_{SDS}\mathcal{L}_{SDS} + \lambda_{BCE}\mathcal{L}_{BCE},$$

with $\lambda_{DC} = 0.005$, $\lambda_{DR} = 10$, $\lambda_{NC} = 1$, $\lambda_N = 0.25$, $\lambda_{SDS} = 0.0001$, $\lambda_{BCE} = 0.75$. The complete Stage I pipeline is summarized in Fig. 2 (left). All hyperparameters are fixed without any scene-specific tuning or prior-fine-tuning.

5 GAINS Stage II: Material Properties

In the second stage, we jointly optimize material properties $\mathcal{M} = \{(a_j, r_j, m_j)\}$ and environment lighting L using the re-rendering loss $\mathcal{L}_{\text{color}}$ over 7000 additional iterations. However, multiple combinations of different $\mathcal{M}$ and L produce similar appearances, leading to an ill-posed and ambiguous material estimation. This problem is further exacerbated when only a sparse set of input views is available. Although rendering under the recovered lighting and material properties appears visually plausible, the material properties are overfitted, noisy, inconsistent, and not suitable for relighting (Fig. 1). To alleviate the impact of overfitting under sparse inputs, we complement the loss with additional learning based priors to better guide the optimization towards more plausible material properties, and thus resulting in more robust relighting. We synergize three complementary priors: segmentation (Sec. 5.1), Intrinsic Image Decomposition (IID) (Sec. 5.2), and diffusion guidance (Sec. 5.3).

5.1 Segmentation Guidance

While diffuse albedo varies significantly due to texture variations, we observe that specular material properties are typically consistent within semantically similar regions of an object or scene. This suggests that segmentation cues can help guide the estimation of specular material properties. As a single semantic

object can contain multiple subparts with distinct materials, we employ subpart segmentation to define a consistency loss. Instead of constraining the material parameters to be identical for each subpart, we minimize their variance to support mixed materials and to account for imperfect segmentation boundaries.

To include segmentation guidance, we extend the Gaussian primitives with a one-hot vector $E_j = \{0,1\}^K$ that encodes each Gaussian's class membership. We employ training-free segmentation mask lifting [2] by rendering SH colored images from $V = 100$ sampled novel view points orbiting the scenery. For each image $\{\hat{I}_v\}_{v=1}^V$ we generate (potentially view-inconsistent) segmentation masks with SAM [23] along with mask-related features $\{f_{v,s} = [g_{v,s}, h_{v,s}]\}_{v=1}^V$ where $f_{v,s}$ is the feature vector for a mask region s of view v and $g_{v,s}, h_{v,s}$ are CLIP [22] and DINOv2 [21] features respectively. For each view, we lift the associated mask and features to the Gaussian that contributes most (α-blending wise) and merge already collected Gaussian objects from previous lifting operations with geometric and feature similarity scores.

Next, we leverage the possibly imprecise segmentation masks to define an intra-class consistency (ICC) on the roughness and metallicity parameters:

$$\mathcal{L}_{ICC} = \frac{1}{|S_i|} \sum_{s_i \in S_i} \gamma(|s_i|) \cdot \Big(\mathcal{L}_{r,m} + \mathcal{L}_a + \mathcal{L}_e \Big) \tag{3}$$

$$\mathcal{L}_{r,m} = Var(R_{s_i}) + Var(M_{s_i}) \tag{4}$$

$$\mathcal{L}_a = \Big((1 - R_{s_i}) + M_{s_i} \Big) \cdot \frac{1}{3} \sum_{c=1}^{3} Var(A_{s_i}^{(c)}) \tag{5}$$

$$\mathcal{L}_e = \Big((1 - M_{s_i}) + R_{s_i} \Big) \cdot E_{s_i} \tag{6}$$

where S_i is the set of all masks rendered from a captured viewpoint i, A_{s_i}, R_{s_i}, and M_{s_i} are the rendered masked albedo, roughness, and metallicity, along with E_{s_i} representing the masked re-render loss. $\gamma(|s_i|) = \exp(25 \cdot \frac{|I_i||s_i|}{|S_i|^2})$ is a scaling function depending on the size of the mask s_i. The first variance reduction term $\mathcal{L}_{r,m}$ reduces irregularities for specular roughness and metallicity across all masks. While in general we attribute texture variations to the diffuse texture, an ambiguity exists in the case of mirror-like materials that reflects texture from the environment lighting. To address this ambiguity, we bias the optimization to attribute texture details to specular reflections in the case of mirror-like materials by adding a loss term $\mathcal{L}_a$ that encourages a variance reduction in diffuse albedo in regions where we observe low roughness and high metallicity. Lastly, regions with high specularity typically suffer from larger geometry errors, and thus higher material estimation errors. We therefore add a loss $\mathcal{L}_e$ to encourage more specular in such areas, albeit with a low weight to avoid attributing all errors to specular.

5.2 Intrinsic Image Decomposition Prior

While segmentation improves consistency in roughness and metallicity over multiple viewpoints, it is not suited for diffuse albedo which often contains high-frequency texture variations that are difficult to estimate from sparse viewpoints.

In the absence of appropriate regularization, the model will overfit diffuse albedo and fail to disentangle lighting effects from intrinsic albedo, which ultimately leads to a subpar relighting performance.

Recent advances in monocular Intrinsic Image Decomposition (IID) [13, 24, 43] enable high quality diffuse albedo decompositions, albeit potentially view-inconsistent. We therefore propose to leverage the IID to regularize the diffuse albedo estimation while accounting for potential view-inconsistencies: $\mathcal{L}_{IID} = \beta(\tau) \cdot \mathcal{L}_2(A_i, \hat{A}_i)$, where $\hat{A}_i$ is the IID diffuse albedo for the captured image I_i obtained using Teamwork [24]. For outdoor data, we opt for RGB2X [43], as we found it to produce more reliable decompositions on outdoor scenes used in our experiments. To mitigate the influence of view-inconsistency in the IID, we weight the loss by $\beta(\tau)$ (a linear decrease) which depends on the current iteration step τ, thereby relaxing the influence of the IID diffuse albedo as the model converges. We opt against using additional IID supervision for specular roughness and metallicity as we found these parameters to be less robust and exhibit more inconsistencies across views.

5.3 Multi-illuminated Score Distillation Sampling

While both segmentation guidance and IID guidance improve the accuracy of the specular and diffuse material properties respectively, they do not necessarily provide good generalization to novel viewpoints or lighting. To address these issues, we take inspiration from our usage of SDS in the first stage to aid in reducing unrealistic artifacts on novel viewpoints, and introduce a modified SDS loss to improve view and lighting-consistency for material properties. To avoid overfitting to the learned environment map L, we render relit images $\mathcal{R}_{mat}(C_j, \mathcal{G}, \mathcal{M}, \mathcal{E}_l)$ from a novel view C_j under randomly selected lighting from a predetermined set $\{\mathcal{E}_l\}_{l=1}^{|\mathcal{E}|}$, and define the SDS loss as: $\mathcal{L}_{MI-SDS} = \mathbb{E}_{t,\epsilon}\Big[w(t)||(\epsilon_\phi(Z_j^l; t) - \epsilon)||_2^2\Big]$, where $Z_j^l = \alpha_t \mathcal{R}_{mat}(C_j, \mathcal{G}, \mathcal{M}, \mathcal{E}_l) + \sigma_t \epsilon$ is the noisy diffusion latent of the rendered scene under novel view and lighting. Similarly as in the geometry reconstruction stage, we start the diffusion guidance once the optimization solution has stabilized (*i.e.*, after 3000 steps). As in Stage I, with a guidance scale of 100, the diffusion prior mainly suppresses high-frequency artifacts , introducing only limited hallucinated details.

5.4 Final Loss

Each loss contributes to a specific enhancement of the material estimation process: parameter accuracy, multi-view consistency, or novel views and lighting generalization. We combine all losses as:

$$\mathcal{L}_{mat} = \mathcal{L}_{color} + \lambda_{ICC}\mathcal{L}_{ICC} + \lambda_{IID}\mathcal{L}_{IID} + \lambda_{MI-SDS}\mathcal{L}_{MI-SDS} + \lambda_{TV}\mathcal{L}_{TV},$$

where $\lambda_{ICC} = 0.1$, $\lambda_{IID} = 2$, $\lambda_{MI-SDS} = 0.0001$ and $\lambda_{TV} = 0.1$. $\mathcal{L}_{TV}$ represents a total variation loss on our reconstructed lighting, acting as a smoothing term. Fig. 2 (right) summarizes the full Stage II pipeline. Reported hyperparameters are again fixed without any scene-specific tuning or prior-fine-tuning.

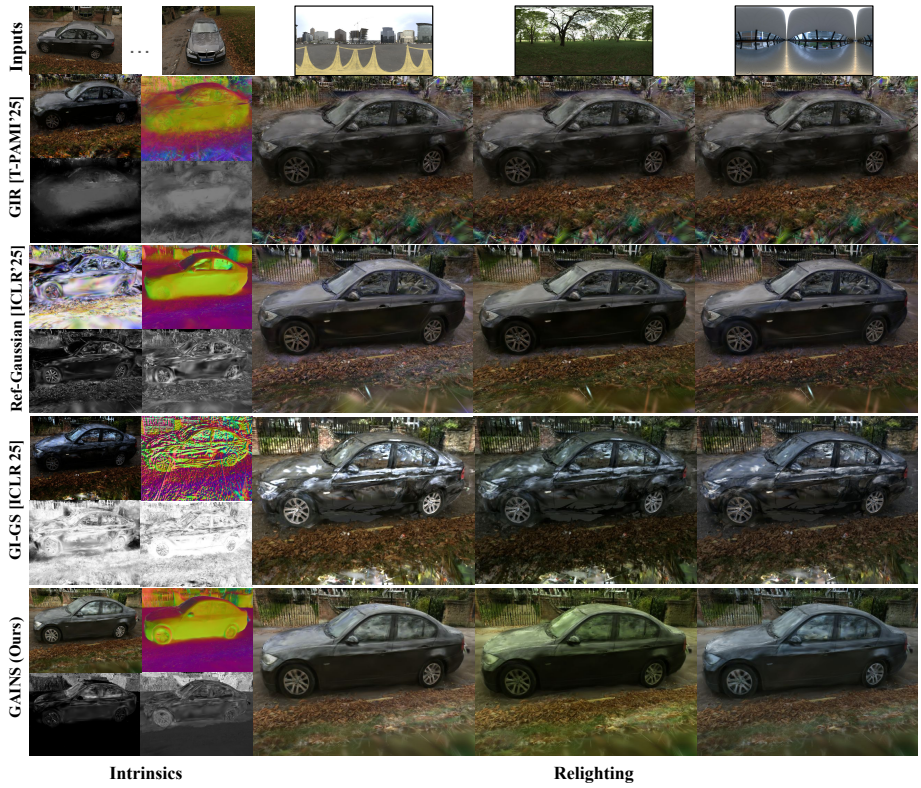

Fig. 3: Qualitative comparison of intrinsic estimation and relighting on the *sedan* scene from the Ref-Real dataset [32] reconstructed from 8 views. Column 1 shows novel-view intrinsic renderings in a 2×2 layout: (top) rendered albedo and surface normals, (bottom) specular roughness and metallicity. Columns 2–4 show relighting results under three different environment maps from novel viewpoints. GAINS recovers significantly more accurate intrinsics, enabling more realistic relighting.

6 Evaluation

Evaluation Framework. We compare GAINS against state-of-the-art IR frameworks that use Gaussian Splatting for both objects and scenes: Ref-Gaussian [39], GI-GS [3], GShader [10], R3DG [5], MaterialRefGS [46] and SVG-IR [31]. We evaluate on two synthetic datasets (*i.e.*, the Shiny Blender dataset [32] modified to include albedo and relighting evaluation and without 'ball' object; and Synthetic4Relight [48] for NVS, albedo, relighting and roughness), and a real-world dataset (Ref-Real [32]), on which we quantitatively evaluate NVS performance only. We also evaluate GIR [28] on real scenes from Ref-Real, but we found that it fails to reconstruct meaningful geometry and reflectance (Fig. 3) while requiring long per-scene training time (over 7 hours). Therefore, we opted not to include GIR in the evaluation over the synthetic datasets. We evaluate NVS,

Table 1: Quantitative evaluation on the synthetic Synthetic4Relight [48] dataset. GAINS estimates better albedo and roughness then baselines leading to better relighting performance. Additionally, GAINS outperforms existing methods in novel-view synthesis. (Red = best, Orange = 2nd best, and Yellow = 3rd best)

Methods	Synthetic4Relight [48] dataset (8 views)										
	NVS			Albedo			Relight			Roughness	Runtime ↓
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	MSE ↓	
GShader [10]	15.11	0.71	0.24	-	-	-	12.58	0.11	0.39	0.09	≈5h
MaterialRefGS [46]	19.10	0.86	0.16	14.56	0.75	0.24	17.95	0.84	0.18	0.76	≈2h30m
SVG-IR [31]	24.21	0.87	0.11	15.89	0.87	0.12	25.22	0.90	0.10	0.03	≈2h
R3DG [5]	26.92	0.90	0.11	22.15	0.88	0.12	27.42	0.90	0.10	0.04	≈1h30m
Ref-Gaussian [39]	24.05	0.83	0.11	17.37	0.83	0.17	24.29	0.85	0.13	0.04	≈45m
GI-GS [3]	27.04	0.91	0.12	20.47	0.88	0.12	23.92	0.82	0.11	0.05	≈1h30m
Ours	29.32	0.94	0.08	22.34	0.91	0.10	28.16	0.93	0.09	0.04	≈1h20m

Table 2: Quantitative evaluation on the Shiny Blender [32] dataset and real Ref-Real [32] dataset. On Shiny Blender, GAINS estimates better albedo and better relighting results compared to existing methods. On the Ref-real and Shiny Blender dataset, GAINS also performs better novel-view synthesis than existing approaches.

Methods	Shiny Blender [32] dataset (8 views)										Ref-Real [32] dataset (8 views)		
	NVS			Albedo			Relight			Normal	NVS		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	MAE ↓	PSNR ↑	SSIM ↑	LPIPS ↓
GShader [10]	10.73	0.52	0.35	-	-	-	6.22	0.13	0.53	55.14	15.26	0.38	0.46
R3DG [5]	20.97	0.83	0.17	16.62	0.79	0.19	12.43	0.69	0.22	24.43	18.67	0.42	0.46
Ref-Gaussian [39]	18.60	0.80	0.18	15.82	0.77	0.22	17.26	0.66	0.20	25.54	19.39	0.45	0.36
GI-GS [3]	20.15	0.79	0.21	21.37	0.87	0.14	15.44	0.57	0.27	63.55	19.64	0.44	0.37
Ours	23.90	0.89	0.12	21.73	0.89	0.12	22.29	0.89	0.12	11.64	20.28	0.56	0.34

albedo, and relighting using three metrics: PSNR, SSIM, and LPIPS, while for roughness we employ MSE. To account for the albedo-lighting ambiguity, we use scale invariant losses to evaluate the albedo for each approach by uniformly scaling each albedo map to minimize MSE w.r.t. ground-truth before applying an error metric. All evaluations are performed on 8 sparse views uniformly sampled from the dense views.

Performance Evaluation. Tab. 1, and 2 quantitatively summarize evaluation on the three datasets. In general, GAINS yields more accurate albedo maps and relighting across both Shiny Blender and Synthetic4Relight compared to all prior methods. Furthermore, GAINS achieves overall better results for NVS than all competing methods. However, error metrics do not always capture important visual differences. Therefore, we also provide qualitative comparisons in Fig. 1, 3, and 4. From the qualitative comparison, especially on the real-world Ref-Real dataset (Fig. 1 and 3), we note less baking of specular reflections, and generally more accurate relighting results, *e.g.*, the reflections of the ground (missing for Ref-Gaussian) and the sky (baking of captured reflection for both Ref-Gaussian and GI-GS) on the ball in Fig. 1. While GI-GS is the 2nd best method in NVS, we observe that it produces less accurate shape and normals (*e.g.*, the noisy normals on the car in the sedan scene in Fig. 3), indicating that the GI-GS NVS performance is mainly due to overfitting. As shown in Fig. 4, our method demonstrates a clear advantage in material estimation under sparse-view settings. While Ref-Gaussian suffers from floaters and GI-GS typically recovers reasonable geometry but fails to produce accurate or stable albedo and roughness maps, our approach, supported by learning-based priors, delivers consistently superior NVS quality and more reliable albedo, roughness, and relighting estimations across all scenes.

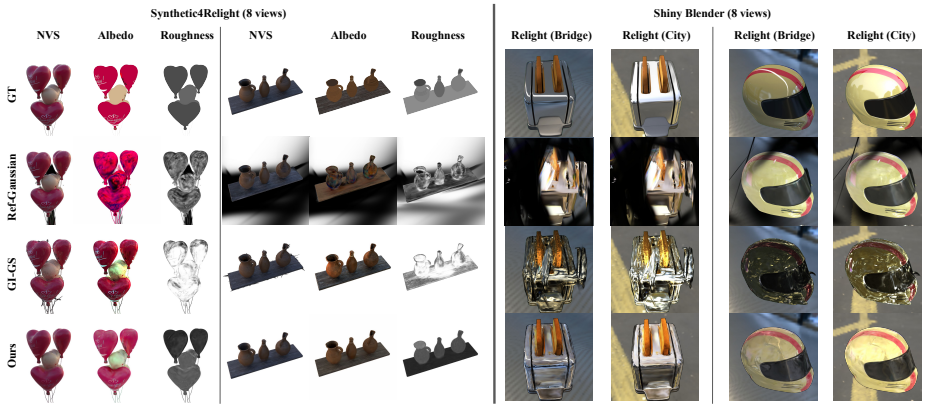

Fig. 4: Qualitative comparison on Synthetic4Relight [48] and Shiny Blender [32] trained with 8 views. While all methods produce reasonable NVS, our method's estimates significantly better albedo, roughness and relighting than GI-GS and Ref-Gaussian that overfit to limited training views and fail to disentangle reflectance from lighting.

Evidence of Directional Relighting In Fig. 6, we provide qualitative results under a novel environment map. Specifically, (a) compares our relit rendering against the ground truth along with the corresponding error map, while (b) visualizes diffuse directional shading evidence by rotating the environment map by 90°. These results highlight our method's ability to accurately reproduce both specular highlights and diffuse directional shading under novel illumination.

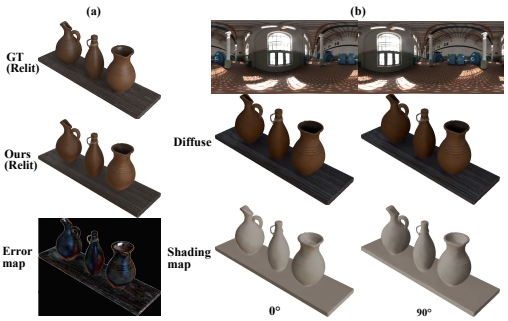

Fig. 6: (a) Relighting comparison against GT with error map (×5). (b) Diffuse directional shading evidence.

Ablation: Number of input views. In addition to our comprehensive quantitative and qualitative evaluation on both synthetic and real datasets, we further analyze the behavior of our method as the number of available input views increases. Fig. 5 summarizes the quantitative comparison for PSNR on NVS, albedo and relighting on Shiny Blender [32]. We observe that in sparse-view scenarios (4-8 views), our method consistently outperforms all baselines across all evaluation scenarios by a significant margin. As the number of views increases (16-32), GI-GS competes in NVS; however, our method continues to produce competitive results within a small margin and continues to surpass in relighting. On denser view counts (64-100), Ref-Gaussian surpasses GAINS in NVS and be-

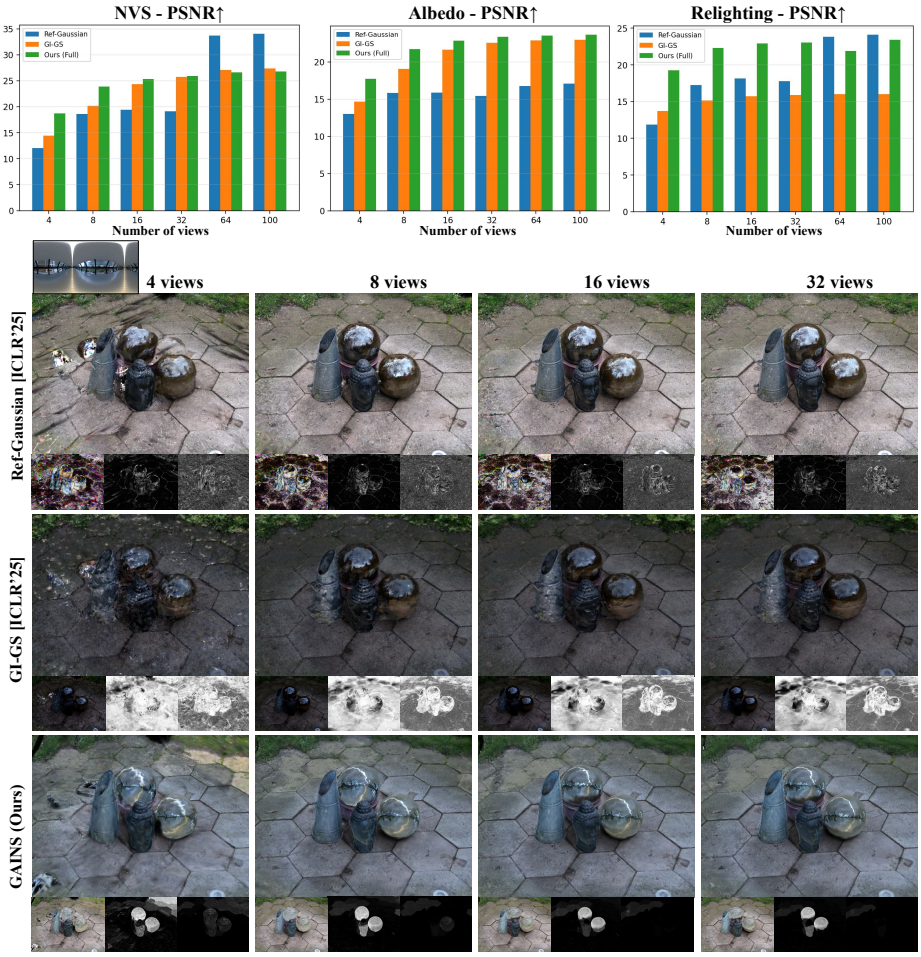

Fig. 5: Comparison of our method with Ref-Gaussian [39] and GI-GS [3] for varying number of input views. Top: Quantitative Evaluation on Shiny Blender [32]. Bottom: Qualitative comparison on the *gardenspheres* scene from Ref-Real [32] – relit images on top and material maps at bottom. GAINS consistently produces robust relighting and material estimation from 4-32 views and remains competitive thereafter.

gins to match our relighting performance. This improvement is largely due to its relaxed optimization objective: instead of explicitly optimizing albedo, it relies on the Stage I SH representation and focuses on appearance fitting. Additionally, it continues scene densification during Stage II, which further improves view synthesis quality as more observations become available. While all methods benefit from additional views, GAINS starts from a more stable and reliable baseline in the extremely sparse-view regime, demonstrating stronger robustness under challenging conditions. Complementing our quantitative experiments, we also

Table 3: Stage I Ablation with and without depth, normal, diffusion (SDS), and outlier removal (OR) on the Shiny Blender dataset [32] without Stage II components.

Shiny Blender [32] Stage I Ablation (8 views)

Method Variant	NVS			Albedo			Relight			Normal
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	MAE ↓
Ours full	23.67	0.88	0.12	21.18	0.88	0.14	21.96	0.86	0.15	11.64
Ours w/o SDS	23.57	0.88	0.13	21.09	0.87	0.14	21.91	0.86	0.15	11.70
Ours w/o SDS, OR	21.75	0.80	0.15	19.98	0.86	0.14	21.26	0.77	0.16	13.79
Ours w/o Normal, SDS, OR	19.42	0.79	0.16	18.79	0.82	0.14	19.41	0.74	0.17	19.38
Ours w/o Depth, Normal, SDS, OR	16.73	0.71	0.21	16.41	0.81	0.17	16.71	0.67	0.21	32.23

provide qualitative relighting and material estimation results (albedo, metallicity, and roughness) in Fig. 5 on the *gardenspheres* scene from the Ref-Real [32] dataset. Across all view counts, GAINS demonstrates strong and stable material reconstruction, producing accurate specular reflections and consequently superior relighting performance. Moreover, our roughness and metallic estimates remain stable and robust. This further demonstrates GAINS' advantage in low-view settings, where competing methods often suffer from noisy, irregular, or unstable material reconstructions. Improvements observed under dense inputs mainly stem from enhanced geometry estimation and improved lighting recovery. In contrast competing methods exhibit pronounced shape degradation in low-view settings (4-8 views).

Ablation: Learning-based priors in Stage I and Stage II. We conduct an ablation study on the Shiny Blender [32] dataset using 8 input views for both Stage I (Tab. 3) and Stage II (Fig. 7). For Stage I, by distilling depth and normal priors into our shape learning stage, we are able to achieve robust normal and geometry reconstruction. Additionally, applying our outlier removal strategy for synthetic data, we effectively remove floaters outside the input view-space. All components enable consistent improvement to achieve the best quality. For Stage II (Fig. 7), we observe that removing the IID prior reduces albedo reconstruction fidelity and leads to inconsistent material separation, resulting in degraded relighting performance. Excluding the segmentation prior produces incorrect object and material boundaries, yielding noisy or unstable specular components and harming novel-view synthesis. Finally, removing the diffusion prior leads to an overall drop in performance across all tasks, demonstrating its critical role in stabilizing optimization and improving generalization under sparse-view settings. More specifically, we notice a drop in reconstruction quality from unobserved views. Overall, the full model consistently achieves the strongest results across NVS, albedo, and relighting.

Limitations. First, GAINS ignores global illumination, which is more challenging for complete scenes than for isolated objects. Second, while we enforce smoothness, the geometric normals can exhibit a slight waviness, which is especially noticeable on very specular objects. As a consequence of wavy normals, some specularity can get baked into the diffuse albedo for very specular materials. Separating geometric and shading normals could potentially alleviate this.

Shiny Blender [32] Stage II Ablation (8 views)

Method Variant	NVS	Albedo	Relight
	PSNR ↑	PSNR ↑	PSNR ↑
Ours full	23.90	21.73	22.29
Ours w/o MI-SDS	23.88	21.66	22.26
Ours w/o Seg	23.75	21.92	22.23
Ours w/o IID	23.84	21.38	22.07
Ours w/o Seg, IID, MI-SDS	23.67	21.18	21.96

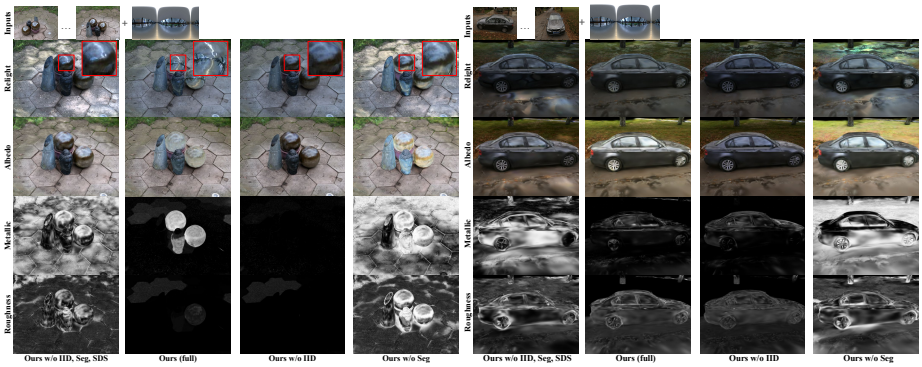

Fig. 7: Stage II Ablation. Top left: Quantitative Evaluation on Shiny Blender [32]. Top right: Qualitative comparison of resulting albedo w/ and w/o MI-SDS with error map (×5). Bottom: Qualitative comparison on the *gardenspheres* and *sedan* scenes from Ref-Real [32]. We analyze three priors: intrinsic components (IID), segmentation (Seg), and diffusion (MI-SDS). Visual results on real data reveal their necessity - without IID material maps degrade and without segmentation multi-view consistency breaks.

7 Conclusion

We introduced GAINS, a Gaussian-based inverse rendering framework tailored for sparse-view settings. By combining segmentation, IID, and diffusion priors within a two-stage optimization pipeline, GAINS effectively stabilizes geometry and material estimation. Extensive experiments on synthetic and real-world datasets show that GAINS achieves improved relighting accuracy and novel-view synthesis compared to prior Gaussian-based IR methods. Qualitative results further demonstrate improved material–lighting separation and reduced reflection baking. GAINS highlights the power of synergizing complementary learning priors for physically consistent inverse rendering under sparse observations.

Acknowledgements

Patrick Noras is supported by the DAAD-PROMOS scholarship during their research stay as a visiting scholar at the University of North Carolina at Chapel Hill. This work is partially supported by a National Institute of Health (NIH) NIBIB project #R21EB035832 and #R21EB037440 and National Science Foundation (NSF) CAREER Award #2543161.

References

1. Burley, B., Studios, W.D.A.: Physically-based shading at disney. In: Acm siggraph. vol. 2012, pp. 1–7. vol. 2012 (2012)
2. Chacko, R., Häni, N., Khaliullin, E., Sun, L., Lee, D.: Lifting by gaussians: A simple, fast and flexible method for 3d instance segmentation. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 3497–3507. IEEE (2025)
3. Chen, H., Lin, Z., Zhang, J.: Gi-gs: Global illumination decomposition on gaussian splatting for inverse rendering. In: ICLR (2025)
4. Debevec, P., Tchou, C., Gardner, A., Hawkins, T., Poullis, C., Stumpfel, J., Jones, A., Yun, N., Einarsson, P., Lundgren, T., et al.: Estimating surface reflectance properties of a complex scene under captured natural illumination (2004)
5. Gao, J., Gu, C., Lin, Y., Zhu, H., Cao, X., Zhang, L., Yao, Y.: Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing. arXiv:2311.16043 (2023)
6. Gu, C., Wei, X., Zeng, Z., Yao, Y., Zhang, L.: Irgs: Inter-reflective gaussian splatting with 2d gaussian ray tracing. In: CVPR (2025)
7. Hasselgren, J., Hofmann, N., Munkberg, J.: Shape, Light, and Material Decomposition from Images using Monte Carlo Rendering and Denoising. arXiv:2206.03380 (2022)
8. Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2d gaussian splatting for geometrically accurate radiance fields. In: SIGGRAPH 2024 Conference Papers. Association for Computing Machinery (2024). https://doi.org/10.1145/3641519.3657428
9. Huang, H., Wu, Y., Deng, C., Gao, G., Gu, M., Liu, Y.S.: Fatesgs: Fast and accurate sparse-view surface reconstruction using gaussian splatting with depth-feature consistency. In: Proceedings of the AAAI Conference on Artificial Intelligence (2025)
10. Jiang, Y., Tu, J., Liu, Y., Gao, X., Long, X., Wang, W., Ma, Y.: Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. arXiv preprint arXiv:2311.17977 (2023)
11. Jin, H., Liu, I., Xu, P., Zhang, X., Han, S., Bi, S., Zhou, X., Xu, Z., Su, H.: Tensoir: Tensorial inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
12. Karis, B., Games, E.: Real shading in unreal engine 4. Proc. Physically Based Shading Theory Practice **4**(3), 1 (2013)
13. Ke, B., Obukhov, A., Huang, S., Metzger, N., Daudt, R.C., Schindler, K.: Repurposing diffusion-based image generators for monocular depth estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
14. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (July 2023), https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/
15. Li, Z., Shafiei, M., Ramamoorthi, R., Sunkavalli, K., Chandraker, M.: Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2475–2484 (2020)
16. Liang, Z., Zhang, Q., Feng, Y., Shan, Y., Jia, K.: Gs-ir: 3d gaussian splatting for inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21644–21653 (2024)

17. Lichy, D., Wu, J., Sengupta, S., Jacobs, D.W.: Shape and material capture at home. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6123–6133 (2021)
18. Liu, R., Wu, R., Hoorick, B.V., Tokmakov, P., Zakharov, S., Vondrick, C.: Zero-1-to-3: Zero-shot one image to 3d object (2023)
19. Liu, Y., Wang, P., Lin, C., Long, X., Wang, J., Liu, L., Komura, T., Wang, W.: NeRO: Neural Geometry and BRDF Reconstruction of Reflective Objects from Multiview Images. ACM Trans. Graph. **42**(4), 114:1–114:22 (Jul 2023). https://doi.org/10.1145/3592134
20. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020)
21. Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Howes, R., Huang, P.Y., Xu, H., Sharma, V., Li, S.W., Galuba, W., Rabbat, M., Assran, M., Ballas, N., Synnaeve, G., Misra, I., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: Dinov2: Learning robust visual features without supervision (2023)
22. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
23. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.Y., Girshick, R., Dollár, P., Feichtenhofer, C.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024), https://arxiv.org/abs/2408.00714
24. Sartor, S., Peers, P.: Teamwork: Collaborative diffusion with low-rank coordination and adaptation. In: Proceedings of the SIGGRAPH Asia 2025 Conference Papers. SA Conference Papers '25 (2025). https://doi.org/10.1145/3757377.3763870, https://doi.org/10.1145/3757377.3763870
25. Sato, Y., Wheeler, M.D., Ikeuchi, K.: Object shape and reflectance modeling from observation. In: Proceedings of the 24th annual conference on Computer graphics and interactive techniques. pp. 379–387 (1997)
26. Sengupta, S., Gu, J., Kim, K., Liu, G., Jacobs, D.W., Kautz, J.: Neural inverse rendering of an indoor scene from a single image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8598–8607 (2019)
27. Sengupta, S., Kanazawa, A., Castillo, C.D., Jacobs, D.W.: SfSNet: Learning Shape, Reflectance and Illuminance of Faces 'in the Wild'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6296–6305 (2018)
28. Shi, Y., Wu, Y., Wu, C., Liu, X., Zhao, C., Feng, H., Zhang, J., Zhou, B., Ding, E., Wang, J.: Gir: 3d gaussian inverse rendering for relightable scene factorization. IEEE Transactions on Transactions on Pattern Analysis and Machine Intelligence (2025)
29. Srinivasan, P.P., Deng, B., Zhang, X., Tancik, M., Mildenhall, B., Barron, J.T.: Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7495–7504 (2021)
30. Sun, C., Cai, G., Li, Z., Yan, K., Zhang, C., Marshall, C., Huang, J.B., Zhao, S., Dong, Z.: Neural-PBIR Reconstruction of Shape, Material, and Illumination. In:

2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 18000–18010. IEEE, Paris, France (Oct 2023). https://doi.org/10.1109/ICCV51070.2023.01654
31. Sun, H., Gao, Y., Xie, J., Yang, J., Wang, B.: Svg-ir: Spatially-varying gaussian splatting for inverse rendering. Proceedings of CVPR 2025 (2025)
32. Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T., Srinivasan, P.P.: Ref-NeRF: Structured view-dependent appearance for neural radiance fields. CVPR (2022)
33. Wang, G., Chen, Z., Loy, C.C., Liu, Z.: Sparsenerf: Distilling depth ranking for few-shot novel view synthesis. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9065–9076 (2023)
34. Wang, H., Hu, W., Zhu, L., Lau, R.W.: Inverse Rendering of Glossy Objects via the Neural Plenoptic Function and Radiance Fields. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19999–20008. IEEE, Seattle, WA, USA (Jun 2024). https://doi.org/10.1109/CVPR52733.2024.01890
35. Wu, J., Hadadan, S., Lin, G., Zwicker, M., Jacobs, D., Sengupta, R.: Gani: Global and near field illumination aware neural inverse rendering. arXiv preprint arXiv:2403.15651 (2024)
36. Xia, R., Dong, Y., Peers, P., Tong, X.: Recovering shape and spatially-varying surface reflectance under unknown illumination. ACM Transactions on Graphics (ToG) **35**(6), 1–12 (2016)
37. Xiong, H., Muttukuru, S., Upadhyay, R., Chari, P., Kadambi, A.: Sparsegs: Real-time 360° sparse view synthesis using gaussian splatting. Arxiv (2023)
38. Yang, C., Li, S., Fang, J., Liang, R., Xie, L., Zhang, X., Shen, W., Tian, Q.: Gaussianobject: High-quality 3d object reconstruction from four views with gaussian splatting. ACM Transactions on Graphics **43**(6) (December 2024)
39. Yao, Y., Zeng, Z., Gu, C., Zhu, X., Zhang, L.: Reflective gaussian splatting. In: ICLR (2025)
40. Yi, T., Fang, J., Wang, J., Wu, G., Xie, L., Zhang, X., Liu, W., Tian, Q., Wang, X.: Gaussiandreamer: Fast generation from text to 3d gaussians by bridging 2d and 3d diffusion models. In: CVPR (2024)
41. Yu, Y., Smith, W.A.P.: Inverserendernet: Learning single image inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
42. Zeng, C., Chen, G., Dong, Y., Peers, P., Wu, H., Tong, X.: Relighting neural radiance fields with shadow and highlight hints. In: ACM SIGGRAPH 2023 Conference Proceedings. pp. 1–11 (2023)
43. Zeng, Z., Deschaintre, V., Georgiev, I., Hold-Geoffroy, Y., Hu, Y., Luan, F., Yan, L.Q., Hašan, M.: Rgb↔x: Image decomposition and synthesis using material- and lighting-aware diffusion models (2024)
44. Zhang, J., Yao, Y., Li, S., Liu, J., Fang, T., McKinnon, D., Tsin, Y., Quan, L.: NeILF++: Inter-Reflectable Light Fields for Geometry and Material Estimation. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3578–3587. IEEE, Paris, France (Oct 2023). https://doi.org/10.1109/ICCV51070.2023.00333
45. Zhang, T., Kuang, Z., Jin, H., Xu, Z., Bi, S., Tan, H., Zhang, H., Hu, Y., Hasan, M., Freeman, W.T., et al.: Relitlrm: Generative relightable radiance for large reconstruction models. arXiv preprint arXiv:2410.06231 (2024)
46. Zhang, W., Tang, J., Zhang, W., Fang, Y., Liu, Y.S., Han, Z.: MaterialRefGS: Reflective gaussian splatting with multi-view consistent material inference. In: Advances in Neural Information Processing Systems (2025)

47. Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T., Barron, J.T.: Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics (ToG) **40**(6), 1–18 (2021)
48. Zhang, Y., Sun, J., He, X., Fu, H., Jia, R., Zhou, X.: Modeling indirect illumination for inverse rendering. In: CVPR (2022)
49. Zhu, Z., Fan, Z., Jiang, Y., Wang, Z.: Fsgs: Real-time few-shot view synthesis using gaussian splatting (2023)