AEGIR: MODELING AREA EMITTERS FOR INDOOR INVERSE RENDERING USING GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review

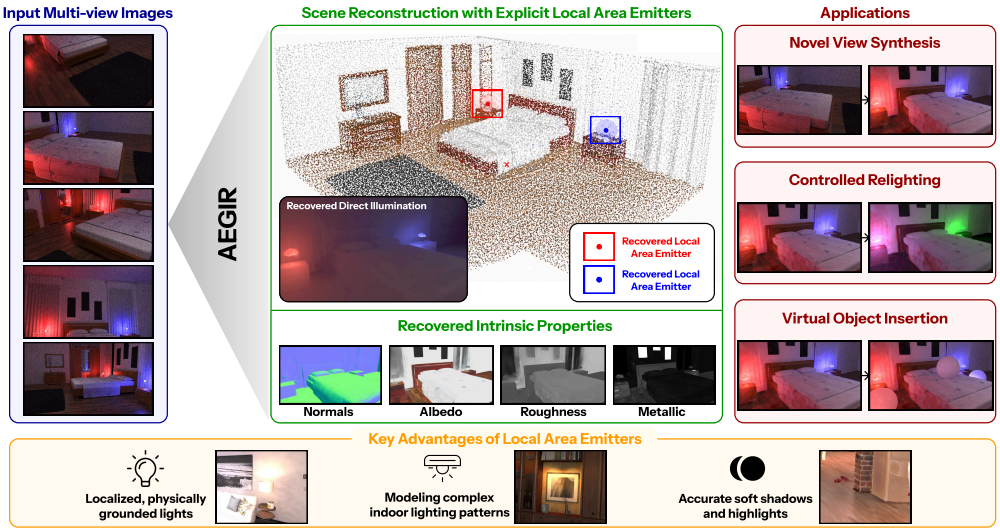


Figure 1: **AEGIR Overview.** From multi-view images, AEGIR jointly models local area emitters, 3D geometry, and materials, capturing accurate shadows and complex lighting patterns through localized, physically grounded illumination for realistic novel view synthesis, relighting, and object insertion.

ABSTRACT

Inverse rendering requires separating illumination from surface materials, which is highly ambiguous due to their tight coupling in observed images. While Gaussian Splatting is efficient for novel view synthesis, existing relightable methods approximate scene lighting using discrete point lights, global environment maps, or implicit representations. By ignoring the physical spatial extent of real-world emitters, these approaches produce inaccurate light attenuation and unrealistic shadows. We present **AEGIR** (**A**rea **E**mitters for **G**aussian **I**nverse **R**endering), a framework that models compact local area emitters with an anisotropic angular emission profile within a relightable Gaussian Splatting representation. Joint optimization of emitters, materials, and geometry is challenging because the flexible emitter parameterization increases both the number of parameters and the ambiguity between illumination and materials. We address this by introducing a differentiable deferred rendering pipeline that integrates multiple importance sampling with targeted regularization. As a result, AEGIR accurately simulates local light transport and achieves more consistent decomposition. Experiments show that explicit area emitters improve illumination reconstruction and enhance downstream tasks, including novel view synthesis, controlled relighting, and virtual object insertion, particularly in scenes with complex local lighting.

1 INTRODUCTION

Methods for reconstructing 3D environments from 2D images have advanced rapidly with Neural Radiance Fields (NeRF) (Mildenhall et al., 2020) and 3D Gaussian Splatting (3DGS) (Kerbl et al.,

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.

2023). Although highly effective for novel view synthesis, these representations remain challenging to extend to inverse rendering. Inverse rendering seeks to decompose a scene into geometry, materials, and illumination, but this problem is inherently ill-posed because appearance depends on the combined effects of surface material and lighting. Nevertheless, this decomposition is essential for downstream tasks such as scene editing and relighting.

A core limitation of point-based illumination models, distant environment maps, and implicit illumination models is their treatment of emitter extent. Global environment maps (Gao et al., 2023; Gu et al., 2025) assume infinite distance, while point lights (Kocsis et al., 2024; Du et al., 2025) lack the physical dimensions needed to cast soft shadows. Real indoor emitters, such as lamps and ceiling panels, instead have finite extent and often lie close to scene surfaces. Implicit light fields (Yao et al., 2022; Zhang et al., 2023) capture spatial variation without explicit source geometry, making geometric attenuation constraints difficult to enforce. Inaccurate emitter extent therefore leads to incorrect attenuation, illumination patterns, and shadow boundaries. Consequently, optimization often absorbs these residual lighting effects into the albedo.

AEGIR (Figure 1) goes beyond finite extent alone, modeling compact area emitter geometry separately from scene 2D Gaussian surface primitives, with independently optimized positions, shapes, and angular emission profiles. Anisotropic scales and Super-Gaussian angular falloff represent sources ranging from compact bulbs to elongated fluorescent tubes, providing an expressive, physically meaningful approximation of local illumination.

However, area emitters are difficult to optimize because their high-dimensional parameterization amplifies the ambiguity between material and illumination. By enforcing radiometric and geometric constraints and introducing a physically grounded, fully differentiable parameterization, our model enables stable optimization while accurately capturing localized light transport. AEGIR jointly optimizes the 2D Gaussian geometry, area emitters, and physically based rendering (PBR) material parameters, specifically diffuse albedo, roughness, and metallic, from multi-view images. Experiments demonstrate that our emitter representation improves illumination estimation and downstream tasks, including novel view synthesis and editing in scenes with complex local lighting. In summary, our core contributions are:

- A novel explicit area emitter formulation that parameterizes localized 3D primitives with shapes and emission profiles. By combining anisotropic scaling with a Super-Gaussian angular falloff, this representation enables flexible modeling of diverse real-world emitters.
- A Gaussian Splatting framework for disentangled recovery of explicit light sources, geometry, and PBR materials from multi-view images, using DN-Splatter (Turkulainen et al., 2025) initialization, soft diffusion material priors, and radiometric and geometric constraints.
- A physically grounded rendering model that evaluates illumination directly from area emitters, enabling physically consistent downstream applications such as novel view synthesis, controlled scene relighting, and virtual object insertion.

2 Related Work

2.1 Inverse Rendering and Intrinsic Decomposition

Data-Driven Inverse Rendering. Data-driven approaches learn priors from large-scale datasets to predict intrinsic scene properties from images. Early neural methods (Yu & Smith, 2019; Li et al., 2019; Zhu et al., 2022b) have been increasingly replaced by diffusion-based models such as IntrinsicDiffusion (Luo et al., 2024), RGB↔X (Zeng et al., 2024), and DiffusionRenderer (Liang et al., 2025a), which estimate intrinsic attributes including albedo, roughness, and surface normals. To improve multi-view consistency, recent methods such as IDArb (Li et al., 2025) and MatSpray (Langsteiner et al., 2025) leverage diffusion priors, while MVInverse (Wu et al., 2025) adopts a feed-forward architecture with explicit consistency constraints. These approaches predict high quality intrinsics but often lack explicit, physically accurate light transport.

Optimization-Based Inverse Rendering. Optimization-based methods enforce multi-view consistency by minimizing photometric error using differentiable rendering. NeRF-based approaches such as NeRD (Boss et al., 2021a), NeRFactor (Zhang et al., 2021b), Neural-PIL (Boss et al., 2021b), NeILF (Yao et al., 2022), and NeILF++ (Zhang et al., 2023) jointly estimate geometry, materials,

and illumination, but are computationally expensive. ESR-NeRF (Jeong et al., 2024) reconstructs emissive sources from paired light-on/light-off low dynamic range (LDR) captures. More recent 3DGS formulations, including Relightable 3D Gaussians (Gao et al., 2023), GS-IR (Liang et al., 2024), SVG-IR (Sun et al., 2025), IRGS (Gu et al., 2025), GeoSplatting (Ye et al., 2025), GS-ID (Du et al., 2025), and EAG-PT (Yang et al., 2026), integrate physically based rendering for efficient relighting. Methods such as FIPT (Wu et al., 2023) and IRIS (Lin et al., 2025) improve indoor scene decomposition using structured lighting representations and stronger geometric and material priors.

Environment maps (Munkberg et al., 2022; Hasselgren et al., 2022; Gao et al., 2023; Yao et al., 2025; Gu et al., 2025), point lights with Spherical Gaussian emission (Kocsis et al., 2024; Du et al., 2025; Niu et al., 2026), and implicit light fields (Yao et al., 2022; Zhang et al., 2023; Coomans et al., 2024; Zhu et al., 2022a) do not explicitly recover finite source geometry. EAG-PT separates emissive surface Gaussians using fixed input emitter masks without optimizing emitter geometry. These approaches introduce ambiguity in inverse rendering, often leading to residual lighting effects being absorbed into the albedo. Motivated by this, AEGIR optimizes compact emitter geometry separately from surface primitives, enabling more accurate illumination decomposition and disentanglement.

2.2 Illumination Modeling in 3D Environments

A common approach is to model lighting using global environment maps, often parameterized by Spherical Gaussians (Zhang et al., 2021a). While effective for distant illumination, these models assume spatially uniform lighting and struggle to capture localized variation and distance attenuation in indoor scenes. Nevertheless, environment maps remain widely used, estimated using feed-forward neural networks (Garon et al., 2019; Li et al., 2023) or, more recently, latent diffusion models (Phongthawee et al., 2023; Tong et al., 2025; Liang et al., 2025b). To introduce spatial locality, methods such as Deep Parametric Indoor Lighting (Gardner et al., 2019) regress discrete 3D light sources from a single image. These spatially localized estimates use isotropic area lights, limiting their ability to represent anisotropic source geometry. Moreover, single-image estimation also does not enforce temporal consistency. Hybrid approaches, such as GS-ID (Du et al., 2025) and SGS-Intrinsic (Niu et al., 2026), model illumination by combining global environment maps with localized point lights represented as Spherical Gaussians. Alternatively, recent generative methods such as LumiNet (Xing et al., 2025) and LuxRemix (Liang et al., 2026) leverage diffusion models to decompose complex indoor lighting into editable components for relighting. Despite these differences, both families of methods lack explicit modeling of spatially extended emitters, which limits their ability to model physically accurate lighting effects. In contrast, AEGIR explicitly parameterizes the spatial extents, shapes, and angular emission profiles of light emitters under proper optimization constraints. This formulation enables more accurate illumination decomposition and improves disentanglement between materials and lighting compared to prior approaches.

3 Methodology

Standard Gaussian Splatting (Kerbl et al., 2023) optimizes volumetric primitives with colors baked into view-dependent Spherical Harmonics (SH). Relightable 3D Gaussians (Gao et al., 2023) decouples illumination from appearance by replacing SH coefficients with microfacet bidirectional reflectance distribution function (BRDF) parameters, typically diffuse albedo and roughness, that respond physically to novel lighting. For stable physically based shading, AEGIR uses 2D Gaussian Splatting (2DGS) (Huang et al., 2024), whose planar primitives provide more reliable surface geometry and normals than volumetric splatting. As illustrated in Figure 2, we couple these relightable Gaussians with a unified area emitter representation that explicitly parameterizes light shape and emission profile to model spatially varying illumination with realistic attenuation and soft shadows. Deferred rendering with multiple importance sampling jointly optimizes geometry, illumination, and intrinsic materials under carefully designed constraints, producing physically grounded illumination and consistent material disentanglement.

3.1 Explicit Local Area Emitters

Anisotropic Super-Gaussian Emitter Formulation. We propose representing explicit area emitters as a dynamic collection of anisotropic 3D ellipsoids with a Super-Gaussian angular falloff

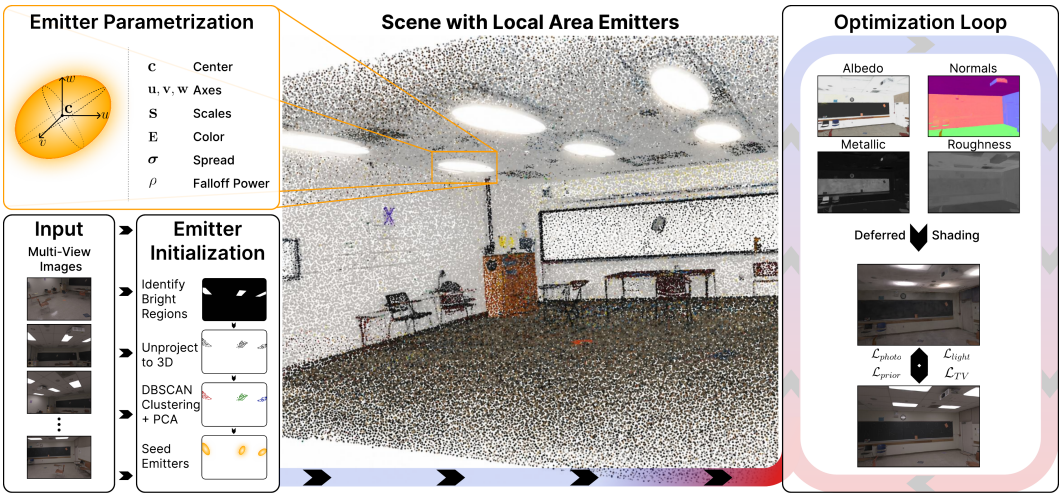


Figure 2: **AEGIR Optimization Framework.** Parameterized area emitters are initialized from multi-view inputs and jointly optimized with 2D Gaussian geometry and PBR materials within a deferred rendering framework, enabling physically accurate scene decomposition.

distribution. Each emitter e is parameterized by its spatial center $\mathbf{c}_e \in \mathbb{R}^3$, local orthonormal axes $\{\mathbf{u}_e, \mathbf{v}_e, \mathbf{w}_e\}$, spatial scales $\mathbf{s}_e = (s_{e,u}, s_{e,v}, s_{e,w}) \in \mathbb{R}^3_+$, RGB emission color $\mathbf{E}_e \in \mathbb{R}^3_+$, axis-aligned angular spread coefficients $\boldsymbol{\sigma}_e = (\sigma_{e,u}, \sigma_{e,v}, \sigma_{e,w}) \in \mathbb{R}^3_+$, and a Super-Gaussian angular falloff power $\rho_e \in \mathbb{R}_+$. For compact notation, $\mathbf{k}_e$ denotes the corresponding emitter axis for $k \in \{u, v, w\}$.

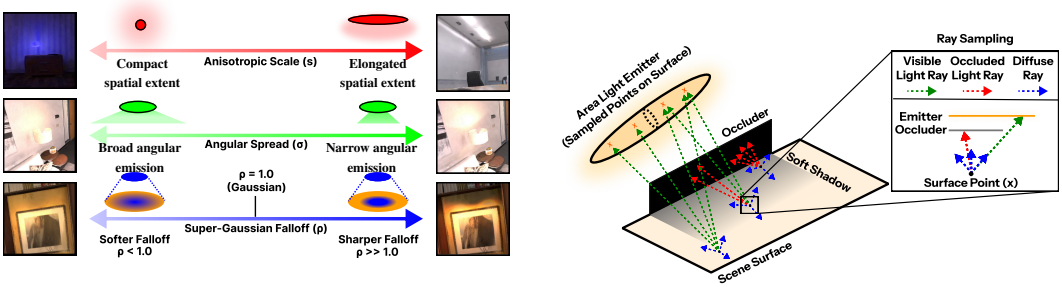
For a given surface point $\mathbf{x}$, let $\boldsymbol{\omega}_{i,e}$ denote the incident light direction pointing from the surface toward emitter e. We formulate the radiance $\mathbf{L}^{\text{emitter}}_{i,e}(\mathbf{x}, \boldsymbol{\omega}_{i,e})$ as an anisotropic Super-Gaussian angular emission profile:

$$\mathbf{L}^{\text{emitter}}_{i,e}(\mathbf{x}, \boldsymbol{\omega}_{i,e}) = \mathbf{E}_e \exp\left(-\left(\sum_{k\in\{u,v,w\}} \frac{(\boldsymbol{\omega}_{i,e} \cdot \mathbf{k}_e)^2}{\sigma^2_{e,k}}\right)^{\rho_e}\right). \tag{1}$$

The roles of spatial extent and angular emission are distinct. $\mathbf{s}_e$ determines the ellipsoid surface used for light sampling, whereas $\boldsymbol{\sigma}_e$ and ρ_e determine the directional emission in Equation 1. We use Monte Carlo integration with multiple importance sampling (MIS), where unit sphere samples are mapped to the emitter surface with an analytic induced area density and converted to a solid-angle density at $\mathbf{x}$. The squared distance factor in this conversion yields the physical inverse-square distance attenuation in the direct light estimator. The complete sampling map, area density, sample allocation across emitters, and MIS mixture are derived in Appendix A.2 (Equations S3-S12).

Figure 3 summarizes the emitter parameterization and its sampling based evaluation. To understand how Equation 1 achieves both complex anisotropy and variable sharpness, we decompose the angular spread mechanism into two components, illustrated in Figure 3a: (1) **Axis-Aligned Anisotropic Spread:** Rather than assuming a uniform angular spread, we project $\boldsymbol{\omega}_{i,e}$ onto the emitter's local coordinate frame $\{\mathbf{u}_e, \mathbf{v}_e, \mathbf{w}_e\}$. Independent coefficients $\sigma_{e,u}$, $\sigma_{e,v}$, and $\sigma_{e,w}$ decouple the angular spread across the three axes, allowing one primitive to represent sources ranging from near-isotropic bulbs to directional elongated fixtures. (2) **Super-Gaussian Angular Falloff:** The learnable power ρ_e generalizes the standard Gaussian profile ($\rho_e = 1$), allowing the model to manipulate the light source falloff. Increasing ρ_e flattens the high emission region and steepens the transition to low emission around the unit anisotropic angular distance defined by $\boldsymbol{\sigma}_e$. Together, $\boldsymbol{\sigma}_e$ and ρ_e represent angular profiles ranging from broad, smoothly varying emission to narrowly concentrated, sharply falling directional emission. The spatial scales $\mathbf{s}_e$, rather than ρ_e, determine the emitter extent that produces soft shadows.

Emitter Initialization and Adaptive Control. To constrain the ill-posed joint optimization, we initialize area emitters by unprojecting pixels that are bright relative to their local neighborhoods



(a) Effect of area emitter parameters.

(b) Area emitter sampling and visibility tracing.

Figure 3: **Area Emitter Formulation and Evaluation.** (a) The spatial scales ($\mathbf{s}_e$) determine the sampled ellipsoid, while the angular parameters ($\boldsymbol{\sigma}_e$ and ρ_e) determine the emitted radiance profile. (b) Illumination is evaluated using emitter surface sampling and visibility tracing, which resolves complex occlusions and generates accurate soft shadows.

into a 3D point cloud. These points are clustered using DBSCAN (Ester et al., 1996). Subsequently, Principal Component Analysis (PCA) of each cluster determines an emitter's initial center, scale, and axes. Initialization does not fix the final positions, scales, emissions, or emitter count. During optimization, all parameters are updated and the emitter set is dynamically split and pruned. Let $k_e^\star = \arg\max_{k\in\{u,v,w\}} s_{e,k}$ and let $\mathbf{k}_e^\star$ denote the corresponding axis. For a split, the child centers are: $\mathbf{c}_{e,a}, \mathbf{c}_{e,b} = \mathbf{c}_e \pm \frac{1}{2} s_{e,k_e^\star} \mathbf{k}_e^\star$. Appendix C.4 specifies splitting/pruning thresholds and schedule.

3.2 Deferred Rendering of Relightable Gaussians and Area Emitters

AEGIR utilizes deferred rendering to decouple geometry and material rasterization from lighting evaluation. We rasterize 2D Gaussians into a G-buffer containing albedo, roughness, metallic, and surface normals, and evaluate the rendering equation with a microfacet BRDF using Monte Carlo integration. To reduce variance, we employ Multiple Importance Sampling (MIS), illustrated in Figure 3b, combining cosine-weighted hemisphere sampling for diffuse transport with direct light sampling for high-frequency shadows and specular highlights (details in Appendix A.2). Rays are then traced through the 2D Gaussian scene to compute a soft differentiable visibility term:

$$V_i(\mathbf{x}, \boldsymbol{\omega}_i) = 1 - \alpha_{\text{trace}}\ \text{sigmoid}(d_{\text{light}} - \epsilon - d_{\text{trace}})\,, \tag{2}$$

where α_{trace} is accumulated ray opacity, d_{light} is the distance to the sampled emitter point, d_{trace} is the normalized transmittance-weighted mean intersection depth, and ϵ is a stability bias. The total incident light combines direct emission $\mathbf{L}_i^{\text{direct}}(\mathbf{x}, \boldsymbol{\omega}_i)$ and the traced indirect light proxy $\mathbf{L}_i^{\text{trace}}(\mathbf{x}, \boldsymbol{\omega}_i)$:

$$\mathbf{L}_i(\mathbf{x}, \boldsymbol{\omega}_i) = V_i(\mathbf{x}, \boldsymbol{\omega}_i)\mathbf{L}_i^{\text{direct}}(\mathbf{x}, \boldsymbol{\omega}_i) + (1 - V_i(\mathbf{x}, \boldsymbol{\omega}_i))\mathbf{L}_i^{\text{trace}}(\mathbf{x}, \boldsymbol{\omega}_i), \tag{3}$$

During optimization, a traced ray that intersects the Gaussian scene queries the frozen SH color features at that intersection. These features already encode much of the observed surface appearance and therefore provide a fixed proxy for indirect illumination while geometry, explicit emitters, and materials are jointly optimized. Freezing the proxy discourages the optimization from absorbing indirect effects into additional local emitters. During controlled relighting, the training time SH proxy is discarded because it is tied to the original illumination. The selected, edited, or replacement emitters are then rendered with their direct contribution plus one path-traced indirect bounce. Appendix A.3 details shadow tracing and indirect illumination. Finally, the incident light gathered from all samples is integrated against the microfacet BRDF to evaluate the pixel color. For completeness, we alternatively integrate a training-time ReSTIR sampler (following (Wang et al., 2023)) into the direct emitter branch of AEGIR. This modifies only the training sampling strategy, leaving the emitter formulation, objectives, relighting renderer, and evaluation unchanged. Appendix A.4 details the alternative sampler. Table S7 and Figure S13 in Appendix D.8 compare it with standard sampling.

3.3 Joint Optimization and Training Curriculum

Decomposing observed radiance into geometry, materials, and lighting is a highly ambiguous problem. This difficulty increases when modeling area light emitters, as they introduce more unknown variables.

Without strict constraints, the optimization tends to bake shadows or highlights into the material representation. We therefore use a progressive curriculum. DN-Splatter (Turkulainen et al., 2025) first provides dense 2D Gaussian geometry and initial Spherical Harmonics. Afterwards, for the first 15,000 optimization iterations, the Gaussian geometry, microfacet materials, and area emitters are jointly optimized. The SH coefficients remain frozen and provide the indirect light proxy. Subsequently, geometry and materials are frozen for the final 5,000 iterations while emitters alone are refined. The overall objective of the joint optimization is:

$$\mathcal{L}_{total} = \mathcal{L}_{photo} + \mathcal{L}_{prior} + \mathcal{L}_{TV} + \mathcal{L}_{light}. \tag{4}$$

(1) **Contrast-Aware Photometric Loss ($\mathcal{L}_{photo}$):** We introduce a focus map W_f that upweights pixels with high local contrast. This encourages the optimization to prioritize the reconstruction of high-frequency lighting cues that standard losses often blur. (2) **Diffusion Priors ($\mathcal{L}_{prior}$):** To reduce global ambiguities such as distinguishing a white wall in shadow from a dark wall, we supervise materials with pseudo ground truth from a pretrained DiffusionRenderer (Liang et al., 2025a). This discourages the model from erroneously adjusting the albedo to compensate for complex lighting. (3) **Edge-Aware Total Variation ($\mathcal{L}_{TV}$):** To mitigate noise in the diffusion priors, we apply a total variation (TV) loss on the material maps guided by ground truth image gradients. It encourages piecewise constant materials on flat surfaces while preserving sharp structural boundaries, encouraging the emitters rather than the materials to explain illumination patterns. (4) **Emitter Regularization ($\mathcal{L}_{light}$):** To keep explicit lights physically plausible, we penalize excessive spatial scaling, encourage color neutrality to reduce illumination leakage into the albedo, and penalize positions outside the observable scene. Detailed mathematical formulations for all loss components are provided in Appendix B, with weights and learning rates in Appendix C.4. The final emitter refinement stage forces the lighting representation to resolve remaining discrepancies, refining soft shadows and structured highlights without baking illumination into the albedo.

4 Results and Discussion

4.1 Experimental Setup and Datasets

We evaluate 24 scenes: 10 synthetic scenes from Hypersim (Roberts et al., 2021) and FIPT-Synthetic (Wu et al., 2023), and 14 real/scanned scenes from ScanNet++ (Yeshwanth et al., 2023), Replica (Straub et al., 2019), and FIPT-Real (Wu et al., 2023) (Table S1). We compare against GS-ID (Du et al., 2025), IRGS (Gu et al., 2025), IRIS (Lin et al., 2025), NeILF++ (Zhang et al., 2023), EAG-PT (Yang et al., 2026), and SGS-Intrinsic (Niu et al., 2026). Since the official GS-ID implementation is not publicly available, we reproduce their hybrid illumination formulation and deshadowing model for evaluation. We report novel view synthesis (NVS) and albedo using PSNR, SSIM, and LPIPS, and roughness using MSE. Metrics use pre-denoising renders. Visualizations follow the published procedure of each method, with OptiX denoising (Parker et al., 2010) used for AEGIR. All methods are timed on a single NVIDIA RTX 4090 GPU. Appendices C.2-C.4 provide baseline adaptations, metric definitions, and training settings. Appendix C.5 reports the runtime breakdown and seed variability.

4.2 Core Contribution: Explicit Local Area Emitters

Table 1: **Quantitative evaluation of pure lighting estimation.** Metrics are computed by re-rendering each scene in Mitsuba (Jakob et al., 2022) using the light sources estimated by each method and comparing against renderings under ground-truth light sources. We report metrics for indoor local lighting and external environment lighting separately.

Method	Indoor Lighting			Environment Lighting		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
GS-ID (Du et al., 2025)	<u>28.35</u>	<u>0.92</u>	<u>0.11</u>	26.35	0.86	0.15
IRGS (Gu et al., 2025)	11.26	0.49	0.25	<u>28.98</u>	<u>0.88</u>	<u>0.14</u>
AEGIR (Ours)	**31.42**	**0.94**	**0.06**	**30.22**	**0.92**	**0.07**



Figure 4: **Qualitative comparison of pure lighting estimation.** All images use Mitsuba (Jakob et al., 2022) with ground-truth geometry and materials, comparing ground-truth lighting (Reference) with estimated lighting of each method. Bottom-left insets show direct illumination (shading). AEGIR uses explicit local area emitters, GS-ID combines point lights and an environment map, and IRGS uses only environment maps.

To rigorously evaluate the estimated lighting, we use a controlled Mitsuba (Jakob et al., 2022) pipeline that isolates illumination errors from geometry and materials. Using synthetic scenes from Bitterli's rendering resources (Bitterli, 2016), we first render ground-truth training images with a simplified BRDF that is consistent with all evaluated methods. After training, the estimated lighting from each method, explicit area emitters (AEGIR), point lights with an environment map (GS-ID), and an environment map alone (IRGS), is imported back into Mitsuba. These baselines support the required lighting export. Therefore, adapting the other representations to this setup was not included. We then re-render each scene with fixed ground-truth geometry and materials, so the metrics reflect lighting errors only. Table 1 and Figure 4 show that the lighting representation strongly affects performance. For indoor lighting, IRGS fails to localize the ceiling light and instead places energy at the windows using environment map radiance, causing severe underexposure (11.26 dB PSNR). GS-ID improves localization (28.35 dB) but produces overly concentrated lighting. AEGIR best matches the ground truth (31.42 dB), recovering smoother indoor shading. For environment lighting, IRGS performs better (28.98 dB) since environment maps naturally capture window illumination. GS-ID performs worse (26.35 dB) because its added point lights introduce localized bright spots. AEGIR again performs best (30.22 dB), with smoother window illumination closest to the ground truth. Across both settings, AEGIR achieves the best PSNR, SSIM, and LPIPS among the evaluated lighting representations. Expanded direct illumination maps are provided in Figure S1 in Appendix D.1.

4.3 Novel View Synthesis and Intrinsic Disentanglement

Table 2: **Quantitative evaluation on Hypersim and FIPT-Synthetic datasets.** We report the performance of novel view synthesis, intrinsic material decomposition, and the optimization time (in minutes). Best per column is bold, second-best is underlined, "–" denotes an unavailable metric.

Method	Hypersim						FIPT-Synthetic							Time (min)
	Novel View Synthesis			Albedo			Novel View Synthesis			Albedo			Roughness	
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	MSE ↓	
IRIS (Lin et al., 2025)	20.01	0.68	0.36	11.29	0.43	0.52	**22.89**	**0.87**	**0.18**	14.51	0.77	0.23	0.096	60.0
GS-ID (Du et al., 2025)	19.47	0.75	0.25	17.67	0.73	0.31	18.26	0.78	0.21	15.04	0.76	0.26	0.124	55.0
IRGS (Gu et al., 2025)	21.22	0.80	0.23	11.78	0.66	0.37	22.27	0.82	0.20	11.39	0.67	0.31	0.149	**50.0**
NeILF++ (Zhang et al., 2023)	20.95	0.79	0.15	14.95	0.66	0.28	18.05	0.73	0.26	12.08	0.70	0.29	0.163	180.0
EAG-PT (Yang et al., 2026)	21.34	0.82	0.32	9.83	0.71	0.44	21.37	0.79	0.49	16.18	0.71	0.42	–	165.0
SGS-Intrinsic (Niu et al., 2026)	18.79	0.76	0.35	11.02	0.69	0.38	18.47	0.69	0.45	12.93	0.76	0.26	0.153	**50.0**
AEGIR (Ours)	**21.50**	**0.83**	**0.12**	**20.55**	**0.78**	**0.19**	22.53	0.86	0.19	**17.33**	**0.80**	**0.21**	**0.092**	55.0

As shown in Table 3, AEGIR achieves the best or tied-best novel view synthesis (NVS) performance across ScanNet++, Replica, and FIPT-Real in PSNR, SSIM, and LPIPS. Our method takes 55 minutes including geometry initialization, approximately three times faster than NeILF++ and comparable to other baselines. Because the real-world datasets lack ground-truth emitter parameters and complete

Table 3: **Quantitative evaluation on real-world datasets.** We report the performance of novel view synthesis and optimization times (in minutes) across the ScanNet++, Replica, and FIPT-Real benchmarks. Best per column is bold, second-best is underlined.

Method	ScanNet++			Replica			FIPT-Real			Time (min)
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	
IRIS (Lin et al., 2025)	15.22	0.80	0.23	17.41	0.75	0.27	<u>21.96</u>	<u>0.84</u>	**0.19**	60.0
GS-ID (Du et al., 2025)	<u>18.02</u>	<u>0.82</u>	<u>0.21</u>	17.05	0.73	0.29	20.31	0.82	<u>0.21</u>	<u>55.0</u>
IRGS (Gu et al., 2025)	17.48	<u>0.82</u>	0.39	13.64	0.63	0.34	21.23	0.81	0.26	**50.0**
NeILF++ (Zhang et al., 2023)	17.54	**0.83**	0.22	<u>17.79</u>	<u>0.77</u>	<u>0.25</u>	20.67	0.82	<u>0.21</u>	180.0
EAG-PT (Yang et al., 2026)	16.34	0.80	0.31	16.91	0.71	0.27	21.87	0.82	0.23	165.0
SGS-Intrinsic (Niu et al., 2026)	15.96	0.79	0.32	13.82	0.65	0.34	19.76	0.79	0.24	**50.0**
AEGIR (Ours)	**18.55**	**0.83**	**0.20**	**17.93**	**0.78**	**0.22**	**22.07**	**0.85**	**0.19**	<u>55.0</u>

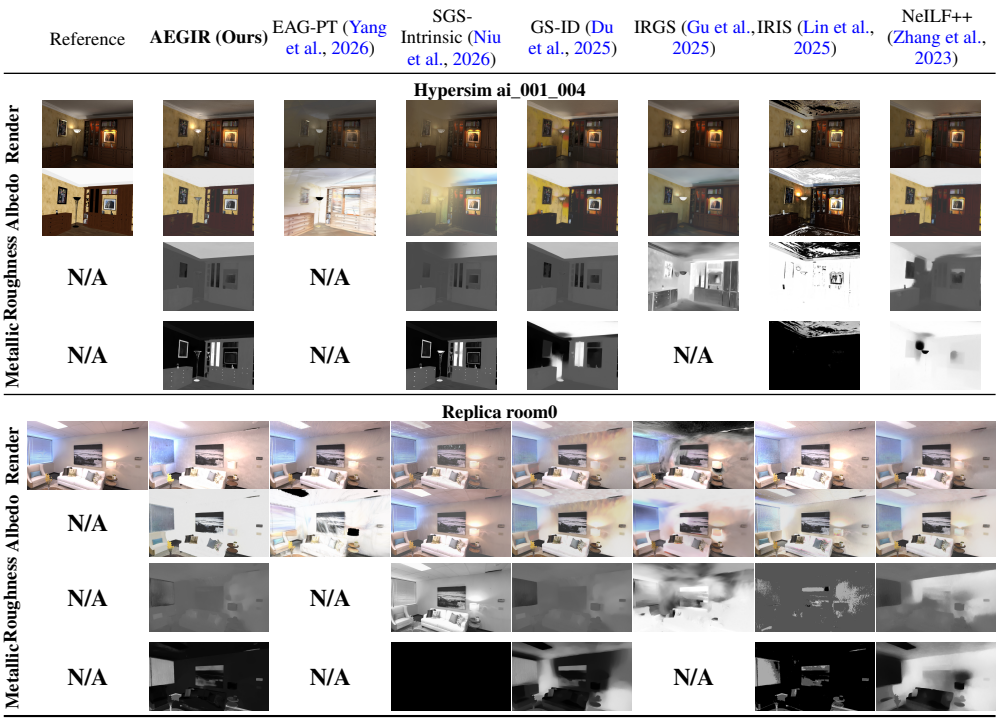


Figure 5: **Qualitative comparison of novel view synthesis and intrinsic decomposition.** Within each scene block, rows show the final render, albedo, roughness, and metallic maps. Columns show the reference and evaluated methods. "N/A" in the reference column indicates missing annotations in the dataset, while "N/A" in the results indicates outputs not produced by the corresponding method.

material annotations, we evaluate them using quantitative NVS and qualitative decomposition rather than absolute emitter errors. On synthetic datasets (Table 2), our method achieves the best overall performance on the Hypersim benchmark in both NVS and albedo recovery, outperforming the next best method by nearly 3 dB in albedo PSNR. On FIPT-Synthetic, our method remains competitive with IRIS, which achieves slightly higher NVS scores on this dataset by baking residual illumination into the surface albedo. In contrast, our approach maintains better material recovery without relying on such artifacts. This trade-off is further illustrated in Figure S4, which shows that baseline methods often trade material accuracy for rendering quality, while our method balances both. Additionally, we compare AEGIR with FIPT on FIPT-Synthetic using LDR views. FIPT-LDR achieves 17.06/0.71/0.34 PSNR/SSIM/LPIPS, whereas AEGIR reaches 22.53/0.86/0.19, despite FIPT-LDR additionally using ground-truth geometry and an emitter mask. Figure 5 shows sharper renderings and cleaner intrinsic maps, avoiding the blur and baked-in illumination artifacts common in other approaches. Appendix D.4 expands these comparisons in Figures S5 and S6. Moreover, Figure S7 shows full

AEGIR decomposition and projected recovered emitters. Geometry also improves (Appendix D.2, Table S2 and Figure S2). Stage-wise results show improvements over the raw material prior and further NVS gains from emitter refinement (Appendix D.3, Table S3), with multi-view albedo consistency illustrated in Figure S3. Controlled relighting evaluates recovered scenes under replacement illumination (Table S4). Figures S8-S10 demonstrate emitter editing, dynamic relighting, and object insertion (Appendix D.5).

4.4 Ablation Study

We conduct two complementary ablations. The **primitive ablation** (Table 4) tests how emitter representation affects NVS, albedo, and light source recovery, comparing points, isotropic area emitters, diffuse disks, anisotropic Gaussians, and full AEGIR representation. Position and Size measure source localization and spatial extent errors in normalized scene coordinates. Flux measures relative emitted power error, and average primitive count measures how compact the representation is. Appendix C.3 defines these metrics, source matching, and point light grouping. Appendix C.4 specifies the representation specific sampling rules with fixed pixel counts and iterations. Meanwhile, the **component ablation** (Table S5) removes individual losses, adaptive control, or emitter initialization to test their contributions to NVS, albedo accuracy, and qualitative illumination estimation and disentanglement.

Table 4: **Quantitative results of emitter primitive ablation.** We evaluate each emitter variant on FIPT-Synthetic scenes using representation specific sampling counts (Appendix C.4). P/S/L denotes PSNR/SSIM/LPIPS. Position and Size errors are measured in normalized scene coordinates, and Flux denotes relative emitted power error (Appendix C.3). Best results in each column are shown in bold.

Emitter	NVS P/S/L ↑ / ↑ / ↓	Albedo P/S/L ↑ / ↑ / ↓	Position ↓	Size ↓	Flux ↓	Avg. count
Point	20.63/0.77/0.36	16.55/0.75/0.28	0.186	0.407	0.691	252.1
Isotropic area	20.89/0.77/0.35	16.98/0.78/0.27	0.178	0.191	0.783	53.4
Diffuse 2DGS disk	20.01/0.73/0.33	14.97/0.71/0.38	0.108	0.173	0.792	1165.9
Anisotropic Gaussian ($\rho = 1$)	21.43/0.80/0.28	17.15/0.79/0.22	0.106	0.072	0.650	14.3
Full AEGIR (learned ρ)	**22.53/0.86/0.19**	**17.33/0.80/0.21**	**0.086**	**0.065**	**0.464**	**12.8**

In the primitive ablation (Table 4), isotropic area emitters improve NVS and albedo PSNR over point lights while requiring fewer primitives. Diffuse 2DGS disks improve localization and size recovery over isotropic emitters but yield worse rendering and albedo quality. Anisotropic Gaussians improve all image metrics and emitter errors over diffuse disks. Full AEGIR further achieves the best NVS, albedo, and emitter recovery using only 12.8 primitives on average. Learning ρ improves NVS by 1.10 dB and albedo by 0.18 dB PSNR over fixed $\rho = 1$, while also improving emitter recovery and compactness. These results show that the overall decomposition gains come from the complete formulation. Appendix D.6 details both ablations, with primitive ablation visuals in Figure S11 followed by the component ablation table (Table S5) and visuals (Figure S12).

5 Conclusion and Future Work

We present AEGIR, an inverse rendering framework that models explicit local area emitters for physically grounded illumination estimation. By modeling the spatial extent, shape, and anisotropic emission of real-world light sources within a 2D Gaussian Splatting representation, AEGIR bridges neural scene modeling with physically based rendering. Combined with differentiable soft shadow tracing, multiple importance sampling, and regularization for materials and lighting, our approach reduces inverse rendering ambiguity and reduces illumination baking into albedo. This results in a more consistent decomposition of geometry, materials, and lighting, with strong performance in material recovery, illumination estimation, and novel view synthesis.

Finally, several technical limitations remain. Individual ellipsoids cannot capture sharp polygonal fixture boundaries, and noisy reconstructed normals can introduce shading artifacts. Runtime increases with emitter count, while the opaque microfacet model excludes transparent and mirror-like transport. Appendix E discusses these limitations and future directions. Despite these limitations, AEGIR demonstrates effective illumination recovery under the complex lighting conditions.

AI USE STATEMENT

AI tools were used for manuscript editing and organization, including clarity and consistency checks of the written methods and results, as well as feedback on the presentation of methodology and experiments. The authors reviewed all AI-assisted outputs and take responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

Appendix A specifies the rendering model, sampling densities, and alternative ReSTIR integration. Appendix B defines the optimization objectives. Appendix C documents the datasets and scene selection, baseline adaptations, evaluation protocols, training settings, and runtime reporting. Appendix D provides additional quantitative and qualitative evaluations supporting the main results. Code, optimization checkpoints, and evaluation scripts will be released upon publication.

REFERENCES

Benedikt Bitterli. Rendering resources, 2016. https://benedikt-bitterli.me/resources/.

Benedikt Bitterli, Chris Wyman, Matt Pharr, Peter Shirley, Aaron Lefohn, and Wojciech Jarosz. Spatiotemporal reservoir resampling for real-time ray tracing with dynamic direct lighting. *ACM Trans. Graph.*, 39(4), August 2020. ISSN 0730-0301. doi: 10.1145/3386569.3392481. URL https://doi.org/10.1145/3386569.3392481.

Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lensch. Nerd: Neural reflectance decomposition from image collections. In *IEEE International Conference on Computer Vision (ICCV)*, 2021a.

Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan T. Barron, and Hendrik P.A. Lensch. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021b.

Arno Coomans, Edoardo Alberto Dominici, Christian Döring, Joerg H. Mueller, Jozef Hladky, and Markus Steinberger. Real-time Neural Rendering of Dynamic Light Fields. *Computer Graphics Forum*, 2024. ISSN 1467-8659. doi: 10.1111/cgf.15014.

Kang Du, Zhihao Liang, Yulin Shen, and Zeyu Wang. Gs-id: Illumination decomposition on gaussian splatting via adaptive light aggregation and diffusion-guided material priors, 2025. URL https://arxiv.org/abs/2408.08524.

Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pp. 226–231. AAAI Press, 1996.

Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing. *arXiv:2311.16043*, 2023.

Marc-Andre Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagne, and Jean-Francois Lalonde. Deep parametric indoor lighting estimation. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-Francois Lalonde. Fast spatially-varying indoor lighting estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

Chun Gu, Xiaofei Wei, Zixuan Zeng, Yuxuan Yao, and Li Zhang. Irgs: Inter-reflective gaussian splatting with 2d gaussian ray tracing. In *CVPR*, 2025.

Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. Shape, Light, and Material Decomposition from Images using Monte Carlo Rendering and Denoising. *arXiv:2206.03380*, 2022.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery, 2024. doi: 10.1145/3641519.3657428.

Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. Mitsuba 3 renderer, 2022. https://mitsuba-renderer.org.

Jinseo Jeong, Junseo Koo, Qimeng Zhang, and Gunhee Kim. Esr-nerf: Emissive source reconstruction using ldr multi-view images. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4598–4609, 2024. doi: 10.1109/CVPR52733.2024.00440.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), July 2023. URL https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

Peter Kocsis, Vincent Sitzmann, and Matthias Nießner. Intrinsic image diffusion for indoor single-view material estimation. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Philipp Langsteiner, Jan-Niklas Dihlmann, and Hendrik P. A. Lensch. Matspray: Fusing 2d material world knowledge on 3d geometry, 2025. URL https://arxiv.org/abs/2512.18314.

Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image, 2019. URL https://arxiv.org/abs/1905.02722.

Zhengqin Li, Li Yu, Mikhail Okunev, Manmohan Chandraker, and Zhao Dong. Spatiotemporally consistent hdr indoor lighting estimation, 2023. URL https://arxiv.org/abs/2305.04374.

Zhibing Li, Tong Wu, Jing Tan, Mengchen Zhang, Jiaqi Wang, and Dahua Lin. IDArb: Intrinsic decomposition for arbitrary number of input views and illuminations. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=uuef1HP6X7.

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 *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2025a.

Ruofan Liang, Kai He, Zan Gojcic, Igor Gilitschenski, Sanja Fidler, Nandita Vijaykumar, and Zian Wang. Luxdit: Lighting estimation with video diffusion transformer. *arXiv preprint arXiv:2509.03680*, 2025b.

Ruofan Liang, Norman Müller, Ethan Weber, Duncan Zauss, Nandita Vijaykumar, Peter Kontschieder, and Christian Richardt. Luxremix: Lighting decomposition and remixing for indoor scenes, 2026. URL https://arxiv.org/abs/2601.15283.

Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering, 2024. URL https://arxiv.org/abs/2311.16473.

Chih-Hao Lin, Jia-Bin Huang, Zhengqin Li, Zhao Dong, Christian Richardt, Tuotuo Li, Michael Zollhöfer, Johannes Kopf, Shenlong Wang, and Changil Kim. IRIS: Inverse rendering of indoor scenes from low dynamic range images. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Jundan Luo, Duygu Ceylan, Jae Shin Yoon, Nanxuan Zhao, Julien Philip, Anna Frühstück, Wenbin Li, Christian Richardt, and Tuanfeng Y. Wang. IntrinsicDiffusion: Joint intrinsic layers from latent diffusion models. In *SIGGRAPH 2024 Conference Papers*, 2024. doi: 10.1145/3641519.3657472. URL https://intrinsicdiffusion.github.io.

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 *ECCV*, 2020.

Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. Extracting Triangular 3D Models, Materials, and Lighting From Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8280–8290, June 2022.

Jiahao Niu, Rongjia Zheng, Wenju Xu, Wei-Shi Zheng, and Qing Zhang. Sgs-intrinsic: Semantic-invariant gaussian splatting for sparse-view indoor inverse rendering, 2026. URL https://arxiv.org/abs/2603.27516.

Steven G. Parker, James Bigler, Andreas Dietrich, Heiko Friedrich, Jared Hoberock, David Luebke, David McAllister, Morgan McGuire, Keith Morley, Austin Robison, and Martin Stich. Optix: a general purpose ray tracing engine. *ACM Trans. Graph.*, 29(4), July 2010. ISSN 0730-0301. doi: 10.1145/1778765.1778803. URL https://doi.org/10.1145/1778765.1778803.

Pakkapon Phongthawee, Worameth Chinchuthakun, Nontaphat Sinsunthithet, Amit Raj, Varun Jampani, Pramook Khungurn, and Supasorn Suwajanakorn. Diffusionlight: Light probes for free by painting a chrome ball. In *ArXiv*, 2023.

Mike Roberts, Jason Ramapuram, Anurag Ranjan, Atulit Kumar, Miguel Angel Bautista, Nathan Paczan, Russ Webb, and Joshua M. Susskind. Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In *International Conference on Computer Vision (ICCV) 2021*, 2021.

Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J. Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, Anton Clarkson, Mingfei Yan, Brian Budge, Yajie Yan, Xiaqing Pan, June Yon, Yuyang Zou, Kimberly Leon, Nigel Carter, Jesus Briales, Tyler Gillingham, Elias Mueggler, Luis Pesqueira, Manolis Savva, Dhruv Batra, Hauke M. Strasdat, Renzo De Nardi, Michael Goesele, Steven Lovegrove, and Richard Newcombe. The Replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019.

Hanxiao Sun, YuPeng Gao, Jin Xie, Jian Yang, and Beibei Wang. Svg-ir: Spatially-varying gaussian splatting for inverse rendering, 2025. URL https://arxiv.org/abs/2504.06815.

Mutian Tong, Rundi Wu, and Changxi Zheng. Spatiotemporally consistent indoor lighting estimation with diffusion priors. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, SIGGRAPH Conference Papers '25, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715402. doi: 10.1145/3721238.3730749. URL https://doi.org/10.1145/3721238.3730749.

Matias Turkulainen, Xuqian Ren, Iaroslav Melekhov, Otto Seiskari, Esa Rahtu, and Juho Kannala. Dn-splatter: Depth and normal priors for gaussian splatting and meshing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025.

Yu-Chen Wang, Chris Wyman, Lifan Wu, and Shuang Zhao. Amortizing samples in physics-based inverse rendering using restir. *ACM Transactions on Graphics (SIGGRAPH Asia)*, 42(6):1–17, December 2023. doi: 10.1145/3618331.

Liwen Wu, Rui Zhu, Mustafa B Yaldiz, Yinhao Zhu, Hong Cai, Janarbek Matai, Fatih Porikli, Tzu-Mao Li, Manmohan Chandraker, and Ravi Ramamoorthi. Factorized inverse path tracing for efficient and accurate material-lighting estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3848–3858, 2023.

Xiangzuo Wu, Chengwei Ren, Jun Zhou, Xiu Li, and Yuan Liu. Mvinverse: Feed-forward multi-view inverse rendering in seconds, 2025. URL https://arxiv.org/abs/2512.21003.

Xiaoyan Xing, Konrad Groh, Sezer Karaoglu, Theo Gevers, and Anand Bhattad. Luminet: Latent intrinsics meets diffusion models for indoor scene relighting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 442–452, 2025.

Xijie Yang, Mulin Yu, Changjian Jiang, Kerui Ren, Tao Lu, Jiangmiao Pang, Dahua Lin, Bo Dai, and Linning Xu. Eag-pt: Emission-aware gaussians and path tracing for diffuse indoor scene reconstruction and editing. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, SIGGRAPH Conference Papers '26,

New York, NY, USA, 2026. Association for Computing Machinery. ISBN 9798400725548. doi: 10.1145/3799902.3811054. URL https://doi.org/10.1145/3799902.3811054.

Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Neilf: Neural incident light field for physically-based material estimation. In *European Conference on Computer Vision (ECCV)*, 2022.

Yuxuan Yao, Zixuan Zeng, Chun Gu, Xiatian Zhu, and Li Zhang. Reflective gaussian splatting. In *ICLR*, 2025.

Kai Ye, Chong Gao, Guanbin Li, Wenzheng Chen, and Baoquan Chen. Geosplatting: Towards geometry guided gaussian splatting for physically-based inverse rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 28991–29000, 2025.

Chandan Yeshwanth, Yueh-Cheng Liu, Matthias Nießner, and Angela Dai. Scannet++: A high-fidelity dataset of 3d indoor scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12–22, 2023.

Ye Yu and William AP Smith. Inverserendernet: Learning single image inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. Rgb↔x: Image decomposition and synthesis using material- and lighting-aware diffusion models. In *ACM SIGGRAPH 2024 Conference Papers*, SIGGRAPH '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705250. doi: 10.1145/3641519.3657445. URL https://doi.org/10.1145/3641519.3657445.

Jingyang Zhang, Yao Yao, Shiwei Li, Jingbo Liu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Neilf++: Inter-reflectable light fields for geometry and material estimation. *International Conference on Computer Vision (ICCV)*, 2023.

Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. PhySG: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021a.

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

Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph.*, 40(6), dec 2021b. ISSN 0730-0301. doi: 10.1145/3478513.3480496. URL https://doi.org/10.1145/3478513.3480496.

Jingsen Zhu, Fujun Luan, Yuchi Huo, Zihao Lin, Zhihua Zhong, Dianbing Xi, Rui Wang, Hujun Bao, Jiaxiang Zheng, and Rui Tang. Learning-based inverse rendering of complex indoor scenes with differentiable monte carlo raytracing. In *SIGGRAPH Asia 2022 Conference Papers*. ACM, 2022a. URL https://doi.org/10.1145/3550469.3555407.

Rui Zhu, Zhengqin Li, Janarbek Matai, Fatih Porikli, and Manmohan Chandraker. Irisformer: Dense vision transformers for single-image inverse rendering in indoor scenes, 2022b. URL https://arxiv.org/abs/2206.08423.

SUPPLEMENT OVERVIEW

- **Appendix A: Rendering model.** A.1 defines the microfacet BRDF. A.2 derives emitter surface sampling and MIS densities. A.3 details shadow tracing and indirect illumination. A.4 specifies the alternative ReSTIR training sampler.
- **Appendix B: Optimization objectives.** B.1 defines the contrast-aware photometric loss. B.2 the diffusion material priors. B.3 edge-aware material TV. B.4 emitter regularization.
- **Appendix C: Experimental protocol.** C.1 datasets. C.2 baselines. C.3 metrics. C.4 training settings. C.5 runtime and seed variability.
- **Appendix D: Additional evaluations.**
  - D.1 direct illumination maps.
  - D.2 geometry refinement.
  - D.3 optimization stages and material priors.
  - D.4 NVS and decomposition.
  - D.5 relighting and insertion.
  - D.6 ablations.
  - D.7 emitter statistics.
  - D.8 ReSTIR comparison.
- **Appendix E: Limitations and future work.** Discusses emitter representation ambiguity, geometry and material transport, indirect illumination, and scalability.

A Rendering Model Details

This section details the deferred rendering pipeline of AEGIR, including the microfacet BRDF, emitter surface sampling density, multiple importance sampling (MIS), differentiable visibility, and final pixel evaluation.

Core rendering notation	Meaning
$\mathbf{x}, \mathbf{n}$	Surface point and surface normal
$\mathbf{v}, \boldsymbol{\omega}$	View and incident directions
$\mathbf{a}, r, m$	RGB albedo, roughness, and metallic
$\mathbf{c}_e, \mathbf{s}_e$	Emitter center and spatial scales
$\mathbf{u}_e, \mathbf{v}_e, \mathbf{w}_e$	Local emitter orthonormal axes
$\mathbf{E}_e, \boldsymbol{\sigma}_e, \rho_e$	RGB emission, angular spread, and angular falloff power
$\mathbf{q}, \mathbf{p}_e$	Unit-sphere sample and mapped emitter surface point
$\mathbf{L}_i, \mathbf{L}_o, \mathbf{f}_r$	Incident radiance, outgoing radiance, and microfacet BRDF
N_{brdf}	Number of cosine-weighted hemisphere samples
$N_{\text{light}}, N'_{\text{light}}$	Requested and actual emitter sample counts

A.1 Microfacet BRDF Model

AEGIR uses a standard RGB microfacet BRDF. The reflectance $\mathbf{f}_r(\mathbf{v}, \boldsymbol{\omega}) = \mathbf{f}_d + \mathbf{f}_s$ for view direction $\mathbf{v}$ and incident direction $\boldsymbol{\omega}$ contains diffuse and specular lobes. The diffuse component is

$$\mathbf{f}_d = \frac{\mathbf{a}(1-m)}{\pi}, \tag{S1}$$

where $\mathbf{a}$ is albedo and m is metallic. The GGX specular component is

$$\mathbf{f}_s = \frac{D(\mathbf{h}, \alpha_r)\, G(\mathbf{v}, \boldsymbol{\omega}, \alpha_r)\, \mathbf{F}(\mathbf{v}, \mathbf{h})}{4(\mathbf{n}\cdot\mathbf{v})_+(\mathbf{n}\cdot\boldsymbol{\omega})_+ + \epsilon}, \tag{S2}$$

where $\mathbf{n}$ is the surface normal at $\mathbf{x}$, $\mathbf{h}$ is the half-vector, $\alpha_r = r^2$ is derived from roughness r, $(\cdot)_+ = \max(\cdot, 0)$, and ϵ is a small constant that prevents division by zero. D, G, and $\mathbf{F}$ are the GGX normal distribution, Smith shadowing-masking, and Schlick Fresnel terms, respectively, with $\mathbf{F}_0 = 0.04(1-m)\mathbf{1} + \mathbf{a}m$.

A.2 Emitter-Surface Sampling and MIS PDFs

Outgoing radiance $\mathbf{L}_o(\mathbf{x}, \mathbf{v})$ at a surface point $\mathbf{x}$ is estimated using cosine-weighted hemisphere samples and explicit emitter samples. Let N_{brdf} be the number of hemisphere samples and N_{light} the requested budget of explicit-emitter samples. For M active emitters, we allocate

$$K = \left\lfloor \frac{N_{\text{light}}}{M} \right\rfloor, \qquad N'_{\text{light}} = MK, \qquad N = N_{\text{brdf}} + N'_{\text{light}} \tag{S3}$$

samples, so that every emitter receives K samples. In our standard AEGIR experiments, $M < N_{\text{light}}$ throughout optimization, as the requested budget was chosen to accommodate the emitter counts.

The Monte Carlo estimator is

$$\mathbf{L}_o(\mathbf{x}, \mathbf{v}) \approx \frac{1}{N} \sum_{j=1}^{N} \frac{\mathbf{f}_r(\mathbf{v}, \boldsymbol{\omega}_j)\mathbf{L}_i(\mathbf{x}, \boldsymbol{\omega}_j)(\mathbf{n}\cdot\boldsymbol{\omega}_j)_+}{p_{\text{mix}}(\boldsymbol{\omega}_j, \mathbf{x})}. \tag{S4}$$

The cosine-weighted hemisphere density is

$$p_{\text{brdf}}(\boldsymbol{\omega}) = \frac{(\mathbf{n}\cdot\boldsymbol{\omega})_+}{\pi}. \tag{S5}$$

Mapping a Unit-Sphere Sample to an Ellipsoid. For emitter e, let $\mathbf{c}_e$ be its center, $\{\mathbf{u}_e, \mathbf{v}_e, \mathbf{w}_e\}$ its local orthonormal axes, and $\mathbf{s}_e = (s_{e,u}, s_{e,v}, s_{e,w})$ its spatial scales. For simple, tractable light sampling, we draw $\mathbf{q} = (q_u, q_v, q_w)$ uniformly on the unit sphere, $\|\mathbf{q}\|_2 = 1$, and map it to the emitter surface:

$$\mathbf{p}_e(\mathbf{q}) = \mathbf{c}_e + \frac{s_{e,u}}{2} q_u \mathbf{u}_e + \frac{s_{e,v}}{2} q_v \mathbf{v}_e + \frac{s_{e,w}}{2} q_w \mathbf{w}_e. \tag{S6}$$

Conversely, for $k \in \{u, v, w\}$ with emitter axis $\mathbf{k}_e$, the corresponding unit-sphere coordinate of a surface point $\mathbf{p}$ is

$$q_{e,k}(\mathbf{p}) = \frac{2(\mathbf{p} - \mathbf{c}_e) \cdot \mathbf{k}_e}{s_{e,k}}. \tag{S7}$$

Induced Ellipsoid-Area Density. Anisotropic scaling makes the mapped samples nonuniform with respect to ellipsoid surface area. The induced density is

$$p_{\text{area},e}(\mathbf{p}) = \frac{1}{\pi s_{e,u} s_{e,v} s_{e,w} \sqrt{\dfrac{q_{e,u}(\mathbf{p})^2}{s_{e,u}^2} + \dfrac{q_{e,v}(\mathbf{p})^2}{s_{e,v}^2} + \dfrac{q_{e,w}(\mathbf{p})^2}{s_{e,w}^2}}}. \tag{S8}$$

This expression is evaluated directly from the mapped point and does not require an explicit matrix inverse or determinant.

Conversion to Solid Angle. At the surface point $\mathbf{x}$, the sampled incident direction is

$$\boldsymbol{\omega}_{i,e} = \frac{\mathbf{p}_e - \mathbf{x}}{\|\mathbf{p}_e - \mathbf{x}\|_2}. \tag{S9}$$

Let $\mathbf{n}_{l,e}(\mathbf{p}_e)$ be the outward geometric normal. We reject back-facing samples with $\mathbf{n}_{l,e}(\mathbf{p}_e) \cdot (-\boldsymbol{\omega}_{i,e}) \leq 0$ without replacement, so they contribute zero but remain in the drawn sample count. The area density is not renormalized over accepted samples. For a retained front-facing point, the conversion from area to solid angle is

$$p_{\text{light},e}(\boldsymbol{\omega}_{i,e}, \mathbf{x}) = p_{\text{area},e}(\mathbf{p}_e) \frac{\|\mathbf{p}_e - \mathbf{x}\|_2^2}{|\mathbf{n}_{l,e}(\mathbf{p}_e) \cdot (-\boldsymbol{\omega}_{i,e})|}. \tag{S10}$$

The squared distance and projected cosine factors convert the probability of sampling a surface region into the probability of sampling its corresponding direction from $\mathbf{x}$. In particular, the squared distance term produces the inverse-square distance attenuation when the estimator divides by the solid-angle density.

For a shading point, the near and far intersections of the same convex ellipsoid share a direction, but only the front-facing intersection is retained. For BRDF-sampling rays, the nearest front-facing emitter intersection supplies the emission and visibility distance. A ray missing all emitters has zero emitter-sampling density, without implying zero BRDF density or indirect illumination.

Because emitters are sampled uniformly and each receives K samples, the emitter-sampling density is the uniform mixture

$$p_{\text{light}}(\boldsymbol{\omega}, \mathbf{x}) = \frac{1}{M} \sum_{e=1}^{M} p_{\text{light},e}(\boldsymbol{\omega}, \mathbf{x}). \tag{S11}$$

The sample-weighted MIS mixture is

$$p_{\text{mix}}(\boldsymbol{\omega}, \mathbf{x}) = \frac{N_{\text{brdf}} p_{\text{brdf}}(\boldsymbol{\omega}) + N'_{\text{light}} p_{\text{light}}(\boldsymbol{\omega}, \mathbf{x})}{N_{\text{brdf}} + N'_{\text{light}}}. \tag{S12}$$

Equations S3-S12 make explicit that the actual number N'_{light}, rather than the requested budget before division across emitters, weights the light sampling strategy.

A.3 Differentiable Shadow Tracing and Illumination

AEGIR traces shadow rays through the 2D Gaussian geometry. Visibility is a soft differentiable mask obtained by comparing the sampled light distance against the traced geometry depth:

$$V_i(\mathbf{x}, \boldsymbol{\omega}_i) = 1 - \alpha_{\text{trace}} \, \text{sigmoid}(d_{\text{light}} - \epsilon - d_{\text{trace}}) \, . \tag{S13}$$

Because the sigmoid steepness depends on scene scale, we normalize scenes before optimization, which is equivalent to using a scene-dependent steepness. Here α_{trace} is accumulated opacity and d_{trace} is the corresponding normalized transmittance-weighted mean intersection depth, not a first-hit depth. This sigmoid is a smooth visibility approximation. The total incident light is

$$\begin{aligned} \mathbf{L}_i(\mathbf{x}, \boldsymbol{\omega}_i) = V_i(\mathbf{x}, \boldsymbol{\omega}_i) \mathbf{L}_i^{\text{direct}}(\mathbf{x}, \boldsymbol{\omega}_i) \\ + \big(1 - V_i(\mathbf{x}, \boldsymbol{\omega}_i)\big) \mathbf{L}_i^{\text{trace}}(\mathbf{x}, \boldsymbol{\omega}_i). \end{aligned} \tag{S14}$$

During optimization, the indirect component follows IRGS (Gu et al., 2025), where frozen Spherical Harmonic (SH) color features are queried at the Gaussian splats intersected by traced rays. This proxy stabilizes optimization and reduces illumination leakage into the explicit emitters. Direct illumination usually dominates except where emitters are fully occluded. We use the fixed proxy because tracing indirect bounces during optimization was computationally expensive and unstable in our exploratory tests.

For controlled relighting and editing, the target emitters are selected or replaced, the frozen SH lighting is discarded, and rendering uses the edited explicit emitters with direct light plus one path-traced indirect bounce. The SH features encode the original observed illumination and therefore are not reused after a lighting edit. Finally, Equation S4 aggregates the incident samples into HDR $\mathbf{L}_o(\mathbf{x}, \mathbf{v})$, which is mapped to LDR for loss calculation.

A.4 ReSTIR-Based Training Sampler

Following (Wang et al., 2023), we integrate ReSTIR (Bitterli et al., 2020) into the training-time direct light branch of AEGIR, between deferred rasterization and ray tracing. It replaces part of explicit emitter next-event estimation (NEE), alongside cosine-weighted diffuse sampling. The emitter representation, G-buffer, and optimization losses are unchanged.

Sampling and differentiation. For an emitter surface candidate $\mathbf{y}_j$ viewed from shading point $\mathbf{x}$, the visibility-free target is

$$\widehat{p}_0(\mathbf{y}_j, \mathbf{x}) = \text{lum}[\mathbf{L}_e(\mathbf{y}_j) \, \mathbf{f}_r(\mathbf{v}, \boldsymbol{\omega}_j) \, (\mathbf{n} \cdot \boldsymbol{\omega}_j)_+] \, , \tag{S15}$$

where $\boldsymbol{\omega}_j$ points toward $\mathbf{y}_j$, $\mathbf{v}$ is the view direction, $\mathbf{L}_e$ is the radiance from Equation 1, and lum denotes the perceptually weighted brightness of linear RGB. Fresh candidates use weights $w(\mathbf{y}_j) = \widehat{p}_0(\mathbf{y}_j, \mathbf{x}) / p_{\text{prop}}(\mathbf{y}_j)$, where p_{prop} is the training proposal solid-angle PDF surrogate, including emitter selection and area to solid angle conversion. A reused sample is reconstructed under current emitter parameters, reweighted at $\mathbf{x}$, and merged using its previous normalization and capped effective count. Resampling the merged set $\mathcal{C}$ gives

$$W(\mathbf{y}) = \frac{\sum_{\mathbf{z} \in \mathcal{C}} w(\mathbf{z})}{M_{\text{eff}} \, \widehat{p}_0(\mathbf{y}, \mathbf{x})}, \tag{S16}$$

where M_{eff} is the effective candidate count, including the capped count of the reused reservoir. The selected sample contributes

$$\widehat{\mathbf{L}}_{\text{res}}(\mathbf{x}) = \text{stopgrad}(W(\mathbf{y})) \, \mathbf{h}(\mathbf{y}, \mathbf{x}), \tag{S17}$$

where $\mathbf{h}$ evaluates AEGIR direct light contribution with soft visibility. Reservoir selection and W are detached, such that reconstructing the selected point from emitter local coordinates preserves gradients to emitter and material parameters. Visibility is traced only for selected candidates.

Reuse and tested settings. A persistent world-space hash table reuses emitter samples across views and iterations, storing emitter indices and local coordinates. We clear it after splitting/pruning and every 1,000 iterations outside adaptive control. Each shading point traces 256 rays: 128 cosine-weighted and 128 direct-light rays, matching the standard requested budget (its realized light-ray count follows Equation S3). The latter comprise 16 reservoir samples and 112 fresh NEE samples. The reservoir pool combines 32 fresh candidates (29 flux-weighted and three uniform over emitters when multiple emitters exist) with one valid previous sample, if available. We draw 16 samples with replacement, capping the previous effective count at 640. Reservoir and fresh NEE means are combined with weights $1/8$ and $7/8$, respectively. Later in optimization, we gradually replace fresh NEE samples with samples resampled from reservoirs, while retaining fresh reservoir candidates for exploration.

Scope and approximations. Indirect contributions retain standard MIS weights despite reservoir sampling, and uniform exploration candidates retain the flux-based surrogate. These approximations mean that we adopt Wang et al. method for interior direct light reuse, using hashed cells instead of nearest neighbor search and differentiating soft visibility directly. Appendix D.8 reports the results of the comparison.

B Loss and Regularization Details

This section presents the optimization objectives used in AEGIR. Inverse rendering is highly ill-posed, as different combinations of geometry, material, and illumination can produce the same image, so carefully designed losses are needed to guide the solution toward physically consistent results. This becomes even more challenging with explicit area emitters, which introduce additional degrees of freedom in shape, size, and emission, making optimization more difficult. The total optimization loss is defined in the main paper.

B.1 Contrast-Aware Photometric Loss

Standard photometric losses often blur high-frequency details. We introduce a local contrast-aware focus map W_f to prioritize the reconstruction of sharp lighting patterns. We define ΔI as the absolute difference between the grayscale ground-truth image I_{gray} and its local average,

$$\Delta I = \left| I_{\text{gray}} - \text{AvgPool}(I_{\text{gray}}) \right|, \tag{S18}$$

and the focus map is

$$W_f = 1 + \lambda_{focus} \frac{\Delta I}{\max(\Delta I) + \epsilon_f}, \tag{S19}$$

where ϵ_f is a small constant. The final objective combines the W_f-weighted ℓ_1 loss with the Structural Similarity Index Measure (SSIM):

$$\mathcal{L}_{photo} = \|W_f \odot (\hat{\mathbf{I}} - \mathbf{I})\|_1 + \lambda_{dssim}\big(1 - \text{SSIM}(\hat{\mathbf{I}}, \mathbf{I})\big), \tag{S20}$$

where $\hat{\mathbf{I}}$ and $\mathbf{I}$ are the rendered and ground-truth RGB images, respectively.

B.2 Material Supervision with Diffusion Priors

To resolve ambiguities, such as distinguishing a white wall in shadow from a dark wall, we use pseudo ground truth maps for albedo $\mathbf{a}$, roughness r, and metallic m extracted using DiffusionRenderer (Liang et al., 2025a). We apply ℓ_1 regularization between our rendered G-buffer maps and these priors:

$$\mathcal{L}_{prior} = \sum_{P \in \{\mathbf{a}, r, m\}} \lambda_P \|P_{\text{render}} - P_{\text{prior}}\|_1. \tag{S21}$$

B.3 Edge-Aware Material Total Variation (TV)

To mitigate multi-view noise introduced by the diffusion priors, we apply an edge-aware total variation (TV) loss. This loss encourages materials to remain constant across flat surfaces while preserving sharp transitions at structural boundaries. Smoothing the material textures discourages them from absorbing illumination gradients and encourages the light emitters to account for these lighting effects:

$$\mathcal{L}_{TV} = \lambda_{TV} \sum_{P \in \{\mathbf{a}, r, m\}} \gamma_P \sum_{i,j} \|\nabla P_{i,j}\| \exp\left(-\beta \|\nabla \mathbf{I}_{i,j}\|\right), \tag{S22}$$

where $P \in \{\mathbf{a}, r, m\}$ denotes a material map. The gradient $\nabla P_{i,j}$ is evaluated at pixel (i, j), γ_P is the material-specific TV weight, and β controls sensitivity to the visual edges $\nabla \mathbf{I}$ in the ground-truth image.

B.4 Area Emitter Regularization

To keep the lighting realistic, we apply three constraints to the explicit emitters. First, the area regularizer ($\mathcal{L}_{area}$) uses the squared norm of each spatial scale vector $\mathbf{s}_e$ as a compact proxy for emitter extent, discouraging excessive growth of the M emitters:

$$\mathcal{L}_{area} = \frac{1}{M} \sum_{e=1}^{M} \|\mathbf{s}_e\|_2^2. \tag{S23}$$

Second, a soft white light prior ($\mathcal{L}_{white}$) penalizes differences across the color channels c of the RGB emission $\mathbf{E}_e$. This discourages the optimization from explaining material colors with unrealistic lighting tints, while still allowing the recovery of actual colored light sources:

$$\mathcal{L}_{white} = \frac{1}{M} \sum_{e=1}^{M} \left(\max_c E_{e,c} - \min_c E_{e,c} \right). \tag{S24}$$

Third, a bounding loss ($\mathcal{L}_{bounds}$) penalizes emitters outside the scene. It penalizes deviations of each center $\mathbf{c}_e$ from the scene bounding box $[\mathbf{b}_{\min}, \mathbf{b}_{\max}]$ across dimensions $d \in \{x, y, z\}$:

$$\begin{aligned} \mathcal{L}_{bounds} = \frac{1}{M} \sum_{e=1}^{M} \sum_{d \in \{x,y,z\}} & \left(\max(0, b_{\min,d} - c_{e,d}) \right. \\ & \left. + \max(0, c_{e,d} - b_{\max,d}) \right). \end{aligned} \tag{S25}$$

Finally, the total emitter regularization loss is a weighted sum of these three constraints:

$$\mathcal{L}_{light} = \lambda_{area}\mathcal{L}_{area} + \lambda_{white}\mathcal{L}_{white} + \lambda_{bounds}\mathcal{L}_{bounds}. \tag{S26}$$

C Implementation and Optimization Details

This section details the experimental setup used to evaluate the AEGIR pipeline, including the selected datasets, evaluation metrics, training schedules, adaptive emitter control strategies, and hyperparameter configurations, to support reproducibility.

C.1 Datasets

We evaluate 24 scenes, including 10 synthetic scenes (6 Hypersim and 4 FIPT-Synthetic) and 14 real/scanned scenes (4 ScanNet++, 8 Replica, and 2 FIPT-Real). The selected Hypersim (Roberts et al., 2021) scenes are `ai_001_004`, `ai_004_002`, `ai_010_006`, `ai_015_003`, `ai_032_001`, and `ai_034_002`. The ScanNet++ (Yeshwanth et al., 2023) scenes are `0a7cc12c0e`, `8b5caf3398`, `a24858e51e`, and `b20a261fdf`. For the controlled lighting only evaluation, we reuse the same four Benedikt Bitterli's scenes (Bitterli, 2016) in the FIPT-Synthetic split (Wu et al., 2023). Table S1 summarizes the protocol.

Table S1: **Datasets and evaluation protocol.** Scene counts, identifiers or sources, and evaluation targets for each benchmark. All methods use identical train/test splits within each dataset.

Dataset	Scenes	Scene identifiers / source	Evaluation target
Hypersim	6	`ai_001_004`, `ai_004_002`, `ai_010_006`, `ai_015_003`, `ai_032_001`, `ai_034_002`	NVS, albedo
FIPT-Synthetic	4	FIPT synthetic split	NVS, albedo, roughness, ablations
FIPT-Real	2	FIPT real split	NVS
Replica	8	Replica indoor scans	NVS, ablations
ScanNet++	4	`0a7cc12c0e`, `8b5caf3398`, `a24858e51e`, `b20a261fdf`	NVS
Bitterli's scenes	4	Same FIPT-Synthetic scenes	Mitsuba lighting only evaluation

C.2 Baselines

Since the official GS-ID implementation was unavailable for these experiments, we reproduce its hybrid illumination formulation and deshadowing model from the published method, using the same DN-Splatter initialization as AEGIR. All methods use identical train/test camera splits. Native FIPT (Wu et al., 2023) assumes known geometry and multi-view HDR observations, whereas AEGIR takes posed LDR images and initializes geometry. We therefore use IRIS (Lin et al., 2025) as the primary matched LDR baseline because it targets multi-view LDR inverse rendering on the same FIPT-Synthetic scenes and its paper reports better results than the FIPT-LDR adaptation. We also run FIPT-LDR on the four FIPT-Synthetic scenes with identical LDR observations, poses, splits, and held-out NVS metrics. Following the IRIS protocol, IRIS estimates its emitter mask from the LDR observations, whereas FIPT-LDR additionally receives ground-truth geometry and a ground-truth emitter mask. AEGIR initializes geometry from the images and uses no emitter mask input.

EAG-PT (Yang et al., 2026) labels diffuse surface 2D Gaussians as emitters using fixed input view emitter masks and estimated normals. These masks are inputs to optimization, not an independently optimized emitter geometry, so the method has no separate angular source profile. AEGIR instead models separate anisotropic emitters and GGX materials, requires no external emitter mask, and uses internally detected bright regions only to initialize freely optimized emitters. Moreover, EAG-PT does not output roughness or metallic maps. SGS-Intrinsic (Niu et al., 2026) uses an environment map with local Spherical Gaussian point lights. Table 2 reports their matched synthetic results, and Table 3 reports their matched real/scanned results on ScanNet++, Replica, and FIPT-Real. ESR-NeRF (Jeong et al., 2024) uses paired light-on/light-off LDR captures, whereas AEGIR recovers finite editable emitters from one lighting condition.

C.3 Evaluation Metrics

Image Quality and Evaluation Conventions. We evaluate novel view synthesis and albedo using peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS) (Zhang et al., 2018), and roughness using mean-squared error (MSE). All methods use consistent evaluation conventions for color space, exposure, tone mapping, masks, crops, and resolution. PSNR is averaged over held-out views within each scene and then across scenes, rather than computed from pooled pixel errors. Any additional cross-dataset averaging is specified with the corresponding table or figure. Quantitative metrics are computed before denoising, however displayed results follow the published procedure of each method, with OptiX denoising (Parker et al., 2010) used for AEGIR visualizations.

Emitter Recovery Metrics. Table 4 in Section 4.4 reports held-out NVS and albedo metrics, direct emitter position/size/flux errors, and average primitive count.

- **Position.** Position is the average nearest-center distance between recovered and ground-truth emitter sets after normalizing each scene to unit scale, so all split primitives contribute. An area primitive's position is its center, while point lights are assigned to the nearest ground-truth source and the emission-weighted group mean is used as the recovered position.
- **Size.** For Size, each ground-truth emitter is paired with the nearest recovered area light source or its associated point light group. Scalar Size is the largest ellipsoid scale, the larger in-plane disk scale, or twice the farthest point distance from a point light group center. We take the absolute Size difference and average across emitters.
- **Flux.** Flux is the relative difference between the total recovered and ground-truth RGB emitted power. For area light sources, power is approximated by sampling the emitter surface and outgoing directions with cosine projection, then summed over all primitives representing a source. However, for point light groups, RGB intensities are summed.
- **Source Correspondences.** Missing or extra source correspondences use nearest emitter assignment.

C.4 Training Schedule and Hyperparameters

Initialization and Training Stages. We initialize the scene geometry from a DN-Splatter (Turkulainen et al., 2025) checkpoint trained for 10,000 iterations. During the first 15,000 AEGIR iterations, positions, scales, rotations, opacities, materials, and emitters are jointly optimized, while DN-Splatter SH coefficients remain frozen and serve only as the indirect light proxy. Geometry and materials are frozen for the final 5,000 iterations, during which only the emitters are refined. For physically based materials, all 2D Gaussians are initialized with a roughness of 0.6 and a metallic value of 0.2.

Optimizer and Learning Rates. We optimize for 20,000 iterations using the Adam optimizer. The initial learning rates are 1×10^{-3} for material and lighting parameters and 1×10^{-5} for geometry parameters. Between iterations 5,000 and 15,000, geometry and material learning rates decay to zero. During the same interval, the emitter learning rate decays to 1×10^{-4}.

Emitter Initialization Ablation. For the ablation without bright-pixel emitter initialization, we instead place 27 white spheres uniformly on a $3 \times 3 \times 3$ grid spanning the scene. Each sphere starts with a spatial scale equal to 5% of the scene diagonal and is then optimized as a trainable ellipsoid. The same losses, schedule, and adaptive splitting and pruning subsequently adjust each emitter's center $\mathbf{c}_e$, scale, emission, orientation, and angular profile, as well as the emitter count.

Rendering and Standard Sampling. For rendering, we use the OptiX (Parker et al., 2010) 2D Gaussian tracer from IRGS (Gu et al., 2025). The bounding volume hierarchy (BVH) is updated as the Gaussian geometry changes during the joint stage and reused during the final emitter-only stage. Ray marching terminates when transmittance falls below 0.05. During optimization, we sample 2^{16} traced rays per iteration. For rendering equation evaluation, we use $N_{\text{brdf}} = 128$ cosine-weighted hemisphere samples and a requested budget of $N_{\text{light}} = 128$ explicit emitter samples, allocated uniformly as described in Equation S3.

Primitive Ablation Sampling. Table 4 holds shaded pixel counts and optimization iterations fixed, not the total ray count. Each point emitter supplies one light sample. When the diffuse-disk count exceeds N_{light}, one sample is drawn per disk, so the actual light ray count exceeds the requested budget. The standard AEGIR allocation is not applied to these cases. This comparison therefore does not assume equal ray budgets or runtimes, as it is not applicable across all primitives. The alternative ReSTIR training configuration is specified separately in Appendix A.4.

Adaptive Emitter Control. To efficiently represent complex lighting without excessive memory consumption or over-exposure, we dynamically adapt the set of area emitters through splitting and pruning. These updates occur every 1,000 iterations between iterations 5,000 and 15,000. We evaluate each emitter e based on its scale $\mathbf{s}_e$, center $\mathbf{c}_e$, and a perceptual energy measure Φ_e. Given a linear RGB emission $\mathbf{E}_e = (E_{e,R}, E_{e,G}, E_{e,B})$, this energy is defined as:

$$\Phi_e = \left(\frac{E_{e,R} + E_{e,G} + E_{e,B}}{3} \right)^{1/2.2}. \tag{S27}$$

- **Pruning:** We remove an emitter if its contribution becomes negligible ($\Phi_e < 0.1$) or physically implausible (drifting outside valid scene bounds).
- **Splitting:** To increase sampling resolution around dominant light sources, we split an emitter if it becomes excessively bright ($\Phi_e > 5.0$) or if its largest scale exceeds 20% of the maximum scene scale. The emitter is divided along its major axis into two children, with each inheriting half the intensity and half the spatial scale of the parent.

Regularization Weights. We balance accurate image reconstruction with realistic physics using the following tuned loss weights:

- **Photometric Loss:** We set the focus map weight to $\lambda_{focus} = 5.0$ within the image reconstruction loss, and set the structural dissimilarity weight to $\lambda_{dssim} = 0.2$.
- **Material Regularization:** For the material guidance from the pretrained DiffusionRenderer, we apply equal weighting to the albedo, roughness, and metallic components ($\lambda_{\mathbf{a}} = \lambda_r = \lambda_m = 1.0$). We regularize these materials using the TV loss ($\lambda_{TV} = 1.0$), scaling the individual components by $\gamma_{\mathbf{a}} = 0.5$, $\gamma_r = 0.25$, and $\gamma_m = 0.25$, respectively. Finally, we set $\beta = 1.0$.
- **Emitter Regularization:** The emitter regularization weights λ_{area} and λ_{bounds} are set to 0.5, while λ_{white} is set to a lower value of 0.1 to give the optimization enough flexibility to recover colored light sources.

C.5 Runtime and Stability

All methods are timed on the same NVIDIA RTX 4090 GPU, including every pipeline step except prior generation. Standard AEGIR takes about 55 minutes, including 15 for DN-Splatter initialization, 30 for joint optimization, and 10 for emitter only refinement. The latter two stages account for the optimization time of 40 minutes. DiffusionRenderer priors take approximately 15 additional minutes, giving about 70 minutes including this preprocessing. ReSTIR 45.5 minutes uses the same timing scope as Standard and the baselines. GS-ID, SGS-Intrinsic, and IRIS also require prior preprocessing. AEGIR and our GS-ID reproduction share the DN-Splatter initialization. Main comparisons report single optimization runs. To evaluate stability, we run 5 random seeds for each of the four FIPT-Synthetic scenes. The reported NVS deviations (± 0.14 PSNR, ± 0.014 SSIM, ± 0.022 LPIPS) are the means of the per-scene standard deviations across those seeds.

D ADDITIONAL RESULTS

Unless stated otherwise, the analyses in this appendix use the same datasets, splits, and evaluation protocol as the main paper.

Main paper topic	Additional results
Area light estimation (Section 4.2)	Appendix D.1: direct illumination maps
NVS and intrinsic decomposition (Section 4.3)	Appendices D.2-D.5: geometry, optimization stages, decomposition, and editing
Ablations (Section 4.4)	Appendices D.6 and D.7: component and primitive ablations, emitter statistics
ReSTIR sampling (Section 3.2)	Appendix D.8: sampling comparison (method in Appendix A.4)

D.1 EXPLICIT AREA LIGHT ESTIMATION

This section expands Section 4.2 with direct illumination maps on synthetic, captured, and scanned scenes (Figure S1).



Figure S1: **Direct illumination maps.** Columns compare the reference images with illumination from AEGIR spatially extended area emitters and GS-ID point lights (Du et al., 2025). Rows show Hypersim `ai_001_004`, FIPT-Synthetic bathroom, Replica room0, and FIPT-Real conference room.

D.2 GEOMETRY REFINEMENT

DN-Splatter initializes the 2D Gaussian geometry. During the first 15,000 AEGIR iterations, geometry, materials, and emitters are optimized jointly. Geometry and materials are then frozen for the final 5,000 iteration emitter-only refinement. Table S2 reports geometry before and after the joint stage.

Table S2: **Geometry before and after joint optimization.** Each arrow shows the DN-Splatter initialization → AEGIR geometry after 15,000 joint optimization iterations, aggregated by benchmark. Real/scanned scenes use provided scanned depth, without normal metrics. RMSE denotes root mean square error, and AbsRel denotes absolute relative error.

Benchmark group	Normal RMSE ↓	Mean angular error (°) ↓	Depth AbsRel ↓	Depth RMSE ↓
Synthetic (Hypersim, FIPT-Synthetic - 10 scenes)	0.203 → 0.187	14.11 → 12.70	0.067 → 0.063	0.241 → 0.224
Real/scanned (ScanNet++, Replica, FIPT-Real - 14 scenes)	–	–	0.087 → 0.078	0.288 → 0.249

AEGIR optimizes 2D Gaussian disks rather than a mesh, so the comparison uses deferred shading depth and world-space unit normals. Using the same DN-Splatter initialization, GS-ID changes synthetic normal RMSE from 0.203 to 0.199 and depth AbsRel from 0.067 to 0.070, whereas AEGIR reaches 0.187 and 0.063, respectively. Figure S2 shows the matched initial and final outputs.

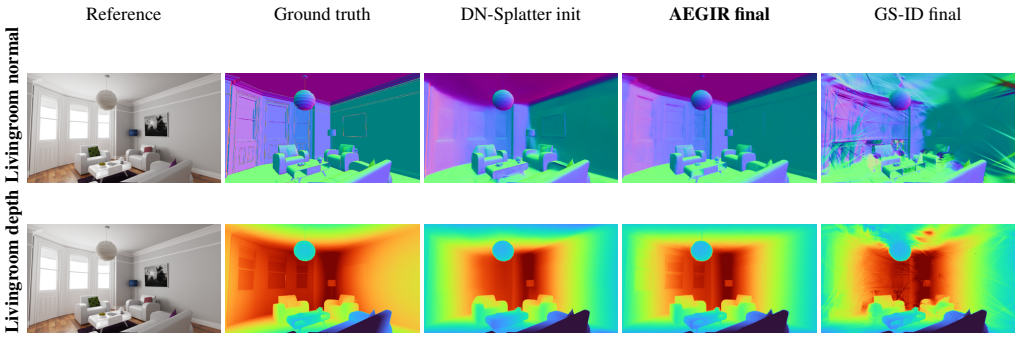


Figure S2: **Geometry refinement from a shared DN-Splatter initialization.** Rows show world-space normals and depth for the livingroom. Columns compare the RGB reference, ground truth, shared initialization, and final AEGIR and GS-ID geometry. All normal maps use the same coordinate frame and color mapping.

D.3 Optimization Stages and the Diffusion Prior

DiffusionRenderer is run once before optimization and enters only through the soft loss $\mathcal{L}_{prior}$, so its prediction is not used as the final material prediction. During the first 15,000 iterations, geometry, materials, and emitters remain learnable under multi-view photometric supervision, the soft diffusion prior, and edge-aware total variation. Geometry and materials freeze only for the final 5,000 emitter refinement iterations. Table S3 reports this analysis over ten synthetic scenes.

Table S3: **Stage-wise analysis on ten synthetic scenes.** Results are scene weighted averages over six Hypersim and four FIPT-Synthetic scenes, computed from per-scene checkpoint evaluations. P/S/L denotes PSNR/SSIM/LPIPS. DiffusionRenderer is an offline prior generation step. Best per column is bold.

Phase/checkpoint	Updated parameters	Time	Normal RMSE/angular error	NVS P/S/L	Albedo P/S/L
DN-Splatter initialization (10,000)	Geometry initialization	15 min	0.203/14.11°	14.79/0.80/0.32	–
DiffusionRenderer prediction	None (soft supervision)	15 min	–	–	14.89/0.75/0.27
Joint AEGIR (0-15,000)	Geometry, materials, emitters	30 min	**0.187/12.70°**	20.93/0.82/0.21	**19.26/0.79/0.20**
Emitter refinement (15,000-20,000)	Emitters only	10 min	unchanged	**21.91/0.84/0.15**	unchanged

The optimized albedo improves from 14.89/0.75/0.27 for the raw prior to 19.26/0.79/0.20 after joint optimization. Emitter only refinement then improves NVS from 20.93/0.82/0.21 to 21.91/0.84/0.15. Figure S3 additionally shows that AEGIR corrects view inconsistencies in prior predictions.

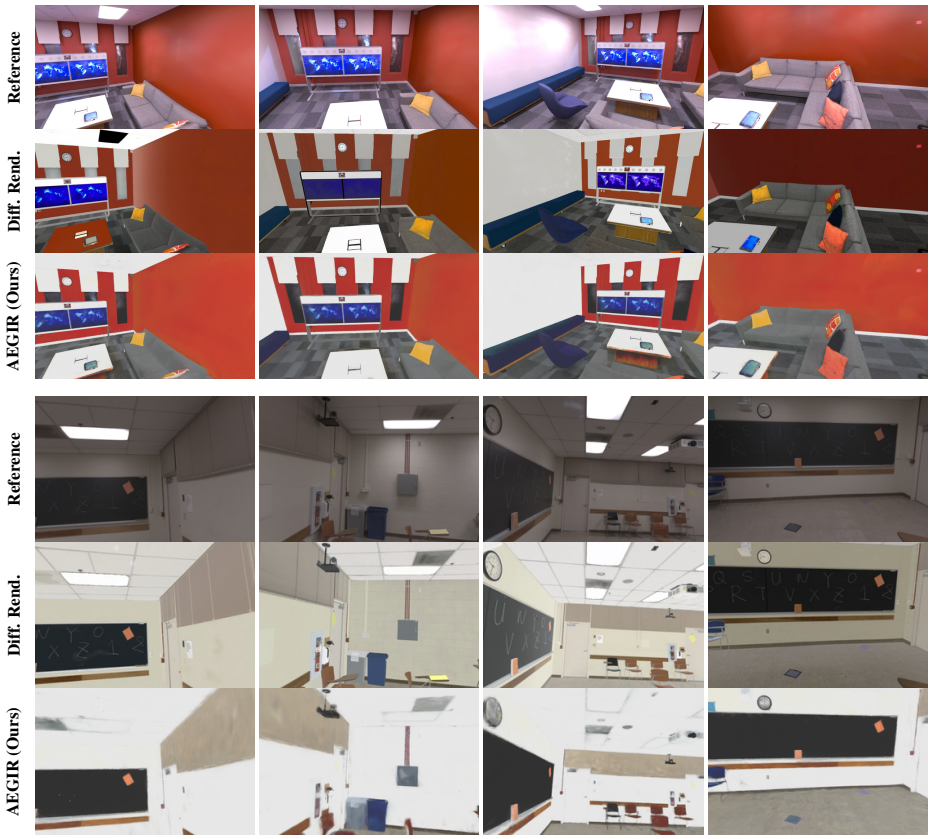


Figure S3: **Multi-view albedo consistency.** Four viewpoints compare RGB references, independent DiffusionRenderer (Liang et al., 2025a) priors, and optimized AEGIR albedo. Multi-view optimization reduces the view-dependent illumination bleeding and inconsistencies in the priors.

D.4 Novel View Synthesis and Intrinsic Decomposition

The following figures expand the NVS and material comparisons in Section 4.3 with additional scenes and views.

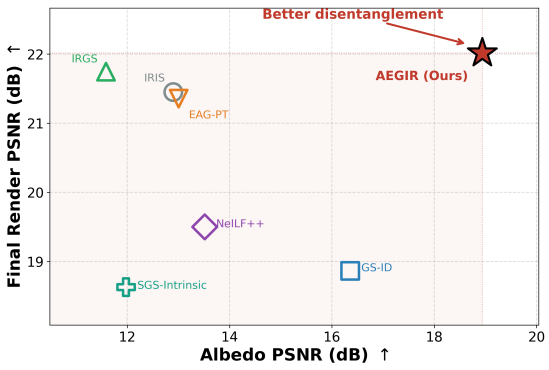


Figure S4: **Disentanglement plot on synthetic datasets.** Albedo vs. render PSNR shows that AEGIR improves both material accuracy and novel view synthesis.

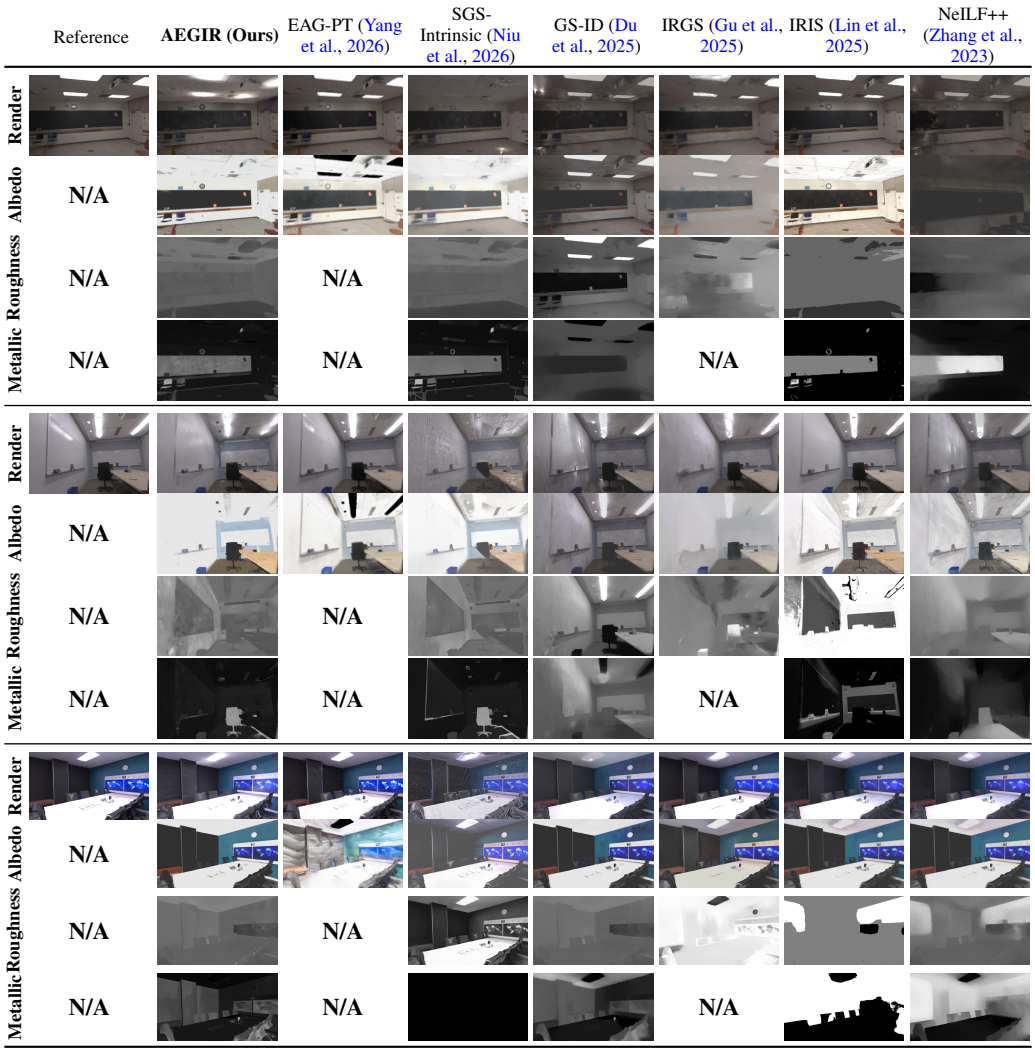


Figure S5: **NVS and intrinsic decomposition on real/scanned indoor scenes.** Rows show render, albedo, roughness, and metallic outputs. Columns compare the reference and all methods. EAG-PT does not output roughness or metallic maps, and IRGS does not output a metallic map. "N/A" denotes an unavailable reference property or method output.

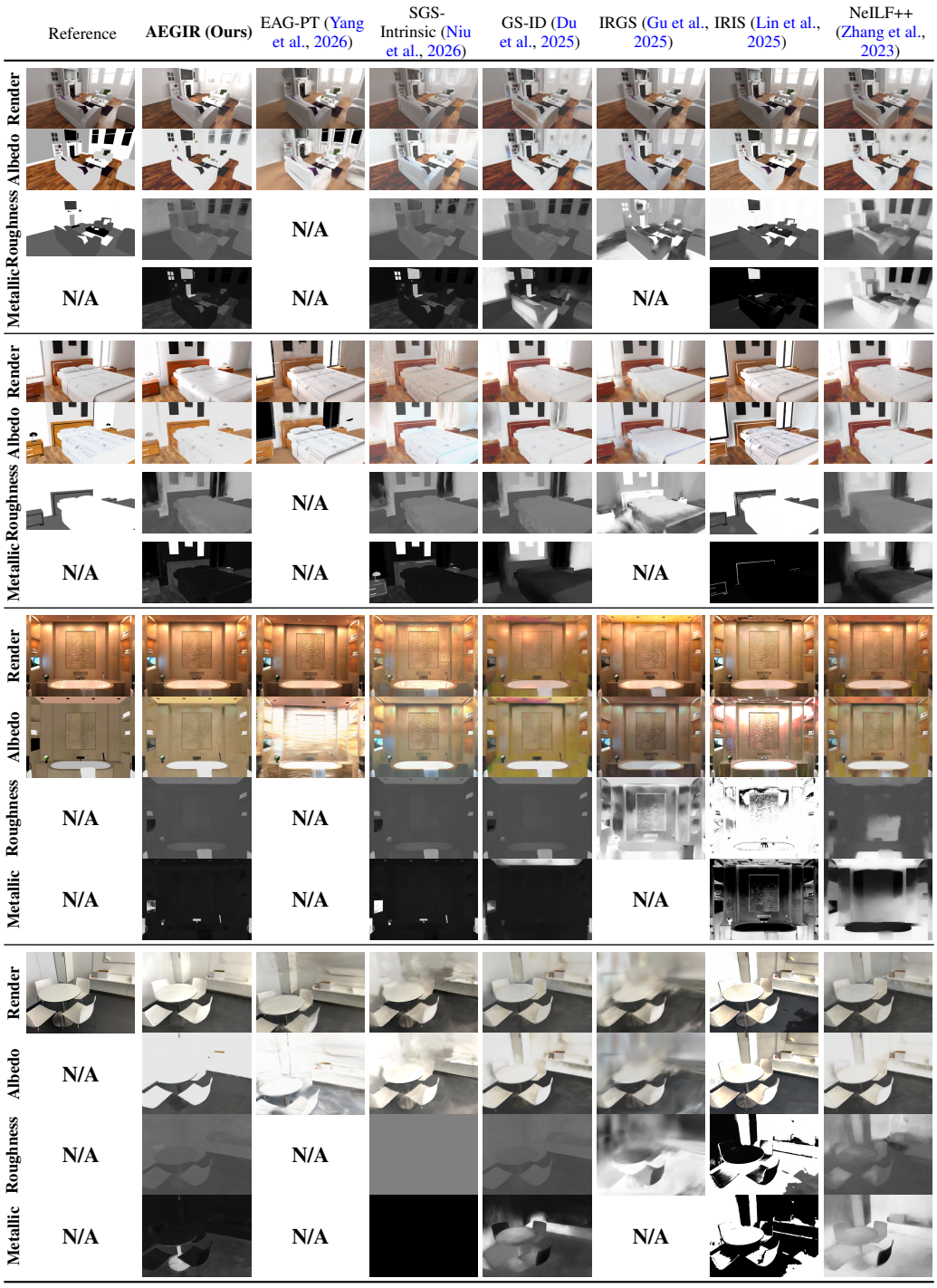


Figure S6: **NVS and intrinsic decomposition on additional scenes.** The four scene blocks cover FIPT-Synthetic livingroom and bedroom, Hypersim `ai_032_001`, and ScanNet++ `8b5caf3398`, with render, albedo, roughness, and metallic outputs for each method. EAG-PT does not output roughness or metallic maps, and IRGS does not output a metallic map. "N/A" denotes an unavailable reference property or method output.

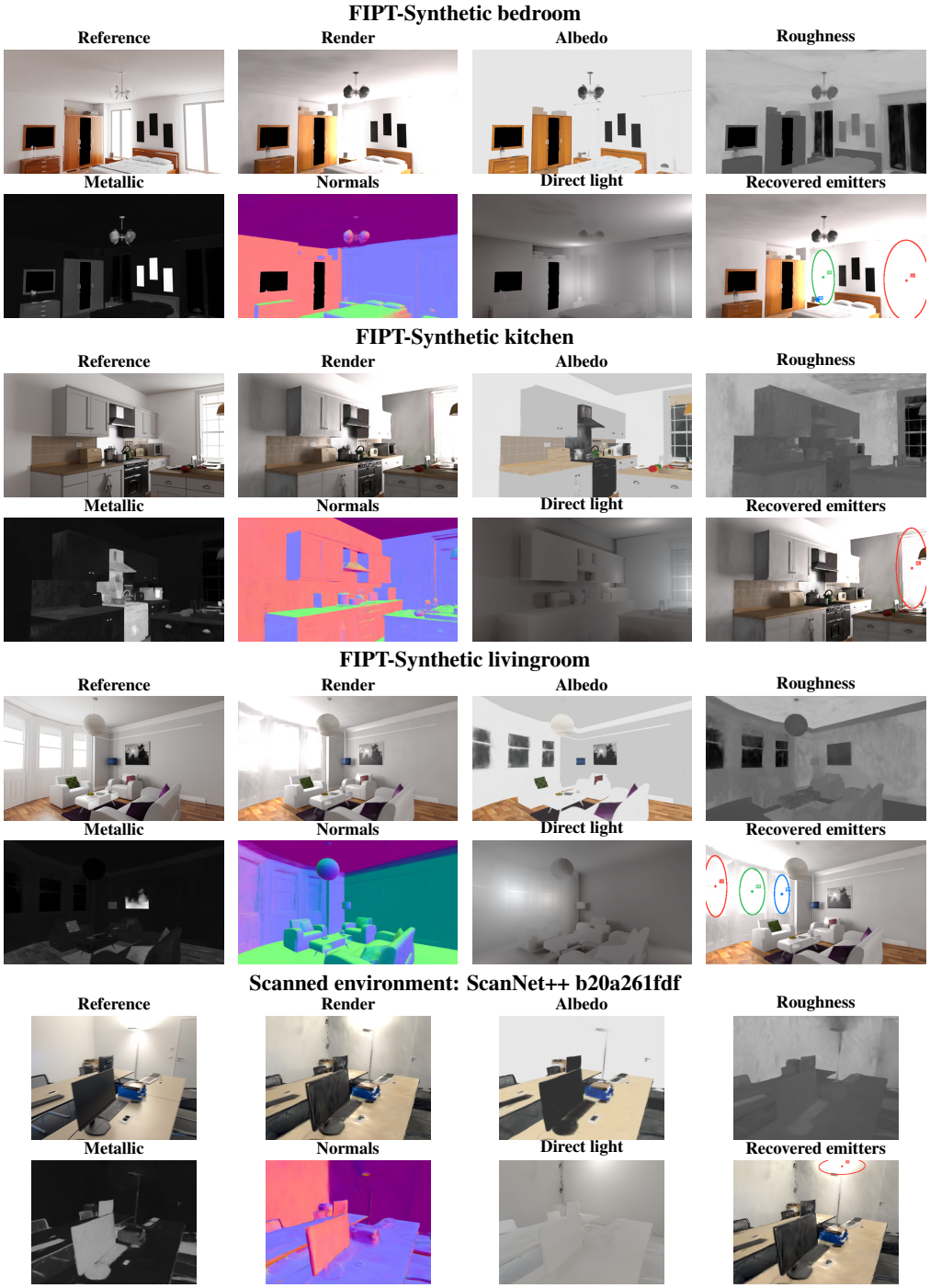


Figure S7: **AEGIR decomposition on FIPT-Synthetic and ScanNet++ scenes.** Each block shows the reference, final render, albedo, roughness, metallic, normals, direct illumination, and projected recovered emitters. The final cell projects the recovered 3D emitters into the image for illustration only.

D.5 Relighting and Object Insertion

Figure S8 illustrates relighting and insertion. Table S4 replaces the recovered emitters with a moving emissive sphere and compares against ground-truth reference images rendered in Mitsuba using the same moving emissive sphere on the FIPT-Synthetic evaluation scenes. We use recovered geometry and one indirect bounce. Camera, sphere trajectory, exposure, frame indices, renderer, and sample budget are identical across methods. This tests the recovered scene under a novel source. Appendix A.3 details how relighting replaces the training time SH proxy with the new illumination.

Table S4: **Controlled novel source relighting.** Aggregate PSNR/SSIM/LPIPS after replacing the recovered illumination with the same moving emissive sphere for every method. Best per column is bold.

Method	Aggregate PSNR/SSIM/LPIPS
IRIS (Lin et al., 2025)	20.12/0.80/0.23
GS-ID (Du et al., 2025)	18.98/0.78/0.26
IRGS (Gu et al., 2025)	11.31/0.66/0.36
NeILF++ (Zhang et al., 2023)	13.01/0.68/0.35
EAG-PT (Yang et al., 2026)	21.54/0.83/0.23
SGS-Intrinsic (Niu et al., 2026)	13.04/0.66/0.37
AEGIR (Ours)	**21.93/0.84/0.20**

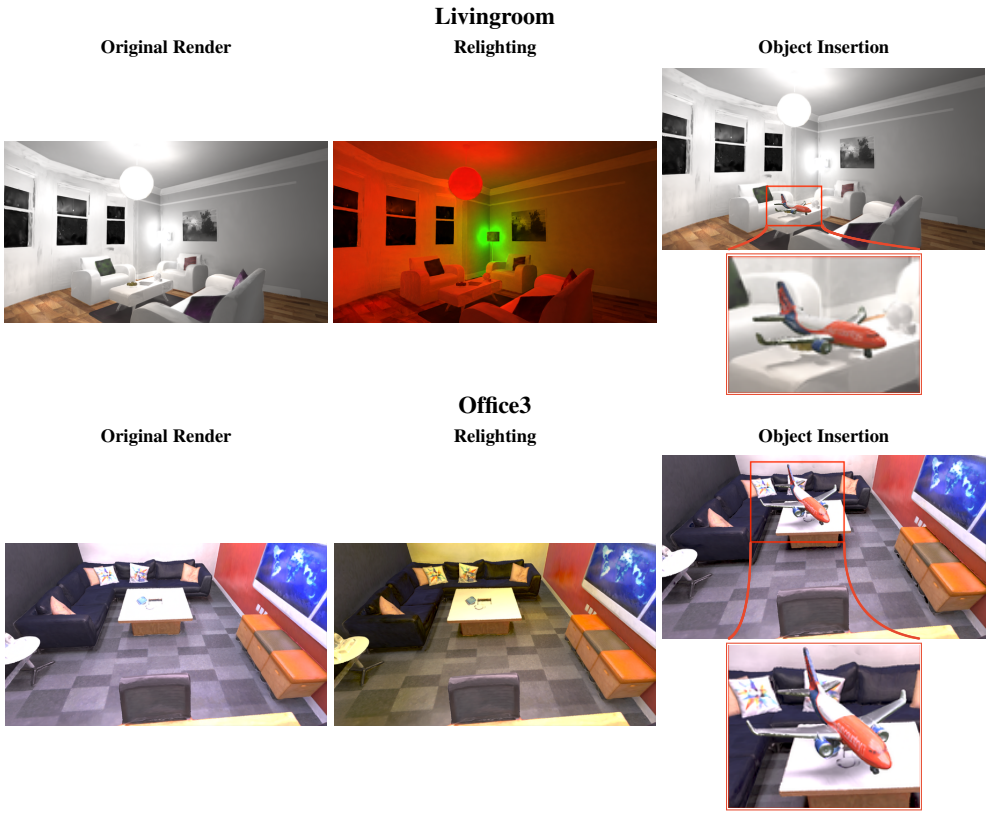


Figure S8: **AEGIR relighting and object insertion.** Each scene shows the original render, qualitative emitter color editing, and virtual object insertion. Relighting produces red and green illumination in the living room and warm yellow illumination in the office. The insertion column shows virtually inserted aircraft. Boxes and connectors locate the enlarged inserted objects. Both edits use the recovered materials and explicit area emitters.

The following results isolate two editing capabilities of the recovered representation. Figure S9 replaces the reconstructed illumination with a moving, color-changing emissive sphere. Figure S10 composites virtual objects into two reconstructed scenes under their recovered illumination.

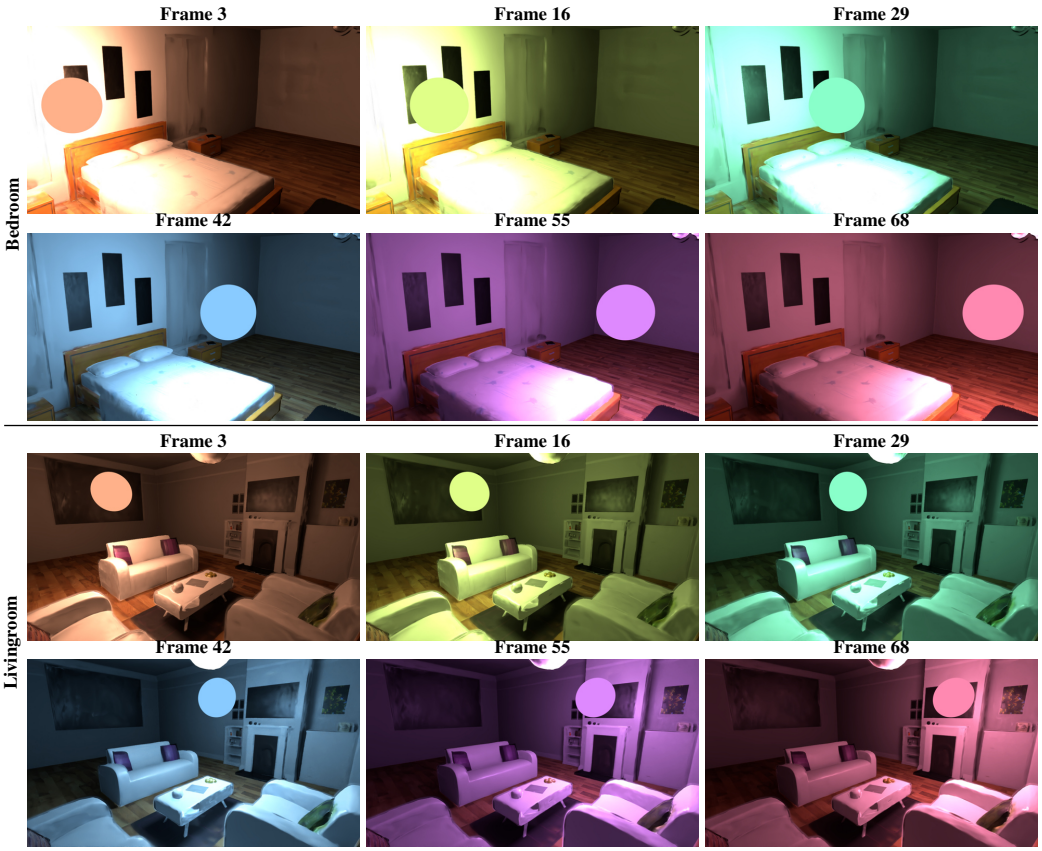


Figure S9: **Moving colored light relighting.** Six evenly spaced frames from each sequence show a moving, color-changing sphere as the only light source after removing the reconstructed emitters. Rendering uses direct illumination and one path-traced indirect bounce.



Figure S10: **Object insertion in two scenes.** Each row shows two matched held-out Clean vs. Inserted view pairs. Inserted diffuse and reflective objects are shaded using the recovered geometry, materials, and explicit area emitters.

D.6 Emitter Primitive and Component Ablations

Table 4 reports quantitative results of the light primitive ablation, with metric definitions and emitter matching in Appendix C.3. Pixel and iteration counts are fixed, although point and disk variants can trace more rays (Appendix C).

To compare emitter parameterizations, we run a same pipeline ablation with the representation specific sampling rules in Appendix C. All variants use the same input images, geometry and material initialization, emitter initialization, renderer, losses, shaded pixel counts, and optimization schedule

on the four FIPT-Synthetic scenes. We compare (i) **Point** emitters with learned positions and RGB intensities but no spatial extent, (ii) **Isotropic area** emitters with equal spatial scales and an isotropic angular profile, (iii) **Diffuse 2DGS disks** with learned in-plane scales and direction independent emission, (iv) **Anisotropic Gaussian** emitters with learned spatial scales and angular spreads but fixed $\rho = 1$, and (v) **Full AEGIR**, which also learns the Super-Gaussian falloff ρ.

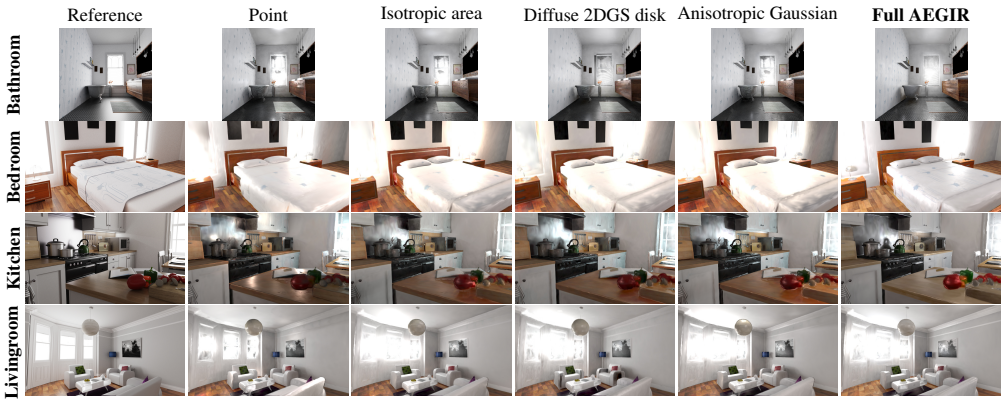


Figure S11: **Emitter primitive ablation on four FIPT-Synthetic scenes.** Rows are scenes and columns compare the reference with five emitter representations.

Diffuse 2DGS disks improve position and size over isotropic area emitters but are worse in NVS, albedo, and Flux. The anisotropic Gaussian improves every image metric and emitter error over the diffuse disk. Full AEGIR further improves NVS, albedo, and all emitter errors while remaining compact, 12.8 primitives on average, versus 14.3 anisotropic Gaussians, 53.4 isotropic area emitters, 252.1 points, and 1165.9 diffuse disks. Figure S11 presents the complete visual comparison.

Table S5: **Quantitative results of component ablation.** We evaluate the contribution of each component to final rendering on Replica and FIPT-Synthetic datasets and albedo quality on FIPT-Synthetic dataset. Best per column is bold, second-best is underlined.

Configuration	NVS (scene-weighted)			Albedo (FIPT-Synthetic)		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
W/o Focus Map (W_f)	16.71	0.72	0.25	17.25	0.78	0.23
W/o Diffusion Priors ($\mathcal{L}_{prior}$)	18.83	0.78	<u>0.22</u>	15.11	0.74	0.26
W/o Material TV ($\mathcal{L}_{TV}$)	18.29	0.76	0.24	15.96	0.75	0.25
W/o Emitter Regularization ($\mathcal{L}_{light}$)	17.43	0.74	0.24	17.24	<u>0.79</u>	0.23
W/o Adaptive Control	17.07	0.72	0.25	17.21	0.77	0.24
W/o Emitter Initialization	<u>19.42</u>	<u>0.80</u>	<u>0.22</u>	<u>17.30</u>	<u>0.79</u>	<u>0.22</u>
AEGIR (Full Method)	**19.46**	**0.81**	**0.21**	**17.33**	**0.80**	**0.21**

We evaluate the contribution of each component to novel view synthesis and albedo accuracy (Table S5), as well as qualitative illumination estimation and intrinsic disentanglement (Figure S12) in scenes with complex lighting from Replica and FIPT-Synthetic. Removing the focus map weakens the detection of localized lighting patterns. Disabling diffusion priors causes a severe drop in the quantitative albedo metrics, as illumination effects become incorrectly embedded into the estimated albedo. Omitting the material TV loss leads to noisy textures due to inconsistencies in DiffusionRenderer outputs. For illumination estimation, removing emitter regularization or adaptive control results in degenerate solutions, where extra light sources are introduced and area light sizes are overestimated. Ablating emitter initialization causes the smallest degradation among the lighting control ablations, especially for albedo, but still reduces NVS quality. However, reaching similar quality without emitter initialization requires roughly 2,000 additional optimization steps. This indicates its role as a useful photometric prior and demonstrates that our method can estimate illumination from shading

cues without requiring every source to be directly visible. The component ablation supports the full method.

Figure S12 shows the qualitative component ablation discussed in Section 4.4. For one Replica scene, each column removes a single optimization component and the rows show the resulting render, albedo, and direct illumination.

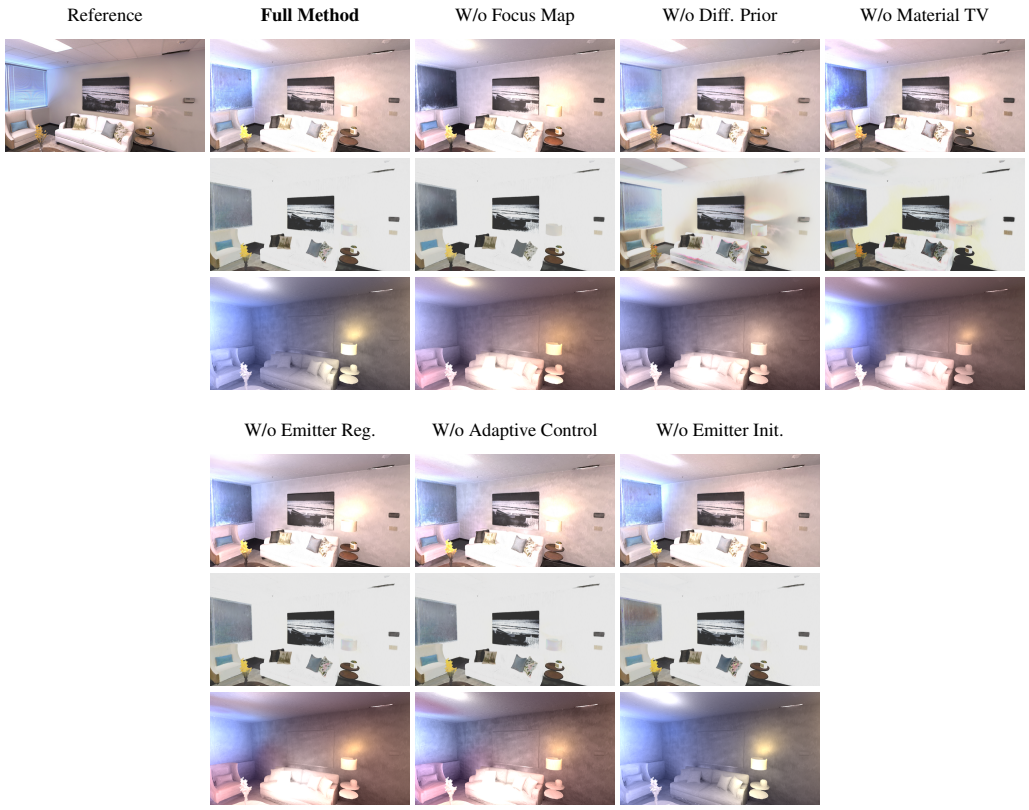


Figure S12: **Qualitative component ablation.** The reference is shown once. Each variant block contains, from top to bottom, the novel view render, estimated albedo, and direct illumination.

D.7 Emitter Count and Scale

The estimated emitter primitives are deliberately sparse and larger than the estimated surface 2D Gaussians. For each scene, Table S6 reports the maximum emitter scale and maximum 2DGS in-plane scale, normalized by the scene diagonal, and then averages these quantities and primitive counts over the dataset. Mean counts are about 9-21 emitter ellipsoids, compared with hundreds of thousands of surface Gaussians. Maximum surface Gaussian scales span 0.160-0.293% of the scene diagonal, whereas final emitter scales span 1.40-4.55%. The maximum emitter scale decreases after refinement for every dataset, while the emitter count grows in three of five datasets because splitting exceeds pruning.

Table S6: **Emitter and surface 2DGS statistics.** Dataset averages of primitive counts and maximum scales, with scale reported as a percentage of the scene diagonal. Initial and final emitter columns denote the sets before and after optimization.

Dataset	Initial emitters		Surface 2DGS		Final emitters	
	Count	Scale	Count	Scale	Count	Scale
FIPT-Synthetic	11.2	4.03%	712,364	0.235%	12.8	3.75%
Hypersim	18.7	2.83%	383,354	0.293%	20.3	2.51%
ScanNet++	19.8	4.91%	950,868	0.264%	21.2	4.55%
Replica	11.8	2.80%	463,370	0.160%	11.5	2.10%
FIPT-Real	10.5	1.75%	551,697	0.283%	9.5	1.40%

D.8 ReSTIR-Based Sampling Results

This section compares standard and alternative ReSTIR-based training sampling. Reported mean time per scene decreases from 55.0 to 45.5 minutes, with dataset dependent quality changes. Both variants use identical standard sampling at inference time.

Both models in Table S7 use the same 24 scenes, held-out views, resolutions, and full standard inference sampler. The comparison evaluates the complete training sampling configurations. The reduction of 9.5 minutes is 17.3% of the reported Standard time. ReSTIR improves all FIPT-Real NVS metrics and Replica NVS PSNR. Standard performs slightly better on all ScanNet++ NVS metrics. On FIPT-Synthetic, ReSTIR improves albedo PSNR by 0.86 dB at equal SSIM, while Standard retains lower albedo LPIPS. On Hypersim, ReSTIR reduces albedo PSNR by 0.03 dB, with albedo SSIM and LPIPS tied at the displayed precision. Standard performs better on all NVS metrics on both synthetic benchmarks. Figure S13 shows representative comparisons from all five datasets. The benefits of ReSTIR sampling vary across datasets, and maintaining the reservoirs requires additional memory. We therefore retain standard sampling for the main evaluation and present ReSTIR as an alternative training-time sampler, highlighting its potential and trade-offs for inverse rendering.

Table S7: **Standard versus ReSTIR-based training sampling.** Both models use standard inference sampling. Reported mean time per scene, 55.0 min for Standard and 45.5 min for ReSTIR, both including all steps except offline priors on the same GPU. Best per dataset and metric is bold.

		NVS			Albedo		
Dataset	Sampler	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
FIPT-Synthetic	Standard	**22.53**	**0.86**	**0.19**	17.33	**0.80**	**0.21**
	ReSTIR	21.59	0.83	0.22	**18.19**	**0.80**	0.22
FIPT-Real	Standard	22.07	0.85	0.19	–	–	–
	ReSTIR	**22.50**	**0.86**	**0.18**	–	–	–
ScanNet++	Standard	**18.55**	**0.83**	**0.20**	–	–	–
	ReSTIR	18.46	0.82	0.21	–	–	–
Replica	Standard	17.93	**0.78**	**0.22**	–	–	–
	ReSTIR	**18.59**	**0.78**	**0.22**	–	–	–
Hypersim	Standard	**21.50**	**0.83**	**0.12**	**20.55**	**0.78**	**0.19**
	ReSTIR	20.98	0.79	0.13	20.52	**0.78**	**0.19**

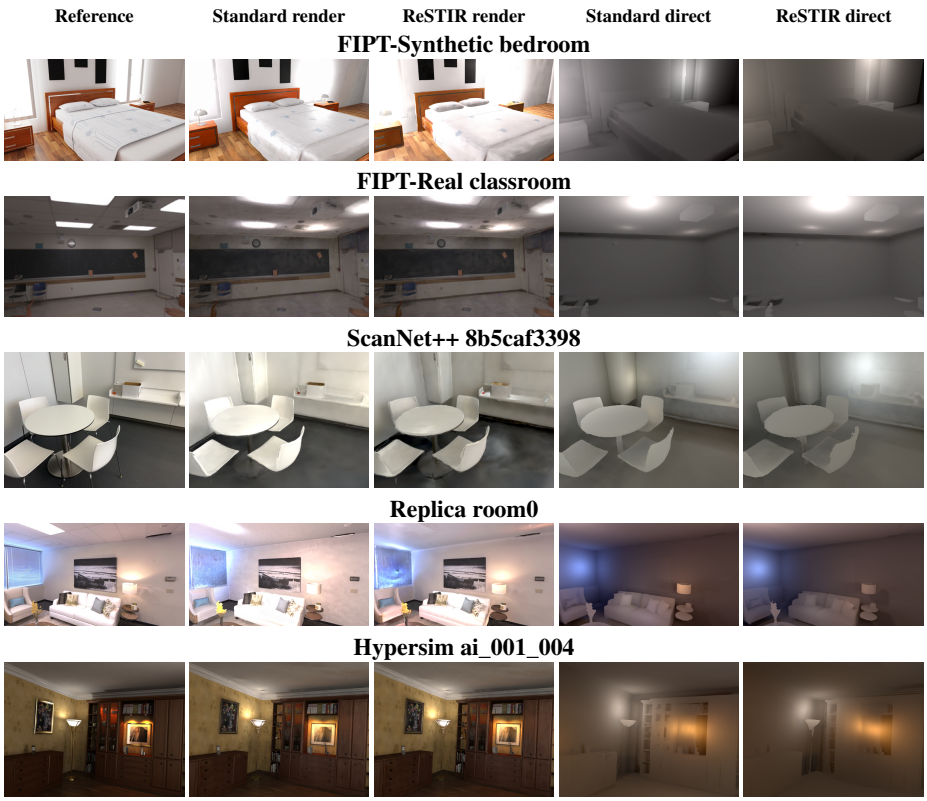


Figure S13: **Qualitative comparison of standard and ReSTIR-based training sampling.** Representative novel views and recovered direct illumination from all five datasets. The two trained models are rendered with identical standard test-time sampling.

E LIMITATIONS AND FUTURE WORK

The remaining limitations of AEGIR fall into four groups.

Emitter Representation and Source Ambiguity. The learned falloff power ρ_e can produce sharp angular illumination transitions, but a single ellipsoid cannot reproduce the corners of a polygonal fixture. Mixtures of split primitives or higher-order emitter shapes could model sharper spatial boundaries, at the cost of additional degrees of freedom. Bright-pixel unprojection initializes the emitters, but their positions, scales, emissions, and count remain learnable. Weak or out-of-view sources may still be missed, and strong reflections may be mistaken for emitters, although we did not observe such failures on the tested scenes. Spawning emitters based on RGB residuals and stronger multi-view consistency checks could help resolve these ambiguities. During surface sampling, we reject samples whose outward normals face away from the shading point (Appendix A.2). This excludes the far side of the ellipsoid for that shading point, but does not make the emitter globally one-sided, because Equation 1 assigns equal emission to opposite directions. Surrounding scene geometry can block emission in unwanted directions. Explicitly one-sided emission would additionally require restricting the angular profile to a selected hemisphere.

Geometry and Material Transport. The deferred renderer depends on the quality of the jointly refined 2D Gaussian geometry. Textureless or heavily occluded regions can retain noisy normals and produce shading artifacts. Stronger geometric priors and surface regularization could improve stability. The current opaque microfacet BRDF also excludes transparent and mirror-like transport. A full bidirectional scattering distribution function (BSDF) with transmission and differentiable refraction would be required for glass and other transmissive materials.

Indirect Illumination. We did not observe indirect illumination leakage as a systematic failure in our experiments, although complete separation cannot be guaranteed in ambiguous cases, such as bright indirect light from an out-of-view source. The frozen DN-Splatter SH features already capture much of the observed surface appearance and provide a fixed proxy for indirect light during optimization, reducing the need to represent indirect effects using additional local emitters. During optimization, more bounces or a learned indirect representation could improve global transport, but would increase cost and may introduce more ambiguity with the explicit emitters.

Scalability. Emitter sampling, tracing, and Monte Carlo integration become more expensive as the number of explicit emitters grows. The alternative ReSTIR-based training sampler evaluated in Appendix D.8 reuses samples across optimization steps and views, lowering the reported mean time from 55.0 to 45.5 minutes. Its quality changes remain dataset dependent, and scaling to scenes with far more emitters may be the scope for future improvements.