IR-HGP: Physically-Aware Gaussian Inverse Rendering for High-Illumination Scenes via Generative Priors

Supplementary Material

6. Details and Mesh Update Strategy for HVD

Periodic Update Mechanism and Efficiency. To prevent specular noise from destabilizing the pipeline, we decouple geometry extraction from per-iteration appearance optimization by updating the TSDF proxy mesh periodically. In our implementation, we set this frequency to every 5,000 iterations. This mesh serves solely as a coarse visibility proxy, not for material estimation.

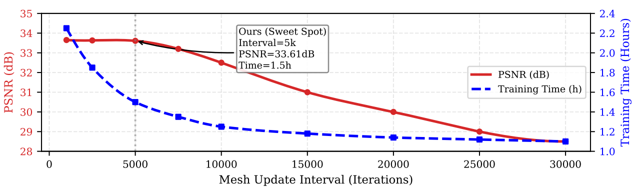

Figure 12. Efficiency trade-off. Updating the proxy mesh every 5k iterations strikes an optimal balance between occlusion reasoning and runtime efficiency.

As illustrated in Fig. 12, this 5k-iteration interval was determined empirically to balance computational cost and visibility fidelity. Increasing the update frequency to every 1k iterations yields negligible quality improvements since the coarse geometry converges early, yet it severely inflates the 18% training time overhead already caused by mesh extraction. Conversely, delaying updates to every 15k iterations leads to outdated visibility proxies, causing shadow inconsistencies against the continuously refined Gaussian surfels.

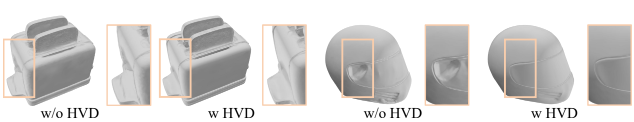

Figure 13. Geometry evolution of toaster and helmet. The periodic proxy mesh ensures robust visibility even under intense illumination.

HVD Ensures Geometric Correctness. Beyond efficiency, HVD prevents specular collapse through a dual strategy: (1) 2DGS normal constraints suppress specular noise; and (2) TSDF acts as a low-pass filter for geometric outliers. As shown in Fig. 13, even an imperfect proxy mesh bounds potential errors to shadow softness, preventing overall geometric degradation under intense illumination.

7. Detailed Analysis of the GIFP

Network Architecture and Condition Encoding. Our conditional diffusion model D_{HDR} employs a latent U-Net architecture adapted for HDR equirectangular images at a 256×512 resolution with 4 sampling stages to ensure computational efficiency. For conditioning, we extract coarse scene illumination features L_{Coarse} via pre-trained ResNet-18. These features are injected into the U-Net's bottleneck and upsampling blocks via cross-attention, providing robust physical conditioning from sparse views.

Score Distillation Sampling Configuration. During optimization, D_{HDR} acts as a frozen critic providing Score Distillation Sampling gradients. We sample timesteps $t \sim \mathcal{U}(0.02, 0.98)$; this range prevents high-frequency artifacts at lower bounds and structural drift at higher bounds. We apply classifier-free guidance with a scale of $w = 7.5$ to enforce lighting plausibility without over-saturating colors. Finally, to allow the underlying geometry to stabilize before applying strong generative regularization, the generative loss weight λ_g is linearly warmed up from 0 to 1.0 over the first 3,000 iterations.

Table 4. Ablation: Full Model vs. w/o GIFP.

		Ours w/o GIFP			Ours Full Model		
Dataset	Object	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
NeRF Synthetic	Chair	33.40	0.9752	0.0351	35.10	0.9819	0.0211
	Drums	28.11	0.9613	0.0412	29.81	0.9725	0.0279
	Ficus	34.77	0.9757	0.0395	36.86	0.9836	0.0102
	Hotdog	35.83	0.9652	0.0453	37.73	0.9776	0.0352
	Lego	32.13	0.9624	0.0481	34.03	0.9714	0.0364
	Material	24.57	0.9581	0.0524	26.07	0.9680	0.0376
Shiny Blender	Car	32.95	0.9698	0.0452	34.63	0.9821	0.0209
	Coffee	33.53	0.9622	0.1012	35.33	0.9764	0.0892
	Helmet	33.20	0.9683	0.0621	35.00	0.9797	0.0379
	Toaster	29.81	0.9512	0.0814	31.57	0.9674	0.0527
Mean		31.83	0.9649	0.0552	33.61	0.9761	0.0369

Quantitative Ablation on GIFP. As detailed in Tab. 4, the synergy between our HVD and PARC modules establishes a highly competitive baseline, consistently outperforming both GS-IR and R3DG. Crucially, the integration of the GIFP module drives substantial gains across all evaluation metrics, firmly solidifying our state-of-the-art performance.

Figure 14. Visual consistency and optimization stability across varying initialization random seeds, demonstrating the robustness enforced by the strong HDR manifold constraints.

Stability and Generalization. As illustrated in Fig. 14, GIFP's strong HDR manifold constraints ensure stable convergence regardless of initialization. Furthermore, rather than memorizing training views, GIFP learns physical HDR statistics. Since our training data includes semi-open scenes, the model learns to treat the sun as an extreme directional light source, effectively constraining optimization to physically plausible states even in unseen environments.

8. Detailed Analysis of the PARC

The PARC module is essential for stabilizing optimization under the extreme dynamic ranges characteristic of high-illumination scenes. Below, we detail its mathematical formulation, optimization strategy, and gradient behavior.

Tone Mapping Formulation. Our learnable tone mapping function $\mathcal{C}_{PARC}$ utilizes a modified ACES curve based on the Narkowicz approximation for its superior differentiability and computational efficiency:

$$\mathcal{C}_{PARC}(x) = \frac{x(2.51x + 0.03)}{x(2.43x + 0.59) + 0.14} \tag{10}$$

where $x = L_{in} \cdot \beta$. In this formulation, β serves as a learnable scaling factor that adaptively calibrates the scene's exposure prior to non-linear compression.

Initialization and Optimization. We initialize the exposure parameter to $\beta_{init} = 1.0$ and optimize it using a dedicated learning rate of 1×10^{-4}. This rate is deliberately set lower than the learning rate for spherical harmonics coefficients to prevent the exposure parameter from prematurely absorbing lighting information that should be attributed to the environment map. To ensure physical consistency, the exposure parameter is constrained to be strictly positive via a softplus activation function during training.

Gradient Analysis. Standard linear photometric losses frequently suffer from exploding gradients in high-luminance regions exceeding standard pixel limits. The derivative of our PARC curve naturally mitigates this by attenuating gradients for intense highlights, acting as an implicit soft-clipping mechanism. Specifically, as input radiance approaches infinity, the corresponding gradient smoothly approaches zero. This behavior prevents the popping artifacts typical of 3DGS optimization against HDR ground truth, effectively balancing the supervisory signal between saturated light sources and subtle shadow regions.