Intrinsic PAPR: Tackling Misattribution in 3D Intrinsic Decomposition via Proximity Attention Point Rendering

Alireza Moazeni[1]*, Shichong Peng[1], Yanshu Zhang[1], Chirag Vashist[1], and Ke Li[1,2,3]

[1] APEX Lab, School of Computing Science, Simon Fraser University, Canada
[2] Alberta Machine Intelligence Institute (Amii), Canada
[3] Canadian Institute for Advanced Research (CIFAR), Canada
{sam62, shichong_peng, yanshu_zhang, chirag_vashist, keli}@sfu.ca

Abstract. Recent point-based intrinsic decomposition and inverse rendering methods have advanced the modelling of the shading and albedo of 3D scenes. However, we identify a fundamental limitation: these methods suffer from a *misattribution issue*, where individual primitives learn incorrect appearance features despite producing correct aggregated renderings. We show that the root cause lies in volume rendering, which aggregates translucent primitives along each ray and only supervises the final colour, preventing direct supervision of individual primitive features. To address this, we propose *Intrinsic PAPR*, a robust intrinsic decomposition framework which leverages Proximity Attention Point Rendering (PAPR) to enable direct per-point supervision. Unlike volume rendering approaches, PAPR eliminates translucent primitives and directly predicts appearance at ray-surface intersections, enabling accurate supervision to the feature of each individual point. Our method incorporates a 2D albedo prior adapted with conditional Implicit Maximum Likelihood Estimation (cIMLE) to handle monocular ambiguities, and employs a space carving loss to ensure multi-view consistency. Extensive evaluations on synthetic and real-world datasets demonstrate that Intrinsic PAPR outperforms point-based inverse rendering, NeRF-based intrinsic decomposition, and diffusion-based PBR methods in novel view synthesis and albedo estimation while resolving the misattribution issue.

Keywords: Neural Rendering · Intrinsic Decomposition · Inverse Rendering · Point-Based Rendering

1 Introduction

Recent advances in point-based neural rendering, such as 3D Gaussian Splatting [23], have achieved impressive quality in novel view synthesis. However, these methods typically encode only the final emitted colour per primitive, limiting editing applications that require modifying scene properties like material

* Corresponding author.

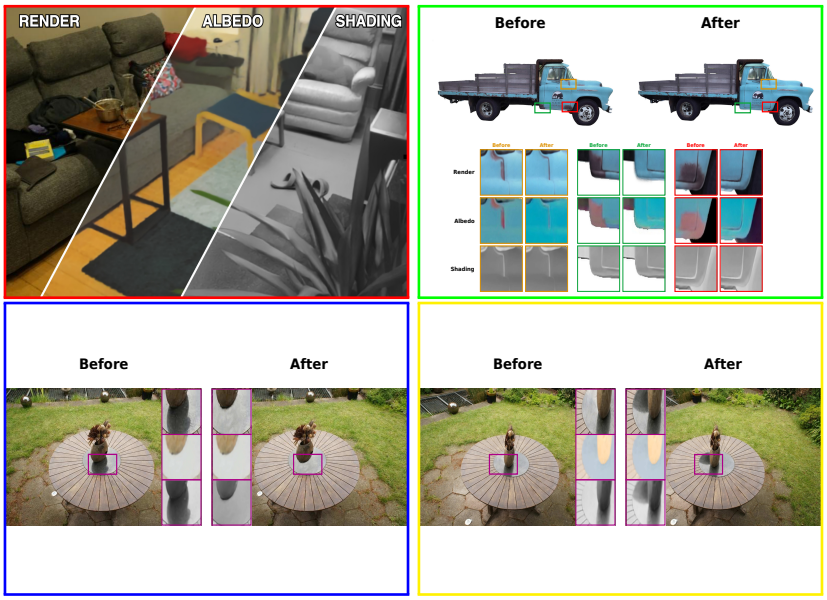

Fig. 1: We present ***Intrinsic PAPR***, which tackles misattribution in 3D intrinsic decomposition and enables robust editing on real-world and synthetic scenes. **(Top-Left)** Our method accurately decomposes a 3D scene into multi-view consistent albedo and shading components. It supports **(Top-Right)** freeform albedo editing (e.g., rust removal on a truck) and allows geometry-aware shading manipulation, seamlessly removing **(Bottom-Left)** or adding **(Bottom-Right)** shadows in user-defined 3D regions while preserving the underlying scene structure. Each editing panel shows the large *before* and *after* images with coloured region boxes, and per-region *before*/*after* render, albedo and shading crops.

albedo or shading. To enable such control, recent works have explored intrinsic decomposition [57] and inverse rendering [31], which learn to represent distinct material and lighting components.

Surprisingly, applying edits to these decomposed models often yields unexpected and inconsistent results. As shown in Figure 2a, if we copy the learnt albedo features from a source region to a target, the rendered target albedo fails to match the source. This indicates that the underlying primitives have learned incorrect albedo features, a phenomenon we term the **misattribution issue**, a critical barrier to creating robust and predictable 3D editing tools.

Our first key contribution is identifying the root cause of misattribution: the aggregation of semi-transparent primitives in volume rendering. In splat-based methods [23, 31], primitives are constrained to ellipsoid shapes, so they must be made translucent and overlaid in various orientations to yield near-opaque, non-ellipsoid regions. Volume rendering then composites the features of all primitives along a ray. This allows ambiguity: a set of primitives can learn incorrect individual features that, when aggregated, still produce the correct final pixel

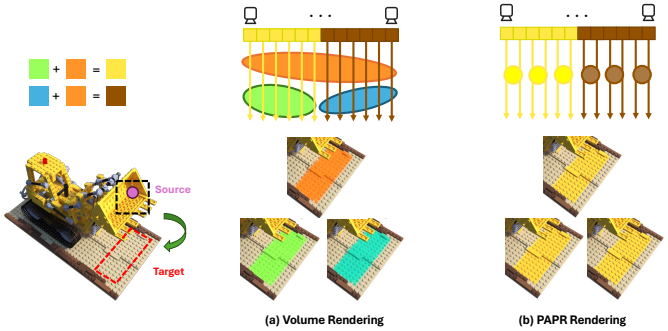

Fig. 2: The misattribution issue in point-based rendering. (a) When using volume rendering with translucent primitives, individual primitives can learn incorrect features while still producing plausible results. (b) Our approach using PAPR ensures accurate per-point feature learning by directly supervising the features at each point.

colour. Since no single primitive is directly supervised by the output pixel, there is no guarantee that individual features are correct. Notably, this ambiguity is not cured by capturing more views—near-coincident primitives render identically from every direction—a prediction we confirm empirically in Sec. 4.

Our second key contribution is the insight that this issue can be resolved by fundamentally changing the rendering paradigm. We propose repurposing Proximity Attention Point Rendering (PAPR) [66], which avoids the pitfalls of volumetric aggregation. Unlike splat-based primitives, PAPR's primitives are points without spatial extent; instead, it forms an opaque surface by connecting nearby points with a learned interpolation kernel, which takes the form of an attention mechanism, thereby eliminating the need for translucency. This approach allows PAPR to predict appearance directly at ray-surface intersections, rather than aggregating features along a ray. Furthermore, because PAPR learns a parsimonious point cloud, each point influences the rendered colour of at least one pixel, so its features are more directly supervised by the image, preventing incorrect, compensatory features and resolving the misattribution issue (Figure 2b).

Building on this insight, we introduce **Intrinsic PAPR** (Figure 1), a novel framework for robust intrinsic decomposition and editable 3D appearance. Our method learns per-point albedo and shading features suitable for high-quality reflectance and illumination editing. Our third contribution is a novel technique to supervise this decomposition robustly. We use a pre-trained 2D intrinsic decomposition model [4] as a prior, but a single image-to-albedo prediction is often ambiguous. We address this by adapting the 2D prior into a probabilistic model with conditional Implicit Maximum Likelihood Estimation (cIMLE) [29,41], generating a distribution of plausible albedos per view. We use a mode selection loss akin to the space carving loss from [49] to distill these multi-view distributions into a single, geometrically consistent 3D albedo representation.

While state-of-the-art inverse rendering methods increasingly rely on complex physically-based rendering (PBR) models to estimate SVBRDF parame-

ters, these formulations are highly ill-posed and often fail to reliably estimate editable parameters in casual multi-view captures. We deliberately adopt a simpler albedo-shading intrinsic decomposition, avoiding the under-constrained nature of full physical inverse rendering. A richer decomposition can represent more complex effects but cannot reliably recover them from observations; by the bias-variance trade-off, for editing it is better to recover a few parameters accurately than many inaccurately. Our decomposition acts as an effective geometric regularizer, forcing the model to explain illumination effects through accurate geometry rather than baking them into the albedo. This not only improves decomposition but also leads to higher-quality novel view synthesis than the original PAPR.

We evaluate Intrinsic PAPR against state-of-the-art methods on synthetic and real-world datasets, demonstrating superior performance in both novel view synthesis and albedo estimation, with a 10.3% and 8.0% PSNR improvement, respectively, over the best inverse-rendering baseline. Quantitative feature-transfer experiments further confirm that our method mitigates the misattribution issue, learning more accurate and consistent per-primitive features and achieving more than a 4× reduction in albedo- and shading-transfer error over a leading 3DGS-based method.

2 Related Work

2.1 Neural Scene Representations

In recent years, there has been a surge in interest in using neural networks for novel view synthesis from multi-view RGB images [6, 17, 23, 34, 35, 59, 66]. Here, we discuss two widely-used scene representations relevant to our setting: NeRF-based and point-based representations.

NeRF [34] achieves impressive rendering quality but is computationally costly, evaluating multiple samples per ray. Precise point-level editing in NeRF is also non-trivial, requiring changes to scene properties at infinitely many points within a volume and careful specification of that volume's boundary.

In contrast, point-based representations render with fewer samples and gain increasing attention for their high rendering quality and efficiency [23, 26, 40, 63, 66, 73]. Because they store local scene properties at each point, precise point-level edits simply modify the properties stored at points within the desired region. Among these, splat-based techniques [23, 26, 63, 73] use radial basis functions to compute point contributions to rays and volume rendering to combine the per-point scene properties into the output. Another line of works uses attention-based model to interpolate properties stored at neighbouring surface points for rendering [40, 66]. In particular, Proximity Attention Point Rendering (PAPR) [66] stands out for its ability to learn a parsimonious point cloud from scratch. In this work, we adopt PAPR as our scene representation and rendering method.

2.2 Intrinsic Decomposition

The problem of image intrinsic decomposition aims to decompose images into reflectance and shading components. Early methods focused on predicting ordinal relationships between the albedos of pixel pairs [37, 69, 72], while later methods directly regress to continuous shading and albedo values [12, 30, 33, 45, 68, 70]. Recent single-image intrinsic methods also leverage diffusion priors for material/intrinsic estimation [25]. For evaluation, the IIW dataset and WHDR metric [2] remain standard for assessing reflectance ordering and can complement our multi-view albedo consistency analysis.

Whereas these methods consider 2D inputs, PoInt-Net [57] performs intrinsic decomposition on RGB-D. A major challenge is generalization to real imagery; Careaga and Aksoy first propose an ordinal-shading model [5] and then a colourful diffuse intrinsic decomposition [4], combining synthetic pretraining with multi-illumination data to generate dense pseudo-ground-truth intrinsic components; we use the latest version as our 2D prior. Multi-view intrinsic transformers such as IDT [13] instead enforce consistency directly in the 2D image domain; in contrast, we distill ambiguity-aware 2D priors into a unified 3D point-based representation.

For 3D settings, intrinsic decomposition of neural scenes [44, 55, 61, 65] estimates albedo and shading jointly with geometry. IntrinsicNeRF [61] uses iterative reflectance clustering; LitNeRF [44] targets faces with few views. In the Gaussian Splatting family, Intrinsic-GS [58] performs multi-view intrinsic decomposition using colour-/shading-invariant priors to stabilize albedo across views. However, NeRF-/GS-based volume aggregation often encodes scenes in networks or splats, making localized edits challenging and susceptible to misattribution. Other works [28, 42, 54, 56] enable colour edits without explicit shading, which can lead to visually inconsistent edits under illumination changes.

A common limitation of these methods is that they either operate in 2D (lacking multi-view geometric consistency) or, when extended to 3D via NeRF or GS, remain susceptible to misattribution from volumetric feature aggregation. Our method bridges this gap with a point-attention framework (PAPR) that stores albedo and shading per point and aggregates them without volumetric blending. We supervise with a 2D intrinsic prior adapted via cIMLE and distill it across views into a single 3D representation, yielding multi-view consistent albedo and shading.

2.3 Inverse Rendering

Inverse rendering recovers geometry, materials, and illumination from images. NeRF-based approaches [7,22,46,50,67] model view-dependent reflectance, sometimes incorporating diffusion priors, while explicit/hybrid methods such as nvdiffrec [36] and Mitsuba 2 [38] optimize meshes, SVBRDFs, and lighting via differentiable path tracing. NeILF/NeILF++ [60, 62] improve disentanglement by explicitly modelling incident light fields.

In point-/Gaussian-based IR, several methods operate in 3DGS/point primitives, including GS-IR [31], DPIR [10], and GaussianShader [21]. Recent works emphasize robust lighting (VMINer [16]), glossy materials [51], and relightable Gaussians with BRDFs [18]. Building on 3DGS, IRGS [20] introduces a differentiable 2D Gaussian ray tracer for full inter-reflections under the complete rendering equation; DiscretizedSDF [71] encodes a discretized signed distance field inside each Gaussian for relightable assets; SVG-IR [47] equips Gaussians with spatially varying material and normal attributes. GAINS [39] leverages intrinsic-image and diffusion priors to stabilize inverse rendering from sparse multi-view captures, and GS-ID [14] focuses on illumination decomposition on Gaussians with diffusion-guided material priors.

Domain-specific extensions include physically-based human IR [53], robust shape/illumination recovery [15], single-shot material estimation [48], and controllable GS editing without explicit decomposition [8, 52].

While these IR approaches enable component-wise editing, the estimation problem is highly ill-posed and typically relies on strong priors, controlled capture (e.g., OLAT), or restrictive assumptions that limit robustness in casual multi-view capture [3]. In contrast, we deliberately target a simpler intrinsic factorization (per-point albedo and shading) with a surface-attentional renderer (PAPR), trading physical completeness for robustness and editability; this rendering shift is not merely a design preference but a structural prerequisite for resolving misattribution. We leverage an ambiguity-aware 2D intrinsic prior distilled into a multi-view-consistent 3D representation that favours stable point-level edits with predictable behaviour across views.

3 Intrinsic PAPR

3.1 Representation and Rendering

We build on Proximity Attention Point Rendering (PAPR) [66]. To enable intrinsic decomposition, we augment PAPR's per-point features by partitioning them into two orthogonal components: an albedo component capturing reflectance and a shading component capturing illumination. Each point thus stores separate albedo and shading information while retaining local scene detail, which is essential for downstream editing. Figure 3 presents an overview of our rendering pipeline.

Specifically, our representation is denoted as $P = \{(\mathbf{p}_i, \mathbf{a}_i, \mathbf{h}_i, \tau_i)\}_{i=1}^{N}$, where $\mathbf{a}_i \in \mathbb{R}^n$ represents the albedo feature vector and $\mathbf{h}_i \in \mathbb{R}^m$ represents the shading feature vector for each point. These two feature vectors serve as inputs to their corresponding value branches in the attention mechanism to produce an albedo feature map $A_{\mathcal{C}} \in \mathbb{R}^{H \times W \times d_{\text{albedo}}}$ and a shading feature map $S_{\mathcal{C}} \in \mathbb{R}^{H \times W \times d_{\text{shading}}}$.

Given a camera $\mathcal{C}$ and pixel (i, j) with ray r_{ij}, we follow PAPR to select a small top-K set of points $\mathcal{N}(r_{ij})$ nearest to the ray, compute proximity-based attention scores, and aggregate features to obtain per-pixel albedo and shading

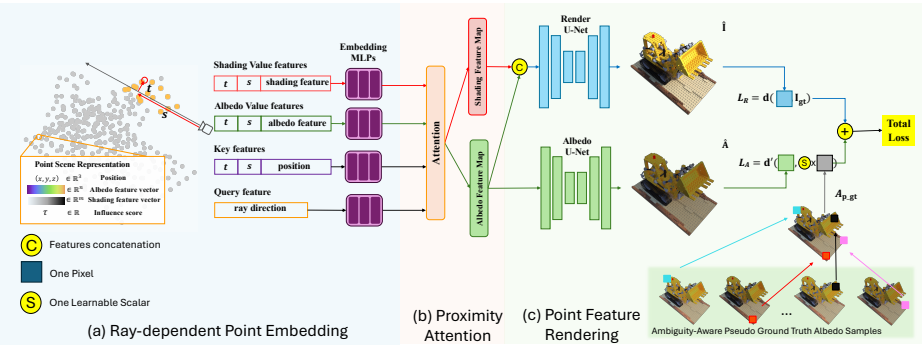

Fig. 3: Rendering pipeline overview. Each scene point stores position, albedo, shading features, and an influence score. (a) Ray-dependent embeddings combine values (albedo, shading), keys (position, t, s), and queries (ray directions) as inputs to the attention model. (b) The attention model selects and aggregates albedo and shading into feature maps. (c) The albedo feature map is rendered (green) to produce the albedo image; both albedo and shading feature maps are concatenated (blue) to generate the final colour image. The model is trained end-to-end with supervision on both albedo and colour outputs.

features:

$$s_k^{ij} = f_\theta\big(\mathbf{p}_k,\, r_{ij},\, \tau_k\big), \quad w_k^{ij} = \mathrm{softmax}_k\Big(s_k^{ij}\Big),$$
$$A_{\mathcal{C}}^{ij} = \sum\nolimits_{k\in\mathcal{N}(r_{ij})} w_k^{ij}\,\mathbf{a}_k, \quad S_{\mathcal{C}}^{ij} = \sum\nolimits_{k\in\mathcal{N}(r_{ij})} w_k^{ij}\,\mathbf{h}_k. \tag{1}$$

Here f_θ is PAPR's learnable scoring function, mapping the point position $\mathbf{p}_k$ and its perpendicular and along-ray distances to the ray r_{ij} to a query–key attention score on positional encodings, re-weighted by the per-point influence τ_k [66]; $\mathcal{N}(r_{ij})$ is the top-K set selected by perpendicular point–ray distance, and the weights are softmax-normalized across it.

Under proximity attention, the score increases as the perpendicular point–ray distance shrinks. Consequently, each surface point admits a non-trivial cone of rays for which its normalized weight dominates ($w_k^{ij} \approx 1$). On those rays, the photometric residual backpropagates primarily to that point, making the features of every point identifiable. In contrast, volumetric splatting distributes transmittance multiplicatively along the ray: supervision constrains only the integrated colour, so the feature of each individual primitive can be incorrect as long as the aggregated colour is correct, since the error in one primitive can compensate for the errors of others and cancel out. The geometric locality of PAPR, together with a parsimonious point set, suppresses such compensations and thereby mitigates misattribution (see Figure 2).

Following the intrinsic factorization $\mathbf{I} \approx \mathbf{A} \odot \mathbf{S}$, we adopt a minimalist supervision strategy: supervise RGB and albedo, and learn shading implicitly. This is sufficient by construction: under the factorization, any two of $(\mathbf{I}, \mathbf{A}, \mathbf{S})$ determine the third, so constraining $\mathbf{I}$ and $\mathbf{A}$ fully determines $\mathbf{S}$ without an additional

shading loss. Moreover, the rendering loss enforces multi-view RGB consistency and the space carving loss enforces multi-view albedo consistency, so the implicit shading is forced to be stable and view-consistent rather than an unconstrained parameter absorbing residual colours. Direct shading supervision would moreover be ill-posed in regions where $\mathbf{A} \approx \mathbf{0}$, because many values of $\mathbf{S}$ yield the same product. To supervise the rendered image, we reuse PAPR's reconstruction objective (MSE+LPIPS) given by:

$$\mathcal{L}_{\text{recon}} = \alpha\,\text{MSE}\Big(\hat{\mathbf{I}}, \mathbf{I}_{gt}\Big) + \beta\,\text{LPIPS}\Big(\hat{\mathbf{I}}, \mathbf{I}_{gt}\Big)\,. \tag{2}$$

To supervise the decomposed reflectance, we augment this objective with a scale-invariant albedo term, introduced next.

3.2 Scale-Invariant Albedo Loss

In image intrinsic decomposition, the albedo is defined only up to a multiplicative scale: a decomposition into albedo A and shading S is as valid as one into γA and S/γ for any positive γ. To make our albedo loss scale-invariant, we introduce a learnable scalar λ per training view. Before computing the loss between the predicted albedo $\hat{\mathbf{A}}$ and the pseudo-ground-truth albedo $\mathbf{A}_{p-\text{gt}}$, we scale $\mathbf{A}_{p-\text{gt}}$ by its λ. Jointly optimizing these scalars with the model parameters aligns the pseudo-ground-truth and predicted scales, ensuring scale-invariant albedo across views.

3.3 Ambiguity-aware Albedo Estimates

In single-view 2D image intrinsic decomposition, the absence of geometric information introduces inherent ambiguities in determining albedo; different reflectance properties can explain the same observed image, leading to multiple plausible albedo interpretations. For instance, the dotted pattern on the baseboard of a Lego bulldozer could be shadows cast by raised dots, or a flat board with a dotted texture.

Since the prior model produces a single prediction per input image, predictions across views may be inconsistent due to per-view ambiguity. A view-consistent solution instead requires a range of plausible predictions per view. Inspired by the ambiguity-aware depth estimation of [49], we adapt the prior model [4] to be ambiguity-aware by training it with conditional Implicit Maximum Likelihood Estimation (cIMLE), avoiding GAN mode collapse [19]. Concretely, we sample M latent codes per view, injecting each via AdaIN to yield M albedo hypotheses; see Appendix for details.

Here the modes represent the plausible albedo configurations consistent with the input image. Since the training dataset provides only one observed albedo per input, the samples for a given input share the scale factor of Sec. 3.2, isolating the effect of scale and input ambiguity on the prediction variations. Further details on ambiguity-aware prior training are in the supplement.

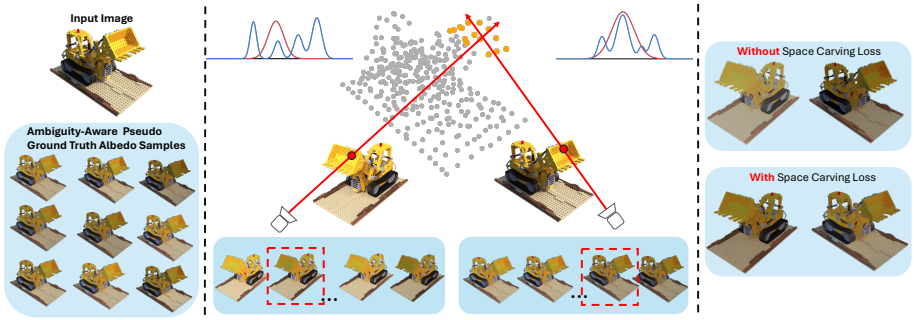

Fig. 4: Illustration of Space Carving (SC) Loss. **Left**: Multiple pseudo-ground-truth albedo samples per view generated by our ambiguity-aware 2D prior model capture estimation ambiguities. **Middle**: SCL selects consistent albedo modes along rays across views to resolve these ambiguities. **Right**: Predicted albedo maps exhibit improved cross-view consistency for 3D rendering.

3.4 Albedo Fusion via Space Carving Loss

To distill the view-wise albedo hypotheses into a single, globally consistent reflectance, we formulate a probabilistic space carving for reflectance, illustrated in Figure 4, in the spirit of ambiguity-aware space carving for depth [49]. For each view v and pixel (i, j), the cIMLE prior provides a finite set of albedo hypotheses $\{p^{ij}_{v,k}\}$. We define the fused albedo $\hat{p}^{ij}$ to be the mode that is jointly consistent across views up to the per-view scale indeterminacy of albedo, i.e., we select one hypothesis per view and minimize $\sum_v \|\hat{p}^{ij} - \lambda_v\, p^{ij}_{v,k_v}\|_2^2$, with λ_v learned to absorb scale. This induces the sample-based objective below, which is the reflectance analogue of SCADE's mode selection loss and supervises modes rather than moments:

$$\mathcal{L}_{\text{SC}} = \sum_{(i,j)\in\Omega} \min_{k\in[M]} \left\| \hat{p}^{ij} - p^{*ij}_{k} \right\|_2^2, \tag{3}$$

where $\hat{p}^{ij}$ is the predicted albedo at pixel (i, j), p^{*ij}_k represents the k-th scaled albedo hypothesis with λ from the ambiguity-aware prior, Ω denotes all pixel positions, and M is the number of albedo samples per image.

Although the objective performs per-pixel selection among hypotheses, the continuity of the albedo decoder preserves local smoothness in practice without requiring additional regularizers.

The total loss to optimize and train our model is the sum of reconstruction loss and weighted space carving loss presented in Eq. 4:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + w_{SC}\mathcal{L}_{\text{SC}} \tag{4}$$

Implementation Details. We provide full implementation and training details, including all hyperparameters, in the supplement.

4 Experiments

Evaluation Setup. We evaluate our method on NeRF Synthetic [34] and TensoIR [22] (synthetic) datasets, as well as Tanks & Temples [24] and Mip-NeRF 360 [1] (real-world). To quantify performance across novel view synthesis (NVS) and albedo reconstruction, we report PSNR, SSIM, and LPIPS [64]. Following prior work (e.g., NeRFactor, GS-IR), we apply per-view scale-and-shift alignment before computing albedo metrics.

Baselines. We compare against state-of-the-art baselines spanning volumetric versus surface representations and physically-based versus intrinsic models: point- and Gaussian-based inverse rendering methods (DPIR [10], GS-IR [31], IRGS [20], DiscretizedSDF [71]), a diffusion-prior PBR approach (MaterialFusion [32]), and a NeRF-based method (Intrinsic-NeRF [61]). We use official implementations and default configurations for all baselines, plus vanilla PAPR [66] as a pure reconstruction baseline.

Results. As reported in Table 1, Intrinsic PAPR delivers higher-fidelity NVS and more accurate albedo reconstructions than all inverse-rendering baselines across both synthetic and real-world scenes; Figure 5 shows this qualitatively on NeRF Synthetic. The supplement provides per-scene breakdowns and also shows that simply adding 2D prior supervision to baselines does not improve their albedo quality, highlighting the role of PAPR's representation. For positioning against scene-level intrinsic decomposition, MAIR++ [9] on our nine synthetic scenes reaches NVS / albedo PSNR of 24.42 / 22.28 dB, trailing our 35.00 / 31.21 dB, a gap due to domain mismatch: MAIR++ targets indoor SVBRDF estimation and excludes object-centric scenes.

Our framework also surpasses vanilla PAPR [66] in NVS, by supervising albedo and final RGB separately. Explicit albedo supervision resolves texture-illumination ambiguities, forcing the model to explain image formation via a more accurate shading component; since shading is tightly coupled to geometry, this disentanglement yields cleaner geometry and thus higher-quality NVS. Further qualitative NVS comparisons against vanilla PAPR and depth visualizations are in the supplement.

4.1 Albedo and Shading Feature Quality Evaluation

To evaluate how our design addresses the misattribution issue, we compare against our point-based inverse rendering baselines, DPIR [10] and GS-IR [31], on per-point intrinsic transfer tasks (albedo and shading).

Protocol. For both albedo and shading transfers, the edited target should match the source primitive's attribute. We render the edited scene for each method, decompose into albedo/shading with [27], project source/target 3D points to image pixels, and compute the transfer error: $\mathrm{TE} = \frac{1}{N}\sum_{i=1}^{N}\|\hat{A}_E(p_t^{(i)}) - A_{\mathrm{GT}}(p_s^{(i)})\|_2^2$, where $A \in \{\text{albedo}, \text{shading}\}$. For synthetic scenes, A_{GT} at p_s is read from Blender. Source points are sampled uniformly at random from the source mask; we report mean ± std over N=100 trials. When used, the decoupling error measures edit locality: $\mathrm{DE} = \frac{1}{|R|}\sum_{p\in R}\|\hat{A}_{\mathrm{post}}(p) - \hat{A}_{\mathrm{pre}}(p)\|_2^2$ in

Table 1: Novel-view synthesis and albedo reconstruction. Left: NVS on synthetic datasets (NeRF Synthetic [34], TensoIR [22]) and real-world datasets (Tanks & Temples [24], Mip-NeRF 360 [1]). Right: Albedo on NeRF Synthetic vs. Blender ground truth. Intrinsic PAPR achieves the best performance across all metrics and datasets.

Novel View Synthesis (Image Quality Metrics)

Method	Synthetic			Real-world		
	PSNR ↑	*SSIM* ↑	$LPIPS_{vgg}$ ↓	*PSNR* ↑	*SSIM* ↑	$LPIPS_{vgg}$ ↓
DPIR [10]	26.47	0.867	0.133	20.17	0.756	0.290
MaterialFusion [32]	28.59	0.902	0.103	25.18	0.801	0.237
Intrinsic-NeRF [61]	29.51	0.936	0.051	23.79	0.768	0.252
GS-IR [31]	30.40	0.945	0.055	26.44	0.851	0.185
IRGS [20]	31.35	0.949	0.048	27.26	0.878	0.157
DiscretizedSDF [71]	31.72	0.959	0.038	27.86	0.876	0.147
PAPR [66]	32.84	0.973	0.035	28.03	0.865	0.159
Intrinsic PAPR (Ours)	**35.00**	**0.978**	**0.021**	**29.99**	**0.912**	**0.094**

Albedo Reconstruction (Synthetic)

Method	*PSNR* ↑	*SSIM* ↑	$LPIPS_{vgg}$ ↓
DPIR [10]	22.89	0.854	0.215
Intrinsic-NeRF [61]	24.86	0.880	0.108
MaterialFusion [32]	25.36	0.899	0.097
GS-IR [31]	26.33	0.877	0.127
IRGS [20]	28.44	0.913	0.095
DiscretizedSDF [71]	28.89	0.921	0.097
Intrinsic PAPR (Ours)	**31.21**	**0.949**	**0.056**

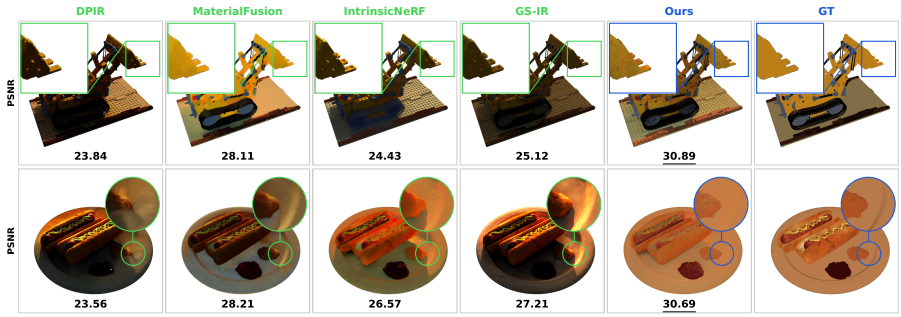

Fig. 5: Qualitative albedo comparison on the NeRF Synthetic dataset [34]. Per-view albedo PSNR is shown per method; Intrinsic PAPR achieves the best average. Ground truth is from Blender. To visualize the albedo maps, we follow the approaches of NeRFactor [65] and GS-IR [31], computing a scaling factor for each RGB channel.

regions R that exclude the edited area. Masks are defined with our UI and shared across methods; per-scene visualizations appear in the supplement.

Albedo Transfer Albedo transfer isolates primitive-level reflectance attribution: the desired outcome is that the target reproduces the source albedo independent of shading. Volumetric baselines, whose features are supervised only after transmittance-weighted aggregation, admit compensatory errors and manifest as colour drift and high variability across transfers (Figure 6, left). In contrast, our surface point renderer assigns high attention to the source primitive along some rays, providing direct per-primitive supervision. The transferred albedo closely matches the source with markedly lower variance (Table 2), evidencing correct attribution rather than aggregation compensation. This advantage is structural, not a matter of training data: sweeping training views from 25 to 200 keeps our per-primitive transfer error 5.04–5.23× below GS-IR at every view count, and at 25 views we already undercut GS-IR's 200-view result (0.078 vs. 0.314). More views cannot help, because near-coincident primitives

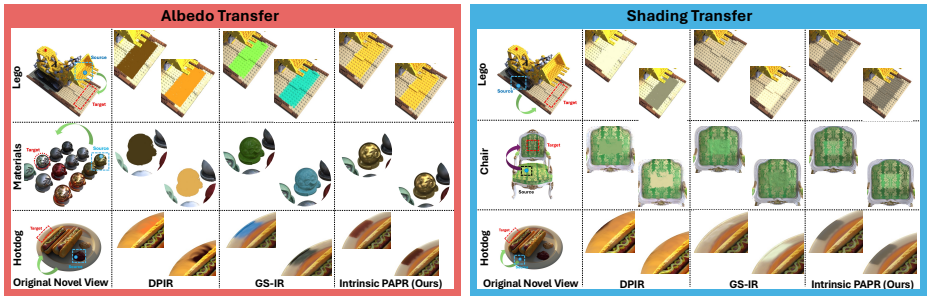

Fig. 6: Per-point intrinsic transfer on NeRF-Synthetic. **Left:** Albedo transfer—our method preserves primitive-level reflectance across random source points, while volumetric baselines show colour drift and high variability. **Right:** Shading transfer—our method produces geometry-consistent illumination tied to the source primitive, whereas DPIR and GS-IR show inconsistent shading and artefacts from mixed attributions.

Table 2: Per-point intrinsic attribution via transfer. Mean and standard deviation of MSE over 100 random transfers per scene, for albedo (left) and shading (right). Our method achieves substantially lower error and variance than volumetric baselines, indicating that per-point albedo/shading features are correctly learned.

Albedo Transfer (MSE: Avg ↓ / STD ↓)						
	Lego		Materials		Hotdog	
Method	Avg	STD	Avg	STD	Avg	STD
DPIR [10]	0.236	0.179	0.221	0.106	0.291	0.135
GS-IR [31]	0.381	0.219	0.318	0.175	0.280	0.141
Ours	**0.053**	**0.016**	**0.067**	**0.026**	**0.071**	**0.033**

Shading Transfer (MSE: Avg ↓ / STD ↓)						
	Lego		Chair		Hotdog	
Method	Avg	STD	Avg	STD	Avg	STD
DPIR [10]	0.328	0.223	0.425	0.135	0.482	0.119
GS-IR [31]	0.414	0.157	0.319	0.101	0.540	0.187
Ours	**0.074**	**0.029**	**0.059**	**0.022**	**0.041**	**0.033**

render identically from every direction; no amount of multi-view supervision can disambiguate features that volumetric aggregation has rendered observationally equivalent (per-view-count breakdown in the supplement).

Shading Transfer Shading transfer probes whether illumination effects are correctly bound to individual primitives and local geometry. DPIR and GS-IR exhibit inconsistencies and artefacts indicative of mixed attributions along the ray, whereas our method produces geometry-consistent shading that tracks the source primitive (Figure 6, right). This successful transfer proves our indirectly supervised shading branch captures genuine illumination rather than absorbing residual errors, which would otherwise yield arbitrary artefacts upon transfer. Quantitatively, the error and variance are lowest (Table 2), confirming precise per-primitive shading attribution.

4.2 Freeform Intrinsic Editing

Precise per-primitive albedo and shading attribution enables practical, freeform editing tools. Users paint arbitrary regions on an input view with our UI (details

in the supplement); we map these pixels to the underlying 3D points and modify their albedo or shading features, so edits live in the 3D representation and remain multi-view consistent, geometry-aware, and localized.

Figure 7 (a) illustrates local albedo and shading edits in user-drawn regions, such as writing letters or drawing circular patterns on surfaces. The modified appearance follows the surface across views and blends smoothly with unedited areas, providing intuitive "paint-like" control of material and shading.

Building on this, we use the same per-primitive shading features to control illumination strength. In Figure 7 (b), we edit shading intensity by scaling the shading feature vectors of selected primitives before transferring them to a target region, yielding predictable changes from subtle brightening to strong darkening. Applying a uniform scale to the shading features of all primitives (Figure 7 (c)) produces coherent scene-level shading adjustments while leaving geometry and albedo unchanged.

Finally, our representation supports more structured edits. Figure 7 (d) demonstrates geometry-aware shadow addition and removal: users select a 2D region, and we adjust shading magnitudes only for primitives projecting into that region, producing shadows that align with object boundaries and occlusions. On the albedo side, disentanglement from shading allows us to synthesize colours not present in the original scene (Figure 7 (e)) by forming convex combinations of existing albedo features; we expose these combinations as a continuous palette in the UI, enabling recolouring with unseen yet realistic hues.

5 Ablation Study

We ablate Intrinsic PAPR along two axes: the contribution of each component, and the choice of decomposition scope. Throughout, we report novel-view-synthesis (NVS) PSNR, albedo PSNR, and cross-view albedo consistency, the last measured by the **mean albedo consistency error (MACE)**: we track K surface points over T consecutive frames of a rendered camera trajectory and measure the drift in their rendered albedo, $\text{MACE} = \frac{1}{K(T-1)} \sum_{i=1}^{K} \sum_{j=1}^{T-1} \|\hat{\mathbf{x}}_{ij} - \hat{\mathbf{x}}_{i,j+1}\|_2^2$, where $\hat{\mathbf{x}}_{ij}$ is the rendered albedo for point i in frame j (lower is better).

Per-component Ablation. Starting from the vanilla PAPR backbone, we add one component at a time—the albedo/shading split with the 2D prior, the cIMLE-adapted prior, the space-carving loss (Sec. 3), and the learnable per-view scalar—with all other settings held fixed (Table 3, top). The improvement is monotone across all rungs—NVS 32.84 → 35.00 dB, albedo 29.11 → 31.21 dB, MACE 0.019 → 0.015—and the space-carving loss contributes the largest single albedo gain (+0.84 dB) and the largest reduction in cross-view drift, confirming that resolving multi-view inconsistencies in the 2D prior—not merely adding supervision—turns albedo supervision into an effective geometric regularizer. Per-scene results, and a per-scene MACE comparison against the raw 2D prior, are in the supplement.

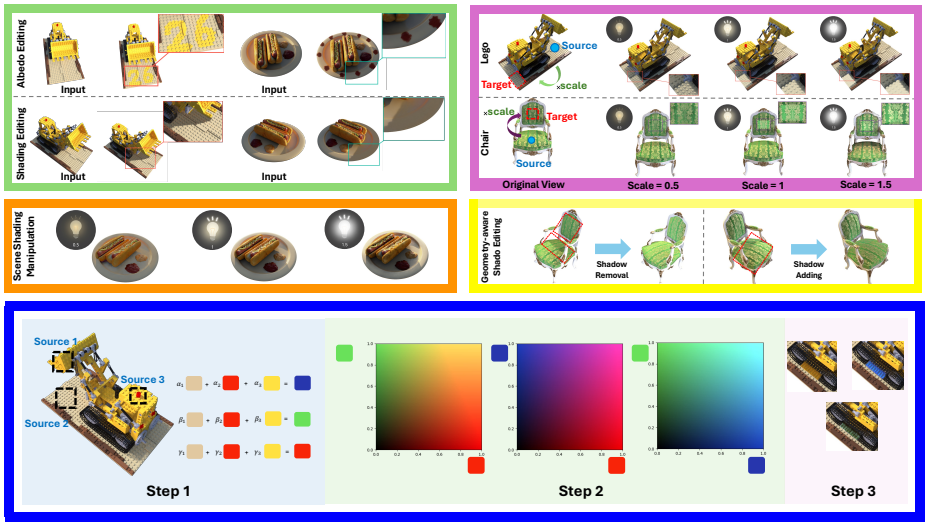

Fig. 7: Freeform intrinsic editing enabled by precise per-primitive attribution. (a) Freeform region edits on *albedo* and *shading* produce 3D-consistent results that adhere to object boundaries and occlusions. (b) Point-level *shading intensity control*: scaling shading features modulates contrast locally in a predictable, geometry-consistent manner. (c) *Scene-level shading manipulation* through uniform scaling of shading features adjusts global contrast while respecting geometry. (d) *Shadow addition/removal*: geometry-aware edits specified in 2D yield coherent shadows that align with scene structure. (e) Illustration of unseen colour generation via albedo feature interpolation: (1) deriving primary colour coefficients, (2) generating colour palettes by convex combinations of different pairs of primary colours, and (3) applying unseen colours.

Decomposition Scope. We next test whether a richer material model helps. Adding a specular branch, or replacing our shading with a Cook-Torrance microfacet BRDF [11] and spherical-harmonic lighting [43], lowers NVS, albedo, and consistency alike (Table 3, bottom). The extra parameters make the already ill-posed decomposition harder to estimate—a bias–variance trade-off—so our simpler factorization is better-conditioned.

6 Discussion and Conclusion

Limitations. Our key design choice of reducing the number of free parameters for more accurate recovery also means that some effects, such as full relighting and subsurface scattering, cannot be modelled; and because we assume surface opacity, highly semi-transparent phenomena (e.g., smoke or clear glass) are likewise not handled. Integrating stronger priors and extending the decomposition to more expressive yet stable appearance models are key future directions.

Conclusion. This work revisits point-based intrinsic decomposition and inverse rendering through the lens of misattribution in volume rendering: aggregating

Table 3: Ablation study on our synthetic scenes (NVS PSNR, albedo PSNR, and albedo consistency, MACE). *Top:* component ablation—starting from vanilla PAPR, we add one component at a time. As a reference, the off-the-shelf 2D prior run per view (no 3D fusion) gives a MACE of 0.020; our Full model lowers this to 0.015, showing that multi-view fusion, not the prior alone, makes the albedo view-consistent. *Bottom:* decomposition scope—adding a specular branch, or replacing our shading with a Cook-Torrance microfacet BRDF and spherical-harmonic lighting, lowers all three metrics, supporting our simpler factorization.

Variant	NVS PSNR↑	Albedo PSNR↑	MACE↓
Vanilla PAPR	32.84	–	–
+ albedo/shading split + 2D prior	33.71	29.11	0.019
+ cIMLE-adapted prior (M=10)	34.36	29.74	0.018
+ space-carving loss	34.83	30.58	0.016
+ per-view scalar (Full)	**35.00**	**31.21**	**0.015**
Decomposition scope (alternatives to Full)			
+ specular branch	33.20	30.95	0.017
+ Cook-Torrance + SH lighting	33.40	29.40	0.021

semi-transparent primitives along rays fundamentally limits reliable supervision of per-primitive appearance, so we repurpose Proximity Attention Point Rendering (PAPR) as a surface-aligned alternative that restores point-wise identifiability. On this basis, Intrinsic PAPR pairs a minimalist albedo–shading factorization with an ambiguity-aware 2D intrinsic prior and a space carving loss that distills multi-view albedo hypotheses into consistent 3D reflectance. Across synthetic and real scenes, it yields more accurate albedo, higher-quality novel view synthesis, and more reliable feature transfer than recent point-based and NeRF-based baselines, a step toward robust, predictable 3D editing tools and toward extending PAPR to richer appearance models and downstream applications.

Acknowledgements

This research was enabled in part by NSERC, the Canada Foundation for Innovation, the Canada CIFAR AI Chairs program, the BC DRI Group and the Digital Research Alliance of Canada.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5460–5469. IEEE (2022)
2. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. ACM Transactions on Graphics **33**(4), 1–12 (2014)
3. Bi, Z., Zeng, Y., Zeng, C., Pei, F., Feng, X., Zhou, K., Wu, H.: GS^3: Efficient relighting with triple gaussian splatting. In: SIGGRAPH Asia 2024 Conference Papers. pp. 12:1–12:12. ACM, New York, NY, USA (2024). https://doi.org/10.1145/3680528.3687576
4. Careaga, C., Aksoy, Y.: Colorful diffuse intrinsic image decomposition in the wild. ACM Transactions on Graphics **43**(6) (2024)
5. Careaga, C., Aksoy, Y.: Intrinsic image decomposition via ordinal shading. ACM Transactions on Graphics **43**(1) (2024). https://doi.org/10.1145/3630750
6. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: TensoRF: Tensorial radiance fields. In: Computer Vision – ECCV 2022. Lecture Notes in Computer Science, vol. 13692, pp. 333–350. Springer, Cham, Switzerland (2022). https://doi.org/10.1007/978-3-031-19824-3_20
7. Chen, X., Peng, S., Yang, D., Liu, Y., Pan, B., Lv, C., Zhou, X.: Intrinsicanything: Learning diffusion priors for inverse rendering under unknown illumination. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) Computer Vision – ECCV 2024. Lecture Notes in Computer Science, vol. 15118, pp. 450–467. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-73027-6_26
8. Chen, Y., Chen, Z., Zhang, C., Wang, F., Yang, X., Wang, Y., Cai, Z., Yang, L., Liu, H., Lin, G.: GaussianEditor: Swift and controllable 3D editing with gaussian splatting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 21476–21485 (2024). https://doi.org/10.1109/CVPR52733.2024.02029
9. Choi, J., Lee, S., Park, H., Jung, S.W., Kim, I.J., Cho, J.: MAIR++: Improving multi-view attention inverse rendering with implicit lighting representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(6), 5076–5093 (2025). https://doi.org/10.1109/TPAMI.2025.3548679
10. Chung, H.G., Choi, S., Baek, S.H.: Differentiable point-based inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4399–4408 (2024). https://doi.org/10.1109/CVPR52733.2024.00421
11. Cook, R.L., Torrance, K.E.: A reflectance model for computer graphics. ACM Transactions on Graphics **1**(1), 7–24 (1982). https://doi.org/10.1145/357290.357293
12. Das, P., Karaoglu, S., Gevers, T.: PIE-Net: Photometric invariant edge guided network for intrinsic image decomposition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19758–19767. IEEE (2022). https://doi.org/10.1109/CVPR52688.2022.01917
13. Du, K., Guan, Y., Wang, Z.: IDT: A physically grounded transformer for feed-forward multi-view intrinsic decomposition. CoRR **abs/2512.23667** (2025). https://doi.org/10.48550/arXiv.2512.23667

14. Du, K., Liang, Z., Shen, Y., Wang, Z.: GS-ID: Illumination decomposition on gaussian splatting via adaptive light aggregation and diffusion-guided material priors. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 26220–26229 (October 2025), https://openaccess.thecvf.com/content/ICCV2025/papers/Du_GS-ID_Illumination_Decomposition_on_Gaussian_Splatting_via_Adaptive_Light_Aggregation_ICCV_2025_paper.pdf, accessed 2026-07-01
15. Engelhardt, A., Raj, A., Boss, M., Zhang, Y., Kar, A., Li, Y., Sun, D., Martin-Brualla, R., Barron, J.T., Lensch, H.P.A., Jampani, V.: SHINOBI: Shape and illumination using neural object decomposition via BRDF optimization in-the-wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19636–19646 (2024). https://doi.org/10.1109/CVPR52733.2024.01857
16. Fei, F., Tang, J., Tan, P., Shi, B.: VMINer: Versatile multi-view inverse rendering with near- and far-field light sources. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11800–11809 (2024). https://doi.org/10.1109/CVPR52733.2024.01121
17. Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: Radiance fields without neural networks. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5491–5500 (2022). https://doi.org/10.1109/CVPR52688.2022.00542
18. Gao, J., Gu, C., Lin, Y., Li, Z., Zhu, H., Cao, X., Zhang, L., Yao, Y.: Relightable 3D gaussians: Realistic point cloud relighting with BRDF decomposition and ray tracing. In: European Conference on Computer Vision (ECCV). pp. 73–89. Springer, Cham, Switzerland (2024)
19. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada. pp. 2672–2680 (2014), https://proceedings.neurips.cc/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html, accessed 2026-07-01
20. Gu, C., Wei, X., Zeng, Z., Yao, Y., Zhang, L.: IRGS: Inter-reflective gaussian splatting with 2D gaussian ray tracing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10943–10952 (2025). https://doi.org/10.1109/CVPR52734.2025.01022
21. Jiang, Y., Tu, J., Liu, Y., Gao, X., Long, X., Wang, W., Ma, Y.: GaussianShader: 3D Gaussian Splatting with Shading Functions for Reflective Surfaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5322–5332 (2024). https://doi.org/10.1109/CVPR52733.2024.00509
22. 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). pp. 165–174 (2023). https://doi.org/10.1109/CVPR52729.2023.00024
23. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4), 139:1–139:14 (2023). https://doi.org/10.1145/3592433

24. Knapitsch, A., Park, J., Zhou, Q.Y., Koltun, V.: Tanks and temples: Benchmarking large-scale scene reconstruction. ACM Transactions on Graphics **36**(4), 78:1–78:13 (2017)
25. Kocsis, P., Sitzmann, V., Nießner, M.: Intrinsic image diffusion for indoor single-view material estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5198–5208 (2024). https://doi.org/10.1109/CVPR52733.2024.00497
26. Lassner, C., Zollhöfer, M.: Pulsar: Efficient sphere-based neural rendering. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1440–1449 (2021). https://doi.org/10.1109/CVPR46437.2021.00149
27. Lee, H.Y., Tseng, H.Y., Lee, H.Y., Yang, M.H.: Exploiting diffusion prior for generalizable dense prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7861–7871 (2024). https://doi.org/10.1109/CVPR52733.2024.00751
28. Lee, J.H., Kim, D.S.: ICE-NeRF: Interactive color editing of NeRFs via Decomposition-Aware weight optimization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3491–3501 (Oct 2023). https://doi.org/10.1109/ICCV51070.2023.00323
29. Li, K., Peng, S., Zhang, T., Malik, J.: Multimodal image synthesis with conditional implicit maximum likelihood estimation. International Journal of Computer Vision **128**(10-11), 2607–2628 (2020)
30. 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: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2472–2481. IEEE (2020). https://doi.org/10.1109/CVPR42600.2020.00255
31. 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 (CVPR). pp. 21644–21653 (2024). https://doi.org/10.1109/CVPR52733.2024.02045
32. Litman, Y., Patashnik, O., Deng, K., Agrawal, A., Zawar, R., la Torre, F.D., Tulsiani, S.: MaterialFusion: Enhancing inverse rendering with material diffusion priors. In: 2025 International Conference on 3D Vision (3DV). pp. 802–812. IEEE (2025). https://doi.org/10.1109/3DV66043.2025.00079
33. Luo, J., Huang, Z., Li, Y., Zhou, X., Zhang, G., Bao, H.: NIID-Net: Adapting surface normal knowledge for intrinsic image decomposition in indoor scenes. IEEE Transactions on Visualization and Computer Graphics **26**(12), 3434–3445 (2020). https://doi.org/10.1109/TVCG.2020.3023565
34. 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: European Conference on Computer Vision (ECCV). Springer, Cham, Switzerland (2020)
35. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics (TOG) **41**(4), 102:1–102:15 (2022). https://doi.org/10.1145/3528223.3530127
36. Munkberg, J., Hasselgren, J., Shen, T., Gao, J., Chen, W., Evans, A., Müller, T., Fidler, S.: Extracting triangular 3D models, materials, and lighting from images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8270–8280 (2022). https://doi.org/10.1109/CVPR52688.2022.00810

37. Narihira, T., Maire, M., Yu, S.X.: Learning Lightness from Human Judgement on Relative Reflectance. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2965–2973. IEEE Computer Society (2015). https://doi.org/10.1109/CVPR.2015.7298915
38. Nimier-David, M., Vicini, D., Zeltner, T., Jakob, W.: Mitsuba 2: A retargetable forward and inverse renderer. ACM Transactions on Graphics **38**(6), 1–17 (2019)
39. Noras, P., Choi, J.M., Stricker, D., Peers, P., Sengupta, R.: GAINS: Gaussian-based inverse rendering from sparse multi-view captures (2025). https://doi.org/10.48550/arXiv.2512.09925
40. Ost, J., Laradji, I.H., Newell, A., Bahat, Y., Heide, F.: Neural point light fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 18398–18408. IEEE (2022). https://doi.org/10.1109/CVPR52688.2022.01787
41. Peng, S., Moazenipourasil, S.A., Li, K.: CHIMLE: Conditional hierarchical IMLE for multimodal conditional image synthesis. Advances in Neural Information Processing Systems **35**, 280–296 (2022)
42. Radl, L., Steiner, M., Kurz, A., Steinberger, M.: LAENeRF: Local appearance editing for neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4969–4978 (2024). https://doi.org/10.1109/CVPR52733.2024.00475
43. Ramamoorthi, R., Hanrahan, P.: An efficient representation for irradiance environment maps. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). pp. 497–500 (2001). https://doi.org/10.1145/383259.383317
44. Sarkar, K., Bühler, M.C., Li, G., Wang, D., Vicini, D., Riviere, J., Zhang, Y., Orts-Escolano, S., Gotardo, P.F.U., Beeler, T., Meka, A.: LitNeRF: Intrinsic radiance decomposition for high-quality view synthesis and relighting of faces. In: SIGGRAPH Asia 2023 Conference Papers. pp. 42:1–42:11. ACM (2023). https://doi.org/10.1145/3610548.3618210
45. 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: 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019. pp. 8597–8606. IEEE (2019). https://doi.org/10.1109/ICCV.2019.00869
46. 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: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7495–7504. IEEE (2021). https://doi.org/10.1109/CVPR46437.2021.00741
47. Sun, H., Gao, Y., Xie, J., Yang, J., Wang, B.: SVG-IR: Spatially-varying gaussian splatting for inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16143–16152 (2025). https://doi.org/10.1109/CVPR52734.2025.01505
48. Tiwari, A., Ikehata, S., Raman, S.: MERLiN: Single-shot material estimation and relighting for photometric stereo. In: Computer Vision – ECCV 2024. Lecture Notes in Computer Science, vol. 15070, pp. 251–269. Springer (2024). https://doi.org/10.1007/978-3-031-73254-6_15
49. Uy, M.A., Martin-Brualla, R., Guibas, L., Li, K.: SCADE: NeRFs from space carving with ambiguity-aware depth estimates. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16518–16527 (2023). https://doi.org/10.1109/CVPR52729.2023.01585

50. Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T., Srinivasan, P.P.: Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5481–5490. IEEE, Los Alamitos, CA, USA (2022). https://doi.org/10.1109/CVPR52688.2022.00541
51. Wang, H., Hu, W., Zhu, L., Lau, R.W.H.: Inverse rendering of glossy objects via the neural plenoptic function and radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19999–20008 (2024). https://doi.org/10.1109/CVPR52733.2024.01890
52. Wang, J., Fang, J., Zhang, X., Xie, L., Tian, Q.: GaussianEditor: Editing 3D gaussians delicately with text instructions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20902–20911 (2024). https://doi.org/10.1109/CVPR52733.2024.01975
53. Wang, S., Antić, B., Geiger, A., Tang, S.: Intrinsicavatar: Physically based inverse rendering of dynamic humans from monocular videos via explicit ray tracing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1877–1888 (2024). https://doi.org/10.1109/CVPR52733.2024.00184
54. Wang, X., Zhu, J., Ye, Q., Huo, Y., Ran, Y., Zhong, Z., Chen, J.: Seal-3D: Interactive pixel-level editing for neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 17637–17647 (2023). https://doi.org/10.1109/ICCV51070.2023.01621
55. Wang, Z., Shen, T., Gao, J., Huang, S., Munkberg, J., Hasselgren, J., Gojcic, Z., Chen, W., Fidler, S.: Neural fields meet explicit geometric representations for inverse rendering of urban scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8370–8380 (2023). https://doi.org/10.1109/CVPR52729.2023.00809
56. Wu, Q., Tan, J., Xu, K.: PaletteNeRF: Palette-based color editing for NeRFs. Communications in Information and Systems **23**(4), 447–475 (2023). https://doi.org/10.4310/CIS.2023.v23.n4.a4
57. Xing, X., Groh, K., Karaoglu, S., Gevers, T.: Intrinsic appearance decomposition using point cloud representation. In: 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). pp. 4234–4238. IEEE, Los Alamitos, CA, USA (2023). https://doi.org/10.1109/ICCVW60793.2023.00457
58. Xing, X., Groh, K., Karaoglu, S., Gevers, T.: Intrinsic-GS: Multi-view intrinsic image decomposition using gaussian splatting and color-invariant priors. Color and Imaging Conference **32**(1), 56–63 (2024). https://doi.org/10.2352/CIC.2024.32.1.12
59. Xu, Q., Xu, Z., Philip, J., Bi, S., Shu, Z., Sunkavalli, K., Neumann, U.: Point-NeRF: Point-Based neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5428–5438 (2022). https://doi.org/10.1109/CVPR52688.2022.00536
60. Yao, Y., Zhang, J., Liu, J., Qu, Y., Fang, T., McKinnon, D., Tsin, Y., Quan, L.: NeILF: Neural incident light field for physically-based material estimation. In: Computer Vision – ECCV 2022. Lecture Notes in Computer Science, vol. 13691, pp. 700–716. Springer (2022). https://doi.org/10.1007/978-3-031-19821-2_40
61. Ye, W., Chen, S., Bao, C., Bao, H., Pollefeys, M., Cui, Z., Zhang, G.: Intrinsic-NeRF: Learning intrinsic neural radiance fields for editable novel view synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 339–351 (2023). https://doi.org/10.1109/ICCV51070.2023.00038

62. 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: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3578–3587 (2023). https://doi.org/10.1109/ICCV51070.2023.00333
63. Zhang, Q., Baek, S.H., Rusinkiewicz, S., Heide, F.: Differentiable point-based radiance fields for efficient view synthesis. In: SIGGRAPH Asia 2022 Conference Papers (SA '22). pp. 7:1–7:12. ACM, New York, NY, USA (2022). https://doi.org/10.1145/3550469.3555413
64. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. pp. 586–595. Computer Vision Foundation / IEEE Computer Society (2018). https://doi.org/10.1109/CVPR.2018.00068
65. 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)
66. Zhang, Y., Peng, S., Moazeni, A., Li, K.: PAPR: Proximity attention point rendering. In: Advances in Neural Information Processing Systems 36 (NeurIPS 2023) (2023), https://proceedings.neurips.cc/paper_files/paper/2023/hash/bda5c35eded86adaf0231748e3ce071c-Abstract-Conference.html, accessed 2026-07-01
67. Zhang, Y., Sun, J., He, X., Fu, H., Jia, R., Zhou, X.: Modeling indirect illumination for inverse rendering. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 18622–18631 (2022). https://doi.org/10.1109/CVPR52688.2022.01809
68. Zhou, H., Yu, X., Jacobs, D.: GLoSH: Global-local spherical harmonics for intrinsic image decomposition. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7819–7828. IEEE (2019). https://doi.org/10.1109/ICCV.2019.00791
69. Zhou, T., Krähenbühl, P., Efros, A.A.: Learning data-driven reflectance priors for intrinsic image decomposition. In: 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015. pp. 3469–3477. IEEE Computer Society (2015). https://doi.org/10.1109/ICCV.2015.396
70. Zhu, R., Li, Z., Matai, J., Porikli, F., Chandraker, M.: Irisformer: Dense vision transformers for single-image inverse rendering in indoor scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2812–2821 (2022). https://doi.org/10.1109/CVPR52688.2022.00284
71. Zhu, Z.L., Yang, J., Wang, B.: Gaussian splatting with discretized SDF for relightable assets. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 25155–25164 (2025). https://doi.org/10.1109/ICCV51701.2025.02333
72. Zoran, D., Isola, P., Krishnan, D., Freeman, W.T.: Learning ordinal relationships for mid-level vision. In: 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015. pp. 388–396. IEEE Computer Society (2015). https://doi.org/10.1109/ICCV.2015.52
73. Zuo, Y., Deng, J.: View synthesis with sculpted neural points. In: The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net (2023), https://openreview.net/forum?id=0ypGZvm0er0, accessed 2026-07-01