Sparse Light Field-Guided 3D Gaussian Relighting with Depth-Semantic Optimization

YukunHuang, YiqiXu, ShaorongWang*

Beijing Forestry University

Abstract. 3D Gaussian relighting is essential for interactive object editing and photorealistic rendering. Current methods depend on dense SfM point clouds for BRDF initialization, but non-uniform distributions in complex material regions yield incomplete geometry and compromise light propagation in occluded areas, causing material decomposition ambiguities and unstable shadows. We propose DSSG, a relighting framework integrating depth-semantic optimization with sparse light field guidance. First, our Sparse Large Variance (SLV) strategy generates Gaussian distributions from random points, covering potential light interaction regions, while progressive filtering suppresses material noise. Second, we introduce depth-semantic optimization: DPT-generated depth maps optimize geometric consistency via Pearson correlation for stable shadow tracing, while DINO-ViT features align cross-view semantics to resolve albedo-shading ambiguity. A geometry-material alternating strategy dynamically schedules constraints across reconstruction phases. Our differentiable rendering framework decouples BRDF attributes from light transport, achieving physically plausible relighting. Experiments show improved novel view synthesis and relighting quality with reduced shadow errors under various lighting conditions.

Keywords: 3D Gaussian Splatting · Relighting · Depth Geometric Constraints · Semantic Consistency · Sparse Light Field Guidance

1 Introduction

3D scene reconstruction and relighting bridge computer vision and graphics to achieve photorealistic rendering under dynamic lighting. Current approaches include polygon mesh-based methods [3] with explicit BRDF but limited by geometric accuracy; NeRF-based methods [1] with high-quality synthesis but opaque material decomposition and real-time bottlenecks; and 3D Gaussian Splatting (3DGS) [2], which offers promising relighting with real-time capabilities through differentiable rasterization.

Current 3DGS relighting methods [4–6] face two challenges. First, SfM dependency creates geometry-material coupling: (1) Non-uniform sampling causes spatial biases in material optimization; (2) Poor Monte Carlo efficiency accumulates errors in indirect illumination; (3) Geometric errors propagate through normals to BRDF, creating detrimental feedback loops. Second, traditional optimization fails to handle nonlinear error propagation, causing cross-view albedo ambiguities. While deep learning methods [26] introduce priors and physics-based approaches [7] achieve high precision, they lack real-time capability. Recent semantic methods [8] inadequately consider dynamic lighting physics, producing shadow artifacts and implausible reflections.

We propose DSSG (Depth-Semantic driven Sparse-light-field Gaussian), leveraging depth and semantic information for stable geometry-material decomposition. We use monocular depth for geometric guidance, enabling accurate surface capture, particularly at boundaries. DINO-ViT features establish semantic correspondences across views, reducing BRDF ambiguities. Our SLV initialization [9] randomly distributes large-variance Gaussians covering light interaction regions, followed by progressive filtering for adaptive convergence while maintaining smooth material transitions.

Our contributions: 1) Depth-semantic optimization using DPT depth maps and DINO-ViT features to resolve cross-view diffuse inconsistencies; 2) SLV-based initialization with progressive frequency filtering for robust BRDF estimation; 3) The DSSG framework achieving real-time relighting with high quality through collaborative geometry-material-lighting optimization.

2 Related Work

2.1 Neural Rendering-based Scene Relighting

Neural rendering methods fall into two categories: NeRF-based implicit representations and explicit 3DGS approaches. NeRF variants model radiance implicitly through MLPs, with PhySG [10] decomposing BRDF, NeRFactor [11] handling global illumination, and Ref-NeRF [12] improving specularity. Despite advances by NeRD [13] and InvRender [14], dense sampling prevents real-time performance.3DGS [2] achieves efficiency through anisotropic Gaussian rasterization. Extensions include Relightable3DGS [15] for BRDF integration, PhysGauss [16] for radiance transfer, and GauFRe [17] for frequency decomposition. GIR [18] and GS-IR [19] further enable real-time global illumination through indirect lighting and incident radiance modeling.

2.2 BRDF Parameter Estimation

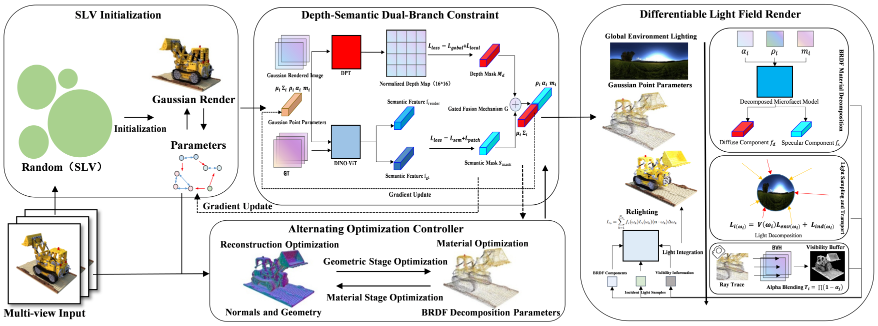
BRDF estimation from images remains challenging due to geometry-material-lighting entanglement. Deep learning approaches predict parameters directly: Zhang et al. [20] leveraged polarization cues. While efficient, these methods suffer from limited generalization.Differentiable rendering offers better accuracy-efficiency trade-offs. Bi et al. [21] jointly optimize geometry and BRDF, while Neural-PIL [22] accelerates computation via pre-integrated lighting. Within 3DGS frameworks, Relightable3DGS [15] achieves real-time relighting and PhysGauss [16] enhances specularity through importance sampling.

2.3 Depth and Semantic Information in 3D Reconstruction

Depth-semantic fusion provides complementary constraints for robust reconstruction. Pre-trained models include depth estimators (DPT [23]) and semantic analyzers (DINO-ViT [24]). Recent neural methods achieve dynamic fusion, with GaussianGroup [25] incorporating semantics into 3DGS for enhanced dynamic reconstruction.

3 Method

Figure 1 illustrates DSSG, our relightable 3D Gaussian Splatting framework, comprising three stages: (1) SLV initialization for uniform Gaussian point cloud generation within camera frustums; (2) depth-semantic optimization using DPT-based depth maps and DINO-ViT features for geometric consistency and material alignment; (3) differentiable rendering for physically realistic dynamic relighting. The pipeline alternates between geometry reconstruction and material optimization phases.



SLV Initialization
Random（SLV）
Initialization
Gaussian Render
Parameters
Gradient Update
Multi-view Input
Depth-Semantic Dual-Branch Constraint
Gaussian Rendered Image
DPT
Normalized Depth Map（16*16）
$L_{loss} = L_{gobal}+L_{local}$
Depth Mask M_d
$\mu_i\ \Sigma_i\ \rho_i\ \alpha_i\ m_i$
Gaussian Point Parameters
GT
DINO-ViT
Semantic Feature f_{render}
Semantic Feature f_{gt}
$L_{loss} = L_{sem}+L_{patch}$
Semantic Mask S_{mask}
Gated Fusion Mechanism G
$\rho_i\ \alpha_i\ m_i$
$\mu_i\ \Sigma_i$
Gradient Update
Alternating Optimization Controller
Reconstruction Optimization
Material Optimization
Geometric Stage Optimization
Material Stage Optimization
Normals and Geometry
BRDF Decomposition Parameters
Differentiable Light Field Render
Global Environment Lighting
Gaussian Point Parameters
Relighting
$L_o = \sum_{k=1}^{N_s} f_r(\omega_k) L_i(\omega_k)(\mathrm{n}\cdot\omega_k)\Delta\omega_k$
Light Integration
BRDF Components
Incident Light Samples
Visibility Information
α_i
ρ_i
m_i
Decomposed Microfacet Model
Diffuse Component f_d
Specular Component f_s
BRDF Material Decomposition
$L_{i(\omega_i)} = V(\omega_i)L_{env(\omega_i)} + L_{ind(\omega_i)}$
Light Decomposition
Light Sampling and Transport
Ray Trace
BVH
Visibility Buffer
Alpha Blending $T_i = \prod(1-\alpha_j)$

Fig. 1. Overview of the proposed DSSG framework for relightable 3D Gaussian Splatting.

3.1 Sparse Light Field Guided Initialization

We introduce a random large-variance strategy for sparse 3D Gaussian initialization with progressive frequency-domain filtering to ensure uniform spatial coverage.

Random Large-Variance Sparse Light Field Initialization Building on RAIN-GS [9], we develop a random large-variance sampling method tailored for relightable rendering. Our approach incorporates three improvements: (1) larger initial variances to expand Gaussian influence for light transport; (2) depth gradient masks guiding variance decay in high-curvature regions; (3) random uniform distributions for complex lighting effects.

We uniformly sample N Gaussian points $\{\mathcal{G}_i\}_{i=1}^N$ within view frustum $\Omega \subset \mathbb{R}^3$, with positions $\mu_i \sim \mathcal{U}(\Omega)$ and isotropic covariance initialization. Unlike SfM-based methods, our strategy effectively covers light-critical regions including specular and indirect lighting areas. Sampling density $\rho = N/|\Omega|$ adapts to scene scale.

The initial Gaussian probability distribution is:

$$P(x) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}(x \mid \mu_i, \Sigma_0) \tag{1}$$

This distribution provides smooth gradients for material decomposition, reducing parameter oscillations common with local over-dense sampling. Gaussian points progressively converge to surfaces through RGB loss, depth constraints, and frequency filtering.

Progressive Frequency-Domain Filtering To prevent high-frequency noise and material oscillations from random initialization, we employ progressive frequency-domain filtering with time-varying low-pass filters.

The Gaussian modulation function is:

$$\Sigma_{t+1} = \mathcal{F}^{-1}[F(\omega, t) \cdot \mathcal{F}(\Sigma_t)] \tag{2}$$

where $F(\omega, t) = \exp(-\|\omega\|^2/2\beta(t))$ with bandwidth $\beta(t) = \gamma^{2t}$.

Using Fourier transform properties, this simplifies to:

$$\Sigma_{t+1} = \gamma \cdot \Sigma_t \tag{3}$$

Expanding yields:

$$\Sigma_t = \gamma^t \Sigma_0 \odot M_d \tag{4}$$

where M_d is the depth gradient mask with $m_{ij} = \exp(-\|\nabla D_{ij}\|_2)$, and D is the DPT-estimated depth map. This preserves details in high-curvature regions while accelerating convergence in planar areas.

Figure 2 shows spectral analysis results. Early optimization ($t < 10^4$) focuses on low frequencies ($\|\omega\| < 0.2$) for coarse material estimation, while later stages ($t > 2 \times 10^4$) capture high-frequency surface details(e). Filtering accelerates error reduction by 20% and reduces final error by an order of magnitude (b). Spectral evolution (c) and material variance decay (d) confirm improved convergence efficiency.

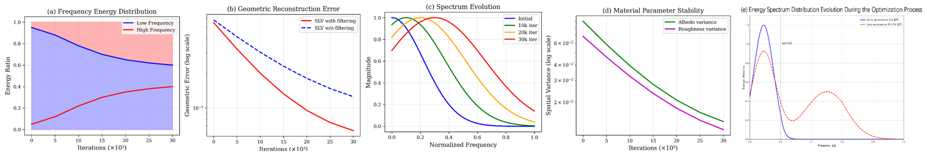


Fig. 2. Spectral Analysis of Progressive Frequency Domain Filtering with SLV Initialization in 3D Gaussian Splatting

3.2 Depth-Semantic Optimization

Depth-Based Geometric Optimization Current relighting methods often neglect local depth consistency, causing surface discontinuities that degrade shadow quality. We propose a geometric optimization strategy enforcing global and local depth constraints, directly coupling them with Gaussian parameter gradients. Unlike prior variational frameworks [27] for static reconstruction, our method targets dynamic shadow quality through Pearson correlation-based local consistency assessment.

We first generate a normalized depth map D_{gt} using the DPT model [23] and establish multi-scale constraints with the rendered Gaussian splatting depth D_{render}. The global depth loss preserves overall scene structure:

$$L_{\text{global}} = \|D_{\text{render}} - D_{\text{gt}}\|_1 \tag{5}$$

To capture high-frequency geometric details, we introduce a local depth correlation loss. We partition the depth map into $K \times K$ patches (with $K = 16$) and compute the Pearson correlation coefficient for each patch:

$$\rho_k = \frac{\text{Cov}(P_k^{\text{render}}, P_k^{\text{gt}})}{\sigma_{\text{render}} \sigma_{\text{gt}}} \tag{6}$$

where Cov denotes covariance and σ represents standard deviation. The local depth loss is then:

$$L_{\text{local}} = 1 - \frac{1}{K^2} \sum_{k=1}^{K^2} \rho_k \tag{7}$$

Through backpropagation, depth gradients flow into the Gaussian parameter space. For each position parameter μ_i, the gradient update becomes:

$$\frac{\partial L_{\text{depth}}}{\partial \mu_i} = \lambda_d \left(\frac{\partial L_{\text{global}}}{\partial \mu_i} + \alpha \frac{\partial L_{\text{local}}}{\partial \mu_i} \right) \odot M_d \tag{8}$$

where M_d serves as a depth gradient mask that balances global and local contributions. This mask amplifies gradient signals in high-curvature regions (e.g., object boundaries), encouraging Gaussian concentration near geometric details.

Rather than directly supervising with DPT depth, we employ it as a differentiable guide, with the local correlation loss enhancing geometric consistency across viewpoints.

Visual Feature-Based Cross-View Semantic Alignment Material decomposition suffers from albedo ambiguity due to view-dependent surface reflectance observationsthe same surface exhibits varying appearance under different viewing angles. To address this challenge, we exploit DINO-ViT's robust view-invariant representations [24]. By establishing material parameter constraints within the DINO-ViT feature space, we achieve semantic consistency across viewpoints.

Given a rendered image I_{render} and its corresponding ground truth I_{gt} in our multi-view setup, we extract global semantic features using DINO-ViT. Specifically, we feed both images through the model and extract the [CLS] token output from the final layer [24], yielding feature vectors $f_{\text{render}}, f_{\text{gt}} \in \mathbb{R}^{768}$. We then formulate a semantic similarity loss:

$$L_{\text{sem}} = 1 - \frac{f_{\text{render}} \cdot f_{\text{gt}}}{\|f_{\text{render}}\| \|f_{\text{gt}}\|} \tag{9}$$

This loss encourages consistent feature space representations across viewpoints, guiding the model toward view-invariant material parameters.

For enhanced local material consistency, we implement a spatial attention-guided pooling strategy. From ViT's final attention map $A \in \mathbb{R}^{H \times W}$, we identify the top-k salient regions $\{R_i\}_{i=1}^{k}$ and compute their feature contrast loss:

$$L_{\text{patch}} = \sum_{i=1}^{k} \left(1 - \cos\left(g(R_i^{\text{render}}), g(R_i^{\text{gt}})\right)\right) \tag{10}$$

where $g(\cdot)$ performs region-wise average pooling and cos computes cosine similarity.

Semantic gradients propagate to the albedo parameter space via differentiable rendering:

$$\frac{\partial L_{\text{sem}}}{\partial \rho_i} = \lambda_s \left(\frac{\partial L_{\text{sem}}}{\partial \rho_i} + \beta \frac{\partial L_{\text{patch}}}{\partial \rho_i}\right) \odot S_{\text{mask}} \tag{11}$$

where $S_{\text{mask}} = \sigma(\|\nabla_x A\|)$ represents an attention gradient-based spatial mask, σ is the sigmoid function, and $\beta = 0.5$ balances global and local contributions. This mask prioritizes updates in semantically salient regions such as material boundaries.

We fuse the semantic mask S_{mask} with the depth mask M_d using a gating mechanism:

$$G = \sigma(W_g[S_{\text{mask}} \| M_d]) \tag{12}$$

where W_g is a learnable weight matrix and $\|$ denotes concatenation. The final gradient update combines both constraints:

$$\frac{\partial L_{\text{total}}}{\partial \theta} = G \odot \frac{\partial L_{\text{sem}}}{\partial \theta} + (1 - G) \odot \frac{\partial L_{\text{depth}}}{\partial \theta} \tag{13}$$

This adaptive fusion allows material and geometric constraints to be weighted according to local feature importance.

3.3 Differentiable Light Field Rendering

To enable photorealistic dynamic relighting, we develop DSSG's differentiable light field rendering framework, inspired by R3DG [15]. Our approach decouples material properties from light transport, enabling efficient real-time ray tracing.

BRDF Parameter Decomposition Building on R3DG's modified Disney BRDF model, we establish a direct coupling between roughness parameters and Gaussian geometric properties. Each Gaussian maintains material parameters $\{\rho_i, \alpha_i, m_i\}$, with roughness α_i derived from covariance eigenvalues:

$$\alpha_i = \sqrt{\frac{\lambda_{\max}(\Sigma_i)}{\lambda_{\min}(\Sigma_i)}} \cdot \gamma(t) \tag{14}$$

where $\gamma(t) = \exp(-0.01t/T)$ introduces temporal decay, complementing the frequency-domain filtering from Section 2.1.2. This formulation exploits the natural relationship between the eigenvalue ratio $\lambda_{\max}/\lambda_{\min}$ and surface anisotropy, eliminating additional parameters while improving spatial coherence.

We reformulate the specular term as:

$$f_s = \frac{D(h)F(v,h)G}{4(n \cdot l)(n \cdot v)} \cdot \mathrm{softplus}(m_i) \tag{15}$$

where $\mathrm{softplus}(x) = \ln(1+e^x)$ ensures non-negative metalness values, addressing gradient instabilities in the original formulation. Within our deferred rendering pipeline, we define surface normals as the eigenvector associated with the minimum eigenvalue of each Gaussian's covariance matrix.

Light Transport Modeling We decompose incident illumination into direct environment lighting L_{env} and indirect local lighting L_{ind}:

$$L_i(\omega_i) = V(\omega_i) L_{\text{env}}(\omega_i) + L_{\text{ind}}(\omega_i) \tag{16}$$

The visibility term $V(\omega_i)$ is computed via ray tracing, while indirect illumination employs third-order spherical harmonics. Using Fibonacci sampling to generate $N_s = 64$ incident directions, we evaluate outgoing radiance through Monte Carlo integration:

$$L_o = \sum_{k=1}^{N_s} f_r(\omega_k) L_i(\omega_k)(n \cdot \omega_k) \Delta\omega_k \tag{17}$$

3.4 Loss Functions

Our optimization employs a two-stage strategy: geometry reconstruction followed by material decomposition. We design a comprehensive loss framework that combines reconstruction objectives, regularization terms, and stage-specific constraints, modulated by a dynamic weighting schedule (detailed in Section 3.5).

Reconstruction and Regularization Losses The core reconstruction losses ensure fidelity between rendered and reference images across both optimization stages:

$$L_{\mathrm{rgb}} = \|I_{\mathrm{render}} - I_{\mathrm{gt}}\|_1, \quad L_{\mathrm{ssim}} = 1 - \mathrm{SSIM}(I_{\mathrm{render}}, I_{\mathrm{gt}}) \tag{18}$$

where L_{rgb} captures pixel-wise differences and L_{ssim} evaluates structural similarity, maintaining consistency at both pixel and perceptual levels.

To mitigate material-lighting ambiguities and ensure physical plausibility, we incorporate regularization terms:

$$\mathcal{L}_{\mathrm{light}} = \sum_{c \in \{R,G,B\}} \left(L^c_{\mathrm{env}} - \frac{1}{3} \sum L^c_{\mathrm{env}} \right)^2 \tag{19}$$

$$\mathcal{L}_{\mathrm{smooth}} = \|\nabla \rho\| e^{-\|\nabla D\|} + \|\nabla \alpha\| e^{-\|\nabla C\|} \tag{20}$$

where $\mathcal{L}_{\mathrm{light}}$ encourages achromatic environment lighting to reduce color bleeding, and $\mathcal{L}_{\mathrm{smooth}}$ adaptively weights material gradients based on depth and color discontinuities from our depth-semantic optimization (Section 2.2).

Stage-Specific Constraints For geometry reconstruction, we enforce depth consistency and shape regularity:

$$L_{\mathrm{depth}} = \|D_{\mathrm{render}} - D_{\mathrm{gt}}\|_1 + \left(1 - \frac{1}{K^2} \sum_{k=1}^{K^2} \rho_k \right) \tag{21}$$

$$L_{\mathrm{var}} = \sum_{i=1}^{N} \left\| \log \left(\frac{\lambda_{\max}(\Sigma_i)}{\lambda_{\min}(\Sigma_i)} \right) \right\|_2 \tag{22}$$

where L_{depth} combines global depth fidelity with local correlation (via Pearson coefficients ρ_k), guided by a time-varying weight $\lambda_d(t)$ that decreases during optimization. L_{var} regularizes Gaussian covariance matrices to prevent degenerate shapes.

During material decomposition, we apply semantic consistency and shadow fidelity constraints:

$$L_{\mathrm{semantic}} = \underbrace{1 - \frac{f_{\mathrm{render}} \cdot f_{\mathrm{gt}}}{\|f_{\mathrm{render}}\| \|f_{\mathrm{gt}}\|}}_{L_{\mathrm{sem}}} + \beta \underbrace{\sum_{i=1}^{k} \left(1 - \cos\left(g(R_i^{\mathrm{render}}), g(R_i^{\mathrm{gt}}) \right) \right)}_{L_{\mathrm{patch}}} \tag{23}$$

$$L_{\mathrm{phys}} = \|\nabla m\|_2^2 + \|\alpha \odot m\|_1, \quad L_{\mathrm{shadow}} = \|M^{\mathrm{render}}_{\mathrm{shadow}} - M^{\mathrm{gt}}_{\mathrm{shadow}}\|_1 \tag{24}$$

where L_{semantic} enforces view consistency in DINO-ViT feature space through global (L_{sem}) and local (L_{patch}) similarity. L_{phys} promotes spatial smoothness and energy conservation for BRDF parameters m. L_{shadow} ensures accurate light occlusion modeling through shadow map comparison.

Optimization Objectives We define separate objectives for each stage that combine the above components:

Geometry reconstruction:

$$L_{\text{geo}} = \lambda_1 L_{\text{rgb}} + \lambda_{\text{ssim}} L_{\text{ssim}} + \lambda_d L_{\text{depth}} + \lambda_{\text{var}} L_{\text{var}} \tag{25}$$

Material decomposition:

$$L_{\text{brdf}} = \lambda_1 L_{\text{rgb}} + \lambda_{\text{ssim}} L_{\text{ssim}} + \lambda_s L_{\text{semantic}} + \lambda_{\text{shadow}} L_{\text{shadow}} + \lambda_{\text{light}} \mathcal{L}_{\text{light}} + \lambda_{\text{smooth}} \mathcal{L}_{\text{smooth}} \tag{26}$$

The weighting coefficients λ are dynamically scheduled to smoothly transition from geometric to material optimization, enabling progressive refinement from coarse structure to fine-grained material properties.

3.5 Geometry-Material Alternating Optimization

Relightable reconstruction faces inherent conflicts: geometric reconstruction requires high point density while material decomposition demands spatial smoothness, with material gradients typically exceeding geometric gradients by an order of magnitude. We propose alternating optimization with dynamic constraint weighting to decouple these conflicting objectives.

We alternate between geometry reconstruction ($t \bmod T < T_{geo}$) and material decomposition phases within period T. Dynamic weighting ensures smooth transitions:

$$\lambda_d(t) = \lambda_d^0 \exp(-\alpha t), \quad \lambda_s(t) = \lambda_s^0 [1 - \exp(-\beta(t - t_{delay}))] \tag{27}$$

where exponentially decaying depth weights provide strong initial geometric guidance, and delayed semantic constraints (activating after t_{delay}) prevent premature material regularization.

Geometry phases optimize positions μ_i and covariances Σ_i, while material phases refine albedo ρ_i, roughness α_i, and metalness m_i. We progressively increase physical regularization weights and employ complexity-adaptive ray sampling to maintain computational efficiency throughout optimization.

4 Experiments

4.1 Experimental Setup

We evaluate on NeRFSynthetic and Synthetic4Relight datasets with parameters: $\{\lambda_d = 0.01, \lambda_s = 0.0001, \gamma = 0.995\}$ from Sections 2.1-2.2.

Geometry optimization: 30,000 iterations with $T_n = 2 \times 10^{-9}$. Loss weights (Eq. 25): $\{\lambda_1 = 0.8, \lambda_{ssim} = 0.2, \lambda_d = 0.01, \lambda_{var} = 0.01\}$.BRDF optimization: 10,000 iterations, 64 rays/Gaussian. Loss weights (Eq. 26): $\{\lambda_1 = 0.8, \lambda_{ssim} = 0.2, \lambda_s = 0.0001, \lambda_{shadow} = 0.01\}$.

Adam optimizer with cosine scheduling, 10% warm-up. Dynamic weighting: $\lambda_d(t) = 0.01e^{-5t/40000}$ (exponential decay), $\lambda_s(t) = 0.0001[1 - e^{-3(t-15000)/20000}]$ (sigmoid growth). Learning rate: $10^{-3} \to 10^{-4}$.

4.2 Performance on View Synthesis

Table 1 presents view synthesis evaluation on the NeRF Synthetic dataset. DSSG achieves comparable rendering quality to view synthesis-specific methods while enabling real-time relighting.

Table 1. Quantitative comparison of view synthesis on NeRFSynthetic dataset

Non-relightable methods					Relightable methods				
Method	**Geom.**	**PSNR↑**	**SSIM↑**	**LPIPS↓**	**Method**	**Geom.**	**PSNR↑**	**SSIM↑**	**LPIPS↓**
NPBG	point	28.10	0.923	0.077	PhysSG	neural	18.91	0.847	0.182
NPBG++	point	28.12	0.928	0.076	NeLF++	neural	26.37	0.911	0.091
FreqPCR	point	31.24	0.950	0.049	Nvdiffrec	mesh	29.05	0.939	0.081
3DGS	point	**33.88**	**0.970**	**0.031**	R3DG	point	31.22	0.959	**0.039**
					DSSG (Ours)	point	**31.67**	**0.961**	0.040

Among relightable methods, DSSG achieves the highest PSNR and SSIM , demonstrating the effectiveness of our depth-semantic constraints. Compared to non-relightable 3DGS [2], DSSG shows only marginal performance gaps while enabling real-time relightinga capability 3DGS lacks entirely.

Figure 3 (left) shows qualitative comparisons. DSSG captures finer details in the chair's velvet texture and microphone's metallic mesh (red box), with better-preserved wire gaps and specular reflections than R3DG.

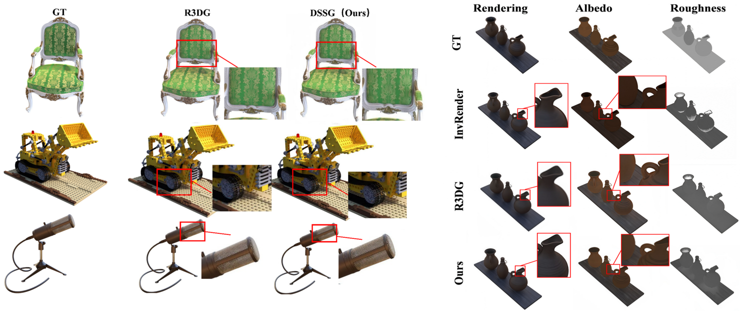


Fig. 3. Qualitative comparison of view synthesis on NeRFSynthetic dataset (left). Visual comparison of material decomposition on Synthetic4Relight dataset (right).

4.3 Performance on Relighting

Table 2 presents relighting results on Synthetic4Relight. DSSG achieves 37.05 dB PSNR (view synthesis) and 31.42 dB (relighting), with superior albedo reconstruction and competitive roughness estimation. The depth-semantic dual constraints and SLV initialization improve point distribution in shadowed regions, enhancing both quantitative metrics and perceptual quality (Figure 3)

Table 2. Quantitative evaluation on Synthetic4Relight dataset

Method	View Synthesis			Relighting			Albedo			Roughness
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓
NerFactor	22.80	0.916	0.150	21.54	0.875	0.171	19.49	0.864	0.206	—
Nvdiffrec-MC	34.29	0.967	0.068	24.22	0.943	0.078	29.61	0.945	0.075	0.009
InvRender	30.74	0.953	0.086	28.67	0.950	0.091	28.28	0.935	0.072	**0.008**
TensorIR	35.80	0.978	0.049	29.69	0.951	0.079	**30.58**	0.946	0.065	0.015
R3DG	36.80	0.982	**0.028**	31.00	0.964	0.050	28.31	0.951	0.058	0.013
DSSG (Ours)	**37.05**	**0.983**	0.030	**31.42**	**0.968**	**0.048**	30.15	**0.955**	**0.054**	0.010

Figure 4 demonstrates DSSG's robustness: (top) accurate rendering under five environmental illuminations; (bottom) real-world applicability with Lego model in custom panoramic environments.

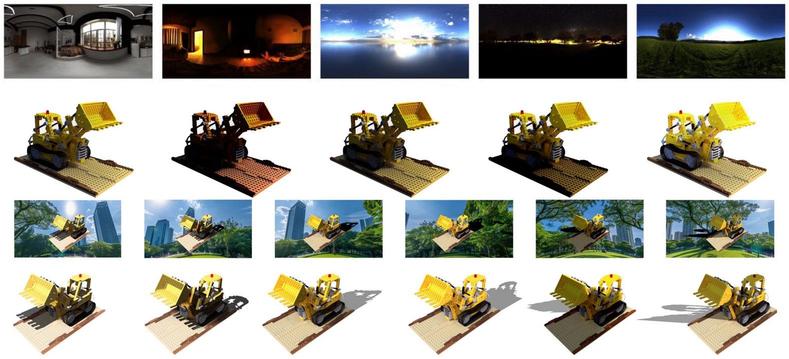


Fig. 4. DSSG relighting results under diverse environmental illumination conditions(top)Scene composition and dynamic relighting results produced by DSSG(bottom)

Additional TNT dataset evaluation (Figure 5) confirms DSSG's superior geometric accuracy, validating its effectiveness on real-world data.

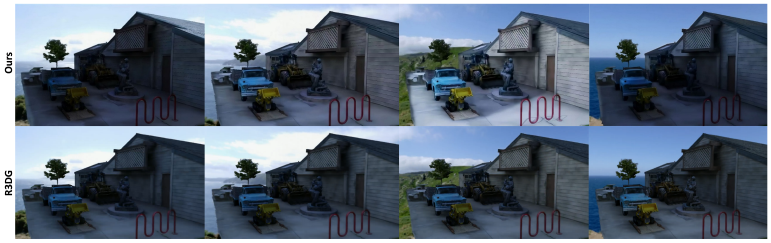


Fig. 5. Comparison of relighting performance on TNT dataset

4.4 Ablation Study

Table 3 validates our design choices. Semantic constraints are crucial for material decomposition, depth constraints improve geometric quality, and SLV initialization enables superior light transport modeling with faster convergence compared to uniform distributions.

Table 3. Ablation study with light transport analysis

Method Variant	View Synthesis			Relighting			Material Decomposition			Geometry Quality		**Light Transport**			**Conv.**
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	Albedo PSNR	Albedo SSIM	Rough. MSE	Normal Cons.	Normal MAE	**Indirect MAE**	**Specular MAE**	**Shadow Acc.**	**Iter.**
DSSG (Full)	**37.05**	**0.983**	**0.030**	**31.42**	**0.968**	**0.048**	**30.15**	**0.955**	**0.010**	**0.912**	**0.062**	**0.086**	**0.102**	**0.891**	**18k**
Core Component Ablation															
w/o Semantic	36.84	0.980	0.035	30.87	0.958	0.056	28.91	0.934	0.014	0.908	0.068	0.095	0.118	0.872	22k
w/o Depth	36.51	0.977	0.038	30.45	0.961	0.053	29.38	0.946	0.012	0.867	0.085	0.104	0.125	0.856	25k
w/o SLV	36.92	0.982	0.032	30.98	0.965	0.050	29.87	0.952	0.011	0.873	0.079	0.112	0.134	0.845	24k
w/o Prog. Filter	36.88	0.981	0.033	31.15	0.966	0.051	29.95	0.953	0.012	0.896	0.071	0.092	0.108	0.878	20k
Initialization Parameter Study															
Small Var. (0.01)	36.42	0.978	0.037	30.23	0.957	0.058	29.12	0.943	0.015	0.876	0.095	0.142	0.168	0.712	28k
Medium Var. (0.05)	36.78	0.981	0.033	30.95	0.963	0.051	29.86	0.951	0.012	0.895	0.078	0.108	0.134	0.823	22k
Large Var. (0.1)	**37.05**	**0.983**	**0.030**	**31.42**	**0.968**	**0.048**	**30.15**	**0.955**	**0.010**	**0.912**	**0.062**	**0.086**	**0.102**	**0.891**	**18k**
Spatial Distribution Strategy															
Random Uniform	36.03	0.974	0.042	29.15	0.945	0.069	27.62	0.921	0.018	0.841	0.115	0.156	0.182	0.698	35k
Grid Uniform	36.48	0.979	0.036	30.34	0.959	0.055	29.23	0.945	0.013	0.882	0.088	0.124	0.145	0.812	26k
SLV (Ours)	**37.05**	**0.983**	**0.030**	**31.42**	**0.968**	**0.048**	**30.15**	**0.955**	**0.010**	**0.912**	**0.062**	**0.086**	**0.102**	**0.891**	**18k**

Figure 6 visualizes component contributions. Without semantic constraints (c), material quality deteriorates. Depth constraints reduce geometric uncertainty at object boundaries (d-f). SLV captures finer surface details with smoother transitions (g-h).

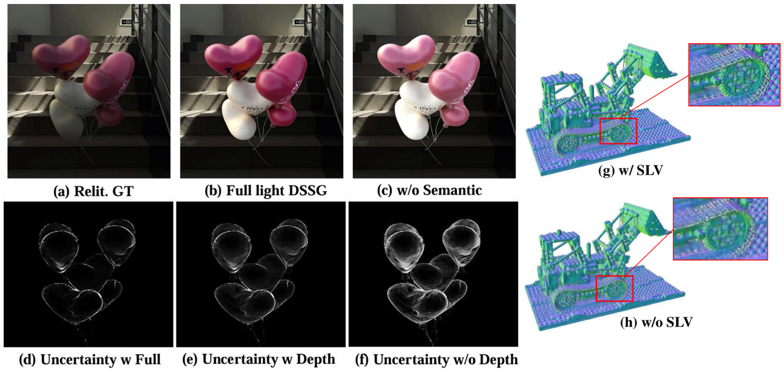


Fig. 6. Visual ablation study of DSSG components

Figure 7 shows SLV achieves 31 dB PSNR in 18k iterations28% faster than SfM initialization. The method maintains smooth convergence with 0.91 normal consistency and rapid variance reduction to 0.01.

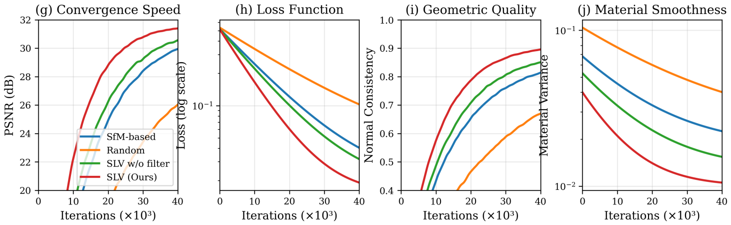


Fig. 7. Convergence analysis of different initialization strategies

5 Conclusion

We presented a depth-semantic-driven sparse light field-guided 3D Gaussian modeling method for high-quality relightable scene reconstruction. By extending traditional 3D Gaussian Splatting to a relightable representation, our SLV initialization strategy eliminates dependency on SfM point clouds while ensuring uniform spatial coverage. To address material decomposition ambiguities, we developed a dual-constraint mechanism that fuses DPT depth estimation with DINO-ViT semantic features, enforcing cross-view material consistency. Experimental results demonstrate that our method achieves accurate material decomposition and superior performance in both view synthesis and scene relighting tasks. The effectiveness of our approach is validated through improved material quality and compelling visual results under diverse environmental lighting conditions.

References

1. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. In: ECCV, LNCS, vol. 12346, pp. 405–421. Springer (2020)
2. Kerbl, B., Kopanas, G., Leimkühler, T.: 3D Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4), 1–14 (2023)
3. Azinovi, D., Li, T.M., Kaplan, A., Niessner, M., Thies, J., Dai, A.: Neural RGB-D surface reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6290–6301 (2022)
4. Liang, Y., Zheng, M., Yu, F., Wetzstein, G., Yuan, J., Liu, X.: GaussianShader: 3D Gaussian splatting with shading functions for reflectance decomposition. arXiv preprint arXiv:2311.17977 (2023)
5. Zhu, T., Zhang, Q., Chen, L., Wang, T., Zhao, J., Ai, Z., Zeng, Y., Sun, Q.: PhysGaussian: Physics-based neural gaussian splatting for joint appearance, geometry and material optimization. arXiv preprint arXiv:2311.16773 (2023)
6. Fan, G., Xu, X., Zhang, J., Chen, H., Lu, Y., Yang, Z., Sun, M., Zeng, H., Wang, Y., Liu, S., Lieng, H., Liu, S., Yuan, C.: Neuralangelo+: Efficiently modeling long videos of dynamic scenes with 3D Gaussian splatting. arXiv preprint arXiv:2312.09069 (2023)

7. Wu, S., Basu, S., Broedermann, T., Van Gool, L.: PBR-NeRF: Inverse rendering with physics-based neural fields. arXiv preprint arXiv:2412.09680 (2024)
8. Ye, M., Zhang, J., Yang, L., Fan, K.J., Xu, J., Han, L., Xiang, L., Chen, X., Cui, S., Wang, X.: Gaussian grouping: Segment and edit anything in 3D scenes. arXiv preprint arXiv:2312.00732 (2023)
9. Jung, J., Han, J., An, H., Kang, J., Park, S., Kim, S.: Relaxing accurate initialization constraint for 3D Gaussian splatting. arXiv preprint arXiv:2403.09413 (2024)
10. Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T.: Modeling indirect illumination for inverse rendering. In: CVPR, pp. 18643–18652 (2021)
11. Bi, S., Xu, Z., Srinivasan, P., Mildenhall, B., Sunkavalli, K., Haan, M., Hold-Geoffroy, Y., Kriegman, D., Ramamoorthi, R.: NeRfactor: Neural factorization of shape and reflectance from an image. In: ICCV, pp. 12682–12691 (2021)
12. Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T., Srinivasan, P.P.: Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In: CVPR, pp. 5481–5490 (2022)
13. Verbin, D., Hedman, P., Mildenhall, B.: NeRD: Neural reflectance decomposition from image collections. In: ICCV, pp. 12684–12694 (2022)
14. Yu, A., Li, R., Tancik, M., Li, H., Ng, R.: Inverse rendering for complex indoor scenes: Shape, materials, and lighting. In: CVPR, pp. 2475–2484 (2022)
15. Gao, J., Gu, C., Lin, Y., Zhu, H., Cao, X., Zhang, L., Yao, Y.: Relightable 3D Gaussian: Real-time point cloud relighting with BRDF decomposition and ray tracing. arXiv preprint arXiv:2311.16043 (2023)
16. Yang, Z., Huang, X., Chen, G., Liu, L., Bao, H.: PhysGauss: Physically-aware Gaussian splatting for dynamic relighting. arXiv preprint arXiv:2403.10242 (2024)
17. Liu, Y., Li, Z., Li, T., Liu, J., Jiang, Y., Wang, H.: GauFRe: Gaussian frequency regularization for relightable 3DGS. arXiv preprint arXiv:2402.01789 (2024)
18. Tang, J., Ren, J., Zhou, H., Liu, Z., Zeng, G.: GIR: 3D Gaussian inverse rendering for relightable scene factorization. In: CVPR, pp. 12947–12956 (2024)
19. Liu, Y., Li, C., Li, K., Fu, Q., Heng, P.A.: GS-IR: Incident radiance fields for 3D Gaussian splatting. arXiv preprint arXiv:2402.15418 (2024)
20. Zhang, X., Ng, R., Chen, Q.: Polarized reflection removal with perfect alignment in the wild. In: CVPR, pp. 1750–1758 (2021)
21. Bi, S., Xu, Z., Sunkavalli, K., Haan, M., Hold-Geoffroy, Y.: Neural reflectance fields for appearance acquisition. In: CVPR, pp. 12116–12125 (2020)
22. Zhang, X., Fanello, S., Tsai, Y.T., Sun, T., Xue, T., Pandey, R., et al.: Neural-PIL: Neural pre-integrated lighting for reflectance decomposition. In: NeurIPS, pp. 10691–10702 (2021)
23. Ranftl, R., Bochkovskiy, A., Koltun, V.: Vision transformers for dense prediction. In: ICCV, pp. 12159–12168 (2021)
24. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: ICCV, pp. 9650–9660 (2021)
25. Liu, Y., Wang, Z., Zhang, S.H.: GaussianGroup: 3D Gaussian splatting with advanced semantic and geometric grouping. arXiv preprint arXiv:2402.12335 (2024)
26. Deng, K., Liu, A., Zhu, J.Y., Ramanan, D.: Depth-supervised NeRF: Fewer views and better quality. In: International Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
27. Yao, Y., Zhang, Z., Zhang, L., Luo, Z., Liu, Z., Shen, Z., Ramamoorthi, R., Huang, J.: SeeDetail: High-fidelity surface details reconstruction from multi-view images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6560–6570 (2022)