F-RNG: FEED-FORWARD RELIGHTABLE NEURAL GAUSSIANS WITH LARGE MODEL PRIORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Capturing relightable 3D assets from real-world objects is a widely researched problem. Several methods based on 3D Gaussian splatting (3DGS) support relighting; however, they usually require dense input views and separate optimization for each scene. In contrast, generalizable feed-forward models can directly reconstruct Gaussians from sparse input views. However, the resulting assets have baked-in illumination and cannot be easily used for relighting. In this paper, we present F-RNG, a feed-forward framework that directly generates relightable 3DGS assets from sparse-view inputs. Training such a model from scratch can require substantial data and computing resources, making feed-forward relightable reconstruction at an acceptable computational cost challenging. We build F-RNG on an existing large reconstruction model (LRM) to extract relightable representations by training only small networks with guidance from an intrinsic decomposition model (IDM), without any fine-tuning or re-training of the large models. Specifically, we aim to improve the underlying geometry and disentangle illumination and material properties. We first introduce a latent-interpolated fine-grained geometry synthesis module to enhance the LRM's geometry representation. Second, we propose a prior-guided relightable appearance distillation module to extract relightable neural representations using IDM priors. Finally, a universal neural renderer enables flexible and high-fidelity relighting under novel environment illumination. Once the asset is constructed, relighting under novel environment illumination no longer requires the large models, avoiding the cost of repeated inference. Compared to the state-of-the-art LRM-based relighting method, F-RNG achieves better visual quality and overall quantitative performance, along with $\sim$25$\times$ faster relighting per additional environment map.

1 INTRODUCTION

Creating relightable 3D assets from sparse images is valuable for digital content creation, but separating illumination, material, and geometry remains ill-posed. Optimization-based methods built on 3D Gaussian splatting (3DGS) Kerbl et al. (2023) can recover relightable assets, but typically require dense views, restrictive priors, controlled illumination, or lengthy per-object optimization Gao et al. (2023); Jiang et al. (2024); Liang et al. (2024); Bi et al. (2024; 2020). Neural appearance models such as RNG Fan et al. (2025) handle complex materials including fur, yet still require per-scene training.

Large reconstruction models (LRMs) instead reconstruct unseen objects in a feed-forward pass from sparse views Zhang et al. (2024); Li et al. (2025); Siddiqui et al. (2024). RelitLRM Zhang et al. (2025), for example, predicts a 3DGS radiance field conditioned on a target environment map. Its spherical harmonics contain baked-in illumination, however, so every new light requires another expensive diffusion pass rather than rendering a reusable asset.

We propose F-RNG, a feed-forward framework that converts LRM predictions into relightable neural Gaussian assets. Instead of training another large model, we keep both the pre-trained RelitLRM and our intrinsic decomposition model (IDM), DiffusionRenderer Liang et al. (2025), frozen and train only compact modules. A latent-interpolated synthesis stage adds fine geometry in salient regions; MaterialFormer distills light-independent neural material features with guidance from IDM

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.



Figure 1: We propose F-RNG, a novel feed-forward relighting framework that directly reconstructs relightable neural Gaussian assets from sparse-view inputs. F-RNG utilizes large-model priors without re-training or fine-tuning them. The framework supports objects from diverse categories with complex appearances such as fur, enabling rendering under novel environment lighting and viewpoints. F-RNG can also reconstruct and relight assets from multi-view images generated by GPT-Image-2 Zewde et al. (2026) (left), as well as photos of a real-world figurine (right).

priors; and a universal neural renderer supports novel environment illumination and views without rerunning either large model.

Across synthetic and real datasets, F-RNG improves visual quality and overall quantitative performance over RelitLRM. After reconstruction, rendering a 512×512 image takes 246 ms, about $25\times$ faster per new environment map than RelitLRM's diffusion stage. Our contributions are:

- a feed-forward framework that produces reusable relightable 3D assets from sparse-view inputs without fine-tuning its large-model priors;
- prior-guided appearance distillation with MaterialFormer and a light-independence regularizer; and
- latent-interpolated geometry synthesis for high-frequency detail.

2 Related Work

Relightable neural representations. NeRF and 3DGS enable high-quality novel-view synthesis Mildenhall et al. (2021); Kerbl et al. (2023), and subsequent inverse-rendering methods separate geometry, material, and illumination Zhang et al. (2021); Liu et al. (2023); Boss et al. (2021). Gaussian approaches use controlled lighting Bi et al. (2020; 2024) or analytical BRDFs Jin et al. (2023); Jiang et al. (2024); Liang et al. (2024); Gao et al. (2023). These assumptions limit complex materials, while RNG Fan et al. (2025) models appearance with neural features but remains a per-scene optimization method. F-RNG retains neural appearance flexibility while generalizing across unseen objects.

Feed-forward 3D reconstruction. LRMs reconstruct radiance fields, Gaussians, meshes, or texture maps from sparse inputs Hong et al. (2023); Xu et al. (2024b); Zhang et al. (2024); Wei et al. (2024); Xu et al. (2024a); Siddiqui et al. (2024); Li et al. (2025). RelitLRM Zhang et al. (2025) is our backbone, but generates a lighting-specific radiance field with a full denoising pass for every target light. F-RNG instead distills its predictions into one reusable relightable asset. Sparse 3DGS methods Chung et al. (2024); Zhu et al. (2024); Li et al. (2024); Feng et al. (2024) and feed-forward NVS models Charatan et al. (2024); Chen et al. (2024) address sparse reconstruction, but retain baked-in lighting.

Image-based relighting and intrinsic decomposition. Recent diffusion methods directly relight images or estimate intrinsic buffers Zeng et al. (2024a); Jin et al. (2024); Kocsis et al. (2024); Luo et al. (2024); Zeng et al. (2024b). DiffusionRenderer Liang et al. (2025) predicts cross-view-

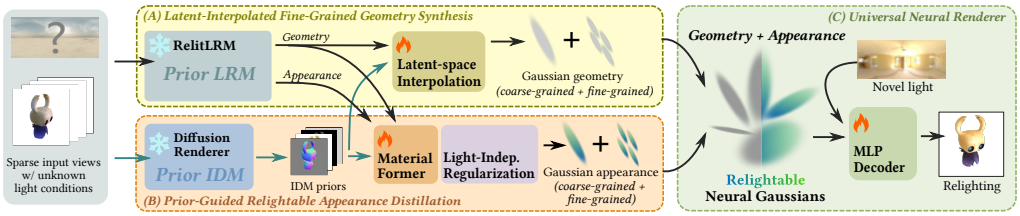
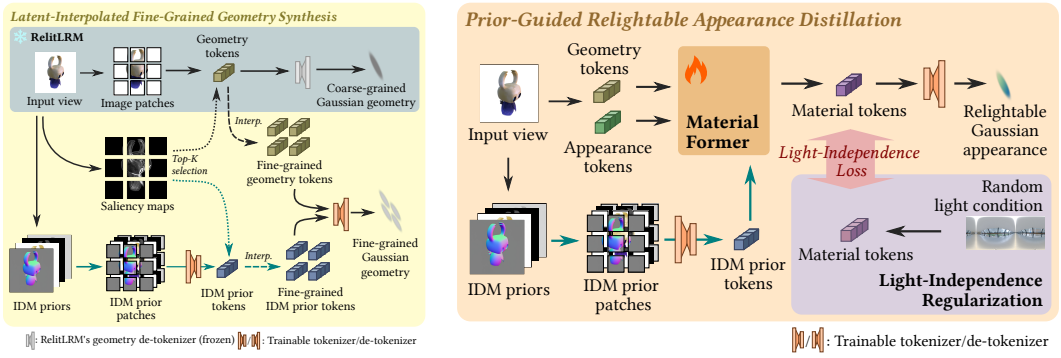


Figure 2: **Overview of F-RNG.** Frozen LRM and IDM priors support (A) latent-interpolated geometry synthesis, (B) relightable appearance distillation with MaterialFormer and light-independence regularization, and (C) a universal neural renderer for novel environment illumination and views.



(a) Fine-grained geometry synthesis.

(b) Prior-guided appearance distillation.

Figure 3: **Details of the two trainable stages.** (a) Salient tokens are interpolated and decoded into additional Gaussians. (b) MaterialFormer fuses LRM tokens with IDM priors and uses a second lighting observation for the light-independence constraint.

consistent shading buffers from a video; we use these buffers only as guidance, allowing the learned neural material to represent appearances that do not follow an analytical BRDF.

3 Method

Given sparse images I_i and Plücker rays $\mathbf{R}_i$, RelitLRM Zhang et al. (2025) predicts geometry tokens $\mathbf{T}_i^{\text{geo}}$ and lighting-specific appearance tokens $\mathbf{T}_i^{\text{app}}$. We also arrange the views as a video for DiffusionRenderer Liang et al. (2025), obtaining normal, depth, albedo, roughness, and metallic buffers. We tokenize these buffers as IDM priors $\mathbf{T}_i^{\text{prior}}$. Both large models remain frozen. As Fig. 2 shows, our trainable components refine the geometry, distill intrinsic neural material, and decode it for relighting. The two trainable stages are detailed in Fig. 3.

3.1 Latent-interpolated fine-grained geometry synthesis

RelitLRM's per-pixel prediction limits Gaussian density and loses fine structure. We locate the top-K patches using local grayscale standard deviation with boundary suppression (defined below), then bilinearly interpolate their geometry and prior tokens. A compact detokenizer maps these tokens to additional sub-pixel Gaussians:

$$\mathbf{T}_i^{\text{prior}} = \text{Tokenizer}(\text{Patchify}(\hat{\mathbf{g}}_i^{n,a,r,m})), \tag{1}$$

$$\mathbf{T}_j^{\text{fine-geo}} = \text{Bilinear}(\mathbf{T}_j^{\text{geo}}), \qquad \mathbf{T}_j^{\text{fine-prior}} = \text{Bilinear}(\mathbf{T}_j^{\text{prior}}), \tag{2}$$

$$\mathbf{G}_j^{\text{fine-geo}} = \text{FineGeoDetokenizer}(\mathbf{T}_j^{\text{fine-geo}} \mid \mathbf{T}_j^{\text{fine-prior}}). \tag{3}$$

Here j indexes selected patches. The original frozen decoder still produces coarse Gaussians, while the new single-layer MLP decodes only interpolated tokens. We use $K = 2\%$ based on the patch-budget analysis in Table 4. Unlike uniformly increasing the Gaussian density, this selection allocates

the additional geometry to regions with high-frequency detail while retaining the coarse structure predicted by the frozen LRM.

Saliency without silhouette bias. We compute local grayscale standard deviation with a box filter B_β of size β. Let D denote the distance from a foreground pixel to the background boundary. A smooth distance weight suppresses the otherwise high response at object silhouettes:

$$S_{\text{raw}} = \sqrt{\max\big(B_\beta(I_{\text{gray}}^2) - B_\beta(I_{\text{gray}})^2, 0\big)}, \tag{4}$$

$$W = \text{GaussianBlur}(\text{Clamp}(\text{Normalize}(D), 0, 1), \sigma)\,, \tag{5}$$

$$S_{\text{final}} = S_{\text{raw}} \cdot W. \tag{6}$$

The boundary buffer smoothly attenuates silhouette responses while retaining interior texture variation. Selecting the top-K patches from this refined map directs the limited synthesis budget toward surface details instead of foreground–background contrast. Geometry and IDM tokens use the same patch selection, preserving their spatial correspondence during interpolation.

3.2 Prior-guided relightable appearance distillation

RelitLRM radiance contains illumination that must be removed. MaterialFormer fuses its geometry and appearance tokens and conditions them on IDM priors:

$$\mathbf{T}_i^{\text{all}} = [\mathbf{T}_i^{\text{geo}} \oplus \mathbf{T}_i^{\text{app}}], \tag{7}$$

$$\mathbf{T}_i^{\text{mat}} = \text{MaterialFormer}(\mathbf{T}_i^{\text{all}} \mid \mathbf{T}_i^{\text{prior}}), \qquad \mathbf{G}_i^{\text{mat}} = \text{MatDetokenizer}(\mathbf{T}_i^{\text{mat}}). \tag{8}$$

The predicted $\mathbf{G}_i^{\text{mat}}$ is a neural material feature for each Gaussian. IDM buffers provide decomposition cues rather than direct targets, so MaterialFormer can correct imperfect priors and represent effects such as fur and subsurface scattering.

Material should remain invariant to illumination. During training, we run the frozen RelitLRM with an additional random environment E_1 for the same object, producing another appearance observation and material token $\mathbf{T}_{i1}^{\text{mat}}$. We enforce

$$\mathcal{L}_{\text{ind}} = \lambda_c \mathcal{C}(\mathbf{T}_i^{\text{mat}}, \mathbf{T}_{i1}^{\text{mat}}) + \lambda_{\text{KLD}} \mathcal{L}_{\text{KLD}}(\mathbf{T}_i^{\text{mat}}, \mathbf{T}_{i1}^{\text{mat}}), \tag{9}$$

where $\mathcal{C}$ is the cosine term and $\mathcal{L}_{\text{KLD}}$ preserves the latent distributions Kullback & Leibler (1951). This discourages baked-in illumination without requiring intrinsic ground truth.

3.3 Universal neural renderer

Following RNG Fan et al. (2025), we avoid a fixed analytical BRDF and attach a neural reflectance feature to each Gaussian. Unlike RNG's per-scene decoder, a single MLP Θ is shared across all objects:

$$\rho(\mathbf{G}_i^{\text{mat}}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) = \Theta(\mathbf{G}_i^{\text{mat}} \mid \boldsymbol{\omega}_o, \boldsymbol{\omega}_i). \tag{10}$$

We implement deferred shading: Gaussian splatting first alpha-blends the 32-channel neural features along each camera ray into an image-space feature; the shared decoder then maps this feature and the incident and outgoing directions to reflectance. This order avoids evaluating the appearance MLP separately for every overlapping Gaussian. The decoder has four layers with 256 hidden units each. We use directions in the common coordinate frame without a local shading-frame conversion, since a unique surface normal is difficult to define for fluffy objects.

For environment illumination, we importance-sample incident light directions and integrate their decoded responses. We use 64 samples per pixel (spp) during training and 128 spp at inference. Only 24 of the 64 training samples contribute gradients; the remaining samples reduce noise in the rendered supervision without storing their backward graphs. Once reconstructed, the asset therefore supports novel environment illumination and views without invoking the LRM or IDM again.

3.4 Architecture and training

MaterialFormer is a four-layer Transformer with hidden dimension 1024, pre-layer normalization, and GeLU activations. Following RelitLRM, we use 8×8 patches for both input images and IDM

Table 1: **Comparison with RelitLRM Zhang et al. (2025).** F-RNG improves quality and relights an existing asset under each new environment map about 25× faster. Best results are shaded.

Dataset	Ours			RelitLRM		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
BLENDERKIT	30.02	0.910	0.083	27.63	0.834	0.123
TENSOIR	30.96	0.948	0.057	28.66	0.936	0.075
OBJAVERSE	31.77	0.945	0.047	30.02	0.950	0.049
STANFORD-ORB	32.24	0.968	0.028	30.72	0.965	0.041
OWL	26.03	0.841	0.244	23.14	0.790	0.284
One-time reconstruction	14.6 s			6.9 s		
Relighting time (per light)	246 ms			6,826 ms		
End-to-end (20 novel lights)	19.52 s			138.39 s		

buffers. The sparse views are organized as a video when extracting DiffusionRenderer priors to encourage cross-view consistency. Only the added tokenizers, detokenizers, MaterialFormer, and universal renderer are trained; both large-model priors remain frozen throughout training.

We supervise the final rendered images with an $\mathcal{L}_2$ loss and a perceptual loss Chen & Koltun (2017), together with the material light-independence regularizer:

$$\mathcal{L} = \lambda_1 \mathcal{L}_2 + \lambda_{\text{per}} \mathcal{L}_{\text{per}} + \mathcal{L}_{\text{ind}}. \tag{11}$$

We set $(\lambda_c, \lambda_{\text{KLD}}, \lambda_1, \lambda_{\text{per}}) = (0.2, 0.2, 1.0, 0.5)$. Training at 512×512 resolution converges in nine days on eight NVIDIA L40 GPUs (48 GB each), or 72 GPU-days. The training set contains 1,000 objects from OBJAVERSE-XL and BLENDERKIT, with approximately 40% diffuse, 40% specular, and 20% fiber appearances. Each object has six sparse input views; supervision uses 100 target viewpoints under each of 100 environment maps, yielding 10 million rendered images. Appendix A.1 gives the data-rendering settings. The compact adaptation thus learns a relightable latent space from existing reconstruction and material priors, instead of learning general 3D reconstruction from these 1,000 objects alone.

4 EXPERIMENTS

Setup. We evaluate on synthetic BLENDERKIT, OBJAVERSE Deitke et al. (2023), TENSOIR Jin et al. (2023), RNA Mullia et al. (2024), and RNG Fan et al. (2025) objects, and real STANFORD-ORB Kuang et al. (2023) and OBJECTS-WITH-LIGHTING Ummenhofer et al. (2024) captures. Unless noted, we use six 512×512 input views and report PSNR, SSIM Wang et al. (2004), and LPIPS Zhang et al. (2018). Training details are given in Sec. 3.4; additional dataset and memory information is in Appendix A.

4.1 RELIGHTING QUALITY AND EFFICIENCY

F-RNG achieves better overall quantitative results than RelitLRM across the evaluated benchmarks (Table 1). On one NVIDIA L40 at 512×512, reconstruction takes 6.9 s for RelitLRM and 14.6 s for F-RNG; relighting takes 6.826 s and 246 ms per environment map, respectively (about 25× faster). Our reconstruction includes IDM extraction, RelitLRM inference, MaterialFormer, and asset assembly. For 20 new environment maps, total times are 19.52 s for F-RNG and 138.39 s for RelitLRM, a 7.1× speedup. Figs. 1, 4, and 5 show clearer textures and more faithful material appearance. We do not compare with LSRM Li et al. (2026) because we have not obtained permission to access its model checkpoint.

With six views, F-RNG outperforms GS-IR Liang et al. (2024) and R3DG Gao et al. (2023), which use 100 views and about 50 minutes per object (Table 2). For image-based comparisons, all methods relight the same single input view under the same environment map. F-RNG produces better visual results and outperforms Neural Gaffer Jin et al. (2024) and DiLightNet Zeng et al. (2024a) on the reported metrics (Table 5). Additional visual comparisons appear in Appendix B.

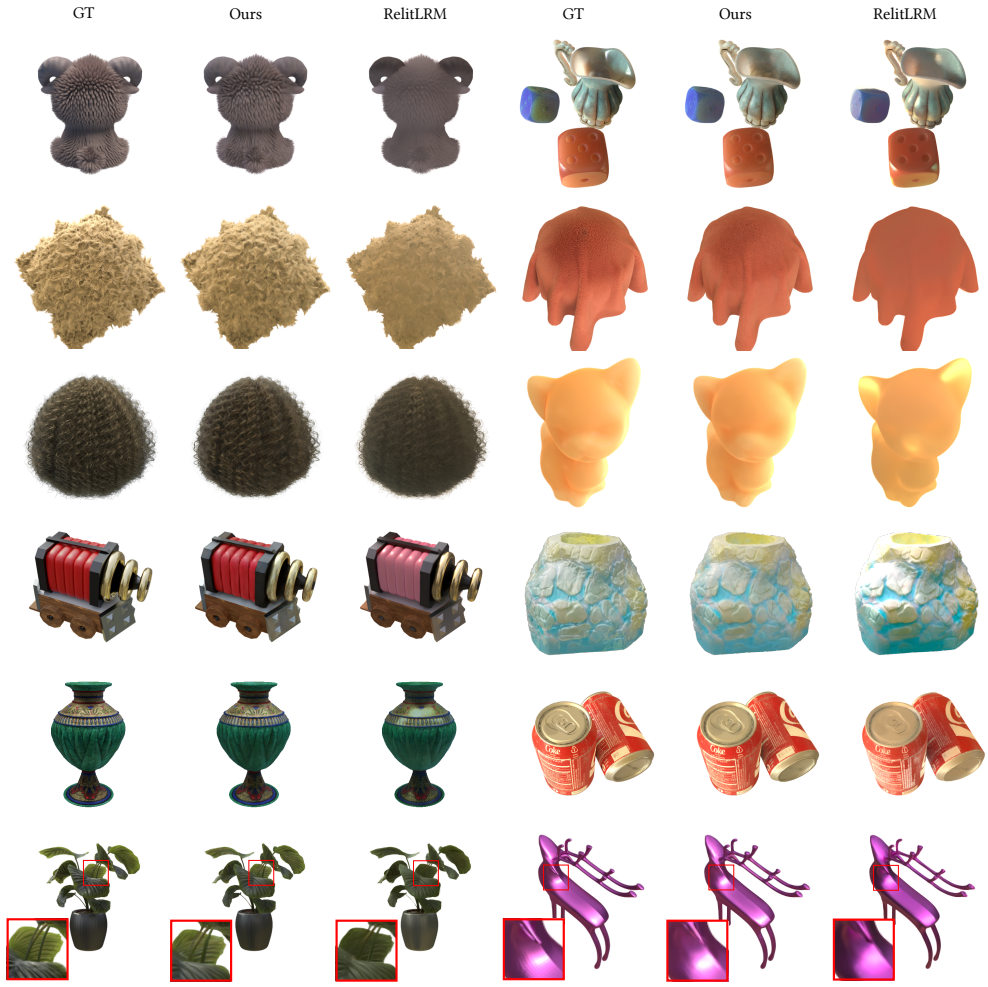


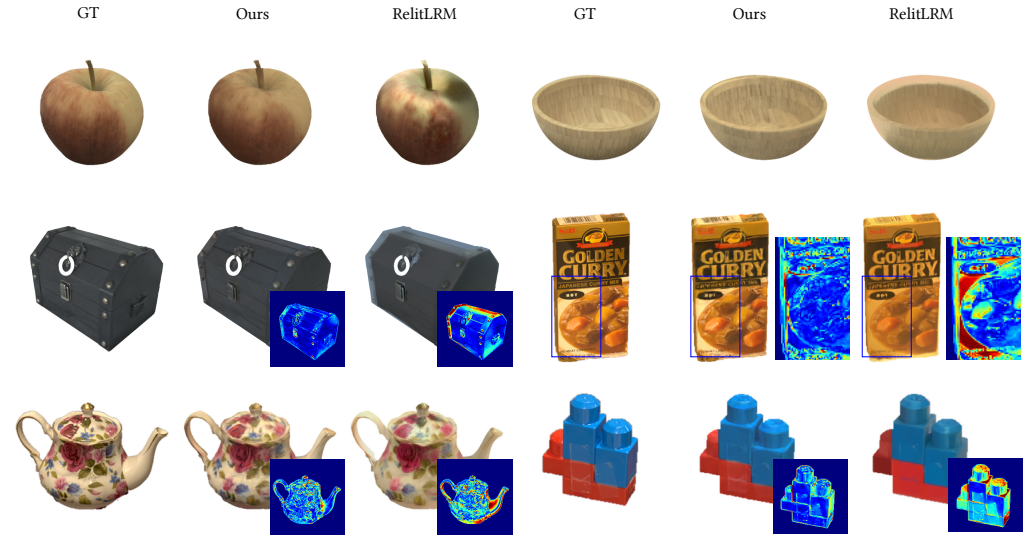
Figure 4: **Synthetic relighting comparison.** F-RNG improves decomposition and preserves high-frequency appearance better than RelitLRM.

Table 2: **Comparison with dense-view 3DGS relighting methods Liang et al. (2024); Gao et al. (2023).** F-RNG uses six views; the baselines use 100. Best and second-best results are shaded; ties at the reported precision share a color.

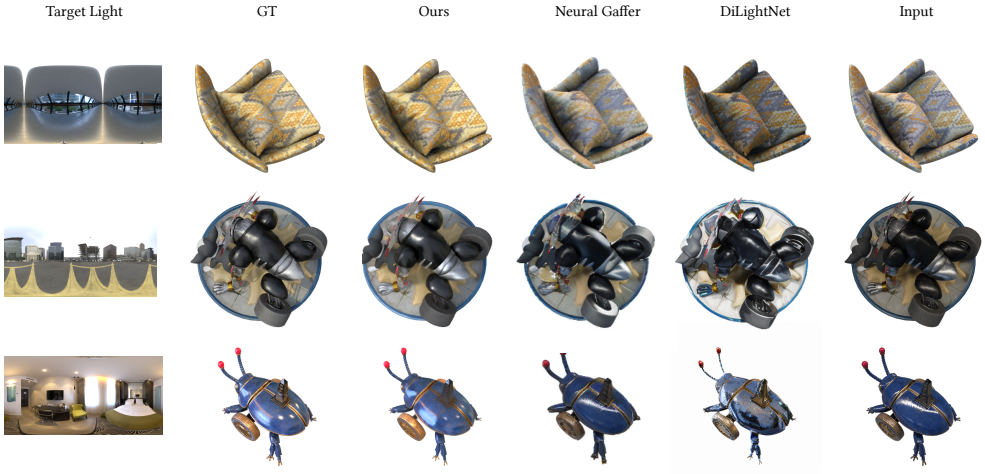
Scene	Ours *(6 views)*			GS-IR *(100 views)*			R3DG *(100 views)*		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Armadillo	34.94	0.983	0.03	27.65	0.908	0.08	30.76	0.953	0.06
Lego	29.65	0.901	0.06	22.88	0.834	0.11	22.49	0.868	0.09
Hotdog	29.78	0.943	0.08	21.51	0.885	0.14	24.59	0.916	0.08
Ficus	29.47	0.965	0.05	23.63	0.866	0.09	27.23	0.964	0.03
TensoIR	30.96	0.948	0.05	23.92	0.873	0.10	26.27	0.925	0.06
Objaverse	31.77	0.945	0.04	20.64	0.825	0.17	26.44	0.907	0.08
Recons. time	14.6 s			∼50 min			∼50 min		

4.2 Ablation Study

Table 3 shows that each component contributes to the final quality. Removing fine-grained Gaussians weakens detail reconstruction, while removing IDM guidance harms appearance decompo-



(a) Real captures.



(b) Matched single-view relighting.

Figure 5: **Real and image-based relighting comparisons.** F-RNG preserves texture better than RelitLRM on real captures and outperforms image-based methods when all methods relight the same single input view.

sition. Removing either normal or albedo priors also hurts performance. The light-independence loss suppresses baked illumination, while neural appearance modeling improves complex materials

Table 3: **Component ablation on OBJAVERSE.** All variants are trained at 256×256 resolution; best results are shaded.

Ablations	PSNR	SSIM	LPIPS
Our full model	32.44	0.957	0.043
w/o fine-grained Gaussians	31.46	0.945	0.047
w/o IDM priors	31.21	0.947	0.049
w/o light-indep. loss	32.07	0.954	0.044
w/o neural appearance	31.72	0.950	0.047

Table 4: **Patch budget and selection** on OBJAVERSE. Random and uniform selection use the same $K = 2\%$ budget.

K	PSNR	SSIM	LPIPS
1%	31.05	0.940	0.052
2%	**31.77**	**0.945**	**0.047**
4%	31.70	0.945	0.046
8%	31.54	0.943	0.047
Random	30.65	0.934	0.056
Uniform	30.77	0.937	0.055

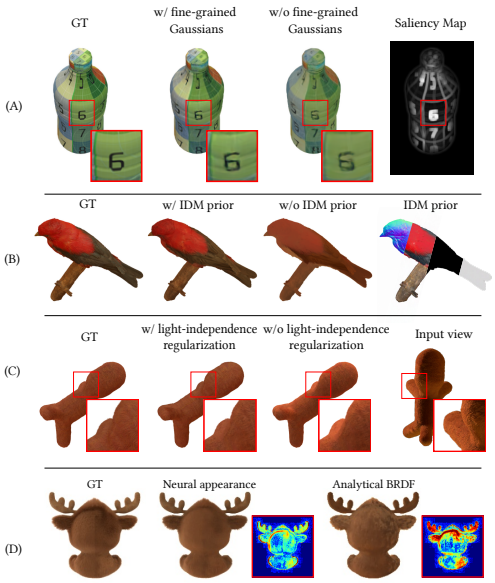


Figure 6: **Visual ablation.** Fine-grained geometry sharpens details; IDM priors and light-independence improve decomposition; neural appearance modeling better represents complex materials.

Table 5: Single-view relighting on OBJAVERSE. All methods receive and relight the same input view; best and second-best results are shaded.

Method	PSNR	SSIM	LPIPS
Ours	28.43	0.93	0.062
Neural Gaffer	26.55	0.90	0.092
DiLightNet	20.34	0.84	0.175

Table 6: **Effect of the number of input views.** Six views provide most of the quality gain; additional views yield diminishing returns.

Dataset	Views	PSNR	SSIM	LPIPS
OBJAVERSE	2	24.84	0.901	0.103
	4	28.31	0.916	0.069
	6	**31.77**	**0.945**	**0.047**
	8	31.83	0.950	0.049
	10	31.95	0.951	0.049
TENSOIR	2	24.53	0.889	0.114
	4	27.75	0.908	0.077
	6	**30.96**	**0.948**	**0.057**
	8	31.14	0.948	0.057
	10	31.28	0.949	0.054

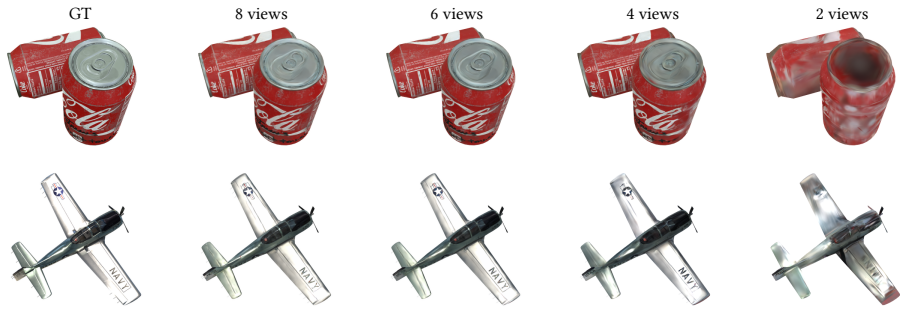


Figure 7: Effect of the number of input views; six views balance quality and memory cost.

compared with analytical shading. Figure 6 illustrates these effects: geometry refinement recovers local texture, the IDM guides appearance decomposition, and the regularizer reduces residual illumination in the learned material. The feature similarities in Appendix D further support the light-independence constraint.

Where to add geometry. Table 4 compares refinement budgets and selection strategies. Increasing K from 1% to 2% improves reconstruction quality, while larger budgets offer little further benefit. At the same budget, saliency-based selection outperforms random and uniform selection. These results support concentrating refinement on patches with useful detail; adding more Gaussians can introduce redundancy and make refinement harder to learn.

4.3 INPUT VIEWS AND PRIOR ROBUSTNESS

Sparse-view quality and memory. Table 6 shows a clear improvement in reconstruction quality as the input increases from two to six views on both datasets. Additional views bring only modest gains, so we use six views to balance quality and memory. The raw per-pixel prediction at this setting contains about 1.57M Gaussians; filtering opacities below 0.01 typically leaves 100K–200K stored Gaussians. Including their 32-channel features, these assets occupy about 50–100 MB of GPU

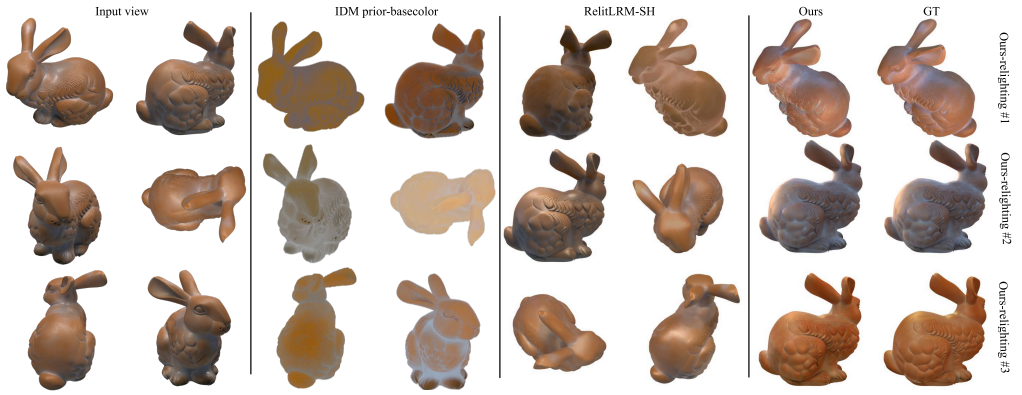


Figure 8: **Correcting imperfect priors.** From left to right: input views, IDM basecolor, RelitLRM radiance, our relighting, and ground truth under three novel lights. The final results recover plausible appearance despite color stains and baked illumination in the priors.

Table 7: **Comparison with image-based relighting models on OBJECTS-WITH-LIGHTING Dataset Ummenhofer et al. (2024).** F-RNG improves quantitative performance and visual quality over LightSwitch Litman et al. (2025) while relighting substantially faster (Fig. 16). The best/second-best results are marked in red / yellow.

Method	PSNR	SSIM	LPIPS	*Relighting time*
Ours	26.03	0.841	0.244	0.3 s
LightSwitch	25.43	0.840	0.297	60 s
DiffusionRenderer	21.57	0.795	0.341	28 s

memory in our implementation. More views increase both the raw Gaussian count and memory demand; Appendix A.3 gives further details and Fig. 7 provides qualitative comparisons.

Imperfect large-model priors. The IDM buffers condition MaterialFormer rather than serving as intrinsic ground-truth targets. This allows the final rendering supervision to correct errors shared by the input priors. Figure 8 compares IDM basecolor and RelitLRM radiance with our outputs under three novel lights: despite color stains in the basecolor and biased baked radiance, the relit results remain closer to the ground truth. This behavior supports using complementary reconstruction and decomposition priors instead of directly treating either prediction as the final material representation.

5 LIMITATIONS AND CONCLUSION

F-RNG can represent layered and scattering appearances through alpha-blended neural features, but highly reflective or transparent materials remain challenging and may yield blurred reflections. The renderer also models visibility implicitly, which limits sharp cast shadows under point lights outside its environment-lit training distribution. Its per-pixel Gaussian parameterization also makes memory grow linearly with input count and resolution; six views balance quality and cost in our experiments. Additional failure cases appear in Appendix C, with memory analysis in Appendix A.3. Rendering also requires Monte Carlo sampling, with 128 spp used for the reported inference results.

We presented a feed-forward framework that converts sparse images into reusable relightable neural Gaussian assets with frozen LRM and IDM priors. Small modules refine geometry, distill light-independent material, and render under novel environment illumination without repeated large-model inference. Compared with RelitLRM, our method improves visual quality and overall quantitative performance, with $25\times$ faster relighting per new environment map. Future work includes explicit visibility for sharp shadows and extension from isolated objects to full scenes.

REFERENCES

Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *ACM TOG*, 2020.

Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. Gs^3: Efficient relighting with triple gaussian splatting. In *SIGGRAPH Asia 2024 Conference Papers*, 2024.

Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems*, 34:10691–10704, 2021.

David Charatan, Sizhe Lester Li, Andrea Tagliasacchi, and Vincent Sitzmann. pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 19457–19467, 2024.

Qifeng Chen and Vladlen Koltun. Photographic image synthesis with cascaded refinement networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 1511–1520, 2017.

Yuedong Chen, Haofei Xu, Chuanxia Zheng, Bohan Zhuang, Marc Pollefeys, Andreas Geiger, Tat-Jen Cham, and Jianfei Cai. MVSplat: Efficient 3D gaussian splatting from sparse multi-view images. In *European Conference on Computer Vision*, pp. 370–386. Springer, 2024.

Jaeyoung Chung, Jeongtaek Oh, and Kyoung Mu Lee. Depth-regularized optimization for 3d gaussian splatting in few-shot images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 811–820, 2024.

Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13142–13153, 2023.

Jiahui Fan, Fujun Luan, Jian Yang, Milos Hasan, and Beibei Wang. RNG: Relightable neural gaussians. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 26525–26534, 2025.

Qijun Feng, Zhen Xing, Zuxuan Wu, and Yu-Gang Jiang. Geogs3d: Single-view 3d reconstruction via geometric-aware diffusion model and gaussian splatting. *arXiv preprint arXiv:2403.10242*, 2024.

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 preprint arXiv:2311.16043*, 2023.

Yicong Hong, Kai Zhang, Jiuxiang Gu, Sai Bi, Yang Zhou, Difan Liu, Feng Liu, Kalyan Sunkavalli, Trung Bui, and Hao Tan. Lrm: Large reconstruction model for single image to 3d. *arXiv preprint arXiv:2311.04400*, 2023.

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

Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. TensoIR: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 165–174, 2023.

Haian Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and Noah Snavely. Neural gaffer: Relighting any object via diffusion. *Advances in Neural Information Processing Systems*, 37:141129–141152, 2024.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

Peter Kocsis, Julien Philip, Kalyan Sunkavalli, Matthias Nießner, and Yannick Hold-Geoffroy. Lightit: Illumination modeling and control for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9359–9369, 2024.

Zhengfei Kuang, Yunzhi Zhang, Hong-Xing Yu, Samir Agarwala, Elliott Wu, Jiajun Wu, et al. Stanford-orb: a real-world 3d object inverse rendering benchmark. In *Advances in Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. DNGaussian: Optimizing sparse-view 3D gaussian radiance fields with global-local depth normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20775–20785, 2024.

Zhengqin Li, Dilin Wang, Ka Chen, Zhaoyang Lv, Thu Nguyen-Phuoc, Milim Lee, Jia-Bin Huang, Lei Xiao, Yufeng Zhu, Carl S Marshall, et al. Lirm: Large inverse rendering model for progressive reconstruction of shape, materials and view-dependent radiance fields. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 505–517, 2025.

Zhengqin Li, Cheng Zhang, Jakob Engel, and Zhao Dong. LSRM: High-fidelity object-centric reconstruction via scaled context windows. *arXiv preprint arXiv:2604.05182*, 2026.

Ruofan Liang, Zan Gojcic, Huan Ling, Jacob Munkberg, Jon Hasselgren, Chih-Hao Lin, Jun Gao, Alexander Keller, Nandita Vijaykumar, Sanja Fidler, et al. Diffusion renderer: Neural inverse and forward rendering with video diffusion models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 26069–26080, 2025.

Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. GS-IR: 3D gaussian splatting for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21644–21653, 2024.

Yehonathan Litman, Fernando De la Torre, and Shubham Tulsiani. Lightswitch: Multi-view relighting with material-guided diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 27750–27759, 2025.

Yuan Liu, Peng Wang, Cheng Lin, Xiaoxiao Long, Jiepeng Wang, Lingjie Liu, Taku Komura, and Wenping Wang. Nero: Neural geometry and brdf reconstruction of reflective objects from multi-view images. *ACM Transactions on Graphics (TOG)*, 42(4):1–22, 2023.

Jundan Luo, Duygu Ceylan, Jae Shin Yoon, Nanxuan Zhao, Julien Philip, Anna Frühstück, Wenbin Li, Christian Richardt, and Tuanfeng Wang. Intrinsicdiffusion: Joint intrinsic layers from latent diffusion models. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–11, 2024.

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. *Communications of the ACM*, 65(1):99–106, 2021.

Krishna Mullia, Fujun Luan, Xin Sun, and Miloš Hašan. Rna: Relightable neural assets. *ACM Transactions on Graphics*, 2024.

Yawar Siddiqui, Tom Monnier, Filippos Kokkinos, Mahendra Kariya, Yanir Kleiman, Emilien Garreau, Oran Gafni, Natalia Neverova, Andrea Vedaldi, Roman Shapovalov, et al. Meta 3d assetgen: Text-to-mesh generation with high-quality geometry, texture, and pbr materials. *Advances in Neural Information Processing Systems*, 37:9532–9564, 2024.

Benjamin Ummenhofer, Sanskar Agrawal, Rene Sepúlveda, Yixing Lao, Kai Zhang, Tianhang Cheng, Stephan R. Richter, Shenlong Wang, and Germán Ros. Objects with lighting: A real-world dataset for evaluating reconstruction and rendering for object relighting. In *3DV*. IEEE, 2024.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 2004.

Xinyue Wei, Kai Zhang, Sai Bi, Hao Tan, Fujun Luan, Valentin Deschaintre, Kalyan Sunkavalli, Hao Su, and Zexiang Xu. Meshlrm: Large reconstruction model for high-quality meshes. *arXiv preprint arXiv:2404.12385*, 2024.

Jiale Xu, Weihao Cheng, Yiming Gao, Xintao Wang, Shenghua Gao, and Ying Shan. Instantmesh: Efficient 3d mesh generation from a single image with sparse-view large reconstruction models. *arXiv preprint arXiv:2404.07191*, 2024a.

Yinghao Xu, Zifan Shi, Wang Yifan, Hansheng Chen, Ceyuan Yang, Sida Peng, Yujun Shen, and Gordon Wetzstein. Grm: Large gaussian reconstruction model for efficient 3d reconstruction and generation. In *European Conference on Computer Vision*, pp. 1–20. Springer, 2024b.

Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. Dilightnet: Fine-grained lighting control for diffusion-based image generation. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–12, 2024a.

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, 2024b. Association for Computing Machinery. ISBN 9798400705250. doi: 10.1145/3641519.3657445. URL https://doi.org/10.1145/3641519.3657445.

Kidus Zewde, Simiao Ren, Xingyu Shen, Jenny Wu, Yuchen Zhou, Tommy Duong, Zikang Zhang, and Ethan Traister. Gpt-image-2 in the wild: A twitter dataset of self-reported ai-generated images from the first week of deployment. *arXiv preprint arXiv:2604.25370*, 2026.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5453–5462, 2021.

Kai Zhang, Sai Bi, Hao Tan, Yuanbo Xiangli, Nanxuan Zhao, Kalyan Sunkavalli, and Zexiang Xu. Gs-lrm: Large reconstruction model for 3d gaussian splatting. In *European Conference on Computer Vision*, pp. 1–19. Springer, 2024.

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.

Tianyuan Zhang, Zhengfei Kuang, Haian Jin, Zexiang Xu, Sai Bi, Hao Tan, He Zhang, Yiwei Hu, Milos Hasan, William T Freeman, et al. RelitLRM: Generative relightable radiance for large reconstruction models. In *International Conference on Learning Representations*, 2025.

Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. Fsgs: Real-time few-shot view synthesis using gaussian splatting. In *European conference on computer vision*, pp. 145–163. Springer, 2024.