GS-PI++: LIGHT-AWARE GENERATIVE PBR DECOMPOSITION OF OBJECTS IN GAUSSIAN SCENES

Anonymous authors
Paper under double-blind review

ABSTRACT

Recovering physically based surface attributes from Gaussian scenes is an ill-posed inverse problem because observed appearance entangles intrinsic reflectance, surface orientation, and illumination. Object-level inference provides a convenient unit for reusable learned priors, but isolating an object removes contextual observations that can help interpret its appearance, including information about the source illumination and the surrounding scene. Uncertainty in surface orientation further compounds this ambiguity, since errors in normals can be partially compensated by changes in predicted material parameters. We introduce GS-PI++, a light-aware generative framework for PBR decomposition of object instances in Gaussian scenes. For each target object, we construct an object-centered scene appearance probe by removing the target instance and rendering outward from its center. The probe composites visible near-field scene content with the source far-field environment, providing a compact directional observation of the conditions under which the object was observed. Conditioned on the object observations and this probe, a point diffusion model jointly predicts base color, roughness, metallicity, and surface normals. The learned attribute predictor avoids per-scene inverse-rendering optimization. Experiments on synthetic assets and real scanned objects show substantial improvements in material recovery over the evaluated baselines, including comparisons in which source illumination is also provided to competing inverse-rendering methods. The recovered attributes further support downstream Gaussian-based relighting and material editing.

1 INTRODUCTION

Editable and relightable Gaussian scene representations require more than reproducing observed radiance: they require explicit estimates of the intrinsic surface attributes that govern light–surface interaction. Gaussian splatting methods (Kerbl et al., 2023) provide efficient scene reconstruction and rendering, but their appearance representations do not directly expose the base color, roughness, metallicity, and surface orientation required for material editing and relighting. Recovering these attributes for object instances embedded in reconstructed Gaussian scenes is therefore an inverse problem in which reflectance, geometry, and illumination must be disentangled from their combined visual effects.

A central difficulty is that intrinsic surface attributes are local, whereas the observations used to infer them are not. An object's appearance depends not only on its own material and geometry, but also on the source illumination and the scene in which it is observed. Nearby geometry can alter visibility, cast shadows, contribute indirect illumination, and appear in specular reflections. Object-level inference is attractive because it enables reusable priors across scenes, but isolating an object also removes contextual observations that can help interpret scene-dependent appearance effects. This suggests separating the spatial scope of the prediction target from the spatial scope of the evidence used for inference.

A second ambiguity arises from the coupling between material and surface orientation. Even when surface positions are available from a reconstructed Gaussian representation, errors in shading normals change how illumination is mapped to observed radiance. A normal error can shift or reshape a specular response, allowing similar image evidence to be explained through compensating changes in roughness or reflectance. Such compensation may reproduce the source observations while yield-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ing intrinsic attributes that behave incorrectly under novel illumination. We therefore jointly reason about material parameters and surface normals rather than estimating them independently. The value of illumination evidence for recovering these quantities is also reflected in recent work on multi-light intrinsic estimation (He et al., 2025).

Optimization-based inverse rendering addresses these coupled factors by repeatedly fitting an explicit image-formation model to each scene (Liang et al., 2024; Gao et al., 2024). This provides flexible scene-specific estimation, but requires iterative rendering and parameter updates for every new scene. Learned generative priors offer a complementary direction: statistical regularities of plausible intrinsic attributes (Liang et al., 2025) can be learned offline and reused at inference time. Recent work has explored diffusion-based intrinsic estimation, feed-forward scene decomposition, and object-level generative PBR prediction. These developments motivate a different question: can a reusable object-level generative prior remain object-centric in its predictions while being conditioned on broader scene observations that help interpret the input appearance?

We address this question with GS-PI++, a light-aware generative framework for PBR decomposition of object instances in Gaussian scenes. Our key premise is that the scope of a reusable prior need not coincide with the scope of its conditioning evidence. The predicted attributes remain object-specific, while their inference can use observations from the broader scene. To connect these two scopes, we construct an object-centered scene appearance probe for each target instance: the target object is removed, and the remaining Gaussian scene is rendered outward from the object center. The resulting probe composites visible near-field scene content with the source far-field environment, yielding a compact directional observation of the conditions under which the object was seen. The probe is used as contextual evidence rather than as an exact measurement of per-surface incident illumination.

Conditioned on the object observations and the scene appearance probe, GS-PI++ jointly denoises material parameters and surface normals in a point-based diffusion model. A shared representation allows reflectance and surface orientation to constrain one another during denoising, while task-specific branches route different appearance and lighting cues to the material and normal predictions. Importantly, the recovered attributes remain intrinsic to the target object: the probe only conditions how the source observations are interpreted. Once recovered, the same attributes can be transferred back to the Gaussian representation and reused for material editing or relighting under novel target illumination. The learned attribute predictor is trained offline, avoiding the scene-specific inverse-rendering optimization.

Our contributions are:

- A scene-conditioned, light-aware formulation for object-level PBR decomposition in Gaussian scenes. Our formulation decouples the spatial scope of the prediction target from that of the conditioning evidence: intrinsic attributes are recovered for individual object instances, while inference can exploit observations from the broader scene. This enables a reusable object-level generative prior to incorporate scene-dependent evidence without making scene illumination part of the predicted intrinsic representation.

- An object-centered scene appearance probe for contextual attribute inference. For each target object, we remove the target instance and render the remaining Gaussian scene outward from its center. The probe combines visible near-field scene content with the source far-field environment, providing a compact directional observation of the conditions under which the object was observed. The construction is not tied to a specific Gaussian representation.

- A joint material–normal point diffusion model for intrinsic attribute recovery. The model jointly denoises material parameters and surface normals through a shared representation, allowing reflectance and surface orientation to constrain one another, while task-specific branches incorporate complementary appearance and lighting cues for material and normal prediction. The learned predictor amortizes attribute inference across scenes without requiring per-scene inverse-rendering optimization.

2 Related Work

Gaussian inverse rendering and PBR decomposition. Gaussian splatting provides explicit, renderable scene representations (Kerbl et al., 2023), but recovering editable PBR attributes from them typically requires separating reflectance, surface orientation, and illumination from observed appearance. Existing Gaussian inverse-rendering methods address this problem by fitting an explicit image-formation model to each scene: GS-IR estimates material and illumination with normal regularization and occlusion baking (Liang et al., 2024); Relightable 3D Gaussians use BRDF decomposition with ray tracing (Gao et al., 2024); IRGS models inter-reflection with 2D Gaussian ray tracing (Gu et al., 2025); GaussianShader introduces shading functions for reflective surfaces (Jiang et al., 2024); and LumiGauss targets relighting of Gaussians reconstructed from in-the-wild photographs (Kaleta et al., 2025). Although these methods differ in their reflectance and light-transport models, they share a reliance on iterative, scene-specific inverse-rendering optimization. In contrast, GS-PI++ starts from an existing Gaussian reconstruction and uses a reusable generative predictor for object-level intrinsic attributes, conditioned on an object-centered scene appearance probe, without solving a new scene-specific inverse-rendering problem for attribute prediction.

Learned and feed-forward scene decomposition. Recent work replaces per-scene fitting with learned intrinsic or relightable scene prediction. InvSplat jointly reconstructs Gaussian scenes and intrinsic attributes from posed images (Karpikova et al., 2026), whereas GS-PI++ operates on an existing Gaussian reconstruction with a reusable object-level generative prior. F-RNG and LightBridge learn relightable appearance representations rather than explicit editable PBR attributes (Fu et al., 2026; Cao et al., 2026), while DiffusionRenderer explores generative inverse and forward rendering with video diffusion (Liang et al., 2025). At the object level, Neural LightRig and Material Anything recover materials from multi-light observations or meshes, respectively (He et al., 2025; Huang et al., 2025). Most closely related, GS-PI (Xu et al., 2026) uses point diffusion to predict PBR materials for isolated Gaussian objects without source-light conditioning or joint normal prediction. GS-PI++ extends this setting to objects embedded in scenes by conditioning joint material–normal generation on an object-centered scene appearance probe.

Instance preprocessing and point-diffusion architectures. Our pipeline treats instance segmentation and object completion as external preprocessing stages. We use SAM 3 (Carion et al., 2026) to obtain object masks from the reconstructed scene and LaS-Comp (Yan et al., 2026) to complete the extracted object observations before attribute prediction. These components provide the object-level interface to our method and are not part of the learned PBR decomposition model. For point-based attribute generation, we build on recent point-transformer designs: Point Transformer V3 introduces serialized local point processing (Wu et al., 2024), SP^{2}T develops sparse spatial-proxy communication (Wan et al., 2025), and DiPT adapts point-transformer processing to diffusion (Bastico et al., 2026). Building on these ideas, our denoiser combines serialized local point attention with spatially anchored proxy communication for fixed-coordinate material–normal diffusion, while scene-appearance conditioning is injected through the proxy pathway. This differs from the two-stream PointInfinity read–compute–write architecture used by GS-PI (Huang et al., 2024b), while retaining object-level point diffusion as the underlying generative formulation.

3 Method

3.1 Problem Formulation

Given a reconstructed Gaussian scene $\mathcal{S}$ and its associated image observations, we recover physically based surface attributes for object instances extracted from the scene. Following GS-PI (Xu et al., 2026), we convert each completed Gaussian instance into a dense surface point cloud through multi-view rendering and depth unprojection. For object instance k, let $\mathbf{X}_k = \{\mathbf{x}_{k,i}\}_{i=1}^{N_k}$ denote its surface points. We predict

$$\mathbf{Y}_k = \{(\mathbf{a}_{k,i}, r_{k,i}, m_{k,i}, \mathbf{n}_{k,i})\}_{i=1}^{N_k}, \tag{1}$$

where $\mathbf{a}_{k,i} \in [0,1]^3$ denotes base color, $r_{k,i} \in [0,1]$ roughness, $m_{k,i} \in [0,1]$ metallic, and $\mathbf{n}_{k,i} \in \mathbb{S}^2$ a shading normal. During diffusion-based attribute prediction, the surface-point coordinates remain fixed. We distinguish target-object observations $\mathcal{C}_k$ from the object-centered scene appearance probe

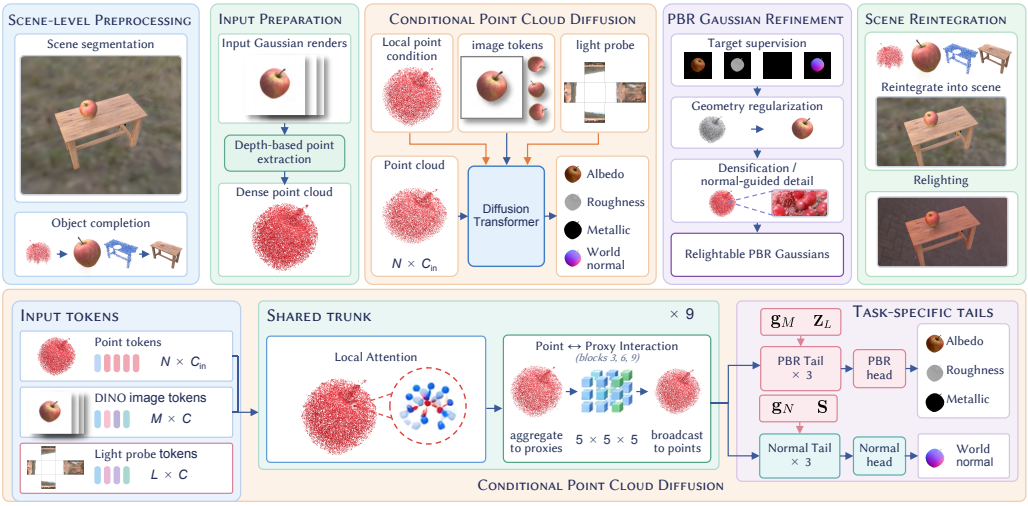


Figure 1: **GS-PI++ pipeline.** An input Gaussian scene is segmented, and its object instances are completed; each completed instance is rendered and unprojected into a dense point cloud. The conditional point-cloud diffusion model predicts albedo, roughness, metallic, and world-space normals from local point conditions, image tokens, and the object-centered light probe. PBR Gaussian refinement transfers the predicted attributes to the Gaussians of each instance, which are reintegrated into the scene and relit under novel illumination. Bottom: the denoiser combines local attention with point–proxy interaction in nine shared blocks, followed by separate PBR and normal tails. Segmentation, completion, point cloud unprojection, and attribute transfer reuse existing components (Carion et al., 2026; Yan et al., 2026; Xu et al., 2026).

$\mathcal{P}_k$. The former describes the target object's observed appearance, while the latter provides directional evidence from the surrounding scene and source far-field environment. Our objective is to learn

$$p_\theta\left(\mathbf{Y}_k \mid \mathbf{X}_k, \mathcal{C}_k, \mathcal{P}_k\right). \tag{2}$$

This formulation decouples the scope of prediction from the scope of conditioning: intrinsic attributes are generated at the object level, while their inference can exploit broader scene observations. It therefore retains the reusability of an object-level generative prior while reintroducing contextual evidence discarded by object isolation. Materials and normals are generated jointly to model their coupled explanations of observed appearance. For clarity, we omit the object index k below.

3.2 Object-Centered Scene Appearance Probes

For each object, we place a probe at its center $\mathbf{c}_k$ and render the reconstructed scene representation along directions $\{\boldsymbol{\omega}_q\}_{q=1}^{Q}$, with the target instance excluded so that its own surfaces do not obstruct the surroundings. The probe thus contains whatever the reconstruction represents as seen from $\mathbf{c}_k$: near-field Gaussians together with the far-field model of the chosen reconstruction, such as distant Gaussians, a background layer, or an environment map. Let $\mathbf{F}_k(\boldsymbol{\omega})$ denote the premultiplied color of the retained near-field content and $A_k(\boldsymbol{\omega})$ its accumulated opacity. The probe color is

$$\mathbf{P}_k(\boldsymbol{\omega}) = \mathbf{F}_k(\boldsymbol{\omega}) + (1 - A_k(\boldsymbol{\omega}))\, E_{\mathrm{src}}(\boldsymbol{\omega}), \tag{3}$$

where E_{src} is the far-field radiance under the source illumination, composited with the near field in linear color space. Because the probe only renders the given representation, its definition does not depend on a particular reconstruction method. Eq. (3) preserves both neighboring-object appearance and the visible far field, so objects at different locations receive different observations even under the same environment.

We represent the probe as

$$\mathcal{P}_k = \{(\mathbf{P}_{k,q}, \boldsymbol{\omega}_q, \Omega_q, v_{k,q})\}_{q=1}^{Q}, \tag{4}$$

where Ω_q is the sample's solid angle and $v_{k,q}$ indicates validity: directions covered by neither the near field nor the far-field model are marked invalid. The same probe definition is used during training and inference.

Because the probe is captured at a single object-centered location, it is a contextual observation rather than a measurement of the incident illumination at every surface point. Probe directions and predicted normals share a common coordinate frame.

3.3 Scene-Conditioned Joint Material–Normal Diffusion

Attribute diffusion. We concatenate the five material channels and three normal channels into an eight-dimensional vector at each point. Material channels are normalized using fixed training-set statistics, while normals retain their signed components. Writing the clean attributes as $\mathbf{Y}_{\text{clean}}$, the forward process is

$$\mathbf{Y}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{Y}_{\text{clean}} + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{5}$$

Only attributes are diffused. By diffusing material and normal channels within the same generative state, the denoiser can model their dependencies throughout sampling. Normals are diffused in Euclidean space and normalized to unit length after inference.

Shared material–normal representation. The denoiser combines local point processing with spatial proxy communication, drawing on point diffusion and sparse proxy attention (Bastico et al., 2026; Wan et al., 2025). Point inputs comprise positions, noisy attributes, observed RGB, and source-view identities. Frozen DINOv2 features provide additional image context (Oquab et al., 2024).

Local attention operates on spatially serialized point windows. For communication across the object, each point exchanges features with its eight neighboring vertices in a regular proxy grid through trilinear associations. At global interaction blocks, proxies aggregate point features, interact with other proxies, and attend to image tokens. Updated proxy features are subsequently propagated back to points.

The shared point and proxy states are

$$(\mathbf{H}, \mathbf{U}) = F_{\text{shared}}\left(\mathbf{Y}_t, \mathbf{X}, \mathbf{C}_{\text{pt}}, \mathbf{Z}_D, \mathbf{e}_t\right), \tag{6}$$

where $\mathbf{C}_{\text{pt}}$ contains point appearance conditions, $\mathbf{Z}_D$ contains image tokens, and $\mathbf{e}_t$ is the timestep embedding. The shared trunk jointly processes the noisy material and normal state. Probe information is introduced in the task-specific branches through three representations at different angular scales: global photometric, low-order angular, and localized directional context.

Global photometric context. We compute the solid-angle-weighted mean RGB and angular coverage:

$$\boldsymbol{\mu} = \frac{\sum_q v_q \Omega_q \mathbf{P}_q}{\max\left(\sum_q v_q \Omega_q, \varepsilon\right)}, \qquad \kappa = \frac{\sum_q v_q \Omega_q}{4\pi}. \tag{7}$$

The global descriptor and material condition are

$$\mathbf{b} = \left[\log(\max(\boldsymbol{\mu}, \varepsilon)), \kappa\right], \qquad \mathbf{g}_M = \rho_M(\mathbf{b}). \tag{8}$$

This representation preserves the observed intensity scale and color balance; $\mathbf{g}_M$ modulates the point and proxy features of the material branch, and the same statistics enter the normal condition.

Low-order angular context. A global average does not describe how appearance varies with direction. We therefore project the probe onto real spherical harmonics through degree two:

$$\mathbf{h}_\ell = \frac{\sum_q v_q \Omega_q B_\ell(\boldsymbol{\omega}_q) \mathbf{P}_q}{\max\left(\sum_q v_q \Omega_q, \varepsilon\right) e}, \qquad \ell = 1, \ldots, 9, \tag{9}$$

where B_ℓ denotes one of the nine basis functions and $e = \max(\mathbf{l}^\top \boldsymbol{\mu}, \varepsilon)$ is the mean probe luminance. Normalization by e separates relative angular structure from the global intensity retained in $\mathbf{b}$.

The normal condition is

$$\mathbf{g}_N = \rho_N\left([\mathbf{b}, \mathbf{h}_1, \ldots, \mathbf{h}_9]\right). \tag{10}$$

This compact descriptor provides broad directional context for interpreting orientation-dependent shading, complemented by local image evidence in the normal branch.

For either task $j \in \{M, N\}$, the global condition modifies the timestep embedding:

$$\mathbf{e}_t^j = \mathbf{e}_t + A_j([\mathbf{e}_t, \mathbf{g}_j]). \tag{11}$$

The resulting embedding controls adaptive normalization and residual gates within the corresponding branch.

Localized directional context. Material inference additionally uses directional tokens $\mathbf{Z}_L$ constructed from $6 \times 8 \times 8$ cubemap bins. Each token combines luminance-normalized mean RGB with peak RGB, which retains concentrated bright features that averaging suppresses, together with an intensity-weighted directional centroid and observation-quality descriptors.

After point aggregation, proxy interaction, and image-context attention, the material proxy state $\overline{\mathbf{U}}_M$ queries the directional tokens:

$$\widetilde{\mathbf{U}}_M = \overline{\mathbf{U}}_M + \Gamma_M([\mathbf{e}_t, \mathbf{g}_M]) \odot \mathrm{Attn}\left(\overline{\mathbf{U}}_M, \mathbf{Z}_L, \mathbf{Z}_L\right), \tag{12}$$

where Γ_M is a learned channel-wise gate. Attention excludes bins without valid observations. Following the proxy feed-forward transformation, the contextual update propagates back to point features through the sparse proxy-to-point associations.

This gives the material branch selective access to directional detail, while the normal branch receives its probe context through the compact global and low-order condition without directional-token cross-attention.

Joint material and normal prediction. Both task branches start from the shared states $(\mathbf{H}, \mathbf{U})$. The normal branch additionally receives source-image features associated with each point's original pixel. A sparse 3×3 RGB neighborhood sampled at four-pixel intervals, its alpha values, and normalized image coordinates form a 38-dimensional feature $\mathbf{s}_i$. A learned adapter adds these observed-image features to the normal branch's point representation.

The noise predictions are

$$\widehat{\boldsymbol{\epsilon}}_M = F_M\left(\mathbf{H}, \mathbf{U}; \mathbf{X}, \mathbf{Z}_D, \mathbf{Z}_L, \mathbf{e}_t^M\right), \tag{13}$$

$$\widehat{\boldsymbol{\epsilon}}_N = F_N\left(\mathbf{H} + \phi(\mathbf{S}), \mathbf{U}; \mathbf{X}, \mathbf{Z}_D, \mathbf{e}_t^N\right), \tag{14}$$

where $\mathbf{S}$ collects the source-image features and ϕ is their adapter. The task branches differ in how probe and image conditions are routed, but they do not define independent generative processes. Material and normal predictions remain coupled through the shared noisy attribute state, the shared trunk, and successive joint denoising updates. This design preserves joint reasoning between reflectance and surface orientation while allowing each task to use its most relevant conditioning cues. The network contains nine shared blocks followed by three blocks in each task branch, with feature width 512 and eight attention heads. Local windows contain 32 points, and the proxy grid contains 5^3 vertices. This design combines joint reasoning between reflectance and surface orientation with task-specific conditioning.

3.4 Training and Inference

We train the denoiser with a Min-SNR-weighted noise-prediction loss (Hang et al., 2023) on the eight attribute channels, restricting normal supervision to valid targets and adding cosine and unit-length penalties on the predicted clean normals at low noise levels; scene appearance probe enters only as conditioning, without a rendering loss (App. B). At inference, we run 50 DDIM steps (Song et al., 2021) from Gaussian noise with point positions and conditions fixed, normalize the predicted normals, and transfer the attributes to the Gaussians by point-to-Gaussian distillation as in GS-PI (Xu et al., 2026) (App. C). The learned attribute-prediction stage therefore avoids per-scene inverse-rendering optimization. The predicted attributes are subsequently transferred to the Gaussian representation using the GS-PI-style refinement procedure, which may update a subset of Gaussian parameter sand is separate from the learned attribute-prediction stage.

4 EXPERIMENTS

4.1 SETUP

Datasets. We evaluate on synthetic and real-world data. For quantitative material evaluation, we use NeRF Synthetic (Mildenhall et al., 2020)(Fig. 4a), game assets, and Objaverse (Deitke et al., 2023). Each asset is rendered from 32 deterministic Fibonacci-hemisphere views, which provide ground-truth PBR attributes and relighting targets under held-out environments. For real-world evaluation, we use the Digital Twin Catalog (DTC) (Dong et al., 2025) and Google Scanned Objects (GSO) (Downs et al., 2022); DTC additionally provides PBR reference materials and multi-illumination observations, which we use for quantitative evaluation on real scanned objects. We train on about 18k objects from Objaverse and DTC, each rendered from 32 views; training and test objects are disjoint. In all quantitative experiments, the far field of the probe (E_{src}) is a given environment map rather than a reconstructed one: the source HDR for synthetic data (an ideal far-field reconstruction) and the DTC environment map for real scans.

Baselines and evaluation protocols. We compare with five optimization-based Gaussian inverse-rendering methods: IRGS (Gu et al., 2025), Relightable 3D Gaussians (Gao et al., 2024), LumiGauss (Kaleta et al., 2025), GaussianShader (Jiang et al., 2024), and GS-IR (Liang et al., 2024). We also compare with the object-level generative baseline GS-PI (Xu et al., 2026), using the same instance preprocessing and attribute-transfer interface as GS-PI++. We omit InvSplat (Karpikova et al., 2026) as its related code remains closed-source.

We report two comparison protocols. In the *native* protocol, each inverse-rendering baseline follows its published setting and estimates illumination jointly with scene attributes (Table 1). Because GS-PI++ receives the source far-field environment through its probe, we additionally report a *provided Source illumination* protocol in which the same source illumination is provided to applicable baselines (Table 3). This second protocol isolates attribute recovery from illumination estimation.

We report PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018) for base color, roughness, metallicity, and novel-view relighting within the ground-truth object mask, together with mean angular error for predicted normals. For coupled relighting, each method is rendered with its own predicted attributes and normals using the same PBR renderer; GS-PI does not predict normals and is therefore evaluated with ground-truth normals.

4.2 COMPARISON WITH BASELINES

Fig. 2 qualitatively compares the recovered attributes and relighting. IRGS retains visible shading in albedo despite reasonable relighting. GS-IR shows visible appearance leakage into its material maps and less accurate relighting. Relightable 3D Gaussians also retain source-dependent variation in the recovered materials. LumiGauss produces illumination-tinted albedo and does not provide the full PBR outputs. GaussianShader shows substantial color and shading bias on several objects and incomplete output coverage. GS-PI produces cleaner object materials, but lacks normal prediction and source-light conditioning. Providing the same source illumination improves several inverse-rendering baselines, but their recovered materials remain less accurate than those of GS-PI++. In contrast, GS-PI++ recovers intrinsic maps closer to the reference and produces more consistent relighting. Additional examples are provided in Appendix F.

Table 1 reports the native-protocol comparison, where GS-PI++ achieves the strongest material accuracy overall, together with the lowest normal error and strong coupled relighting. Table 3 further supports the same trend under controlled source illumination: providing applicable baselines with the same source illumination improves their results, but a substantial material-recovery gap remains, while relighting becomes comparable. Together, these results indicate that the material gains of GS-PI++ are not explained solely by access to the source far-field environment.

4.3 ABLATION STUDY

Fig. 3 evaluates probe conditioning, the joint material–normal design, and the point–proxy backbone, with representative predictions shown in Fig. 3b.

Table 1: **PBR decomposition under native protocols.** Cells report PSNR↑ / SSIM (%)↑ / LPIPS (%)↓; normal error is in degrees. Baselines follow their published light protocols, while GS-PI++ receives the source far field through its probe. Table 3 controls this difference. "/" denotes unavailable outputs; † denotes methods without relighting evaluation. Runtime is reported per object.

	Synthetic Scenes					Real Scanned Scenes					
Method	Albedo	Material Rough.	Metallic	Normal ↓ Deg.	Relighting Coupled	Albedo	Material Rough.	Metallic	Normal ↓ Deg.	Relighting Coupled	Runtime (min) ↓
IRGS	18.98 / 78.23 / 2.66	11.47 / 59.83 / 4.54	/	9.21	18.72 / 78.10 / 2.18	22.13 / 86.11 / 3.80	10.35 / 66.33 / 6.84	/	6.62	20.24 / **86.88** / 3.44	34.3
Relightable3DGaussian	13.54 / 67.61 / 4.20	14.97 / 33.80 / 5.99	/	10.91	15.00 / 72.88 / 3.28	12.49 / 72.85 / 6.45	14.60 / 27.65 / 8.03	/	6.87	12.55 / 73.91 / 6.62	45.4
LumiGauss†	14.24 / 64.74 / 4.32	/	/	20.53	/	14.88 / 68.31 / 6.40	/	/	15.03	/	79.3
GaussianShader†	9.55 / 35.18 / 5.66	13.09 / 42.02 / 5.39	/	14.90	/	7.80 / 27.52 / 9.42	15.40 / 38.66 / 7.24	/	15.18	/	26.9
GS-IR	12.91 / 63.72 / 3.74	9.14 / 51.21 / 6.34	28.18 / 62.31 / 7.32	17.00	14.34 / 61.56 / 3.60	13.58 / 69.19 / 5.50	11.09 / 52.05 / 8.58	24.88 / 43.63 / 11.16	14.39	13.62 / 66.52 / 6.68	28.2
GS-PI	17.54 / 71.48 / 3.52	14.85 / 58.78 / 4.39	21.34 / 13.95 / 9.51	/	16.94 / 61.96 / 3.68	19.42 / 79.34 / 4.80	21.21 / 68.95 / 4.50	26.98 / 31.30 / 11.78	/	18.31 / 71.80 / 5.09	**5.5**
GS-PI++ (ours)	**30.65 / 88.92 / 1.74**	**23.66 / 79.40 / 2.78**	**41.27 / 92.51 / 2.04**	**8.70**	**25.60 / 83.84 / 2.07**	**25.72 / 88.65 / 2.77**	**22.95 / 85.63 / 4.29**	**42.95 / 94.90 / 0.92**	**4.71**	**23.23** / 84.68 / **2.96**	7.2



Figure 2: **Qualitative baseline comparison.** For each object and method, we show albedo(large), roughness, metallic, and normal (right), together with PBR relighting under the native protocol and relighting using PBR attributes and normals recovered with provided source illumination (bottom). Crossed boxes denote unavailable outputs.

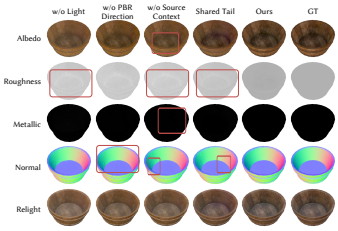
Backbone (A1, A2). Unlike GS-PI, A1 additionally predicts normals. With the RCW backbone, adding the normal channels severely degrades albedo. With our point–proxy backbone, joint prediction instead improves albedo over separate models (A7: 26 dB vs. 29 dB). The point–proxy backbone is therefore what makes joint material–normal prediction beneficial, and the probe improves either backbone.

Conditioning (A3–A5). Removing the probe (A3) in the training stage confines inference to the object's own appearance, so source illumination is baked into the intrinsic channels (albedo 29→20 dB). In Fig. 3b, its albedo is brighter and more saturated than the ground truth and its roughness is overestimated and inherits the albedo texture; these errors carry over to a washed-out relit appearance. Dropping the directional tokens (A4) leaves only global and low-order statistics, which cannot represent concentrated highlights; metallic accuracy drops most (30→15 dB), and the unexplained highlight energy again brightens albedo and roughness. Removing the source-image features from the normal branch (A5) raises the normal error from 4.3° to 8.3°, with visible errors on the inner wall, and also washes out the interior albedo, since orientation errors propagate to materials through the shared denoising state. Perturbing the probe with a mismatched, rotated, or averaged environment map further shifts the recovered albedo (Appendix D), indicating that the model relies on the directional content of the probe rather than its mere presence.

Probe components. Unlike the training-time ablation A3, Table 2 tests the trained full model with different probe components. "Env-only" retains only the source far field, "Near-only" retains only valid near-field observations, "Full" uses their composition in Eq. (3), and "No probe" removes both. Env-only retains much of the probe benefit, while Full further improves material recovery, indicating additional useful information from the near-field scene.

Variant	Alb. PSNR / SSIM / LPIPS	Rough. PSNR / SSIM / LPIPS	Met. PSNR / SSIM / LPIPS	Normal (°) ↓	Relight PSNR / SSIM / LPIPS
A1:RCW w/o probe	8 / 50 / 39	17 / 62 / 34	28 / 78 / 6.9	7.7	14 / 65 / 35
A2:RCW w/ probe	15 / 61 / 21	23 / 73 / **15**	28 / **80** / 7.8	6.5	19 / 73 / 27
A3:w/o probe	20 / 66 / 19	19 / 73 / 28	16 / 54 / 27	8.2	21 / 81 / 17
A4:w/o directional attn.	23 / 73 / 15	18 / 57 / 29	15 / 56 / 34	8.1	22 / 75 / 19
A5: w/o normal image context	24 / 80 / 10	20 / 72 / 22	19 / 62 / 29	8.3	24 / 79 / 13
A6:shared tail	25 / 81 / 16	22 / 74 / 21	22 / 66 / 20	9.5	22 / 75 / 23
A7:Separate M/N	26 / 73 / 6	**24** / 72 / 16	25 / 77 / 11	6.3	24 / 77 / 10
Ours	**29 / 93 / 4.0**	23 / **84** / 18	**30** / 66 / **3.6**	**4.3**	**26 / 89 / 8.4**

(a)

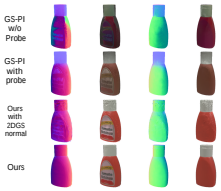


(b)

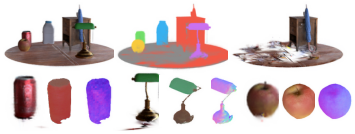
Figure 3: **Model ablation.** (a) Quantitative study. (b) w/o Light, w/o PBR Direction, w/o Source Context, and Shared Tail are A3–A6.



(a) Relighting



(b) Normals study



(c) Failure case

Figure 4: **Relighting and robustness.** (a) Recovered attributes relit under two target HDR environments. (b) Rows: RCW backbone w/o and w/ probe, our attributes with 2DGS normals, ours; columns: normal and relit views. (c) Failure case discussed in Sec. 5.1.

Joint material–normal design (A6, A7). Using a shared task tail for material and normal prediction (A6) increases normal error from 4.3° to 9.5°, while training fully separate material and normal models (A7) increases it to 6.3° and lowers base-color accuracy. Together, these comparisons support a shared denoising representation followed by task-specific material and normal branches. The qualitative examples in Fig. 3b show corresponding differences in normal, material, and relighting.
2DGS reconstruction normals versus predicted normals. Fig. 4b compares four normal variants. The top two rows use matched RCW-backbone models without and with probe conditioning, whose predicted normals retain different amounts of label texture. The bottom two rows fix our recovered PBR attributes and vary only the shading normals between 2DGS reconstruction normals (Huang et al., 2024a) and ours. 2DGS normals contain high-frequency noise and label imprints that produce bumpy relighting, whereas our predicted normals are smoother and yield cleaner shading.

5 Limitations and Conclusion

5.1 Limitations

GS-PI++ inherits errors from Gaussian reconstruction and external instance preprocessing. In the cluttered tabletop failure case in Fig. 4c, instance masks leak onto the table, the extracted can and lamp are blurred or incomplete, and removing the instances leaves holes in the unprocessed table surface; these errors propagate to predicted albedo and normals. The scene appearance probe also inherits reconstruction and exposure errors and assumes a provided source far-field environment; mismatched far fields shift the recovered reflectance, while a single object-centered probe may be insufficient near local emitters or for large objects. Predicted normals need not be integrable, and our material model is limited to an opaque metallic–roughness BRDF. Finally, GS-PI++ predicts extracted object instances rather than all scene surfaces, so full-scene relighting remains incomplete.

5.2 Conclusion

We presented GS-PI++, a light-aware generative framework for PBR decomposition of object instances in Gaussian scenes. An object-centered scene appearance probe combines visible near-field scene content with the source far-field environment to condition a reusable object-level prior, while joint material–normal diffusion couples reflectance and surface orientation during prediction. Experiments on synthetic and real scanned objects show improved material recovery over object-only inference and better normal estimation than decoupled prediction. The learned predictor amortizes attribute inference without per-scene inverse-rendering optimization, and the recovered attributes support downstream Gaussian-based material editing and relighting.

AI USE STATEMENT

We used generative AI tools to help organize, draft, and edit the manuscript text, including revisions for clarity and length; to search for, verify, and format related literature and references; to inspect code; and to check consistency between the text, tables, and figures, including correcting a typo in one figure label. Some code was written with Codex. We reviewed all AI-assisted content and take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Our experiments use publicly available object datasets (NeRF Synthetic, Objaverse, Google Scanned Objects, and the Digital Twin Catalog), each under its respective license.

REPRODUCIBILITY STATEMENT

Section 3 and Appendix A describe the appearance probe and the network architecture; Appendices B and C give the training objective and the inference procedure; Section 4.1 describes the datasets, baselines, and metrics.

REFERENCES

Matteo Bastico, David Ryckelynck, Laurent Corté, Yannick Tillier, and Etienne Decencière. Rethinking metrics and diffusion architecture for 3d point cloud generation. In *2026 International Conference on 3D Vision (3DV)*, pp. 1380–1391. IEEE, 2026.

Hezhi Cao, Panhao Cheng, Huangsheng Du, Qibiao Li, Youcheng Cai, and Ligang Liu. LightBridge: Feed-Forward Generative Relighting for 3D Gaussian Splatting. *arXiv preprint arXiv:2609.02543*, 2026. URL https://arxiv.org/abs/2609.02543.

Nicolas Carion, Laura Gustafson, Yuan-Ting Hu, Shoubhik Debnath, Ronghang Hu, Didac Suris Coll-Vinent, Chaitanya Ryali, Kalyan Vasudev Alwala, Haitham Khedr, Andrew Huang, Jie Lei, Tengyu Ma, Baishan Guo, Arpit Kalla, Markus Marks, Joseph Greer, Meng Wang, Peize Sun, Roman Rädle, Triantafyllos Afouras, Effrosyni Mavroudi, Katherine Xu, Tsung-Han Wu, Yu Zhou, Liliane Momeni, Rishi Hazra, Shuangrui Ding, Sagar Vaze, Francois Porcher, Feng Li, Siyuan Li, Aishwarya Kamath, Ho Kei Cheng, Piotr Dollár, Nikhila Ravi, Kate Saenko, Pengchuan Zhang, and Christoph Feichtenhofer. Sam 3: Segment anything with concepts. In *International Conference on Learning Representations*, 2026.

Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13142–13153, 2023.

Zhao Dong, Ka Chen, Zhaoyang Lv, Hong-Xing Yu, Yunzhi Zhang, Cheng Zhang, Yufeng Zhu, Stephen Tian, Zhengqin Li, Geordie Moffatt, Sean Christofferson, James Fort, Xiaqing Pan, Mingfei Yan, Jiajun Wu, Carl Yuheng Ren, and Richard Newcombe. Digital twin catalog: A large-scale photorealistic 3d object digital twin dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B. McHugh, and Vincent Vanhoucke. Google scanned objects: A high-quality dataset of 3d scanned household items. *arXiv preprint arXiv:2204.11918*, 2022.

Guangming Fu, Jiahui Fan, Jian Yang, Miloš Hašan, and Beibei Wang. F-RNG: Feed-Forward Relightable Neural Gaussians. *arXiv preprint arXiv:2605.25975*, 2026. URL https://arxiv.org/abs/2605.25975.

Jian Gao, Chun Gu, Youtian Lin, Zhihao Li, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3D Gaussians: Realistic Point Cloud Relighting with BRDF Decomposition and Ray Tracing. In

European Conference on Computer Vision, 2024. URL https://arxiv.org/abs/2311.16043.

Chun Gu, Xiaofei Wei, Zixuan Zeng, Yuxuan Yao, and Li Zhang. IRGS: Inter-reflective Gaussian splatting with 2D Gaussian ray tracing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10943–10952, 2025.

Tiankai Hang, Shuyang Gu, Chen Li, Jianmin Bao, Dong Chen, Han Hu, Xin Geng, and Baining Guo. Efficient Diffusion Training via Min-SNR Weighting Strategy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7441–7451, 2023.

Zexin He, Tengfei Wang, Xin Huang, Xingang Pan, and Ziwei Liu. Neural LightRig: Unlocking Accurate Object Normal and Material Estimation with Multi-Light Diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26514–26524, 2025. URL https://arxiv.org/abs/2412.09593.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2D Gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 Conference Papers*, 2024a.

Xin Huang, Tengfei Wang, Ziwei Liu, and Qing Wang. Material Anything: Generating Materials for Any 3D Object via Diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26556–26565, 2025. URL https://arxiv.org/abs/2411.15138.

Zixuan Huang, Justin Johnson, Shoubhik Debnath, James M. Rehg, and Chao-Yuan Wu. PointInfinity: Resolution-Invariant Point Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10050–10060, 2024b.

Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. GaussianShader: 3D Gaussian splatting with shading functions for reflective surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5322–5332, 2024.

Joanna Kaleta, Kacper Kania, Tomasz Trzcinski, and Marek Kowalski. LumiGauss: Relightable Gaussian splatting in the wild. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2025.

Polina Karpikova, Wenjing Bian, Haofei Xu, Hendrik P. A. Lensch, and Andreas Geiger. InvSplat: Inverse Feed-Forward Scene Splatting. *arXiv preprint arXiv:2607.02301*, 2026. URL https://arxiv.org/abs/2607.02301.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4), 2023. URL https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

Ruofan Liang, Zan Gojcic, Huan Ling, Jacob Munkberg, Jon Hasselgren, Zhi-Hao Lin, Jun Gao, Alexander Keller, Nandita Vijaykumar, Sanja Fidler, and Zian Wang. DiffusionRenderer: Neural Inverse and Forward Rendering with Video Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. URL https://arxiv.org/abs/2501.18590.

Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. 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. URL https://arxiv.org/abs/2311.16473.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 405–421, 2020. doi: 10.1007/978-3-030-58452-8_24.

Maxime Oquab, Timothee Darcet, Theo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research*, 2024.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.02502.

Jiaxu Wan, Hong Zhang, Ziqi He, Yangyan Deng, Qishu Wang, Ding Yuan, and Yifan Yang. SP^{2}T: Sparse Proxy Attention for Dual-stream Point Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025. URL https://arxiv.org/abs/2412.11540.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point Transformer V3: Simpler, Faster, Stronger. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4840–4851, 2024. URL https://arxiv.org/abs/2312.10035.

Jieting Xu, Rengan Xie, Zijian Huang, Zehui Jin, Rui Wang, and Yuchi Huo. GS-PI: An optimization-decoupled appearance decomposition approach for generating PBR gaussian assets. In *SIGGRAPH Asia 2026 Conference Papers*, 2026. URL https://arxiv.org/abs/2609.19907. To appear; arXiv:2609.19907.

Weilong Yan, Haipeng Li, Hao Xu, Nianjin Ye, Yihao Ai, Shuaicheng Liu, and Jingyu Hu. LaS-Comp: Zero-shot 3d completion with latent-spatial consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7588–7599, June 2026.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

Table 2: Test-time decomposition of probe conditioning. All rows use the same fully trained GS-PI++ model; probe components are masked only at inference. The source environment accounts for most of the improvement, while adding the visible near field yields a further consistent gain in albedo and relighting. Albedo and relighting benefit more consistently from the additional near-field context.

Method	Albedo PSNR ↑	Relight PSNR ↑
Test-time no probe	5.33	9.10
Env-only	26.87	26.06
Near-only	12.93	15.09
Full (ours)	**27.51**	**26.47**

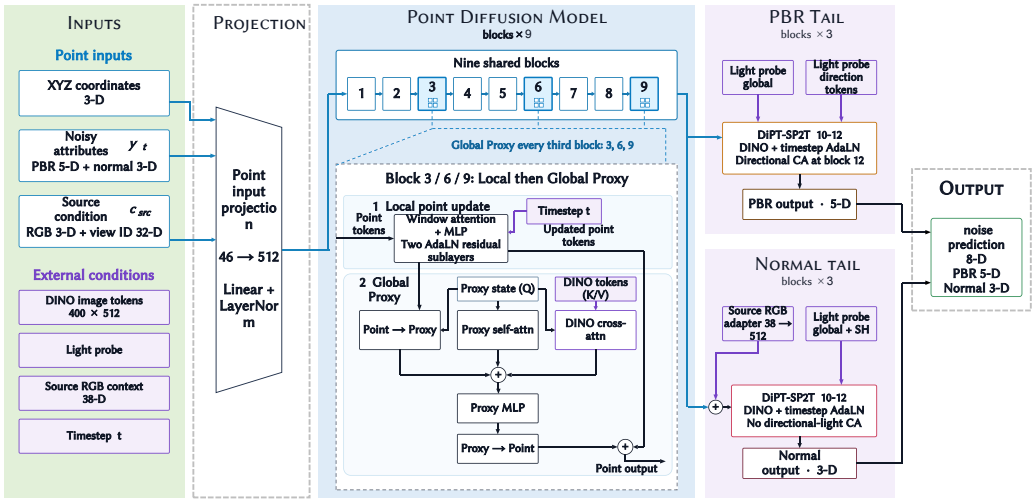


Figure 5: **Network overview.** Nine shared point–proxy blocks, with global proxy exchange at blocks 3, 6, and 9, are followed by a PBR tail and a normal tail of three blocks each. The PBR tail receives the global probe condition and the directional probe tokens; the normal tail receives the global and spherical-harmonic probe condition and the source-RGB features.

A Network Overview

Our network combines object observations and scene appearance probes through a shared backbone and task-specific branches to jointly predict PBR material attributes and surface normals (Fig. 5).

B Learning Objective

Training samples contain surface points, aligned material and normal targets, associated image observations, and scene probes captured under the source lighting. The training set contains about 18k objects from Objaverse and DTC, each rendered from 32 views, and is disjoint from all test objects. We sample 2,048 points per object and use a 1,000-step cosine diffusion schedule.

Let $\langle f_i \rangle_{\mathcal{V}}$ denote the average over minibatch points selected by mask $\mathcal{V}$, with an empty mask contributing zero. For task $j \in \{M, N\}$, the noise-prediction objective is

$$\mathcal{L}_\epsilon^j = \left\langle \frac{w(t)}{2d_j} \left\| \boldsymbol{\epsilon}_{j,i} - \widehat{\boldsymbol{\epsilon}}_{j,i} \right\|_2^2 \right\rangle_{\mathcal{V}_j}, \tag{15}$$

where $d_M = 5$ and $d_N = 3$. Material supervision uses all selected material targets, while normal supervision is restricted to valid normal targets. We adopt Min-SNR weighting (Hang et al., 2023):

$$w(t) = \frac{\min(\mathrm{SNR}(t), \gamma)}{\mathrm{SNR}(t)}, \qquad \mathrm{SNR}(t) = \frac{\bar{\alpha}_t}{1 - \bar{\alpha}_t}, \qquad \gamma = 5. \tag{16}$$

To supervise surface orientation explicitly, we reconstruct the clean attribute estimate:

$$\widehat{\mathbf{Y}}_{\mathrm{clean}} = \frac{\mathbf{Y}_t - \sqrt{1 - \bar{\alpha}_t}\,\widehat{\boldsymbol{\epsilon}}}{\sqrt{\bar{\alpha}_t}}. \tag{17}$$

Let $\widetilde{\mathbf{n}}_i$ denote its normal component, clamped componentwise to $[-4, 4]$ for numerical stability, and define

$$\overline{\mathbf{n}}_i = \frac{\widetilde{\mathbf{n}}_i}{\max(\|\widetilde{\mathbf{n}}_i\|_2, \varepsilon)}. \tag{18}$$

We apply direction and unit-length penalties:

$$\mathcal{L}_{\cos} = \left\langle w(t) \left(1 - \overline{\mathbf{n}}_i^{\top} \mathbf{n}_i\right) \right\rangle_{\mathcal{A}}, \tag{19}$$

$$\mathcal{L}_{\mathrm{unit}} = \left\langle w(t) \left(\|\widetilde{\mathbf{n}}_i\|_2 - 1\right)^2 \right\rangle_{\mathcal{A}}. \tag{20}$$

The mask $\mathcal{A}$ selects valid normal targets with $t \leq 250$ under zero-based timestep indexing, restricting these auxiliary losses to lower-noise estimates.

The complete objective is

$$\mathcal{L} = \mathcal{L}_{\epsilon}^{M} + 0.5\,\mathcal{L}_{\epsilon}^{N} + 0.2\,\mathcal{L}_{\cos} + 0.05\,\mathcal{L}_{\mathrm{unit}}. \tag{21}$$

The model is trained through direct attribute supervision. Scene context enters as conditioning evidence rather than through an explicit differentiable rendering loss.

C Inference and Scene Integration

At inference, each object is accompanied by its appearance observations and scene probe. We initialize the eight attribute channels with Gaussian noise and perform 50 DDIM steps with $\eta = 0$ (Song et al., 2021). Point positions and conditioning observations remain fixed. The full object point set is processed as one diffusion state, and predictions preserve the input point ordering.

After denoising, material normalization is inverted and predicted normals are normalized to unit length. The recovered attributes remain associated with explicit 3D locations, providing an interface for transfer to the Gaussian representation. Material transfer can follow point-to-Gaussian distillation as in GS-PI (Xu et al., 2026). Normal transfer requires consistent coordinate transformations and normalization after interpolation. Instance processing, attribute transfer, and rendering are treated as external components of the scene application.

The source scene probe conditions the interpretation of the input observations. Once attributes have been recovered, a target environment E_{tgt} is supplied to the downstream renderer while those attributes are held fixed. Changing target illumination therefore does not require replacing the source probe or repeating attribute prediction.

The learned predictor uses fixed parameters and requires no scene-specific gradient-based fitting. Its inference cost consists of the denoising evaluations. Scene reconstruction and downstream attribute transfer incur separate costs, and transfer by distillation may itself involve optimization. The amortization claim applies to the learned attribute-prediction stage.

D Probe Perturbation

We investigate the role of the object-centered scene appearance probe (Fig. 6) by perturbing the environment map used for its construction while keeping the observed object appearance fixed. The probe constructed under the original illumination yields accurate material predictions, whereas

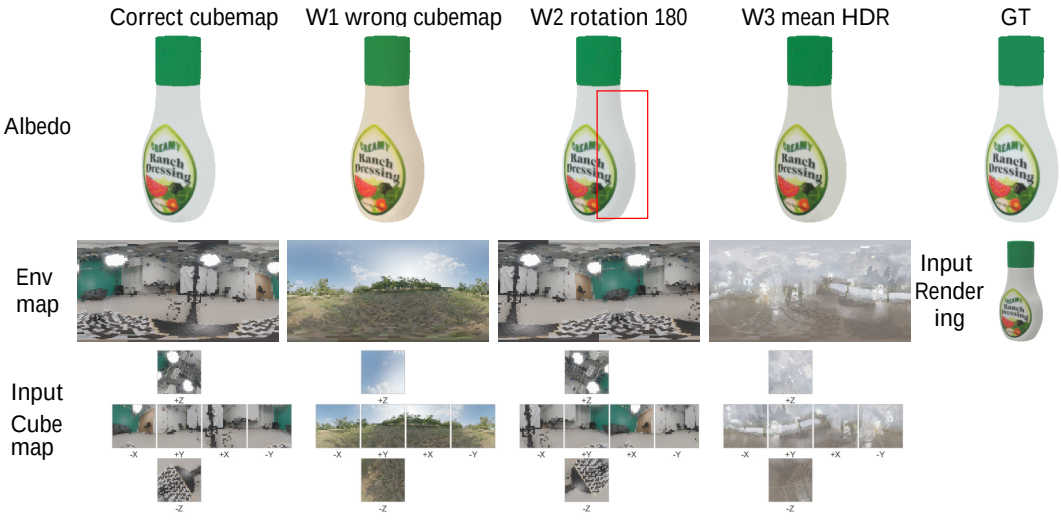


Figure 6: **Probe perturbation.** The observed object appearance (input rendering, right) is kept fixed while the environment map used to construct the probe is changed: the correct environment, an unrelated environment (W1), the correct environment rotated by 180° (W2), and its spatially averaged HDR (W3). Rows show the recovered albedo, the environment map, and the resulting probe cubemap.

Table 3: **Provided source illumination versus predicted illumination input.** Every cell reports PSNR ↑ / SSIM (%) ↑ / LPIPS (%) ↓. Provided Source Illumination: the source illumination is given to the method (for GS-PI++, as the far field of its probe); Predicted: the method estimates the illumination or, like GS-PI, does not use it. '/' marks outputs that a method does not produce or settings that do not apply; † is as in Table 1.

	Provided Source Illumination				**Predicted Illumination**			
Method	Albedo	Roughness	Metallic	Relight	Albedo	Roughness	Metallic	Relight
IRGS	24.06 / 82.51 / 2.80	13.00 / 65.76 / 5.51	/	23.76 / **84.25** / **2.08**	**19.88** / **80.48** / **2.98**	11.15 / **61.68** / 5.20	/	**19.15** / **80.61** / **2.54**
Relightable3DGaussian	21.74 / 74.17 / 3.46	13.46 / 30.15 / 6.96	/	21.35 / 76.61 / 2.85	13.24 / 69.11 / 4.84	14.86 / 32.04 / 6.57	/	14.30 / 73.17 / 4.24
LumiGauss†	13.74 / 66.87 / 4.14	/	/	/	14.42 / 65.76 / 4.91	/	/	/
GaussianShader†	7.35 / 11.50 / 8.07	9.11 / 43.90 / 6.95	/	/	9.05 / 33.00 / 6.73	13.75 / 41.06 / 5.92	/	/
GS-IR	21.89 / 76.07 / 3.59	9.23 / 56.29 / 6.63	29.65 / 71.34 / 7.76	20.73 / 76.17 / 2.99	13.10 / 65.28 / 4.24	9.70 / 51.45 / 6.98	**27.24** / **56.98** / **8.42**	14.13 / 62.98 / 4.48
GS-PI	/	/	/	/	18.08 / 73.73 / 3.89	**16.67** / **61.68** / **4.42**	22.95 / 18.90 / 10.16	17.34 / 64.77 / 4.09
GS-PI++ (ours)	**29.24** / **88.85** / **2.03**	**23.46** / **81.18** / **3.21**	**44.61** / **93.19** / **1.72**	**24.92** / 84.08 / 2.32	/	/	/	/

substituting an unrelated environment map produces pronounced color shifts in the recovered reflectance. Similar deviations arise when the original environment map is rotated or spatially averaged before probe construction. Rotation preserves global color statistics but alters the directional alignment of illumination, while averaging removes angular variation. The resulting degradation suggests that material recovery depends on both the photometric compatibility and the directional structure of the supplied context. These observations support the role of the light probe in resolving material–illumination ambiguity: consistent contextual evidence helps attribute observed appearance to illumination, whereas mismatched or simplified evidence can cause illumination effects to be absorbed into the predicted reflectance.

E Provided source illumination versus Predicted Illumination

Table 3 compares the methods when the source illumination is given to them (provided source illumination) and when it is estimated or unused (Predicted). GS-PI++ always receives the source environment map through its probe and therefore appears only in the Fixed block, whereas GS-PI does not take illumination as input and appears only in the Predicted block.

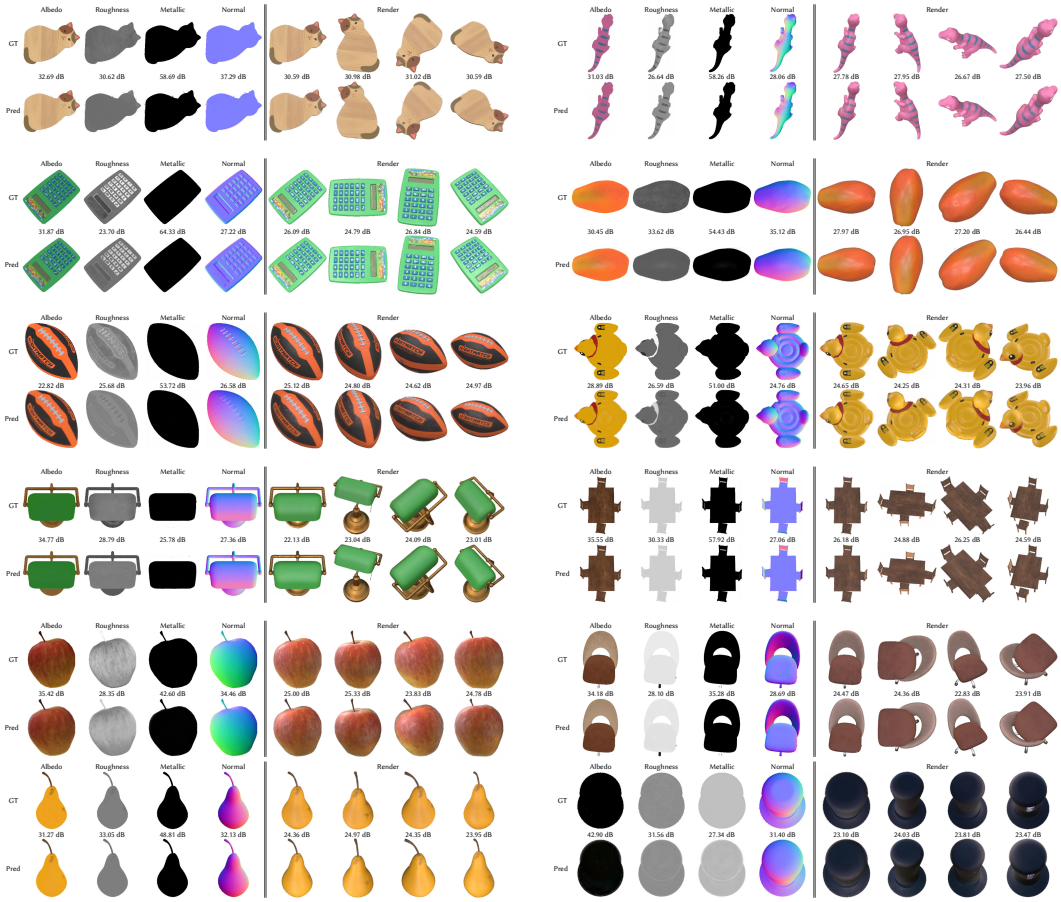


Figure 7: **Additional qualitative results.** For each object, ground truth (GT, top) and our prediction (bottom) are shown for albedo, roughness, metallic, normal, and four rendered views; the number under each map is the PSNR of the prediction with respect to GT.

F Additional Qualitative Results

Fig. 7 shows further predictions of GS-PI++ with per-map PSNR.