Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# GARD: Geometry-Anchored Reflectance Decoupling for Joint Reconstruction of Highly Reflective Surfaces

**Shaofeng Guo**
Northeastern University
**Lijin Fang**
ljfang@mail.neu.edu.cn
Northeastern University
**Qiankun Zhao**
Northeastern University
**Zhehao Lin**
Northeastern University

**Research Article**

**Keywords:**

**Posted Date:** November 25th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11588626/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# GARD: Geometry-Anchored Reflectance Decoupling for Joint Reconstruction of Highly Reflective Surfaces

Shaofeng Guo[1], Lijin Fang[1*], Qiankun Zhao[1], Zhehao Lin[1]

[1]Faculty of Robot Science and Engineering, Northeastern University, Shenyang, 110819, China.

*Corresponding author(s). E-mail(s): ljfang@mail.neu.edu.cn;

**Abstract**

Reconstructing highly reflective surfaces with hybrid Gaussian and signed distance field (SDF) representations remains difficult because view-dependent highlights violate multi-view photometric consistency and can be baked into static appearance or geometry. We present Geometry-Anchored Reflectance Decoupling (GARD), a framework that jointly reconstructs surface structure and view-dependent appearance while explicitly addressing this geometry–appearance entanglement. GARD organizes appearance into a diffuse-oriented base component and an additive reflective residual under a continuous SDF-guided geometry scaffold. Physically based rendering (PBR) material cues provide an appearance prior, while roughness-aware gating and density control regulate the residual contribution and its spatial capacity. In addition, stop-gradient on the shared Gaussian feature input reduces feature-mediated feedback from the residual branch without suppressing directional cues from surface normals and reflected viewing directions. Experiments on challenging real-world reflective scenes show that GARD preserves competitive rendering quality while producing a more controlled base/residual separation and reducing reflection-induced normal artifacts in representative regions. These results demonstrate a practical approach to mitigating geometry–appearance entanglement without assuming uniquely recoverable physical materials or explicitly modeling global light transport.

**Keywords:** 3D Gaussian Splatting, Surface Reconstruction, Reflective Surfaces, Reflectance Decoupling, Neural Rendering

## Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62273081.

## Author Contributions

Shaofeng Guo: Methodology, software, analysis, and original draft. Lijin Fang: Supervision, funding, and review and editing. Qiankun Zhao: Validation and visualization. Zhehao Lin: Data curation and validation.

## 1 Introduction

Recent advancements in hybrid neural-explicit representations, particularly the integration of 3D Gaussian Splatting with Signed Distance Fields, such as GSDF, have set new benchmarks in novel view synthesis and geometry reconstruction [1–4]. By jointly optimizing the highly efficient rasterization of explicit Gaussian primitives and the topological guarantees of implicit surfaces, these hybrid methods achieve an impressive balance

between rendering fidelity and geometric accuracy. However, this joint optimization paradigm remains notoriously fragile when applied to highly reflective surfaces, revealing a critical bottleneck in current inverse rendering pipelines.

The core difficulty stems from a fundamental geometry-appearance entanglement. Most 3D reconstruction frameworks rely on multi-view photometric consistency, which assumes that a surface point yields stable color observations across varying viewpoints. Strong specular reflections severely violate this assumption [5–7]. Because highlights shift dynamically with the viewing angle, they present inconsistent multi-view color cues to the optimization process. Standard photometric losses consequently force baseline methods to improperly bake these view-dependent highlights into static geometry or view-independent diffuse appearance. As illustrated in Figure 1, this specular baking inevitably leads to characteristic failure modes: noisy and distorted surface normals, floating artifacts masking the true geometry, and diffuse albedo maps that are heavily contaminated by highlight artifacts.

Addressing this entanglement is not as trivial as simply increasing the representational capacity of the appearance model. A naive expansion of view-dependent modeling, such as increasing spherical harmonic degrees or enlarging MLP capacity, often backfires [1, 8–10]. The central challenge is that highly reflective surfaces violate multi-view photometric consistency because specular highlights move with viewpoint. Without proper constraints, these inconsistent photometric residuals may be absorbed into static base appearance or even into geometry, producing geometry–appearance entanglement, specular baking, and unstable normals. A weak view-dependent model fails to capture sharp specularities, forcing moving highlights to leak into the diffuse-oriented base appearance or to deform local geometry. Conversely, an overly expressive specular branch may become an unconstrained loss sink that overfits residual photometric errors and weakens geometry supervision. Therefore, a robust reflective reconstruction framework must not only model specular effects, but also regulate where, when, and how much view-dependent information participates in joint optimization.

To address these limitations, we propose Geometry-Anchored Reflectance Decoupling, termed GARD, which introduces a unified reconstruction framework tailored for highly reflective surfaces. The overall pipeline is overviewed in Figure 1. GARD uses an appearance formulation that encourages a diffuse-oriented base component and a separate, additive view-dependent residual component. Rather than attempting to recover unique physical materials under unknown lighting, GARD uses PBR-inspired regularization to provide a controlled representation of appearance [11–14].

Crucially, to limit unchecked residual fitting, GARD introduces roughness-related control of the residual gate and Gaussian growth/pruning criteria. This rule favors residual fitting in regions with lower predicted roughness-like values and suppresses it in other regions, but does not by itself prove that more Gaussian primitives are retained in glossy regions. The control is therefore interpreted as an optimization prior rather than a guarantee of physical material allocation or complete prevention of specular baking.

Furthermore, to evaluate practical behavior, we focus on complex real-world scenes rather than idealized synthetic datasets. Synthetic datasets may hide difficulties caused by uncontrolled illumination and material heterogeneity. Our evaluation across 15 scenes from NeRO and Shiny uses fixed hyperparameters without per-scene tuning, providing evidence of cross-scene robustness under the specified protocol rather than feed-forward generalization to unseen scenes. GARD maintains rendering quality comparable to leading hybrid baselines and yields explicitly separated base/residual representations together with qualitative evidence of smoother normal maps in selected reflective regions. These observations support a controlled representation, but do not establish unique physical decomposition or universal geometric improvement.

In summary, our main contributions are:

- We identify and analyze geometry-appearance entanglement as a primary failure mode in Gaussian-SDF reconstructions, explicitly detailing how specular baking compromises both normal estimation and appearance decomposition.

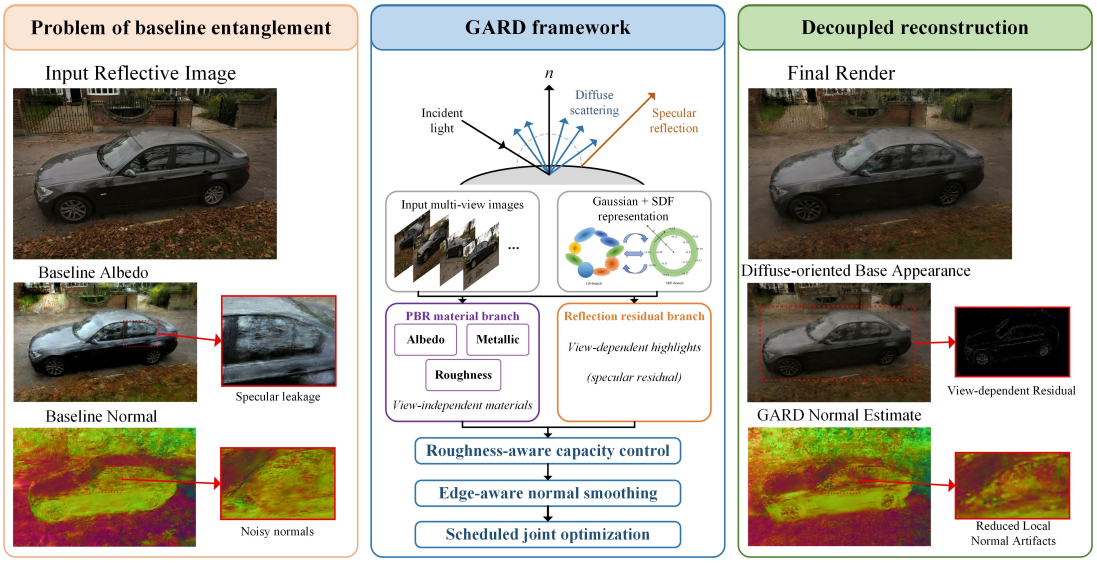



**Fig. 1** ***Motivation and overview of the GARD framework.*** Standard Gaussian-SDF reconstructions may absorb view-dependent highlights into static appearance and generate noisy surface normals in highly reflective regions. GARD encourages a separation between a diffuse-oriented base representation and an additive, view-dependent residual through joint optimization. The resulting base and residual assets are explicitly separated, but are not claimed to be unique physical diffuse/specular layers

- We propose GARD, a unified framework that explicitly separates diffuse and specular appearance features during joint geometry-appearance optimization, guided by physically-inspired material priors.
- We introduce roughness-aware dynamic capacity control, a novel regularization strategy that spatially modulates the specular branch to prevent uncontrolled residual overfitting and preserve geometric integrity.
- We demonstrate cross-scene robustness across 15 challenging scenes from two real-world reflective datasets without per-scene tuning, yielding explicitly separated base/residual representations and qualitative evidence of reduced reflection-induced normal artifacts.

# 2 Related Work

## 2.1 Neural Surface Reconstruction and Hybrid Representations

Neural surface reconstruction has significantly advanced by combining differentiable rendering with geometric priors [2–4, 15, 16]. Pure implicit methods enforce continuous surfaces and regularized normals, while recent hybrid architectures integrate these geometric priors with the efficient rasterization of 3D Gaussian Splatting. More recent inverse-rendering systems also introduce explicit geometry guidance into Gaussian representations. GeoSplatting grounds Gaussians on an optimizable mesh to obtain well-defined normals and occlusion-aware light transport [17], whereas Gaussian Splatting with Discretized SDF encodes SDF samples directly on Gaussian primitives to improve geometry–material decomposition without maintaining an additional continuous SDF network [18]. These methods primarily target physically based inverse rendering and relighting. In contrast, GARD retains a continuous Gaussian–SDF scaffold and focuses on reducing feature-mediated coupling between a learned view-dependent residual and the shared reconstruction representation during joint optimization.

Despite their geometric regularization, Gaussian–SDF systems still rely heavily on multi-view photometric consistency. This assumption degrades on highly reflective surfaces, where shifting highlights provide inconsistent color cues. Photometric errors may consequently be absorbed into spatial distributions, surface normals, or static appearance, producing specular baking and local geometric distortion. GARD addresses this failure mode through a feature-detached residual

pathway and roughness-conditioned capacity control rather than treating the SDF prior alone as sufficient to resolve reflectance ambiguity.

Here, the SDF should be understood as providing a continuous implicit-surface representation and geometry-related regularization; we do not assume that it automatically guarantees correct topology.

## 2.2 View-Dependent Appearance Modeling

Modeling non-Lambertian effects has traditionally relied on increasing the expressiveness of view-dependent rendering branches, such as employing high-degree spherical harmonics or expanding multi-layer perceptron capacities [1, 8–10]. Within Gaussian Splatting, SpecGaussian replaces standard spherical harmonics with anisotropic spherical Gaussians to represent high-frequency, anisotropic view dependence [19]. GaussianShader instead combines a simplified reflective shading function, learned normals, and a residual color term to improve rendering of reflective surfaces [20]. Both approaches enhance the representation of specular appearance, but their primary objective is accurate view-dependent rendering. GARD differs in that its residual pathway is explicitly separated from the diffuse-oriented base representation, applies stop-gradient to the shared Gaussian feature input, and is spatially regulated by a roughness-conditioned gate and density-control rule.

Methods that model non-local reflections provide another complementary direction. SpecTRe-GS traces secondary rays through a Gaussian scene to model reflections of nearby objects on highly specular surfaces [21], while IRGS introduces 2D Gaussian ray tracing for inter-reflective inverse rendering [22]. These ray-tracing formulations can represent light-transport effects that a local residual predictor cannot explicitly recover, but they target a different accuracy–complexity trade-off. GARD does not attempt explicit secondary-ray transport; it uses a lightweight local residual model to reduce the leakage of view-dependent transients into a Gaussian–SDF geometry scaffold. PolGS further demonstrates that polarimetric observations can separate diffuse and specular components and constrain reflective-surface geometry [23]. Unlike PolGS, GARD assumes only calibrated RGB images and requires no specialized polarization acquisition.

Although expressive view-dependent models improve image fitting, they may also become unconstrained error absorbers. A weak model fails to capture sharp specularities, forcing highlights to leak into the base color or geometry, whereas an overly expressive model may fit arbitrary photometric residuals and weaken geometry supervision. GARD therefore regulates not only how specular effects are represented, but also where the residual branch is active and which geometry-related variables it can update.

## 2.3 Inverse Rendering and Material Estimation

Inverse rendering aims to explain complex appearance by estimating environmental illumination, surface normals, and material properties such as albedo and roughness [11–14, 24, 25]. GS-IR transfers this objective to 3D Gaussian Splatting through staged normal reconstruction, occlusion modeling, and physically based material–lighting optimization [26]. GeoSplatting strengthens light-transport estimation through mesh-grounded geometry [17], while the discretized-SDF formulation couples Gaussian opacity and an embedded SDF prior for efficient relightable-asset reconstruction [18]. These approaches pursue material–lighting decomposition and relighting as primary outputs.

Recovering absolute physical materials under unknown, unconstrained illumination nevertheless remains ill posed. GARD consequently uses its predicted material variables and environment map as an appearance-organization prior rather than claiming unique BRDF or illumination recovery. Relative to GS-IR and the geometry-guided relighting methods above, the distinguishing objective of GARD is to isolate additive, high-frequency view-dependent residuals from a diffuse-oriented base while limiting their gradient and spatial capacity within a continuous Gaussian–SDF reconstruction. The resulting components are therefore interpreted as controlled appearance representations, not as uniquely identified physical diffuse and specular layers.

### 2.4 Regularization and Capacity Control

To mitigate ambiguities in inverse rendering and neural reconstruction, existing methods use normal smoothing, depth or surface priors, density regularization, and staged optimization [15, 17, 18, 26, 27]. Such constraints improve geometry or material decomposition, while reflective-appearance methods commonly increase directional capacity or explicitly trace secondary rays [19–22]. GARD addresses a complementary question: how to prevent a local residual branch from spending unrestricted capacity over the entire scene. Its roughness-conditioned score jointly gates the residual contribution and modulates Gaussian growth/pruning criteria, while stop-gradient blocks the feature-mediated gradient path from the residual branch into the shared Gaussian feature. The normal and reflected-direction inputs remain differentiable, so this operation reduces one source of coupling rather than fully isolating residual optimization from geometry. This design does not replace explicit global light transport, but provides a lightweight spatial and gradient constraint for ordinary RGB reconstruction under fixed hyperparameters.

## 3 Method

We propose Geometry-Anchored Reflectance Decoupling (GARD), a unified reconstruction framework tailored for highly reflective real-world surfaces. Given a set of multi-view images with known camera poses, our objective is to jointly reconstruct stable surface geometry and interpretable appearance while preventing view-dependent specular highlights from being erroneously baked into static geometry or diffuse albedo. The overview of our proposed framework is illustrated in Figure 2. Our design leverages a dual-stream architecture that separates the stable base geometry from complex light transport residuals, converting the ill-posed reflective reconstruction problem into a controlled optimization process.

### 3.1 Hybrid Representation and Material Modeling

Our base representation follows a hybrid Gaussian-SDF design [2, 3]. The scene is represented by a set of anisotropic Gaussian primitives, each associated with position, covariance, opacity, and a learnable feature vector. For each camera ray, Gaussian primitives are projected and alpha-composited to produce the baseline rendered color:

$$C = \sum_{i=1}^{N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \tag{1}$$

where $c_i$ and $\alpha_i$ denote the color and opacity of the $i$-th Gaussian primitive. In parallel, the SDF branch represents the underlying surface as a continuous signed distance field to provide depth and normal constraints for geometric stability.

Building upon this, we introduce a PBR-inspired material branch [14, 28, 29]. Given the Gaussian feature $f$, the material branch predicts albedo, roughness $r$, and metallicity $m$. These factors impose a physically-inspired organization on the latent feature space, encouraging the model to encode stable surface appearance separately from view-dependent reflections. We further use an optimizable environment map $L_{\text{env}}$ to compute an auxiliary PBR shading result:

$$C_{\text{pbr}} = \text{Render}_{\text{PBR}}(\text{albedo}, r, m, L_{\text{env}}, n, v), \tag{2}$$

where $n$ represents the surface normal and $v$ denotes the viewing direction. This auxiliary shading serves as an appearance-organization prior. Because it is supervised against the captured RGB image, it may still encode low-frequency shading or reflectance effects and is not assumed to prevent all reflection information from entering the predicted material variables.

The auxiliary PBR shading is supervised against the captured RGB image and may therefore encode low-frequency shading or reflectance effects. It is used as an appearance-organization prior, not as a guarantee that reflection information is excluded from the predicted material variables.

**Terminology clarification.** We use the term *base appearance* to denote the view-consistent color component optimized by the Gaussian representation. This component is diffuse-oriented, but it should not be interpreted as a strict physical

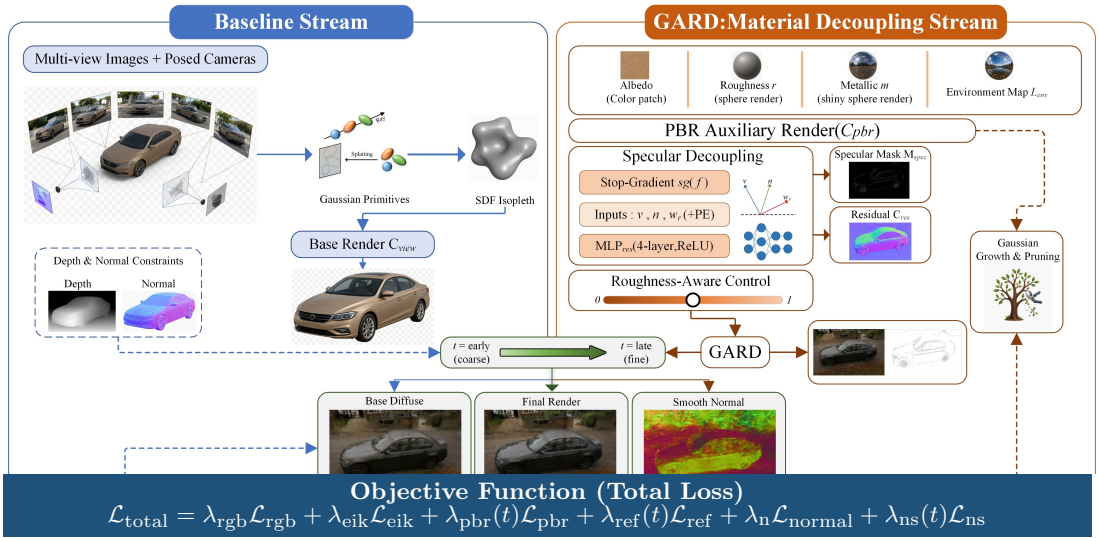



**Fig. 2** ***Detailed architecture of the GARD framework.*** The proposed pipeline separates geometry reconstruction from view-dependent appearance modeling to reduce specular baking. The baseline stream jointly optimizes explicit Gaussian primitives and an implicit surface representation to establish the geometry scaffold and base appearance. Concurrently, the material-decoupling stream organizes appearance using a physically inspired formulation and a dedicated specular residual network. A stop-gradient operation on the shared Gaussian feature input reduces feature-mediated feedback from the residual branch, while a roughness-aware control mechanism modulates its spatial contribution and Gaussian growth/pruning criteria. Surface normals and reflected directions remain differentiable inputs to the residual network. The entire framework is trained through a scheduled coarse-to-fine joint optimization strategy, yielding a diffuse-oriented base representation, view-dependent residuals, and the final mixed render

albedo because it may still include low-frequency shading and illumination effects under unconstrained real-world lighting. The predicted albedo, roughness, metallicity, and environment lighting in our material branch are therefore used as a PBR-inspired structural prior rather than as a complete inverse-rendering solution with unique material recovery. In the following, *base/diffuse* refers to this stable appearance component, while *specular residual* refers to the view-dependent high-frequency component separated from it.

## 3.2 Decoupled Appearance Modeling

The core of our method is an appearance formulation that encourages a stable base component and a view-dependent reflective residual component. The additive reconstruction and sparsity terms do not make this decomposition identifiable; they only impose an optimization preference. As shown in the material decoupling stream of Figure 2, for each Gaussian feature $f$, the reflection residual network predicts a specular mask logit and an RGB residual:

$$(\hat{m}_{\text{spec}}, \hat{C}_{\text{res}}) = \text{MLP}_{\text{res}}(\text{sg}(f), v, n, \omega_r), \qquad (3)$$

where $\text{sg}(\cdot)$ is the stop-gradient operator. In implementation, stop-gradient is applied only to the shared Gaussian feature $f$. The surface normal $n$ and reflected direction $\omega_r$ remain differentiable inputs to the residual branch. Consequently, this operation blocks residual gradients through the feature-input pathway but does not completely isolate residual optimization from geometry-related variables; the remaining coupling through $n$ and $\omega_r$ is constrained by the SDF-guided normal objectives described below. The reflection residual network consists of a 4-layer MLP with ReLU activations. To capture high-frequency specular transients, we apply positional encoding to the view direction $v$ and the reflected direction $\omega_r$, providing the network with sufficient angular bandwidth to model sharp glossy effects [1, 8, 9]. Here, $\omega_r = 2(n^T v)n - v$ denotes the mirror-reflected viewing direction computed from the normal $n$ and view direction $v$.

The specular mask and non-negative residual are computed as follows:

$$M_{\text{spec}} = \sigma(\hat{m}_{\text{spec}}), \tag{4}$$

$$C_{\text{res}} = \text{Softplus}(\hat{C}_{\text{res}}). \tag{5}$$

The gated reflection residual and final decoupled color are obtained by:

$$R_{\text{spec}} = G_{\text{cap}}(r) \odot M_{\text{spec}} \odot C_{\text{res}} \cdot s_{\text{reflection}}, \tag{6}$$

$$C_{\text{decoupled}} = \text{Clamp}\big(C_{\text{base}} + R_{\text{spec}}, 0, 1\big). \tag{7}$$

Here, $C_{\text{base}}$ is the diffuse-oriented base color decoded from the Gaussian feature, while $C_{\text{view}}$ denotes the original view-dependent Gaussian rendering used during the warm-start stage. The scalar $s_{\text{reflection}}$ is a global residual strength coefficient shared across all scenes. $G_{\text{cap}}(r)$ is the roughness-aware residual gate defined below; it makes the residual branch active primarily on locally glossy regions instead of allowing it to absorb arbitrary photometric errors over the whole image. Because $C_{\text{res}}$ is constrained to be non-negative, this formulation is intended to suppress the formation of additional highlight residues in the base appearance; it does not claim to uniquely remove all highlight energy already encoded in $C_{\text{base}}$.

**Roughness-Aware Capacity Control.** To limit the residual branch from becoming an unconstrained error absorber, we introduce a roughness-related control score $s_r$ to modulate the residual gate and Gaussian growth/pruning criteria. Since the predicted variable is not directly supervised by physical roughness labels, $s_r$ should be interpreted as a reflectance-control variable rather than a calibrated material measurement:

$$s_r = \text{Clamp}\left(\frac{\tau_r - r}{\tau_r - r_{\min}}, 0, 1\right)^{\gamma}. \tag{8}$$

The score is not concatenated as an additional network feature. Instead, it enters the optimization as a spatial gate and capacity controller:

$$G_{\text{cap}}(r) = s_r, \tag{9}$$

$$\tau'_{\text{densify}}(x) = \tau_{\text{densify}}\left(1 + \alpha_{\text{d}}(1 - s_r(x))\right), \tag{10}$$

$$\tau'_{\text{prune}}(x) = \tau_{\text{prune}}\left(1 + \alpha_{\text{p}} s_r(x)\right). \tag{11}$$

The first equation gates the residual contribution in the decoupled color equation, so rough diffuse-dominant regions receive little view-dependent residual capacity. The second and third equations adjust the Gaussian densification and pruning criteria during training. Their precise effect depends on the threshold conventions implemented in the Gaussian optimizer; therefore, we do not interpret the equations alone as guaranteeing more primitives in glossy regions. Instead, the complete mechanism should be understood as a spatial capacity-control rule that combines the residual gate with growth and pruning criteria. In our implementation, the adaptive-density module is activated after the residual branch starts, with $\tau_r = 0.20$, $r_{\min} = 0$, $\gamma = 1.0$, $\alpha_{\text{d}} = 2.2$, and $\alpha_{\text{p}} = 1.4$. The residual branch starts at iteration 30000 with a 6000-step warmup, $s_{\text{reflection}} = 0.18$, and a maximum blend weight of 0.65. These parameters are fixed for all scenes and are not tuned per scene.

## 3.3 Geometric Regularization and Optimization

Our geometric regularization relies on multi-source supervision. We jointly constrain Gaussian normals and SDF gradients through depth and normal consistency, ensuring that explicit Gaussian primitives remain anchored to the implicit SDF zero-level set. Let $g = \nabla F(x)$ denote the SDF gradient and $n_{\text{gs}}$ the rendered Gaussian normal. The normal consistency term is:

$$L_{\text{normal}} = \sum_x \left(1 - \left\langle n_{\text{gs}}(x), \frac{g(x)}{\|g(x)\|_2 + \epsilon} \right\rangle\right). \tag{12}$$

To further stabilize geometry, inspired by image-guided geometric regularization in prior work [27], we apply edge-aware normal smoothing with color-aware weights $w_{ij} = \exp(-\frac{\|I_i - I_j\|_1}{\sigma_c})$:

$$L_{\text{ns}} = \sum_{(i,j)\in\mathcal{N}} w_{ij}(1 - n_i^T n_j). \tag{13}$$

This weighting preserves geometric boundaries and material discontinuities while effectively suppressing noise in locally coherent areas.

**Optimization Strategy.** The total objective is optimized via a scheduled coarse-to-fine warmup

strategy:

$$L_{\text{total}} = \lambda_{\text{rgb}} L_{\text{rgb}} + \lambda_{\text{eik}} L_{\text{eik}} + \lambda_{\text{pbr}}(t) L_{\text{pbr}} + \lambda_{\text{ref}}(t) L_{\text{ref}} + \lambda_{\text{n}} L_{\text{normal}} + \lambda_{\text{ns}}(t) L_{\text{ns}}. \tag{14}$$

Here, $L_{\text{ref}}$ denotes the reflection loss and $L_{\text{ns}}$ denotes the edge-aware normal-smoothing loss. The image reconstruction term is:

$$L_{\text{rgb}} = \|C_{\text{final}} - I\|_1, \tag{15}$$

and the SDF regularization follows the Eikonal constraint:

$$L_{\text{eik}} = \sum_x (\|\nabla F(x)\|_2 - 1)^2. \tag{16}$$

The PBR branch is supervised as an auxiliary appearance-organization prior rather than as a complete physical decomposition:

$$L_{\text{pbr}} = \|C_{\text{pbr}} - I\|_1 + \lambda_{\text{brdf}} TV(\text{albedo}, r, m) + \lambda_{\text{env}} TV(L_{\text{env}}). \tag{17}$$

The reflection branch is constrained by sparse residual and mask regularization, together with a base-appearance smoothness prior:

$$L_{\text{ref}} = \|C_{\text{decoupled}} - I\|_1 + \lambda_{\text{mask}} \|M_{\text{spec}}\|_1 + \lambda_{\text{res}} \|C_{\text{res}}\|_1 + \lambda_{\text{base}} TV(C_{\text{base}}) + \lambda_{\text{nh}} \|(1 - H) \odot (C_{\text{base}} - I)\|_1, \tag{18}$$

Here, $C_{\text{pbr}}$ is still fitted to the captured RGB image and may therefore contain low-frequency shading or reflectance effects. We consequently do not interpret the PBR branch as preventing all reflection information from entering the predicted material variables; its role is to provide a structured auxiliary signal, while the residual branch is intended to capture additive high-frequency view-dependent effects. The highlight reliability map is

$$H = \text{Clamp}\left(\frac{\|I - C_{\text{base}}\|_1 - \tau_h}{1 - \tau_h}, 0, 1\right)^{\gamma_h}. \tag{19}$$

The blending weight $w(t)$ increases over time:

$$w(t) = \min\left(1, \frac{t - t_{\text{start}} + 1}{t_{\text{warmup}}}\right) \cdot w_{\max}. \tag{20}$$

The final rendered color is

$$C_{\text{final}} = (1 - w(t)) C_{\text{view}} + w(t) C_{\text{decoupled}}. \tag{21}$$

This schedule allows the model to first establish a geometry scaffold and coarse appearance, then specialize to reflective residuals in the later training stages. At inference, $w$ is fixed to the reported terminal value $w_{\max} = 0.65$, unless otherwise stated. Thus, all reported RGB metrics and rendered comparisons use the scheduled mixture $C_{\text{final}} = (1 - w_{\max}) C_{\text{view}} + w_{\max} C_{\text{decoupled}}$. The components $C_{\text{base}}$ and $R_{\text{spec}}$ are exported separately for decomposition analysis. For all scenes, we use $\lambda_{\text{rgb}} = 1.0$, $\lambda_{\text{eik}} = 0.1$, $\lambda_{\text{n}} = 0.01$, and $\lambda_{\text{ns}} = 0.0015$ after iteration 30000 with $\sigma_c = 0.06$. The PBR branch starts at iteration 28000 with a 5000-step warmup and uses $\lambda_{\text{pbr}} = 0.1$, $\lambda_{\text{brdf}} = 1.5$, and $\lambda_{\text{env}} = 0.02$. The reflection branch starts at iteration 30000 with a 6000-step warmup and uses $\lambda_{\text{mask}} = 0.006$, $\lambda_{\text{res}} = 0.012$, $\lambda_{\text{base}} = 0.035$, $\lambda_{\text{nh}} = 0.045$, $\tau_h = 0.62$, and $\gamma_h = 1.8$. These weights are kept fixed across the evaluated scenes.

# 4 Experiments

## 4.1 Experimental Setup

We evaluate the proposed GARD framework on 15 challenging real-world scenes from two datasets. This includes 12 scenes from the NeRO real-world reflective-object dataset [6]: Sedan, Bear, Bunny, Coral, Maneki, Vase, Cactus, Cup, Grogu, Pepsi, Pitcher, and Teapot, as well as 3 highly reflective scenes from the Shiny dataset [30]: Seasoning, Pasta, and Lab. These scenes cover a comprehensive spectrum of reflective appearances, including glossy coatings, metallic surfaces, curved specular objects, and mixed diffuse-specular materials. Crucially, we restrict our evaluation entirely to real-world captures rather than synthetic datasets. Real-world reflective surfaces present highly unconstrained illumination, unpredictable environmental reflections, and complex material heterogeneity that synthetic environments often fail to replicate, making them the optimal testbed for evaluating multi-view consistency failures.

We compare GARD against representative baselines, including 3DGS, 2DGS, Ref-GS,

Scaffold-GS, GSDF, and Mat-GSDF, our PBR-augmented GSDF baseline [2, 3, 31–33]. Mat-GSDF denotes our controlled baseline that augments GSDF with the same PBR material branch but does not use the proposed reflection residual branch or roughness-aware capacity control. All methods are evaluated at 1/4 of the original image resolution. For NeRO, we use the LLFF hold-8 split; for Shiny, we follow the dataset-specific split. GARD uses one set of hyperparameters across all 15 scenes, with no scene-specific tuning. This setting evaluates robustness of a fixed optimization strategy across varied geometries and materials; it does not constitute feed-forward generalization to unseen scenes.

We report PSNR, SSIM, and LPIPS for novel-view rendering quality. PSNR measures pixel-level fidelity, while SSIM and LPIPS better reflect structural similarity and perceptual quality [34, 35]. For highly reflective objects, LPIPS is particularly informative because small angular deviations may shift highlights spatially, causing severe pixel-wise penalties in PSNR even when the rendered appearance remains perceptually plausible and physically consistent.

For each scene, all compared methods receive the same input images, camera poses, preprocessing, and held-out test views. The NeRO scenes follow the LLFF hold-8 split, while the Shiny scenes follow their dataset-specific split rather than being assumed to share identical view counts. The reported sparse-input observation for Cactus, Cup, and Grogu refers to their dataset-specific setting of 9 input images at $512 \times 512$ resolution; it is not a separate GARD-specific training protocol. The resulting claim concerns robustness under a fixed optimization and evaluation protocol, not feed-forward generalization to unseen scenes.

**Efficiency and model complexity.** Since GARD introduces a material branch, a reflection residual MLP, and an optimizable environment lighting prior, we additionally report computational overhead in Table 1. Following common practice, the comparison is measured on the representative *Sedan* scene under the same resolution, train/test split, and 45k-iteration budget. The table reports both runtime cost and representation scale, so the benefit of reflectance decoupling can be interpreted together with its additional computation.

The results indicate that the proposed decoupling adds moderate runtime overhead mainly during rendering and training, while keeping the learned MLP parameter count nearly unchanged relative to Mat-GSDF. The increased cost is primarily caused by evaluating the reflection residual branch and maintaining the geometry-anchored material structure, rather than by a large increase in network size.

## 4.2 Quantitative Results

Tables 2 and 3 detail the NeRO scene-level quantitative results, Table 4 reports the additional Shiny evaluation, and Table 5 summarizes the average over the 12 NeRO scenes. Compared with earlier standard and view-dependent Gaussian baselines such as 3DGS and Ref-GS, GARD is substantially stronger on several NeRO scenes, although it does not dominate every metric. Ref-GS, for instance, obtains an average of 15.46 PSNR, 0.488 SSIM, and 0.497 LPIPS on NeRO. These results indicate that naively expanding view-dependent modeling within a standard Gaussian pipeline is not sufficient for stable real-world reflective reconstruction. By separating diffuse-oriented base appearance from specular residuals and regulating capacity using roughness cues, our method provides a more controlled alternative.

The most critical comparison lies against the strong geometry-aware hybrid baselines, GSDF and Mat-GSDF. As shown in Table 5, against the formidable GSDF baseline, GARD achieves comparable quantitative performance rather than a massive rendering-metric margin. Specifically, using unrounded averages, GSDF obtains 19.935 PSNR, 0.696 SSIM, and 0.324 LPIPS, while GARD achieves 19.929 PSNR, 0.693 SSIM, and an improved 0.320 LPIPS. The PSNR difference is less than 0.01 and should be regarded as negligible. Therefore, the primary advantage of our approach is not a naive gain in standard image-space metrics, but the ability to produce an interpretable Diffuse/Specular decomposition at essentially equivalent rendering quality. In reflective reconstruction, existing baselines often act as "error absorbers" by baking moving specular highlights directly into the static diffuse color. GARD explicitly reduces this pathological behavior.

Scene-level results reveal a mixed but highly informative pattern. On *Bunny*, GARD obtains

**Table 1 Efficiency and representation complexity on *Sedan*.** MLP params are counted from the saved checkpoint. "Gauss." denotes final Gaussian primitives/anchors, "Mesh V." denotes extracted mesh vertices, and wall FPS is computed from the rendering progress log over the held-out test views.

| Method | MLP Params | Gauss. | Mesh V. | Train time | Wall FPS |
|---|---|---|---|---|---|
| GSDF | 8.2K | 801.2K | 5.44M | 1.37h | 8.44 |
| Mat-GSDF | 14.5K | 481.1K | 5.51M | 1.35h | 5.69 |
| GARD (Ours) | 14.5K | 482.5K | 5.93M | 2.53h | 3.99 |

an LPIPS of 0.134, significantly improving over 3DGS (0.166) and Ref-GS (0.433), though GSDF remains the best overall (0.124). This indicates that while our specular modeling easily surpasses weak view-dependent baselines, the original GSDF geometric prior is already highly effective for this specific geometry.

Interestingly, *Cactus*, *Cup*, and *Grogu* are the only three scenes where GARD ranks first across PSNR, SSIM, and LPIPS simultaneously. These scenes share a relatively sparse input setting (9 images at $512 \times 512$ resolution). This pattern suggests that our diffuse/specular decoupling and capacity control are particularly advantageous in low-view or low-resolution scenarios, where preventing highlight leakage into base appearance and geometry is critical.

Conversely, the failure cases further clarify the scope and architectural trade-offs of the method. On *Pepsi*, GARD is quantitatively weaker than Mat-GSDF (PSNR lower by 1.19, LPIPS worse by 0.024). On *Pitcher*, our LPIPS is comparable to the best result (0.435 vs. 0.436 for Mat-GSDF), but weaker in PSNR. On *Teapot*, Scaffold-GS achieves the best LPIPS (0.429). *Pepsi* contains printed textures with strong local highlights and complex material transitions, where a purely PBR-augmented baseline may better preserve pixel-level appearance by entangling the radiance. *Pitcher* and *Teapot* feature curved reflective structures where highlight motion and self-occlusions are difficult to perfectly localize using a residual branch. In such cases, explicit decomposition introduces a conscious trade-off: it improves interpretability and reduces visible specular baking, but may sacrifice strict image-space alignment when the estimated specular residual is not perfectly localized. These results underscore that the core contribution of GARD is controlled and interpretable reflective reconstruction, rather than universal numerical superiority across all geometries.

Furthermore, Table 4 presents the quantitative comparison on the Shiny dataset (Seasoning, Pasta, and Lab). These scenes provide a complementary testbed with denser real captures, reflective food packages, laboratory objects, and complex indoor illumination. GARD demonstrates highly competitive performance, particularly on the Seasoning scene, where it achieves the best PSNR (28.50), SSIM (0.887), and LPIPS (0.102), indicating that the proposed residual decoupling can preserve sharp reflective structures while keeping the diffuse base less contaminated by highlights. On Pasta and Lab, the results are more mixed: GARD remains competitive in LPIPS on Pasta, while GSDF or Scaffold-GS can be stronger on individual image-space metrics. This behavior is consistent with our central claim that GARD prioritizes controlled appearance decomposition and reduced specular baking rather than universal dominance in every photometric score.

## 4.3 Geometry Consistency Evaluation

Photometric metrics alone cannot fully reflect the behavior of reflective reconstruction methods, since view-dependent highlights may be incorrectly absorbed into static appearance or geometry while still producing plausible RGB renderings. To provide a more geometry-oriented analysis, we additionally evaluate surface consistency using bidirectional nearest-neighbor distances and F-score. Since the real captured scenes do not provide watertight ground-truth meshes, we use the available object point clouds when present and otherwise use SfM point clouds as proxy references. The reconstructed meshes and proxy references are first aligned by bounding-box similarity, and 10k points are sampled for evaluation. These results should therefore be interpreted as supplemental proxy-consistency evidence rather than strict ground-truth surface accuracy. They do not establish a universal geometric improvement, but

**Table 2 Comprehensive quantitative comparison (Part 1).** Comparison across Sedan, Bear, Bunny, Coral, Maneki, and Vase datasets. Best results are **bold**, and second-best are underlined.

| Method | Sedan | | | Bear | | | Bunny | | | Coral | | | Maneki | | | Vase | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ |
| 3DGS | 16.26 | 0.423 | 0.429 | 20.04 | 0.739 | 0.273 | 25.50 | 0.887 | 0.166 | 20.29 | 0.757 | 0.248 | 22.35 | 0.845 | 0.208 | 24.09 | 0.841 | 0.194 |
| 2DGS | 14.28 | 0.365 | 0.512 | 20.24 | 0.760 | 0.257 | 25.00 | 0.865 | 0.212 | 20.56 | 0.777 | 0.232 | 23.71 | 0.850 | 0.227 | 23.12 | 0.823 | 0.227 |
| Ref-GS | 16.74 | 0.348 | 0.507 | 17.36 | 0.515 | 0.436 | 20.68 | 0.630 | 0.433 | 13.24 | 0.491 | 0.471 | 19.33 | 0.606 | 0.450 | 17.39 | 0.571 | 0.423 |
| Scaffold-GS | 17.42 | 0.354 | 0.451 | 19.58 | 0.724 | 0.274 | 24.11 | 0.878 | 0.153 | 19.45 | 0.725 | 0.260 | 23.22 | 0.861 | 0.182 | 23.03 | 0.835 | 0.179 |
| GSDF | **20.44** | **0.510** | **0.343** | 21.03 | 0.758 | 0.246 | **26.56** | **0.902** | **0.124** | 21.02 | 0.780 | 0.223 | 25.08 | 0.881 | **0.151** | 25.42 | 0.862 | **0.151** |
| Mat-GSDF | 20.16 | 0.504 | 0.344 | **21.22** | 0.762 | 0.246 | 26.10 | 0.901 | 0.132 | **21.54** | **0.791** | **0.221** | **25.15** | 0.883 | 0.160 | 25.47 | **0.865** | 0.156 |
| GARD (Ours) | 20.04 | 0.502 | 0.346 | 21.18 | **0.763** | **0.244** | 26.10 | 0.899 | 0.134 | 21.38 | 0.788 | 0.222 | 25.12 | **0.884** | 0.156 | **25.48** | 0.863 | 0.159 |

**Table 3 Comprehensive quantitative comparison (Part 2).** Comparison across Cactus, Cup, Grogu, Pepsi, Pitcher, and Teapot datasets.

| Method | Cactus | | | Cup | | | Grogu | | | Pepsi | | | Pitcher | | | Teapot | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ | P↑ | S↑ | L↓ |
| 3DGS | 15.25 | 0.574 | 0.485 | 16.99 | 0.645 | 0.417 | 13.81 | 0.534 | 0.497 | 16.92 | 0.514 | 0.483 | 16.13 | 0.571 | 0.455 | 14.16 | 0.528 | 0.512 |
| 2DGS | 15.00 | 0.573 | 0.481 | 16.34 | 0.634 | 0.435 | 14.47 | 0.562 | 0.488 | 14.13 | 0.463 | 0.562 | 14.40 | 0.487 | 0.516 | 14.25 | 0.515 | 0.515 |
| Ref-GS | 14.64 | 0.540 | 0.528 | 15.32 | 0.582 | 0.493 | 13.65 | 0.503 | 0.531 | 11.08 | 0.266 | 0.583 | 13.58 | 0.424 | 0.554 | 12.50 | 0.378 | 0.560 |
| Scaffold-GS | 16.85 | 0.679 | 0.423 | 18.07 | 0.684 | 0.390 | 14.67 | 0.594 | 0.439 | 17.12 | **0.598** | 0.463 | 14.99 | 0.565 | 0.466 | 15.53 | **0.627** | **0.429** |
| GSDF | 17.54 | 0.676 | 0.421 | 18.27 | 0.716 | 0.372 | 14.68 | 0.573 | 0.461 | 16.29 | 0.494 | 0.488 | **17.21** | **0.609** | 0.445 | **15.69** | 0.593 | 0.460 |
| Mat-GSDF | 16.59 | 0.647 | 0.429 | 17.53 | 0.669 | 0.401 | 15.26 | 0.602 | 0.443 | **17.69** | 0.535 | **0.461** | 16.23 | 0.546 | 0.436 | 14.99 | 0.534 | 0.483 |
| GARD (Ours) | **17.99** | **0.689** | **0.409** | **18.41** | **0.718** | **0.357** | **15.34** | **0.609** | **0.433** | 16.50 | 0.492 | 0.485 | 16.30 | 0.559 | **0.435** | 15.31 | 0.553 | 0.467 |

assess whether GARD maintains comparable surface consistency while reducing residual-induced artifacts in selected reflective regions. For fair comparison, all method-level averages are computed on the common valid scenes shared by all compared methods.

Table 6 complements the image-space evaluation with geometry-oriented proxy metrics. The results show that GARD maintains comparable proxy surface consistency to the geometry-aware baselines while introducing explicit reflectance decoupling. In the direct comparison with Mat-GSDF, GARD reduces CD-L1 on 4 out of 14 common valid scenes and improves F@2% on 6 out of 14 scenes. These results should be discussed together with the normal-map and surface visualizations. The main benefit of GARD is the reduction of residual-induced appearance and normal artifacts in representative reflective regions, rather than a universal improvement in every scalar geometry metric.

## 4.4 Qualitative Results & Material Decomposition

Qualitatively, the main observed advantage of GARD is a visually more controlled separation of base and residual appearance. As illustrated in Figures 3, 4, and 5, the compared methods show different tendencies in representing view-dependent highlights. GARD encourages high-frequency view-dependent signals to be represented in the residual branch, while the resulting visual differences should not be interpreted as a direct measurement of reduced physical specular baking. This representation can preserve local details in selected reflective regions, although the result depends on the learned residual localization.

The decomposed outputs provide qualitative evidence for the controlled behavior of our design. The Diffuse/Base maps produced by our method appear to contain fewer high-frequency highlight residues in representative glossy regions, while the Specular Residual Component is concentrated around selected highlight regions. Because no ground-truth diffuse/specular layers are available, these observations support a more interpretable representation but do not establish accurate intrinsic decomposition or directly measure reduced physical specular baking.

We further note that the decomposition cannot be justified solely by the additive relation $C_{\text{final}} = C_{\text{base}} + R_{\text{spec}}$. Since the real captured scenes do not provide ground-truth diffuse/specular layers, we do not claim supervised intrinsic decomposition accuracy. Instead, we inspect whether the base

**Table 4** Quantitative comparison on the Shiny dataset (Seasoning, Pasta, and Lab). Best results are bold, and second-best are underlined.

| Method | Seasoning | | | Pasta | | | Lab | | |
|---|---|---|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| 3DGS | 22.92 | 0.776 | 0.252 | 15.09 | 0.380 | 0.369 | <u>28.77</u> | 0.917 | 0.108 |
| 2DGS | 25.96 | 0.825 | 0.226 | 15.47 | 0.404 | 0.417 | 28.09 | 0.906 | 0.122 |
| Ref-GS | 22.23 | 0.658 | 0.437 | 6.55 | 0.061 | 0.752 | 22.82 | 0.714 | 0.304 |
| Scaffold-GS | 26.25 | 0.844 | 0.117 | 16.07 | 0.404 | **0.310** | 26.86 | 0.893 | **0.070** |
| GSDF | <u>27.88</u> | <u>0.884</u> | <u>0.103</u> | **16.14** | **0.419** | 0.372 | **30.17** | **0.942** | <u>0.071</u> |
| Mat-GSDF | 27.34 | 0.869 | 0.113 | 15.33 | 0.366 | 0.327 | 22.97 | 0.807 | 0.177 |
| GARD (Ours) | **28.50** | **0.887** | **0.102** | <u>15.53</u> | <u>0.375</u> | <u>0.326</u> | 23.65 | 0.821 | 0.164 |

**Table 5 Average metrics over 12 real-world reflective scenes.** GARD maintains competitive rendering quality; the table should be interpreted together with the qualitative decomposition and supplemental proxy-geometry analyses.

| Method | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| 3DGS | 18.48 | 0.655 | 0.364 |
| 2DGS | 17.96 | 0.640 | 0.389 |
| Ref-GS | 15.46 | 0.488 | 0.497 |
| Scaffold-GS | 18.67 | 0.677 | 0.342 |
| GSDF | **19.94** | **0.696** | <u>0.324</u> |
| Mat-GSDF | 19.83 | 0.686 | 0.326 |
| GARD (Ours) | <u>19.93</u> | <u>0.693</u> | **0.320** |

component contains fewer sharp view-dependent highlight residues and whether the residual component is localized around reflective regions. This provides qualitative evidence that GARD reduces the tendency of high-frequency view-dependent effects to appear in the static appearance representation, while avoiding unsupported claims about unique material recovery.

The decoupled representation also affects geometry-related visual artifacts, as evidenced by Figure 6 and Figure 7. In reflective regions, conventional baselines may back-propagate photometric errors from shifting highlights into the geometry, resulting in rippled normal artifacts and local surface distortions. By isolating specular residuals and applying SDF-guided geometric regularization, GARD reduces the tendency to explain highlights through surface deformation. The resulting normal maps (Figure 6) appear smoother in locally coherent regions while preserving major object boundaries, enabling cleaner surface visualizations (Figure 7).

## 4.5 Ablation Studies

We conduct comprehensive core ablation studies on the highly challenging *Sedan* scene, which contains massive metallic car bodies and glass regions flooded with strong environmental reflections. To rigorously evaluate component contributions, this ablation utilizes a strictly distinct train/test split from the unified protocol used in Tables 2–5. Therefore, absolute values in Table 8 are independent and highlight the relative component trends.

**Geometry anchor and feature stop-gradient.** The reflection residual branch is constrained to explain view-dependent appearance while reducing its coupling to the shared reconstruction representation. In practice, only the learned Gaussian feature $f$ is detached before entering the residual MLP; the surface normal $n$, reflected direction $\omega_r$, and known camera direction $v$ remain differentiable inputs. This design blocks the residual loss from updating the shared Gaussian feature through the residual feature-input pathway, while still allowing angular appearance cues to be learned through $n$ and $\omega_r$. Because these directional inputs remain differentiable, the stop-gradient operation does not by itself prevent all residual gradients from reaching geometry-related variables. The SDF-guided normal consistency and smoothing objectives therefore remain necessary to constrain the geometry scaffold during joint optimization.

This mechanism is consistent with the trend observed in the ablation and geometry-related visualizations: adding a reflection branch alone improves appearance fitting only weakly, whereas the full GARD model combines residual decoupling with a geometry anchor and yields cleaner

**Table 6 Supplemental geometry consistency evaluation.** CD-L1, accuracy, and completeness are normalized by the proxy-reference bounding-box diagonal. F-score is computed at the indicated diagonal threshold. For the pairwise comparison, negative ΔCD-L1 and positive ΔF@2% indicate better geometry consistency for GARD.

*(a) Method-level average geometry consistency*

| Method | Common valid scenes | | | | | | Object-reference subset | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Scenes | Acc. ↓ | Comp. ↓ | CD-L1 ↓ | F@1% ↑ | F@2% ↑ | Scenes | Comp. ↓ | CD-L1 ↓ | F@2% ↑ |
| 2DGS | 14 | **0.1029** | **0.0620** | **0.1648** | 0.0641 | 0.1549 | 5 | 0.0503 | 0.1208 | 0.1917 |
| Ref-GS | 14 | 0.1065 | 0.0762 | 0.1827 | 0.0667 | 0.1549 | 5 | 0.0544 | 0.1307 | 0.2024 |
| GSDF | 14 | 0.1213 | 0.0830 | 0.2043 | **0.0702** | **0.1671** | 5 | **0.0261** | **0.1207** | **0.2477** |
| Mat-GSDF | 14 | 0.1186 | 0.0647 | 0.1832 | 0.0611 | 0.1492 | 5 | 0.0303 | 0.1249 | 0.2068 |
| GARD (Ours) | 14 | 0.1245 | 0.0758 | 0.2003 | 0.0558 | 0.1363 | 5 | 0.0310 | 0.1314 | 0.1935 |

*(b) Pairwise comparison between GARD and Mat-GSDF*

| Scene | Ref. | ΔCD-L1 ↓ | ΔF@2% ↑ | Scene | Ref. | ΔCD-L1 ↓ | ΔF@2% ↑ |
|---|---|---|---|---|---|---|---|
| Bear | Obj. | +0.0023 | -0.0482 | Pepsi | SfM | +0.0047 | **+0.0001** |
| Bunny | Obj. | +0.0048 | -0.0168 | Pitcher | SfM | **-0.0146** | **+0.0177** |
| Cactus | SfM | **-0.0275** | **+0.0665** | Sedan | SfM | +0.0090 | -0.0700 |
| Coral | Obj. | +0.0153 | -0.0269 | Teapot | SfM | +0.0051 | -0.0349 |
| Cup | SfM | **-0.0083** | -0.0320 | Vase | Obj. | +0.0018 | **+0.0192** |
| Grogu | SfM | +0.0193 | +0.0000 | Lab | SfM | +0.2963 | -0.0621 |
| Maneki | Obj. | +0.0085 | **+0.0064** | Pasta | SfM | **-0.0783** | **+0.0007** |

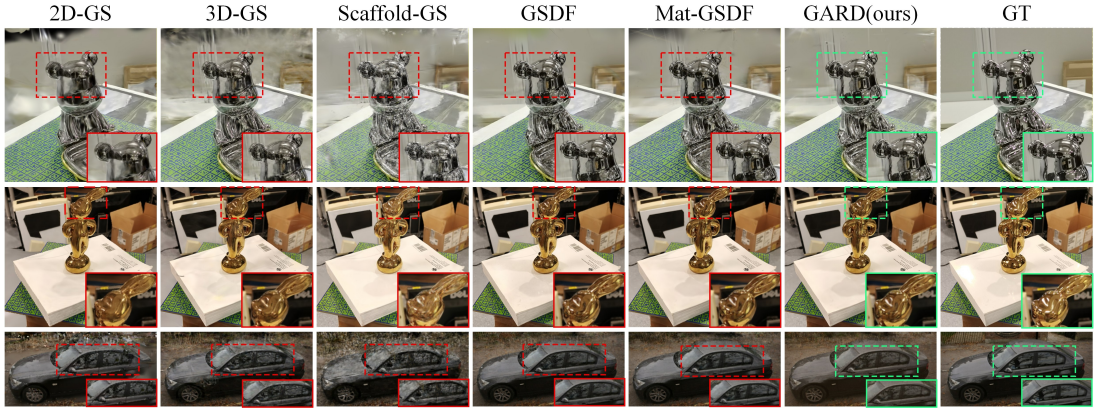



**Fig. 3 Novel-view rendering comparison on representative reflective scenes.** Columns compare baselines and GARD (ours) against the ground truth. Red boxes highlight artifacts such as specular baking and blurred structures. By explicitly decoupling reflection from base color, GARD reduces visible highlight contamination

normal/surface behavior in representative reflective regions. Because the current table does not isolate every stop-gradient choice as an independent row, this result should be interpreted as evidence for the combined design rather than as a standalone causal measurement of the stop-gradient operation.

Overall, the geometry anchor and feature stop-gradient target feature-mediated leakage of reflective residuals into the shared representation, while the controlled residual branch encourages view-dependent transients to remain separate from the stable base appearance. They reduce, rather than eliminate, geometry–appearance coupling because gradients may still propagate through the differentiable normal and reflected-direction inputs.

As shown in Table 8, the vanilla Mat-GSDF baseline obtains the highest PSNR, reaching 23.5716. However, as previously established, this photometric advantage is a symptom of aggressive specular baking into the Diffuse/Base color. In contrast, our full decoupled model (Ours) successfully reallocates this capacity, achieving a slightly better SSIM (0.455219 vs. 0.451791) and improved LPIPS (0.493838 vs. 0.496168), alongside a significantly cleaner decomposed asset representation

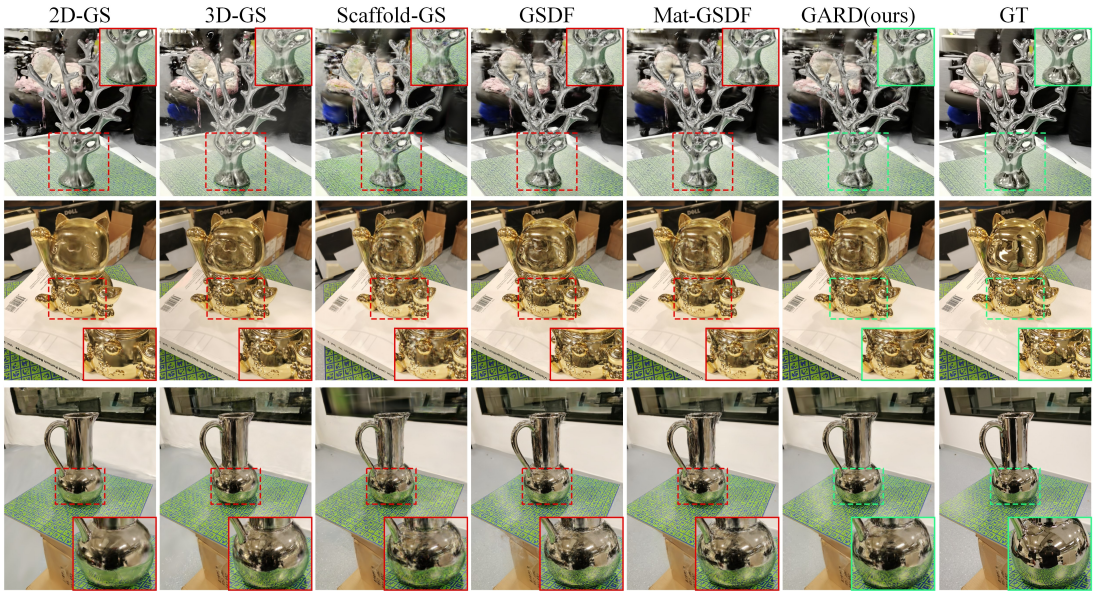



**Fig. 4 Qualitative comparison on complex glossy and metallic curvatures.** Baselines fail to untangle extreme specular shifts from intrinsic color. Our method produces visibly cleaner local details and closer appearance to the ground truth

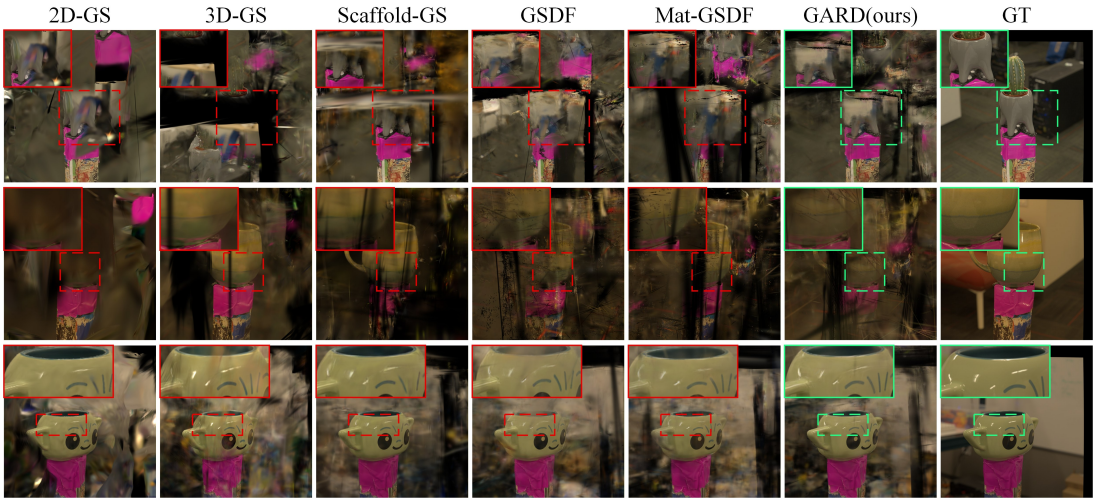



**Fig. 5 Comparison under challenging low-texture and highly reflective regions.** GARD reduces the local floating artifacts and noisy surface projections commonly seen in purely photometric-driven methods by assigning view-dependent effects to the controlled specular residual branch

(as visualized in Figure 8). Removing the PBR formulation (Ours w/o PBR) or utilizing reflection-only variants fails to provide consistent structural improvements, underscoring the necessity of the complete hybrid architecture.

**Roughness-aware capacity control.** GARD does not simply increase the capacity of the view-dependent branch. Instead, it estimates a roughness-guided reflective score $s_r = \text{Clamp}((\tau_r - r)/(\tau_r - r_{\min}), 0, 1)^{\gamma}$, where $r$ is the predicted roughness and $\tau_r$ is the roughness threshold. This score modulates Gaussian growth and pruning, so low-roughness regions are treated as more reflection-sensitive while rough regions are more strongly discouraged from using the residual branch as a generic error absorber.

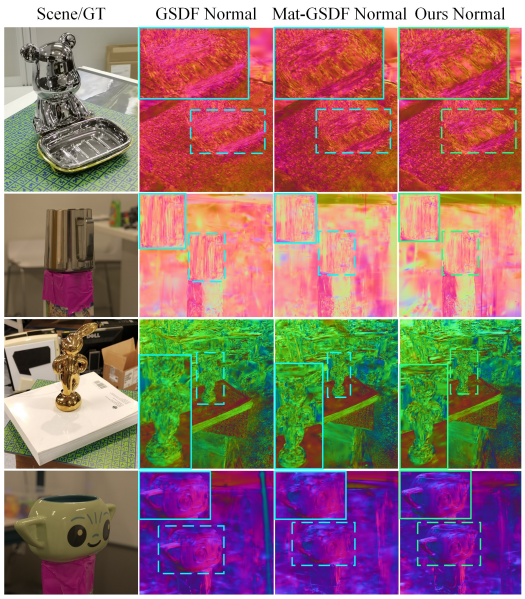



**Fig. 6 Normal-map comparison on reflective regions.** GARD visually reduces high-frequency normal artifacts by assigning view-dependent photometric inconsistency to a controlled residual branch instead of the static geometry representation

Together with residual and mask sparsity losses, this mechanism encourages the specular branch to explain localized reflective transients rather than diffuse texture, shading, or geometry errors.

This mechanism complements the geometry anchor: stop-gradient prevents the residual loss from propagating through the detached feature-input path, while roughness-aware control limits where the residual branch is allowed to spend its view-dependent modeling capacity. The normal and reflected-direction paths remain differentiable; therefore, the base, PBR, residual, and geometry objectives remain jointly coupled, and the operation does not guarantee the absence of all geometry–appearance interactions.

Compared with the tested strong-pruning variant, our *roughness-aware dynamic capacity control* provides a better trade-off on the Sedan ablation. As detailed in Table 8, the strong-pruning variant obtains a lower PSNR (23.2603), lower SSIM (0.449539), and higher LPIPS (0.499225) than the full model. This result indicates that overly aggressive pruning can remove useful representational capacity in this scene, whereas the proposed control directs more residual capacity toward regions with lower predicted roughness-like

**Table 7 Exploratory ablation of the DA3 monocular depth prior on the *Sedan* scene.** Higher PSNR is better; lower L1 error and fewer vertices are better.

| Variant | PSNR | L1 | Vertices |
|---|---|---|---|
| *Mat-GSDF+DA3* | | | |
| Soft $w = 0.0045$ | **23.3871** | **0.042493** | 561027 |
| Soft $w = 0.006$ | <u>23.3287</u> | 0.043770 | 560229 |
| Soft Default | 23.2863 | <u>0.042836</u> | 561232 |
| Conf. = 2.5 | 23.2701 | 0.043473 | 560401 |
| Standard | 23.2298 | 0.045072 | <u>462859</u> |
| Cap 480k | 23.2115 | 0.043327 | 481897 |
| Fused Cap | 22.9989 | 0.044137 | 481134 |
| Soft $w = 0.001$ | 22.9438 | 0.044067 | 561245 |
| *Alternative base* | | | |
| GSDF+DA3 w/o PBR | 19.7583 | 0.069669 | **365464** |

values. Because this comparison is conducted on one ablation scene, it should not be interpreted as proof of universal superiority over every pruning strategy.

We additionally explore the integration of external monocular depth priors (DA3) in Table 7, with visual qualitative comparisons provided in Figure 9 [36]. While a mild soft depth supervision variant ($w = 0.0045$) achieves the best PSNR (23.3871) and L1 error (0.042493) among DA3 configurations, all tested DA3 variants underperform the baseline Mat-GSDF in terms of PSNR. This experiment evaluates the compatibility of an external monocular depth prior with the Mat-GSDF geometry scaffold, rather than providing a complete ablation of DA3 combined with full GARD. The result suggests that, in this setting, enforcing an external monocular depth prior can conflict with high-frequency photometric gradients induced by specular highlights. DA3 should therefore be regarded as an exploratory module rather than a mandatory component of GARD.

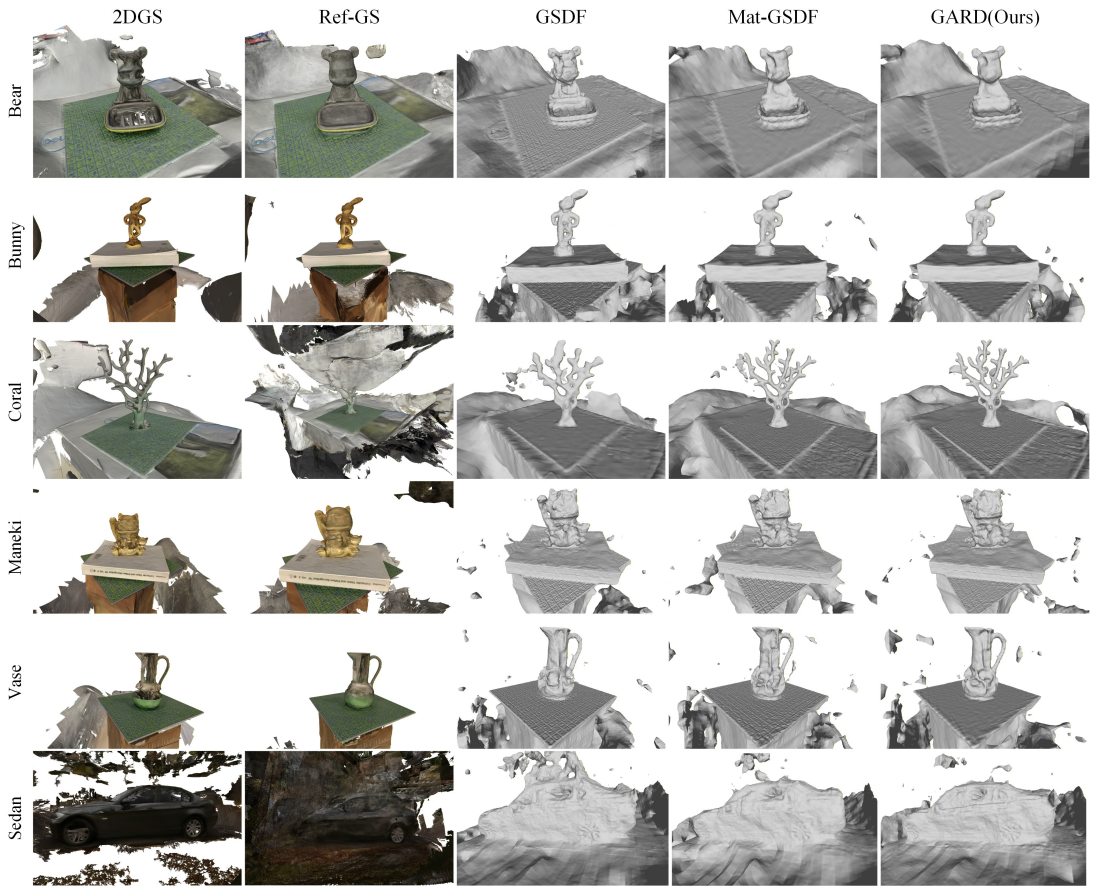



**Fig. 7 Surface reconstruction comparison.** Compared to standard view-dependent baselines that produce incomplete or distorted proxy geometries around reflective objects, GARD limits specular leakage into the SDF optimization, resulting in cleaner surface visualizations in representative reflective regions

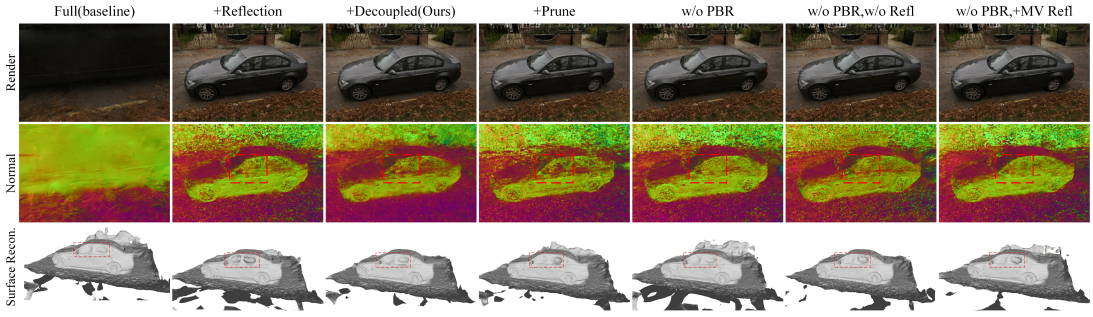



**Fig. 8 Visual ablation of main components on the *Sedan* scene.** The full GARD model provides a favorable trade-off between appearance fitting, local normal coherence, and surface completeness in this example. The strong-pruning variant (column 4) exhibits less complete structural recovery

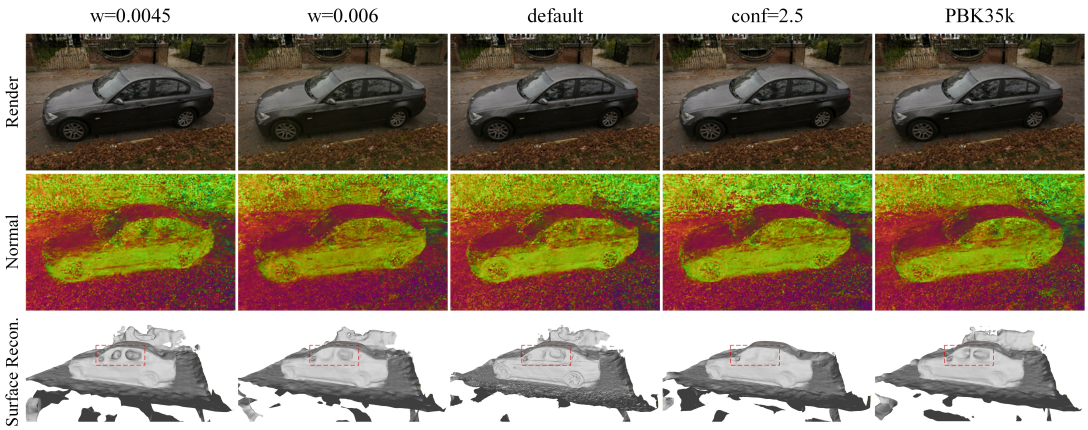



**Fig. 9 Visual comparison of DA3 monocular-depth-prior variants on the *Sedan* scene.** From top to bottom, rows present the rendered appearance, normal maps, and extracted proxy surfaces across various depth-prior configurations. While a moderate soft depth supervision weight balances macro structural alignment with appearance optimization, enforcing rigid external depth constraints invariably conflicts with high-frequency specular gradients. This visual trade-off underscores why monocular depth priors remain an exploratory validation rather than a performance-driving component in unconstrained reflective environments

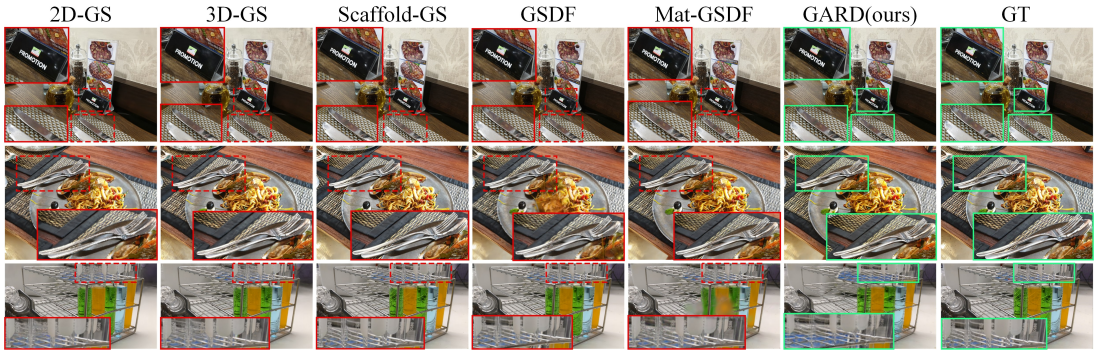



**Fig. 10 Novel-view rendering on the Shiny dataset.** GARD preserves competitive reflective details and exhibits a more controlled base/residual representation on Seasoning, Pasta, and Lab

**Table 8 Core ablation study on the *Sedan* scene.** Evaluated under a specialized split. Higher PSNR and SSIM are better; lower LPIPS and fewer vertices are better. The metrics emphasize the perceptual trade-off between unconstrained PSNR overfitting and true structural decoupling.

| Method | Components | | | Metrics | | | |
|---|---|---|---|---|---|---|---|
| | PBR | Refl. | Cap / Prune | PSNR | SSIM | LPIPS | Vertices |
| *Component Addition* | | | | | | | |
| Mat-GSDF | ✓ | | | **23.5716** | 0.451791 | 0.496168 | 482541 |
| Mat-GSDF w/ Refl. | ✓ | ✓ | | 23.5169 | 0.449007 | 0.498514 | 481058 |
| GARD (Ours) | ✓ | ✓ | ✓ | 23.2620 | **0.455219** | **0.493838** | 481012 |
| *Alternative Settings / Variants* | | | | | | | |
| Ours w/ Strong Pruning | | – | | 23.2603 | 0.449539 | 0.499225 | 481192 |
| Ours w/ Refl. v2 | | – | | 23.3680 | 0.450422 | 0.495927 | **480207** |
| Ours w/o PBR (Cap 480k) | | – | | 23.3650 | 0.449247 | 0.498442 | 480320 |
| Ours w/o PBR w/ Refl. | | – | | 23.4475 | 0.451669 | 0.496625 | 481779 |

### 4.6 Limitations and Future Work

While GARD mitigates specular leakage through feature decoupling, roughness-aware dynamic capacity control, and SDF-guided geometric regularization, it fundamentally operates as a local approximation of complex light transport. Because the specular component is modeled by a residual branch rather than secondary-ray tracing, the framework cannot explicitly resolve deep multi-bounce transport, mutual reflections, or extreme inter-reflections in the manner of SpecTRe-GS or IRGS [21, 22]. In highly specific scenarios—such as densely printed reflective textures on the *Pepsi* scene or hollow reflective curvatures in *Teapot*—this localized approximation may result in imperfect localization of specular components, occasionally affecting the nearby geometric proxy and causing minor photometric misalignments.

Furthermore, maintaining an SDF scaffold and training decoupled residual components alongside PBR materials inevitably increases optimization memory and training time compared to vanilla 3DGS, particularly for high-resolution real-world captures. Future works should explore incorporating lightweight global reflection cues—such as environment-map-guided transport or ray-guided reflection constraints—and accelerating the hybrid architecture using fast-SDF primitives, sparse material-feature caching, or custom CUDA kernel optimizations. These extensions would scale the rigorous physical interpretability of GARD to even more expansive and complex real-world environments.

## 5 Conclusion

We presented GARD, a hybrid reconstruction framework explicitly designed for unconstrained real-world reflective surfaces. Our work addresses a critical pathology in explicit 3D Gaussian and Gaussian-SDF pipelines: under uncontrolled illumination, view-dependent specularities are routinely baked into the static diffuse appearance or force local geometry distortions, leading to severe geometry-appearance entanglement and unstable surface normals. By coupling an SDF-guided geometric scaffold with physically inspired PBR material regularization, GARD provides a rigorous, structured foundation for jointly optimizing geometry and appearance without falling into photometric traps.

The core contribution of our framework is an explicit separation between a diffuse-oriented

base representation and a high-frequency, view-dependent residual representation. Crucially, we introduce roughness-related capacity control and feature stop-gradient in the residual branch to encourage volatile reflective residuals to be represented separately from the static base representation. Comprehensive experiments across 15 challenging real-world scenes show competitive perceptual quality and qualitatively more controlled decomposition in representative reflective regions, without relying on synthetic datasets or per-scene hyperparameter tuning. These results should be interpreted as evidence of cross-scene robustness under the specified protocol, not as proof of identifiable physical material decomposition.

The goal of GARD is not to dominate every image-space metric, but to maintain competitive RGB rendering while reducing the tendency toward geometry–appearance entanglement. Standard hybrid baselines may obtain slightly higher PSNR on scenes with dense textures or complex reflective curvatures. Therefore, the experimental results should be interpreted jointly: PSNR/SSIM/LPIPS measure rendering quality; decomposition analysis checks whether high-frequency residuals are separated from the base appearance; normal and surface visualizations provide qualitative evidence about reflective artifacts; and proxy Chamfer/F-score metrics provide supplemental, rather than definitive, evidence about surface consistency. GARD is intended to encourage cleaner diffuse-oriented base and residual representations while retaining comparable rendering quality, rather than guaranteeing universal improvements in all image-space or geometry metrics.

## Statements and Declarations

**Funding.** This work was supported by the National Natural Science Foundation of China under Grant 62273081.

**Competing interests.** The authors have no relevant financial or non-financial interests to disclose.

**Data availability.** The input images analyzed in this study are available from the NeRO project (https://github.com/liuyuan-pal/NeRO) and the NeX Shiny dataset project (https://nex-mpi.github.io/). Aggregate quantitative results are reported in the tables of this article. No repository is provided with this submission for per-view predictions, trained model checkpoints, or training logs.

**Code availability.** The implementation code is not publicly available at the time of submission.

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

[2] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., *et al.*: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)

[3] Yu, M., Lu, T., Xu, L., Jiang, L., Xiangli, Y., Dai, B.: Gsdf: 3dgs meets sdf for improved neural rendering and reconstruction. Advances in Neural Information Processing Systems **37**, 129507–129530 (2024)

[4] Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In: Advances in Neural Information Processing Systems, vol. 34 (2021)

[5] Yamashita, K., Enyo, Y., Nobuhara, S., Nishino, K.: nlmvs-net: Deep non-lambertian multi-view stereo. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3037–3046 (2023)

[6] Liu, Y., Wang, P., Lin, C., Long, X., Wang, J., Liu, L., Komura, T., Wang, W.: Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images. ACM Transactions on Graphics (ToG) **42**(4), 1–22 (2023)

[7] Yan, Z., Li, C., Lee, G.H.: Nerf-ds: Neural radiance fields for dynamic specular objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8285–8295 (2023)

[8] Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T., Srinivasan, P.P.: Ref-nerf: Structured view-dependent appearance for neural radiance fields. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(11), 9426–9437 (2024)

[9] Wu, L., Bi, S., Xu, Z., Luan, F., Zhang, K., Georgiev, I., Sunkavalli, K., Ramamoorthi, R.: Neural directional encoding for efficient and accurate view-dependent appearance modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 21157–21166 (2024)

[10] Zhu, M., Liu, Y.: Rendering 3d gaussian splatting in web browser: Impacts of spherical harmonics and gaussian counts. IEEE Internet Computing (2025)

[11] Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T., Barron, J.T.: Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics (ToG) **40**(6), 1–18 (2021)

[12] Boss, M., Braun, R., Jampani, V., Barron, J.T., Liu, C., Lensch, H.: Nerd: Neural reflectance decomposition from image collections. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 12684–12694 (2021)

[13] Jin, H., Liu, I., Xu, P., Zhang, X., Han, S., Bi, S., Zhou, X., Xu, Z., Su, H.: Tensoir: Tensorial inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 165–174 (2023)

[14] Zhang, K., Luan, F., Wang, Q., Bala, K., Snavely, N.: 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)

[15] Xiang, H., Li, X., Cheng, K., Lai, X., Zhang, W., Liao, Z., Zeng, L., Liu, X.: Gaussianroom: Improving 3d gaussian splatting with sdf guidance and monocular cues for indoor scene reconstruction. In: 2025 IEEE International Conference on Robotics and Automation (ICRA), pp. 2686–2693 (2025). IEEE

[16] Kim, M., Ali, U.: Gausdf: Signed distance embedded gaussian surfels for 3d reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 33–42 (2026)

[17] Ye, K., Gao, C., Li, G., Chen, W., Chen, B.: Geosplatting: Towards geometry guided gaussian splatting for physically-based inverse rendering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 28991–29000 (2025)

[18] Zhu, Z.-L., Yang, J., Wang, B.: Gaussian splatting with discretized sdf for relightable assets. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 25155–25164 (2025)

[19] Yang, Z., Gao, X., Sun, Y., Huang, Y., Lyu, X., Zhou, W., Jiao, S., Qi, X., Jin, X.: Spec-gaussian: Anisotropic view-dependent appearance for 3d gaussian splatting. In: Advances in Neural Information Processing Systems, vol. 37, pp. 61192–61216 (2024)

[20] Jiang, Y., Tu, J., Liu, Y., Gao, X., Long, X., Wang, W., Ma, Y.: Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5322–5332 (2024)

[21] Tang, J., Fei, F., Li, Z., Tang, X., Liu, S., Chen, Y., Huang, B., Chen, Z., Wu, X., Shi, B.: Spectre-gs: Modeling highly specular surfaces with reflected nearby objects by tracing rays in 3d gaussian splatting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16133–16142 (2025)

[22] Gu, C., Wei, X., Zeng, Z., Yao, Y., Zhang, L.: Irgs: Inter-reflective gaussian splatting with 2d gaussian ray tracing. In: Proceedings

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10943–10952 (2025)

[23] Han, Y., Tie, B., Guo, H., Lyu, Y., Li, S., Shi, B., Jia, Y., Ma, Z.: Polgs: Polarimetric gaussian splatting for fast reflective surface reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 28073–28082 (2025)

[24] Wang, J., Li, H., Chen, Y.: Seg-invrender: fusing semantic segmentation based on nerf for inverse rendering considering shadows. The Visual Computer **41**(7), 4851–4864 (2025)

[25] Wu, H., Hu, Z., Li, L., Zhang, Y., Fan, C., Yu, X.: Nefii: Inverse rendering for reflectance decomposition with near-field indirect illumination. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4295–4304 (2023)

[26] Liang, Z., Zhang, Q., Feng, Y., Shan, Y., Jia, K.: Gs-ir: 3d gaussian splatting for inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21644–21653 (2024)

[27] Turkulainen, M., Ren, X., Melekhov, I., Seiskari, O., Rahtu, E., Kannala, J.: Dn-splatter: Depth and normal priors for gaussian splatting and meshing. In: Proceedings of the Winter Conference on Applications of Computer Vision, pp. 2421–2431 (2025). https://doi.org/10.1109/WACV61041.2025.00241

[28] Cook, R.L., Torrance, K.E.: A reflectance model for computer graphics. ACM Transactions on Graphics (ToG) **1**(1), 7–24 (1982)

[29] Walter, B., Marschner, S.R., Li, H., Torrance, K.E.: Microfacet models for refraction through rough surfaces. Rendering techniques **2007**, 18 (2007)

[30] Wizadwongsa, S., Phongthawee, P., Yenphraphai, J., Suwajanakorn, S.: Nex: Real-time view synthesis with neural basis expansion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8534–8543 (2021)

[31] Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2d gaussian splatting for geometrically accurate radiance fields. In: ACM SIGGRAPH 2024 Conference Papers, pp. 1–11 (2024)

[32] Zhang, Y., Chen, A., Wan, Y., Song, Z., Yu, J., Luo, Y., Yang, W.: Ref-gs: Directional factorization for 2d gaussian splatting. In: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 26483–26492 (2025)

[33] Lu, T., Yu, M., Xu, L., Xiangli, Y., Wang, L., Lin, D., Dai, B.: Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20654–20664 (2024)

[34] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)

[35] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595 (2018)

[36] Lin, H., Chen, S., Liew, J.H., Chen, D.Y., Li, Z., Zhao, Y., Peng, S., Guo, H., Zhou, X., Shi, G., Feng, J., Kang, B.: Depth anything 3: Recovering the visual space from any views. In: International Conference on Learning Representations (2026)