

GlassGS: Geometry and Concept-Aware 3D Gaussian Splatting for Reflective Enclosures

Mingxuan Cui^{1,7}, Yunrui Zhu^{1,2}, Wuqi Wang^{2,3}, Di Lin⁴, Jianhua Zhang⁵, Jie Zhang¹, Ming-Ming Cheng^{6,7}, Shengyong Chen⁵, and Qing Guo^{6,7*}

¹ Agency for Science, Technology and Research (A*STAR), Singapore, 138632

mxuangcui@gmail.com, zhang_jie@cfar.a-star.edu.sg

² Shandong Normal University, Jinan 250358, Shandong, China

zyr@sdu.edu.cn

³ Chang'an University, Xi'an 710064, Shaanxi, China

wuqiwan@chd.edu.cn

⁴ Tianjin University, Tianjin 300072, China

di.lin@tju.edu.cn

⁵ Tianjin University of Technology, Tianjin 300384, China

zjh@ieee.org, sy@ieee.org

⁶ NKIARI, Shenzhen Futian, China

⁷ AAIS & VCIP, Nankai University, China

{cmm, qing.guo}@nankai.edu.cn

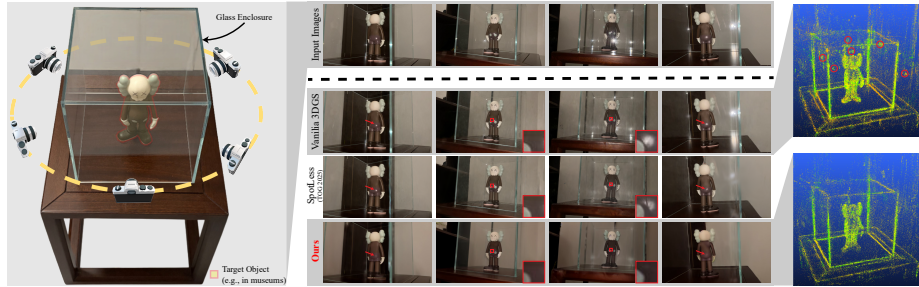


Fig. 1: Data acquisition and reconstruction under reflection contamination : 2D renderings and 3D point clouds. **Left:** data capture with the target inside a glass enclosure and reflective lights; cameras circle the object to collect input images. **Middle:** 2D renderings from Vanilia 3DGS, Spotless (TOG 2025) [32] and our method; **red arrows** indicate specular-flare artifacts degrading reconstruction, **red boxes** mark geometry inconsistencies introduced by reflections. **Right:** corresponding 3D point clouds. **Red circles** show the flare artifacts in 3D.

Abstract. Reconstructing 3D scenes of objects displayed in glass enclosures remains a significant challenge for 3D Gaussian Splatting (3DGS), particularly in museum and exhibition hall settings where transparent protective cases are ubiquitous. The high specular reflections (HSR) generated by glass surfaces severely degrade reconstruction quality, yet their

* Corresponding author.

underlying properties and impact on 3DGS remain poorly understood. Through systematic investigation, we reveal two insights into HSR behavior in 3DGS: (1) Geometry-inconsistency property: HSR causes 3D points to exhibit drastically different appearances across viewpoints, fundamentally violating the geometric consistency assumption that underpins 3DGS; (2) Concept-preservation property: Despite corrupting local appearance, HSR has remarkably limited influence when learning high-level object concepts, as the true object structure remains intact beneath the reflective layer. Motivated by these insights, we propose GlassGS, a geometry and concept-aware 3DGS framework with an iterative two-stage optimization. First, we compute geometry-consistency maps for each input view to evaluate the appearance variations across viewpoints, explicitly identifying HSR-affected pixels by measuring the appearance variations of 3D points across views. Second, we introduce a geometry-inconsistency-guided object concept learning module that leverages these maps to learn robust object representations while actively suppressing reflection interference. The two stages mutually reinforce each other through iterative refinement, improving both inconsistency detection and concept learning. For this new task, we collect the first dataset specifically designed for 3DGS reconstruction of glass-enclosed objects. Extensive experiments demonstrate that GlassGS significantly outperforms state-of-the-art methods, enabling high-fidelity digitization of glass-enclosed scenes and opening new possibilities for virtual museum experiences.

Keywords: 3D Gaussian Splatting · High specular reflections · Geometry consistency · Concept Learning

1 Introduction

3D reconstruction is a fundamental task in augmented reality (AR), autonomous driving, robotic perception, and cultural heritage preservation [19; 28]. Recent advances in Neural Radiance Fields (NeRF) [23] and 3D Gaussian Splatting (3DGS) [15] have significantly improved scene modeling and novel view synthesis. NeRF leverages volumetric rendering to learn radiance fields with high-quality view interpolation, while 3DGS achieves real-time rendering by optimizing 3D Gaussian primitives. However, objects in glass enclosures remain challenging: transparent cases produce high specular reflections (HSR) that corrupt multi-view consistency. As shown in Figure 1, the same reflective region can appear bright, dark, or contain virtual objects across views, violating 3DGS’s per-point appearance consistency and causing reflections to be mistaken for geometry, resulting in severe reconstruction artifacts.

Although reflection-aware extensions have been explored [22; 47], their limitations become evident under strong or dynamic reflections. NeRFReN [10] decomposes scenes into transmission and reflection fields, but struggles with flash-induced artifacts. NeuS-HSR [29] incorporates physical priors with planar reflection assumptions, yet fails under complex non-planar surfaces. Editing-oriented

approaches such as GaussianEditor [41] enable object-level manipulation but rely on explicit semantic masks and cannot resolve ambiguities introduced by virtual reflections. Meanwhile, 2D reflection removal methods [12; 31; 42; 43; 51] leverage diffusion models or flash cues to suppress reflections in single images, but lack multi-view geometric reasoning and thus remain unsuitable for 3D reconstruction. As shown in Figure 1, even dynamic-artifact suppression methods such as SpotLessSplats [32] fail under strong or view-dependent reflections, leaving residual artifacts that still mislead multi-view reconstruction.

Through systematic investigation, we identify two fundamental properties of HSR in the context of 3DGS as detailed in Sec. 3 : **①** Geometry-inconsistency property: HSR causes the same 3D point to exhibit drastically different appearances across viewpoints, fundamentally violating the multi-view consistency assumption that underpins 3DGS. This leads to reflections being misinterpreted as geometry, producing spurious structures and distortions. **②** Concept-preservation property: Despite corrupting local appearance, HSR has limited influence on high-level concept learning. Pretrained vision models tend to capture the underlying object structure rather than its reflective artifacts, suggesting that semantic priors can help disambiguate reflections from true geometry. These findings reveal why geometry-only formulations struggle: they neither exploit cross-view conflicts specific to reflections nor leverage concept-level priors that can disambiguate real objects from reflective artifacts. Moreover, one-shot removal without feedback fails to correct initial errors or progressively improve reliability estimation. To address these gaps, we propose GlassGS, a geometry- and concept-aware 3DGS framework that incorporates an iterative two-stage optimization, tightly coupling cross-view inconsistency analysis with object-centric representation learning.

GlassGS is a geometry- and concept-aware framework that alternates two stages: a geometry-consistency map detects and downweights reflection-contaminated pixels to protect 3D geometry, and concept-guided learning uses these cues to restore object structure and suppress reflection ambiguity. Iterating these stages aligns 2D appearance with 3D geometry, producing geometrically consistent, semantically faithful reconstructions. To evaluate this setting, we collect the dataset specifically tailored to 3DGS reconstruction of glass-enclosed objects, encompassing challenging conditions such as dynamic lighting. Extensive experiments show that GlassGS substantially outperforms state-of-the-art baselines, enabling high-fidelity digitization of glass-enclosed scenes and opening avenues for realistic virtual museum experiences. The main contributions of this paper are as follows:

- We present **GlassGS**, a geometry- and concept-aware 3DGS designed for reflection-heavy glass-enclosed scenarios, grounded in the identified HSR properties.
- We introduce a geometry-consistency map to measure the appearance variations that quantitatively capture reflection contamination and guide optimization in 3DGS.

- We leverage a concept-guided learning module that leverages inconsistency signals to maintain true object structure while suppressing reflective artifacts.
- We design a closed cycle that couples inconsistency detection and concept learning, progressively improving stability and fidelity.
- We establish a dataset targeting 3DGS reconstruction of glass-enclosed objects, enabling rigorous evaluation under complex reflections and dynamic illumination.

2 Related work

3D Reconstruction and Gaussian Splatting. Multi-view 3D reconstruction is fundamental [45]. NeRF and its extensions (*e.g.*, Mip-NeRF, RegNeRF) advanced photorealistic neural implicit reconstruction [3; 23; 26], but volumetric rendering is costly and prone to multi-view inconsistency—especially under high specular reflections (HSR), which distort geometry. Explicit representations like 3D Gaussian Splatting (3DGS) enable real-time rendering and compact models by optimizing Gaussian primitives [15], and works such as GaussianEditor extend 3DGS to editing and texture refinement [41]. Yet Gaussian-based methods still misinterpret specular regions as fragmented geometry; improving 3DGS robustness to HSR motivates our work.

High Specular Reflection (HSR) Removal. Existing HSR removal methods fall into three paradigms: physically-based modeling, appearance-level decomposition, and image-domain reflection removal. Physically-based methods [6; 20; 37] model reflections via geometric symmetry or ray tracing but treat them as visual effects rather than geometry, limiting 3D consistency. Appearance-decomposition approaches [35; 46; 48; 49] integrate reflection–transmission disentanglement into Gaussian-based 3D representations (deferred shading, parameter extension, or flash cues) and aid appearance editing but often leave residuals due to weak geometric supervision. Image-domain methods [7; 9; 29; 44] suppress reflections in 2D using priors or learning but lack multi-view coherence, making them unsuitable for geometry-consistent 3D reconstruction. In contrast, we tackle specular artifacts directly in 3D Gaussian space, enabling reflection-aware reconstruction with explicit geometric consistency.

Semantic and Diffusion Priors in 3D Vision. Semantic priors improve interpretability and structural coherence in 3D reconstruction [36; 40]. Semantic NeRF and Semantic SLAM integrate category-level semantics [17; 38; 50; 53], and GSNeRF embeds semantic features into NeRFs [4]. Vision–language models like CLIP have been extended to 3D understanding [18; 27; 30; 39], and Feature 3DGS fuses CLIP/SAM into Gaussian Splatting, though reflection reasoning remains unexplored [52]. Diffusion models provide strong generative priors for restoration [8], and their 3D variants (RenderDiffusion, RefFusion, 3D-Consistent Inpainting, D3DR) enable multi-view, illumination-consistent reconstruction [1; 2; 11; 24; 33; 34]. Inspired by these trends, we leverage semantic and diffusion priors to guide reflection disentanglement and strengthen 3DGS under specular conditions.

3 Empirical Study and Motivation

While multi-view reconstruction methods have achieved remarkable progress in recent years, their assumptions are fundamentally challenged in museum and exhibition settings where objects are displayed behind transparent glass enclosures. The primary difficulty stems from **High specular reflections (HSR)**, which introduce severe inconsistencies across viewpoints and corrupt the reconstruction pipeline in two distinct ways.

View-dependent appearance variation. Reflections on glass surfaces exhibit dramatic changes when observed from different viewpoints. As highlighted in the red rectangle of the Figure 1, the same reflective region may appear bright in one view due to flash, dark in another due to light-source movement, or even contain virtual objects reflected from the surrounding environment. These variations break the geometric consistency assumption that underpins 3D Gaussian Splatting (3DGS), where each 3D point is expected to maintain a stable appearance across views. Consequently, reflection artifacts are mistakenly interpreted as genuine geometry, leading to spurious Gaussian expansion, distorted surfaces, and degraded reconstruction fidelity.

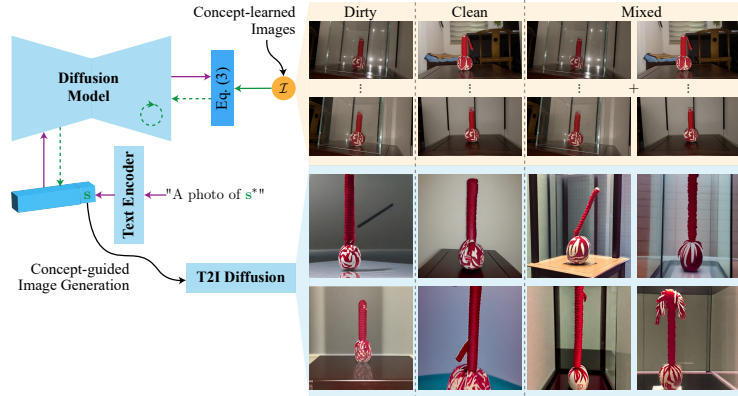


Fig. 2: Concept-preserving stability under reflection contamination. Comparison of textual-inversion results under no, partial, and heavy reflection contamination: example training images (top-right) and corresponding generated outputs (bottom-right)

Concept-preserving stability of the target object. Despite the disruptive local signals introduced by reflections, the true object structure remains intact beneath the glass. To validate it, we conduct a textual inversion experiment under three conditions: training on images without reflections, with partial reflections, and with heavy reflection contamination. As illustrated in Figure 2, all three settings yield generated results that consistently preserve the core object concept—capturing its global shape, topology, and material identity—even

when local appearance is severely corrupted. In other words, while reflections distort pixel-level geometry alignment and introduce spurious highlights, they exert limited influence on the learning of object-level representations. This dichotomy suggests that reflections and genuine object features can be distinguished by their behavior: reflections are highly inconsistent across views, whereas genuine object surfaces maintain a stable conceptual identity.

Together, these observations highlight the need for a reconstruction framework that is both **geometry-aware** and **concept-aware**. By explicitly modeling reflection-induced inconsistencies and simultaneously leveraging the stability of object concepts, it becomes possible to distinguish between reflection artifacts and authentic geometry. This motivates our proposed framework, *GlassGS*, which integrates geometry-inconsistency detection with concept-guided learning in an iterative optimization process.

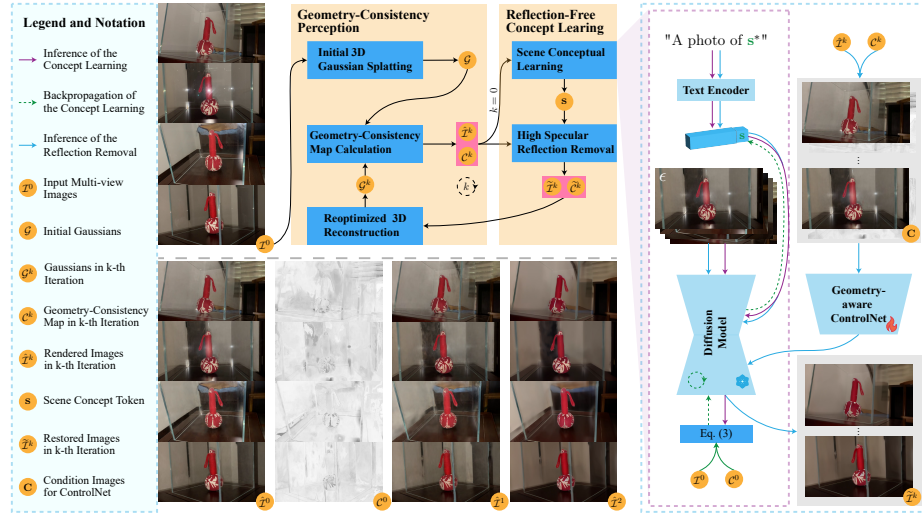


Fig. 3: Pipeline of our GlassGS and examples. Top-left: two key modules along with their operational workflow. **Bottom-left:** rendered results and corresponding geometry-consistency maps across different optimization iterations, demonstrating the progressive removal of reflections. **Right:** pipeline for reflection-free concept learning. The definitions of Arrows are shown on the left side.

4 Methodology

4.1 Overview

As illustrated in Figure 3, GlassGS operates in two mutually reinforcing stages:

❶ **Geometry-inconsistency perception.** For each input view, we explicitly

measure the inconsistency of 3D points across viewpoints utilizing geometry-consistency maps that identify HSR-affected pixels and provide a reliability signal that prevents reflection-dominated observations from steering 3DGS optimization toward non-physical geometry. **② Concept-guided learning under geometry-consistency maps.** By leveraging geometry-consistency maps to denote the geometry inconsistencies, we develop robust object representations that effectively mitigate the impact of reflective interference. This process emphasizes concept-level cues to maintain the integrity of true structural information while suppressing ambiguities introduced by reflections. Consequently, this approach ensures that semantic understanding harmoniously integrates with geometric reasoning for enhanced interpretability.

The two stages are integrated in an iterative refinement loop: improved inconsistency detection enhances concept learning, and enhanced concept learning, in turn, refines inconsistency estimation. This loop progressively aligns 2D appearance and 3D geometry, stabilizes optimization, and reduces artifact accumulation, yielding reconstructions that are both geometrically consistent and semantically faithful.

4.2 Geometry-Consistency Perception

Specular reflection regions introduce geometric ambiguity in multi-view reconstruction. Reflection artifacts produce inconsistent 3D projection signals across viewpoints (*e.g.*, flash, virtual objects reflected on glass), causing 3D Gaussians to converge to non-existent geometric structures incorrectly. This results in model expansion and surface distortion, compromising the accuracy of the reconstructed scene. Moreover, reflective regions exhibit significant color variance across viewpoints, whereas real surfaces maintain consistent color appearance. To address this, we incorporate a geometry-consistency perception (GCP) mechanism in reconstruction, which effectively distinguishes reflections from actual objects.

First of all, given multi-view images $\mathcal{I} = \{\mathbf{I}_i\}_{i=1}^N$ and corresponding camera poses $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^N$, we compute a pixel-wise geometry-consistency map $\mathbf{Cons}_i = \{\text{cons}_x\}_{x \in \mathbf{I}_i}$ to quantify reflection contamination, ensuring precise identification of specular regions. For a pixel x in the image $\mathbf{I}_i$ at the viewpoint $\mathbf{p}_i$, we project it into 3D world coordinates $\mathbf{w}_x$ using depth estimation $\mathbf{D}_i$ from the 3D Gaussian scene. We then track its visibility across adjacent camera poses $\mathcal{P}' \subset \mathcal{P}$ and compute its multi-view color variance as reflection geometry-consistency:

$$\text{cons}_x = \frac{1}{|\mathcal{P}'|} \sum_{\mathbf{p}_j \in \mathcal{P}'} \text{Similarity}(c(\mathbf{w}_x, \mathbf{p}_j), c(\mathbf{w}_x, \mathbf{p}_i)), \quad (1)$$

where $c(\mathbf{w}_x, \mathbf{p}_j)$ is the projected color in $\mathbf{w}_x$ observed from the view $\mathbf{p}_j$. A lower cons_x value indicates strong multi-view inconsistency, suggesting the presence of reflection artifacts. We aggregate cons_x across pixels to construct the geometry-consistency map $\mathbf{Cons}_i$.

Following the 3DGS reconstruction framework, we dynamically refine 3D Gaussian parameters to suppress reflection artifacts while preserving geometric integrity. Let $\mathcal{G} = \{\mathbf{g}_k\}_{k=1}^M$ denote 3D Gaussians with parameters $g_k = \{\mu_k, \Sigma_k, \alpha_k, \mathbf{c}_k\}$, where μ_k represents the Gaussian center in 3D space, Σ_k is the covariance matrix controlling shape and orientation, $\alpha_k \in [0, 1]$ is the opacity parameter, and $\mathbf{c}_k$ denotes the view-dependent color coefficients.

To improve robustness against reflections, we integrate reflection-aware geometry-consistency maps into the optimization process. The objective function is as follows:

$$\mathcal{L}_{geo} = \lambda_1 \sum_{i=1}^N \|\mathbf{Cons}_i \odot (\mathbf{I}_i - \hat{\mathbf{I}}_i)\|_1 + \lambda_2 \sum_{i=1}^N \text{D} - \text{SSIM}(\mathbf{I}_i, \hat{\mathbf{I}}_i, \mathbf{Cons}_i), \quad (2)$$

where $\hat{\mathbf{I}}_i = \text{Render}(\mathcal{G}, \mathbf{p}_i)$ denotes the rendered image from the 3D Gaussian representation $\mathcal{G}$ at the viewpoint $\mathbf{p}_i$. The weights λ_1, λ_2 balance detail preservation (L1 loss) and structural consistency (D-SSIM loss) following vanilla 3DGS. Both the L1 and D-SSIM losses suppress reflections by down-weighting low-geometry-consistency regions, focusing optimization on reliable areas while allowing visual clues from other viewpoints in uncertain zones.

4.3 Reflection-Free Concept Learning

To disentangle genuine object features from reflection artifacts, we introduce a reflection-free concept learning (RFCL) framework. Motivated by our finding that object concepts remain stable even under heavy reflection contamination (*i.e.* Figure 2), we first learn a reflection-free scene embedding $\mathbf{s}_{\text{scene}}$ via textual inversion, capturing the semantic identity of the object independent of local noise. We then apply a geometry-aware ControlNet suppression module that leverages reflection-inconsistency maps to filter spurious geometric signals. Together, these two stages—concept learning and geometry suppression—complement each other, enabling robust reflection removal of the HSR images for glass-enclosed scenes.

Scene Conceptual Learning. Given a small set of reflection-contaminated patches $\mathcal{I}$, we optimize a reflection-free text embedding $\mathbf{s}_{\text{scene}}$ to encode the concept of the scene. The optimization objective is formulated as:

$$\mathbf{s}_{\text{scene}} = \arg \min_{\mathbf{s}^*} \mathbb{E}_{\mathbf{I}_i \in \mathcal{I}, \mathbf{z} = \phi(\mathbf{I}), \epsilon \in \mathcal{N}(0,1), t} [\|\mathbf{Cons}_i \odot (\epsilon - \epsilon_{\theta}^{SD}(\mathbf{z}_t, t, \mathcal{Y}(\phi, \mathbf{s}^*)))\|_2^2], \quad (3)$$

where ϵ_{θ}^{SD} is a frozen Stable Diffusion denoiser, which predicts the noise ϵ added to the latent representation $\mathbf{z}_t$ at timestep t in the forward diffusion process. The function $\phi(\mathbf{I})$ maps the input image $\mathbf{I}$ into latent space, typically via a variational autoencoder (VAE), and produces the corresponding latent variable $\mathbf{z}$. $\mathbf{z}_t$ represents the noisy latent representation at step t . To integrate the reflection-free embedding into the text-conditioning process, $\mathcal{Y}(\phi, \mathbf{s}^*)$ replaces a placeholder token with the learned embedding $\mathbf{s}^*$, ensuring that the model incorporates reflection-free features in the denoising process under the guidance of $\mathbf{Cons}_i$.

Geometry-aware ControlNet. After obtaining $\mathbf{s}_{\text{scene}}$, we integrate it into a ControlNet-based suppression network. To enable precise geometry reflection control, we design a 4-channel conditional input consisting of: Rendered Image $\hat{\mathbf{I}}^{p_i}$, which is contaminated by reflections. And a geometry-consistency map $\mathbf{Cons}_i$, which quantifies the reflection intensity. The conditional input is represented as:

$$\mathbf{C} = \text{Concat}(\hat{\mathbf{I}}^{p_i}, \mathbf{Cons}_i) \in \mathbb{R}^{H \times W \times 4}, \quad (4)$$

where $\text{Concat}(\hat{\mathbf{I}}^{p_i}, \mathbf{Cons}_i)$ preserves geometry structures of the viewpoint while providing probabilistic suppression signals through $\mathbf{Cons}_i$.

To utilize the consistency map as a geometric guide for reflection removal, we trained a ControlNet model on a large-scale synthetic dataset to regulate the diffusion process. The input to our ControlNet consists of a four-channel image: three channels representing the contaminated image and one corresponding to the geometry-consistency map.

However, effectively training a model with limited 3D data presents a significant challenge. In our work, we leverage each clean image and its corresponding synthetically reflection-contaminated image to simulate the computation process of cons_i . We employ the pixel from an artificially perturbed and Gaussian-blurred image of the clean image to get the approximation of cons_x :

$$\begin{aligned} \text{cons}'_x &= \text{Similarity}(\bar{c}(\mathbf{w}_x, \mathbf{p}_j), c(\mathbf{w}_x, \mathbf{p}_i)) \\ &\approx \frac{1}{|\mathcal{P}'|} \sum_{\mathbf{p}_j \in \mathcal{P}'} \text{Similarity}(c(\mathbf{w}_x, \mathbf{p}_j), c(\mathbf{w}_x, \mathbf{p}_i)) \\ &= \text{cons}_x. \end{aligned} \quad (5)$$

Gaussian blur can be viewed as a weighted average over neighboring sample points, *i.e.*, $\bar{c}(\mathbf{w}_x, \mathbf{p}_j)$. In essence, Equation (1) performs a similar operation—averaging over a point’s neighborhood and comparing it to the point itself via a similarity measure. This analogy allows us to use a 2D Gaussian blur at a pixel to approximate 3D neighborhood sampling. In both cases, the geometry-consistency estimation provides a coarse assessment of visibility in the surrounding region of a pixel, making the two approaches partially equivalent and thereby helping to bridge the gap between 2D and 3D data.

We adopt a **geometry-consistency-aware training strategy** for ControlNet. Given a clean image $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$, its latent $\mathbf{z}_0$ is obtained via a variational autoencoder ϕ . A random sample timestep $t \sim \mathcal{U}\{0, T-1\}$ is sampled from the T timesteps, and Gaussian noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ is added to $\mathbf{z}_0$ according to the parameter $\bar{\alpha}_t$:

$$\mathbf{z}_t = \sqrt{\bar{\alpha}_t} \mathbf{z}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \mathbf{z}_0 \sim \phi(\mathbf{I}). \quad (6)$$

The text encoder produces the text-guidance feature h_{text} . The control image $\mathbf{C}$ contains 4 channels: three RGB channels and one geometry-consistency channel $\mathbf{C}^{(4)}$. The predicted noise $\hat{\epsilon}$ from the ControlNet-augmented diffusion U-Net is given by:

$$\hat{\epsilon} = \text{UNet}_\theta(\mathbf{z}_t, t, h_{\text{text}}, \mathbf{C}). \quad (7)$$

The predicted latent $\hat{\mathbf{x}}_0$ is computed to align with $\mathbf{z}_0$:

$$\hat{\mathbf{x}}_0 = \frac{(\mathbf{z}_t - \sqrt{1 - \bar{\alpha}_t} \hat{\epsilon})}{\sqrt{\bar{\alpha}_t}}. \quad (8)$$

Finally, the overall training loss is:

$$\mathcal{L} = \mathcal{L}_{\text{noise}}(\hat{\epsilon}, \epsilon) + \mathcal{L}_{\mathbf{x}_0}(\hat{\mathbf{x}}_0, \mathbf{z}_0). \quad (9)$$

For inference, the ControlNet-denoised latent representation at diffusion step t is computed as:

$$\mathbf{z}_{t-1} = \text{ControlNet}_\epsilon \left(\mathbf{z}_t, t, \underbrace{\mathbf{s}_{\text{scene}}}_{\text{concept}}, \underbrace{\mathbf{C}}_{\text{geometry_condition}} \right). \quad (10)$$

Here, $\mathbf{s}_{\text{scene}}$ provides high-level scene cues to the generation model, and $\mathbf{C}$ supplies local geometry context along with geometry-consistency-guided suppression of undesired reflections. $\text{ControlNet}_\epsilon$ denotes the ControlNet-augmented denoiser, which jointly conditions on semantic scene embedding and geometry consistency information.

4.4 Self-Cycling Refinement

To ensure reflection-free 3D reconstruction while preserving tight geometry–texture consistency, we propose Self-Cycling Refinement—an iterative feedback mechanism that progressively suppresses residual reflections. The procedure alternates between 3D Gaussian refinement and reflection-free image synthesis, repeating until convergence. The pipeline comprises four steps:

1. Initialization with reflection-aware 3D reconstruction. Given the reflection-contaminated images $\mathcal{I}^0 = \{\mathbf{I}_i^0\}_{i=1}^N$, we estimate the initial geometry-consistency maps $\mathcal{C}^0 = \{\mathbf{Cons}_i^0\}_{i=1}^N$ and optimize the 3DGS parameters by reflection-aware loss $\mathcal{L}_{\text{geo}}$, *i.e.*

$$\mathcal{G}^k = \arg \min_{\mathcal{G}} \mathcal{L}_{\text{geo}}(\mathcal{G}, \mathcal{I}^0, \mathcal{C}^0). \quad (11)$$

We then render the corresponding multi-view images $\hat{\mathcal{I}}^0$.

2. Reflection-free image generation. From the rendered inputs $\hat{\mathcal{I}}^{k-1}$, we generate de-reflected images $\tilde{\mathcal{I}}^k = \{\tilde{\mathbf{I}}_i^k\}_{i=1}^N$ using ControlNet-augmented Diffusion.
3. Reoptimize 3D Gaussian Reconstruction. To emphasize newly reliable view-points, we assign $\tilde{\mathcal{C}}^k = \{1 - \mathbf{Cons}_i^{k-1}\}_{i=1}^N$ as the geometry-consistency to $\tilde{\mathcal{I}}^k$. We then refine the 3DGS by minimizing the reflection-aware loss over the concatenation of the original and de-reflected observations:

$$\mathcal{G}^k = \arg \min_{\mathcal{G}} \mathcal{L}_{\text{geo}}(\mathcal{G}, [\tilde{\mathcal{I}}^k; \mathcal{I}^0], [\tilde{\mathcal{C}}^k; \mathcal{C}^0]), \quad (12)$$

where $[\cdot; \cdot]$ denotes image concatenation over views.

4. Rendering and geometry-consistency update. We render multi-view images $\hat{\mathcal{I}}^k$ from $\mathcal{G}^k$ and recompute the geometry-consistency maps $\mathcal{C}^k$.

We iterate steps 2-4 until either the maximum number of iterations is reached or the geometry-consistency exceeds a threshold, *i.e.*, $\text{Mean}(\mathcal{C}^k) > \tau$. Upon convergence, we output the final reflection-free 3D Gaussians and images.

5 Experiment

To evaluate the effectiveness of our GlassGS in 3D Gaussian reconstruction under strong specular reflections, we conduct comprehensive experiments on the GlassFlash3D Dataset we collected and the Real-World Dataset from [29]. Besides, we synthesized the GlassFlash2D dataset to train the geometry-aware-guided ControlNet.

5.1 Datasets

High specular reflections (HSR) pose severe challenges in 3D reconstruction, resulting in inaccurate geometry, distorted textures, and incorrect depth estimations. Despite the rapid progress in neural reconstruction methods, existing datasets lack dedicated benchmarks that systematically evaluate reflection-aware 3D reconstruction. To bridge this gap, we collect GlassFlash2D and GlassFlash3D, two datasets specifically designed for reflection handling in single-view and multi-view restoration.

Synthetic Training Data in GlassFlash2D. Background: 20K indoor scenes randomly sampled from NYU Depth V2 [25], covering 464 distinct room layouts with depth maps. Foreground: 20K reflective objects (glassware, screens, metallic items) from Flickr24K [16]. Specular Flares: 1,924 flare patterns from Flare7K++ [5], including diffuse glows and sharp caustics.

Each synthetic contaminated image is generated by compositing a background, a reflective foreground object, and a flare pattern. This process yields a total of 100,000 paired samples, consisting of clean background images and their corresponding contaminated counterparts. These pairs are used to train a robust ControlNet model capable of precisely editing low geometry-consistency regions while preserving high geometry-consistency areas.

Real-World Reflection Data in GlassFlash3D. GlassFlash3D consists of 23 sets of real-world objects, each captured in two reflection conditions:

- With Glass Obstruction (HSR Scenario): The object is placed behind a glass barrier, introducing real-world reflection artifacts such as specular highlights, virtual objects, and transparency distortions. Each scene contains 80 HSR images for training.
- Without Glass (Reflection-Free Scenario): The object is imaged under the same conditions, free of glass-induced reflections, serving as a clean reference. Each scene contains 20 clean images and corresponding masks indicating the target objects for testing.

Real-World Dataset from NeuS-HSR. We also used the NeuS-HSR Real-World dataset, but because it contains no true reflection-free images, we performed qualitative analysis on it.



Fig. 4: Examples of the synthetic dataset.

5.2 Baselines and Evaluation Metrics

Baselines. We compare our approach against several representative baselines. **Vanilla 3DGS**[15] directly applies the original 3D Gaussian Splatting framework for scene reconstruction without any reflection-specific processing, serving as a fundamental reference. **DSRNet+3DGS**[14] leverages the single-image reflection removal model DSRNet to preprocess input views, aiming to suppress reflection artifacts at the 2D level before performing 3DGS reconstruction. Similarly, **YTMT+3DGS** employs YTMT[13] as a 2D reflection separation method, followed by 3DGS reconstruction, thereby testing the effectiveness of advanced 2D reflection removal in supporting 3D scene recovery. Finally, **Spot-LessSplats**[32] represents a recent dynamic interference removal strategy within 3DGS, originally designed to ignore transient distractors such as moving objects. Since reflections can be interpreted as semi-transparent dynamic artifacts, we adapt SpotLessSplats to this setting to evaluate its ability to mitigate reflection-induced inconsistencies during reconstruction. It is worth noting that, due to the absence of per-pixel masks for each input image, conventional inpainting-based approaches such as Infusion[21] and RefFusion[24] do not apply to our task.

Evaluation Metrics. On GlassFlash3D, we evaluate reconstruction with PSNR (pixel accuracy), SSIM (structural consistency), and MAE (pixel-wise error). To isolate object fidelity from background reflections, we also report Mask PSNR, Mask SSIM, and Mask MAE computed only within object regions. These

masked metrics ensure improvements reflect object recovery rather than background artifacts. Combined, the global and masked metrics provide a balanced, rigorous evaluation protocol for glass-enclosed scene reconstruction.

5.3 Quantitative and Qualitative Results

Table 1: Quantitative results of different methods on GlassFlash3D Dataset.

Method	PSNR $\uparrow$	SSIM $\uparrow$	MAE $\downarrow$	Mask PSNR $\uparrow$	Mask SSIM $\uparrow$	Mask MAE $\downarrow$
Vanilla 3DGS	15.38	0.62	36.05	17.17	0.65	32.59
DSRNet+3DGS	15.90	0.65	33.30	16.91	0.68	32.67
YTMT+3DGS	15.89	0.64	33.32	17.10	0.66	33.21
SpotLessSplats	15.90	0.66	33.92	17.75	0.69	30.82
Ours	15.90	0.68	31.99	18.94	0.71	24.68

Quantitative results on GlassFlash3D Dataset. As shown in Table 1, our method achieves the best results on all metrics, with particularly large gains in Mask MAE and Mask PSNR. These object-centric improvements demonstrate that our geometry- and concept-aware design faithfully reconstructs target objects while suppressing reflection artifacts. 2D removal baselines (DSRNet+3DGS, YTMT+3DGS) reduce some interference and outperform vanilla 3DGS, but remain well below pipelines tailored for 3D reconstruction, highlighting the need for reflection-aware strategies integrated into the 3DGS optimization. SpotLessSplats performs relatively strongly by treating reflections as dynamic distractors, yet reflective highlights differ from moving objects—being partially transparent and altering texture and illumination—so it cannot fully remove reflection artifacts, resulting in inferior reconstructions compared to GlassGS.

Table 2: Ablation study of several key designs in our method.

Method	PSNR $\uparrow$	SSIM $\uparrow$	MAE $\downarrow$	Mask PSNR $\uparrow$	Mask SSIM $\uparrow$	Mask MAE $\downarrow$
Vanilla 3DGS	15.38	0.62	36.05	17.17	0.65	32.59
Ours wo/ GCP	15.90	0.62	33.44	17.86	0.65	28.22
Ours wo/ RFCL	15.89	0.64	33.88	17.62	0.68	30.31
Ours wo/ Cons in Equation (2)	15.29	0.62	36.73	17.44	0.65	30.15
Ours wo/ Cons in Equation (3)	15.90	0.67	32.03	18.78	0.70	24.95
Ours wo/ $\mathcal{L}_{x_0}$ in Equation (9)	15.75	0.62	34.33	17.70	0.65	28.81
Ours	15.90	0.68	31.99	18.94	0.71	24.68

Ablation study of GCP and RFCL module. As shown in Table 2, removing the geometry-consistency perception module reduces our framework to a 2D reflection-removal strategy and causes a marked performance drop because cross-view geometric consistency is lost. Nonetheless, the iterative refinement

still yields results better than vanilla 3DGS, highlighting the value of iteration even without geometry perception. Removing the reflection-free concept learning module produces only a moderate decline: the system can extract partial semantic cues from rendered images, but without explicit concept guidance, restoration becomes less controllable and more ambiguous, resembling a generic 3D pipeline. Overall, the ablations confirm both components are essential—geometry-aware inconsistency analysis suppresses reflection artifacts, while concept-aware learning provides semantic stability; their synergy is crucial for high-fidelity reconstructions in glass-enclosed scenes.

The effect of iteration. Figure 5 illustrates the progression of our primary evaluation metrics — MAE, SSIM, Mask MAE, and Mask SSIM — across iterations for a representative scene. Significant changes are observed during the initial three iterations, as the Reflection-Free Concept Learning module actively refines the images, a process also depicted in the bottom-left corner of Figure 3. In subsequent iterations, as cross-view consistency is progressively enhanced, the reflection-removal module ceases to modify regions that already exhibit high consistency, resulting in stabilization of the metrics. At this stage, the mean value of $\mathbf{Cons}_i$ stabilizes at 0.92, signifying that both geometric and conceptual consistency have been achieved, thereby ensuring the reliability of the reconstruction.

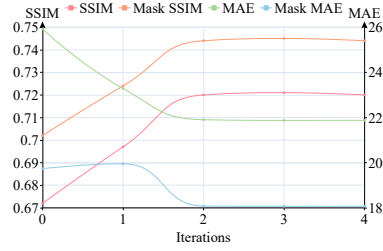


Fig. 5: Iteration convergence curves: MAE, SSIM, Mask MAE, and Mask SSIM across four iterations.

More visualization results. We also evaluated our method on the Real-World scenes from NeuS-HSR; because ground-truth images are unavailable, the comparison is qualitative only. As shown in Figure 6, our approach achieves the best suppression of the blue specular spots. Although SpotLessSplats removes some reflective regions, it does not explicitly enforce cross-view image consistency, which fails to eliminate all reflections and introduces additional noise. Overall, our method outperforms the baselines on these scenes, but some areas remain only partially restored due to a substantial domain gap between these reflections and the data used during our diffusion training.

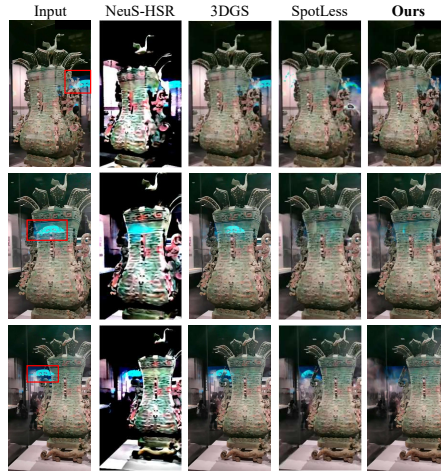


Fig. 6: Results on the Real-world Dataset from NeuS-HSR

Limitations and future work. Our method achieves promising results in removing reflections that interfere with object reconstruction, yet it also has certain limitations. First, as discussed in the Figure 6, the diversity of real-world reflections cannot be fully covered by artificially synthesized data. When encountering reflections beyond the scope of our design, our approach—though superior to existing solutions—still exhibits imperfections. In addition, consistency evaluation in 3DGS remains an open research direction. Although our current strategy can effectively identify reflective regions, some false detections persist, leading to diffusion to repair areas that should not be altered and thereby introducing uncertainty into the reconstruction. In future work, we will further refine the GlassFlash2D Dataset to ensure greater diversity. Moreover, integrating image semantics—by combining geometric consistency with conventional 2D reflection removal techniques—to improve the accuracy of the consistency map represents another promising avenue for exploration.

6 Conclusion

This work addresses the challenge of reconstructing glass-enclosed objects with high specular reflection by introducing GlassGS, a geometry- and concept-aware framework for 3D Gaussian Splatting. By analyzing the properties of high specular reflections, we demonstrate that concept-guided learning incorporated with geometry information provides a principled solution. Our iterative two-stage optimization method effectively mitigates reflection artifacts while maintaining semantic integrity, resulting in geometrically consistent and perceptually accurate reconstructions. Coupled with the dataset specifically designed for glass-enclosed scenes, GlassGS establishes a new standard for reflection-aware 3D reconstruction, paving the way for realistic digitization applications in cultural heritage preservation, augmented reality (AR), virtual reality (VR), and other domains.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant No. 62225604, the Shenzhen Science and Technology Program under Grant No. JCYJ20240813114237048, and the Fundamental Research Funds for the Central Universities (No. XXX-63263254). We also acknowledge the support from our collaborating institutions and thank all colleagues for their valuable discussions and assistance during this work. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

Bibliography

- [1] Anciukevičius, T., Xu, Z., Fisher, M., Henderson, P., Bilen, H., Mitra, N.J., Guerrero, P.: Renderdiffusion: Image diffusion for 3d reconstruction, inpainting and generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12608–12618 (2023)
- [2] Antsfeld, L., Chidlovskii, B.: 3d-consistent image inpainting with diffusion models. arXiv preprint arXiv:2412.05881 (2024)
- [3] Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5855–5864 (2021)
- [4] Chou, Z.T., Huang, S.Y., Liu, I., Wang, Y.C.F., et al.: Gsnerf: Generalizable semantic neural radiance fields with enhanced 3d scene understanding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20806–20815 (2024)
- [5] Dai, Y., Li, C., Zhou, S., Feng, R., Luo, Y., Loy, C.C.: Flare7k++: Mixing synthetic and real datasets for nighttime flare removal and beyond. arXiv preprint arXiv:2306.04236 (2023)
- [6] Fan, C.D., Chang, C.W., Liu, Y.R., Lee, J.Y., Huang, J.L., Tseng, Y.C., Liu, Y.L.: Spectromotion: Dynamic 3d reconstruction of specular scenes. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 21328–21338 (2025)
- [7] Fan, Q., Yang, J., Hua, G., Chen, B., Wipf, D.: A generic deep architecture for single image reflection removal and image smoothing. In: Int. Conf. Comput. Vis. pp. 3238–3247 (2017)
- [8] Fu, Y., Ding, J., Wang, R., Fu, Q., Tsang, I.W., Cheng, M.M., Guo, Q.: Benchmarking drag★ for eye direction transformation and beyond. Visual Intelligence **3**(1), 29 (2025)
- [9] Gao, W., Jin, L., Akoudad, Y., Yang, Y.: Aps-neus: Adaptive planar and skip-sampling for 3d object surface reconstruction in high-specular scenes. Image and Vision Computing p. 105665 (2025)
- [10] Guo, Y.C., Kang, D., Bao, L., He, Y., Zhang, S.H.: Nerfren: Neural radiance fields with reflections. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18409–18418 (2022)
- [11] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Adv. Neural Inform. Process. Syst. **33**, 6840–6851 (2020)
- [12] Hong, Y., Zhong, H., Weng, S., Liang, J., Shi, B.: L-differ: Single image reflection removal with language-based diffusion model. In: European Conference on Computer Vision. pp. 58–76. Springer (2024)
- [13] Hu, Q., Guo, X.: Trash or treasure? an interactive dual-stream strategy for single image reflection separation. Advances in Neural Information Processing Systems **34**, 24683–24694 (2021)

- [14] Hu, Q., Guo, X.: Single image reflection separation via component synergy. In: *Int. Conf. Comput. Vis.* pp. 13138–13147 (2023)
- [15] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139–1 (2023)
- [16] Kotp, Y., Torki, M.: Flare-free vision: Empowering uformer with depth insights. In: *ICASSP 2024*. pp. 2565–2569. IEEE (2024)
- [17] Li, B., Cai, Z., Li, Y.F., Reid, I., Rezatofighi, H.: Hier-slam: Scaling-up semantics in slam with a hierarchically categorical gaussian splatting. In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 9748–9754. IEEE (2025)
- [18] Liao, G., Zhou, K., Bao, Z., Liu, K., Li, Q.: Ov-nerf: Open-vocabulary neural radiance fields with vision and language foundation models for 3d semantic understanding. *IEEE Transactions on Circuits and Systems for Video Technology* (2024)
- [19] Liu, J., Liu, R., Chen, K., Zhang, J., Guo, D.: Collaborative visual inertial slam for multiple smart phones. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 11553–11559 (2021). <https://doi.org/10.1109/ICRA48506.2021.9561946>
- [20] Liu, J., Tang, X., Cheng, F., Yang, R., Li, Z., Liu, J., Huang, Y., Lin, J., Liu, S., Wu, X., et al.: Mirrorgaussian: Reflecting 3d gaussians for reconstructing mirror reflections. In: *European Conference on Computer Vision*. pp. 377–393. Springer (2024)
- [21] Liu, Z., Ouyang, H., Wang, Q., Cheng, K.L., Xiao, J., Zhu, K., Xue, N., Liu, Y., Shen, Y., Cao, Y.: Infusion: Inpainting 3d gaussians via learning depth completion from diffusion prior. *arXiv preprint arXiv:2404.11613* (2024)
- [22] Mao, S., Wu, C., Shen, Z., Wang, Y., Wu, D., Zhang, L.: Neus-pir: Learning relightable neural surface using pre-integrated rendering. *Computational Visual Media* (2025)
- [23] 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)
- [24] Mirzaei, A., De Lutio, R., Kim, S.W., Acuna, D., Kelly, J., Fidler, S., Gilitschenski, I., Gojcic, Z.: Reffusion: Reference adapted diffusion models for 3d scene inpainting. *arXiv preprint arXiv:2404.10765* (2024)
- [25] Nathan Silberman, Derek Hoiem, P.K., Fergus, R.: Indoor segmentation and support inference from rgb-d images. In: *ECCV* (2012)
- [26] Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S., Geiger, A., Radwan, N.: Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 5480–5490 (2022)
- [27] Peng, S., Genova, K., Jiang, C., Tagliasacchi, A., Pollefeys, M., Funkhouser, T., et al.: Openscene: 3d scene understanding with open vocabularies. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 815–824 (2023)

- [28] Peng, Z., Zhou, K., Shao, T.: Gaussian-plus-sdf slam: High-fidelity 3d reconstruction at 150+ fps. *Computational Visual Media* (2025)
- [29] Qiu, J., Jiang, P.T., Zhu, Y., Yin, Z.X., Cheng, M.M., Ren, B.: Looking through the glass: Neural surface reconstruction against high specular reflections. In: *IEEE Conf. Comput. Vis. Pattern Recog.* pp. 20823–20833 (2023)
- [30] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Aspell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: *Inter. conf. on machine learning*. pp. 8748–8763. *PmLR* (2021)
- [31] Rosh, G., Prasad, B.P., Boregowda, L.R., Mitra, K.: Deep unsupervised reflection removal using diffusion models. In: *2023 IEEE International Conference on Image Processing (ICIP)*. pp. 2045–2049. *IEEE* (2023)
- [32] Sabour, S., Goli, L., Kopanas, G., Matthews, M., Lagun, D., Guibas, L., Jacobson, A., Fleet, D., Tagliasacchi, A.: Spotlessplats: Ignoring distractors in 3d gaussian splatting. *ACM Transactions on Graphics* **44**(2), 1–11 (2025)
- [33] Skorokhodov, V., Durasov, N., Fua, P.: D3dr: Lighting-aware object insertion in gaussian splatting. *arXiv preprint arXiv:2503.06740* (2025)
- [34] Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020)
- [35] Song, J., Ye, Z., Zhou, Q., Yang, W., Fei, B., Xu, J., He, Y., Ouyang, W.: Reflections unlock: Geometry-aware reflection disentanglement in 3d gaussian splatting for photorealistic scenes rendering. *arXiv preprint arXiv:2507.06103* (2025)
- [36] Sun, J.M., Wu, T., Gao, L.: Recent advances in implicit representation-based 3d shape generation. *Visual Intelligence* **2**(1), 9 (2024)
- [37] 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 Computer Vision and Pattern Recognition Conference*. pp. 16133–16142 (2025)
- [38] Vora, S., Radwan, N., Greff, K., Meyer, H., Genova, K., Sajjadi, M.S., Pot, E., Tagliasacchi, A., Duckworth, D.: Nesf: Neural semantic fields for generalizable semantic segmentation of 3d scenes (2022)
- [39] Wang, C., Chai, M., He, M., Chen, D., Liao, J.: Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 3835–3844 (2022)
- [40] Wang, C., Peng, H.Y., Liu, Y.T., Gu, J., Hu, S.M.: Diffusion models for 3d generation: A survey. *Computational Visual Media* **11**(1), 1–28 (2025)
- [41] Wang, J., Fang, J., Zhang, X., Xie, L., Tian, Q.: Gaussianeditor: Editing 3d gaussians delicately with text instructions. In: *IEEE Conf. Comput. Vis. Pattern Recog.* pp. 20902–20911 (2024)
- [42] Wang, T., Lu, W., Zhang, K., Luo, W., Kim, T.K., Lu, T., Li, H., Yang, M.H.: Promptrr: Diffusion models as prompt generators for single image reflection removal. *arXiv preprint arXiv:2402.02374* (2024)

- [43] Wang, T., Xie, M., Cai, H., Shah, S., Metzler, C.A.: Flash-split: 2d reflection removal with flash cues and latent diffusion separation. arXiv preprint arXiv:2501.00637 (2024)
- [44] Wang, X., Zhang, Y., Chen, Y.: Polarized reflection removal with dual-stream attention guidance. *Pattern Recognition* **157**, 110945 (2025)
- [45] Wu, T., Yuan, Y.J., Zhang, L.X., Yang, J., Cao, Y.P., Yan, L.Q., Gao, L.: Recent advances in 3d gaussian splatting. *Computational Visual Media* **10**(4), 613–642 (2024)
- [46] Xie, M., Cai, H., Shah, S., Xu, Y., Feng, B.Y., Huang, J.B., Metzler, C.A.: Flash-splat: 3d reflection removal with flash cues and gaussian splats. In: *European Conference on Computer Vision*. pp. 122–139. Springer (2024)
- [47] Xu, J., Zhu, Z., Bao, H., Xu, W.: Hybrid mesh-neural representation for 3d transparent object reconstruction. *Computational Visual Media* **11**(1), 123–140 (2025)
- [48] Ye, K., Hou, Q., Zhou, K.: 3d gaussian splatting with deferred reflection. In: *ACM SIGGRAPH 2024 Conference Papers*. pp. 1–10 (2024)
- [49] Zhang, R., Luo, T., Yang, W., Fei, B., Xu, J., Zhou, Q., Liu, K., He, Y.: Re-fgaussian: Disentangling reflections from 3d gaussian splatting for realistic rendering. arXiv preprint arXiv:2406.05852 (2024)
- [50] Zhi, S., Laidlow, T., Leutenegger, S., Davison, A.J.: In-place scene labelling and understanding with implicit scene representation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 15838–15847 (2021)
- [51] Zhong, H., Hong, Y., Weng, S., Liang, J., Shi, B.: Language-guided image reflection separation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 24913–24922 (2024)
- [52] Zhou, S., Chang, H., Jiang, S., Fan, Z., Zhu, Z., Xu, D., Chari, P., You, S., Wang, Z., Kadambi, A.: Feature 3dgs: Supercharging 3d gaussian splatting to enable distilled feature fields. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 21676–21685 (2024)
- [53] Zhu, S., Wang, G., Blum, H., Liu, J., Song, L., Pollefeys, M., Wang, H.: Sni-slam: Semantic neural implicit slam. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 21167–21177 (2024)