4DREFGS: REFLECTIVE 4D GAUSSIAN SPLATTING FOR MULTI-VIEW DYNAMIC SCENE RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Reconstructing dynamic reflective scenes from multi-view videos is challenging because physical motion and reflected appearance are strongly entangled. Our observation is that surface-attached and localized virtual reflections exhibit different dynamic behaviors: the former remain tied to physical surface, while the latter may exhibit independent apparent motion. Modeling both with the same dynamic primitives can therefore entangle geometry, motion and reflected appearance. We introduce 4DRefGS, a role-aware framework for reconstructing dynamic reflective scenes. First, to disentangle physical surface motion from reflection-induced apparent motion, we employ two asymmetric 4D Gaussian fields: primary planar Gaussians reconstruct the physical surface, whereas auxiliary Gaussians independently follow localized virtual-reflection cues without contributing to geometry or diffuse appearance. Second, we model reflective appearance through a reflection-aware deferred shading framework, where scene-shared directional lighting and time-varying surface-attached features predict the global reflection, while auxiliary features provide spatially localized corrections. To facilitate evaluation, we further collect SelfCap, a synchronized multi-view dataset featuring fast motion and substantial time-varying reflections. Experiments on SelfCap and Neural3DV demonstrate that 4DRefGS faithfully reconstructs dynamic reflective appearance, substantially improving rendering quality on challenging reflective scenes while remaining competitive on general dynamic scenes.

1 INTRODUCTION

High-quality reconstruction and photorealistic rendering of dynamic scenes are important for applications such as free-viewpoint video, AR/VR, and 4D content creation. Despite recent progress in Neural Radiance Fields (Mildenhall et al., 2020) and Gaussian Splatting (Kerbl et al., 2023), reconstructing dynamic reflective scenes remains particularly challenging. In such scenes, observed image changes arise not only from physical object motion, but also from view-dependent reflections and changing reflected content. Specular highlights and reflected virtual images may exhibit apparent motion that differs substantially from the motion of the underlying surface. As a result, a reconstruction model may incorrectly explain reflection changes through geometric deformation or time-varying surface appearance, entangling physical motion with reflection.

Existing dynamic reconstruction methods (Wang et al., 2025; Li et al., 2024; Liao et al., 2026) mainly focus on modeling scene motion, while representing view-dependent appearance with spherical harmonics or learned per-Gaussian features. Although effective for general dynamic scenes, such representations are often insufficient for sharp specular reflections and structured reflected content. Conversely, reflective-scene reconstruction methods (Tang & Cham, 2024; Jiang et al., 2024; Fan et al., 2026) explicitly model surface materials and illumination, but most assume static geometry or slowly varying appearance. Recent methods (Fan et al., 2025; 2026) begin to address dynamic reflective scenes, yet typically rely on a single surface representation or a globally shared illumination model, making it difficult to simultaneously preserve physical geometry and capture localized virtual reflections under rapid motion. This raises a key question: how can we model reflection dynamics without allowing reflected appearance to corrupt the reconstructed physical surface?

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Our insight is that surface-attached and localized virtual reflections obey different dynamics. Surface-attached reflections remain coupled to the physical surface and should therefore share its geometry and motion, whereas localized virtual reflections may exhibit independent apparent motion. Representing both with the same dynamic primitives can consequently entangle physical motion and reflection. Based on this observation, we introduce 4DRefGS, a role-aware dual 4D Gaussian representation for reconstructing dynamic reflective scenes.

We design 4DRefGS around two dynamic Gaussian sets with asymmetric roles. Primary planar Gaussians establish the physical surface and encode geometry, diffuse albedo, roughness, and global reflection features, while auxiliary Gaussians are independently optimized to follow localized reflected content, but contribute only reflection features, preventing them from altering geometry or diffuse appearance. This constrained decomposition encourages the primary representation to preserve a coherent physical surface, while reserving auxiliary capacity for near-field reflections and virtual-image structures. Separating the two representations alone, however, is not sufficient to model reflective appearance. We further develop a reflection-aware deferred shading framework that further decomposes reflection according to its spatial structure. A scene-shared directional representation provides common illumination context, which is combined with time-varying surface-attached features from the primary Gaussians to predict the global reflection. Auxiliary Gaussian features then provide spatially localized corrections for near-field and virtual-image effects that cannot be explained by the shared directional component. Temporal appearance capacity is further adapted to rapid motion and reflection changes while preserving a stable physical-surface representation.

To evaluate dynamic reflective reconstruction under the targeted conditions, we collect SelfCap, a synchronized multi-view dataset featuring objects with diverse materials, rapid motion, and substantial illumination changes. Experiments on SelfCap and public Neural3DV (Li et al., 2022) show that 4DRefGS faithfully reconstructs dynamic reflective appearance and substantially improves novel-view rendering in challenging reflective scenes. Our main contributions are summarized as follows:

- We introduce a role-aware dual 4D Gaussian representation that separates physical-surface motion from localized virtual-reflection motion through asymmetric Gaussian roles, preventing flexible reflection modeling from corrupting the reconstructed geometry.
- We develop a reflection-aware global–local shading formulation that combines scene-shared directional information, time-varying surface-attached reflection, and localized auxiliary corrections for structured modeling of dynamic reflective appearance.
- We collect SelfCap, a multi-view dataset with rapid reflective objects and substantial illumination changes, and demonstrate clear improvements on challenging reflective scenes.

2 Related Work

Dynamic Scene Reconstruction. Dynamic scene reconstruction aims to recover scene geometry and temporal motion from RGB observations. Existing NeRF-based (Mildenhall et al., 2021) methods typically learn deformation fields that warp observations to a canonical radiance field, condition radiance fields on time (Sun et al., 2022; Chen et al., 2021; Park et al., 2021; Song et al., 2023; Aliev et al., 2020), or use factorized spatio-temporal structures to improve efficiency (Cao & Johnson, 2023). Despite high reconstruction quality, these approaches often require expensive optimization and dense neural rendering, making real-time high-resolution dynamic rendering difficult.

Recent dynamic Gaussian methods improve efficiency by extending 3D Gaussian Splatting (Kerbl et al., 2023) to time-varying scenes. Some methods deform canonical Gaussians with learned motion fields (Yang et al., 2024b; Bae et al., 2024; Qingming et al., 2025; Guo et al., 2024; Lu et al., 2024; Shaw et al., 2024; Zhu et al., 2024; Liang et al., 2025; Xu et al., 2024a; Kim et al., 2024), while others represent dynamics with 4D Gaussians, spatio-temporal features, sparse control points, or streaming updates (Li et al., 2024; Yang et al., 2024a; Lee et al., 2024; Luiten et al., 2024; Xu et al., 2024b; Wang et al., 2025; Duan et al., 2024; Gao et al., 2025). These methods enable efficient dynamic novel-view synthesis, but are primarily designed to model scene motion and and generic appearance variation rather than the distinct dynamics of reflected content.

Reflective Scene Reconstruction and Rendering. Reflective appearance depends jointly on surface geometry, material properties, illumination, and viewing direction, making accurate recon-

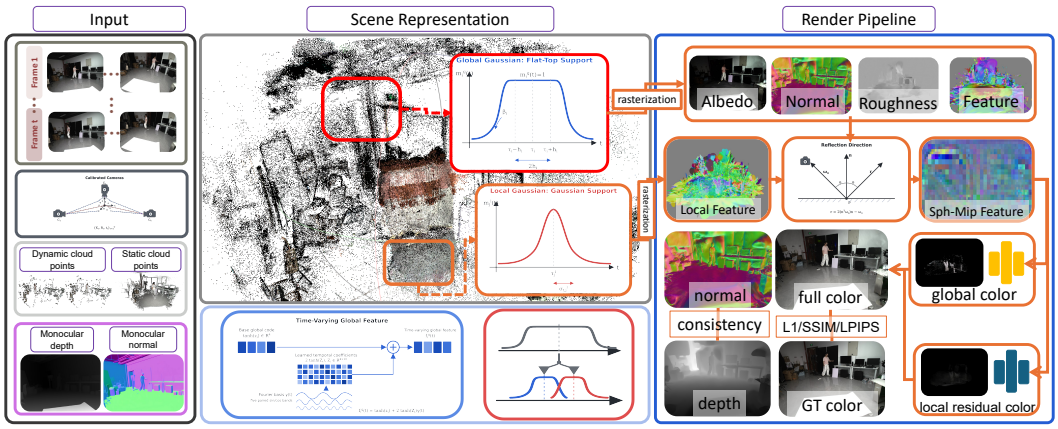

Figure 1: Overview of 4DRefGS. We separate the degrees of freedom used to represent the physical surface from those used to explain reflected appearance. Primary Gaussians form the geometry-bearing dynamic surface, while auxiliary Gaussians provide additional appearance capacity only through the reflection pathway. Both fields are coupled by a scene-shared directional representation, and temporal surface support is separated from reflected-appearance dynamics so that the surface can remain stable while its reflection evolves.

struction and novel-view rendering challenging. Recent Gaussian-based methods improve reflective rendering through illumination and material modeling, deferred shading, or directional factorization (Tang & Cham, 2024; Jiang et al., 2024; Keyang et al., 2024; Zhang et al., 2025). More recent works further explore residual reflection modeling (Kouros et al., 2026), diffuse-specular separation (Fan & Wang, 2026), and joint reflection-transmission modeling (Shi et al., 2026). While these methods substantially improve reflective appearance modeling, they are formulated for static-scene reconstruction and do not address time-varying surface motion or reflection dynamics.

Among the most relevant works, Ref-DGS (Fan et al., 2026) spatially decouples geometry from near-field reflections using complementary geometry and reflection Gaussians, but is restricted to static scenes and therefore does not model reflection motion over time. SpectroMotion (Fan et al., 2025), in contrast, addresses dynamic specular scenes through surface deformation and a time-varying environment representation, but does not provide independent motion degrees of freedom for localized reflected structures. Existing methods therefore address either spatial reflection decomposition or dynamic specular appearance, but not their joint spatiotemporal disentanglement. Our method bridges this gap by allowing localized virtual reflections to evolve independently of the underlying physical surface.

3 Method

3.1 Spatiotemporal Structure of Dynamic Reflections

Dynamic reflective reconstruction is challenging because the physical surface and its reflected appearance need not obey the same spatiotemporal constraints. A surface follows the geometry and motion of the underlying object, whereas its reflected appearance may change substantially with viewpoint, surface orientation, and illumination. Localized reflected structures can further slide across a surface, appear or disappear, or exhibit apparent motion different from the physical motion itself. Consequently, under image reconstruction supervision, a homogeneous dynamic representation may explain reflection changes by modifying surface motion, temporal visibility, or appearance, even when the underlying geometry should remain unchanged. The problem is therefore not simply insufficient representation capacity, but how that capacity is allocated: *appearance flexibility should not automatically introduce additional freedom in geometry or temporal surface support*.

Reflected appearance itself also requires different forms of spatial capacity. Direction-dependent effects contain structure that can be shared across the scene, whereas sharp near-field or virtual-image-like structures may require additional localized flexibility. Encoding all such variation inde-

pendently at each surface primitive discards useful shared structure, while forcing it entirely into a scene-shared representation cannot adequately capture localized effects.

These observations motivate 4DRefGS to separate the spatial and temporal degrees of freedom used for reflected appearance from those required by the physical surface. We use an asymmetric dual-4D-Gaussian representation to isolate geometry-bearing and reflection-only capacity, model shared directional structure with auxiliary residual flexibility, and decouple persistent surface support from time-varying reflected appearance.

3.2 Asymmetric Dual 4D Gaussian Representation

We represent the scene with a primary Gaussian field $\mathcal{G}^p(t)$ and an auxiliary Gaussian field $\mathcal{G}^a(t)$. Both use the linear trajectory model of FreeTimeGS (Wang et al., 2025). For Gaussian i, its center evolves as $\boldsymbol{\mu}_i(t) = \boldsymbol{\mu}_i + \mathbf{v}_i(t - \tau_i)$, where $\boldsymbol{\mu}_i$, $\mathbf{v}_i$, and τ_i denote its reference position, velocity, and temporal center.

The two fields therefore share the same basic kinematics but obey different representational constraints: only the primary field may define the reconstructed physical surface, whereas the auxiliary field can affect the image only through the reflection pathway.

Primary surface field. Primary Gaussians are the only geometry-bearing primitives. Following the planar parameterization of PGSR (Chen et al., 2024), each primary Gaussian represents a locally planar surface element and its normal is obtained from the rotated smallest-scale axis. In addition to its geometric parameters, a primary Gaussian stores base color, roughness, and a compact surface-attached reflection feature.

The primary and auxiliary rasterizers produce

$$\begin{aligned}(\mathbf{c}_{\text{base}}, \mathbf{f}_s, \mathbf{n}, D, \rho)(p,t) &= \mathcal{R}_p(\mathcal{G}^p(t)),\\ \mathbf{f}_a(p,t) &= \mathcal{R}_a(\mathcal{G}^a(t)).\end{aligned} \tag{1}$$

Here $\mathbf{c}_{\text{base}}$ denotes learned base color, $\mathbf{f}_s$ the surface-attached reflection feature, $\mathbf{n}$ the surface normal, D plane depth, and ρ roughness.

We use the term *base color* rather than assuming a uniquely identifiable intrinsic albedo. Our goal is not to recover a unique physical decomposition of image formation from RGB supervision, but to explicitly control which representation is allowed to modify the reconstructed surface.

Auxiliary reflection field. Auxiliary Gaussians carry only compact reflection features and are rasterized independently. They have no direct rendering path to depth, normals, roughness, or base color. Additional appearance capacity can therefore explain reflection variation without directly becoming a second geometry representation.

Crucially, auxiliary Gaussians still retain their own 4D trajectories. They can therefore follow reflected structures whose apparent motion differs from that of the underlying surface, while their restricted rendering pathway prevents such motion from being interpreted directly as physical deformation.

The contribution is thus not merely the presence of two Gaussian sets. It is the asymmetric assignment of representational freedom: the primary field owns the physical surface, while the auxiliary field is granted independent spatiotemporal appearance capacity without geometry ownership. The complete Gaussian, opacity, and initialization parameterization is given in the supplementary material.

3.3 Shared Directional Reflection with Auxiliary Residual Capacity

The two fields provide different spatial capacity, but they should not form two unrelated appearance models. We first explain what can be shared through the physical surface and a common directional context, and introduce auxiliary capacity only for spatial variation that is difficult to represent in this shared form.

For surface normal $\mathbf{n}$ and viewing direction $\boldsymbol{\omega}_o$, we compute the reflected direction $\mathbf{r} = 2(\mathbf{n}^\mathsf{T}\boldsymbol{\omega}_o)\mathbf{n} - \boldsymbol{\omega}_o$. Following Ref-GS (Zhang et al., 2025), a learnable spherical Mip representation provides a scene-shared directional feature $\mathbf{e} = E_{\mathrm{sph}}(\mathbf{r}, \rho)$.

The surface-conditioned and auxiliary-conditioned components are

$$\begin{aligned} \mathbf{u}_s &= \phi_s([\mathbf{e}; \mathrm{vec}(\mathbf{e} \otimes \mathbf{f}_s); \chi]) \,, \\ \mathbf{u}_a &= \phi_a([\mathbf{e}; \mathbf{f}_a; \rho; \chi]) \,, \end{aligned} \tag{2}$$

where χ is a stop-gradient angular cue detailed in the supplementary material.

The first branch combines scene-shared directional information with features attached to the geometry-bearing surface. The second branch receives the same directional context but obtains additional spatial conditioning from the independently moving auxiliary field. This shared conditioning is intentional: the auxiliary branch is not an independent appearance field, but a spatially flexible correction expressed under the same directional context as the surface-conditioned prediction.

We combine both terms in log-radiance space,

$$\mathbf{L}_{\mathrm{ref}} = \exp(\mathbf{u}_s + \mathbf{u}_a), \qquad \mathbf{I}_{\mathrm{sRGB}} = \mathrm{lin2sRGB}(\mathbf{c}_{\mathrm{base}} + \mathbf{L}_{\mathrm{ref}}). \tag{3}$$

This formulation is a structured appearance factorization rather than a claim of uniquely recovering physical diffuse and specular transport. The surface-conditioned branch provides the shared explanation of view-dependent appearance, while the auxiliary branch supplies additional residual capacity where greater spatial flexibility is needed.

Since the auxiliary pathway is more flexible, we discourage unnecessary residual corrections using $\mathcal{L}_{\mathrm{aux}} = \mathbb{E}_{p,d}[|u_{a,d}(p,t)|]$. This implements a preference for shared directional explanations before allocating additional auxiliary capacity.

3.4 Decoupling Surface Support from Reflection Dynamics

The previous sections determine *where* reflected-appearance capacity resides. Dynamic reconstruction additionally requires deciding *what is allowed to change over time*. A physical carrier may remain valid for many frames while its reflected appearance changes rapidly. Using the same temporal degree of freedom for both quantities would unnecessarily turn appearance variation into surface-visibility variation.

We therefore allocate temporal flexibility conservatively: we first keep the physical carrier stable, then vary reflected appearance while preserving that carrier, and only increase the primitive representation itself when the existing appearance capacity becomes insufficient.

Stable surface support. For primary Gaussians, we adopt the flat-top temporal support of SharpTimeGS (Liao et al., 2026):

$$m_i^p(t) = \exp\left[-\frac{1}{2}\left(\frac{\max(|t-\tau_i| - h_i, 0)}{\beta_i}\right)^2\right], \tag{4}$$

where h_i controls the plateau width and β_i the differentiable boundary. Inside $|t-\tau_i| \le h_i$, the gate remains constant. The same primary primitive can therefore maintain stable support throughout the interval in which it represents the same physical surface.

Auxiliary Gaussians instead retain Gaussian temporal support, $m_j^a(t) = \exp[-\frac{1}{2}((t-\tau_j^a)/\sigma_{t,j}^a)^2]$, allowing their reflection capacity to be more temporally localized. The corresponding opacity parameterization and alpha-compositing analysis are provided in the supplementary material.

Time-varying reflection on stable geometry. Stable temporal support should not imply constant reflected appearance. We therefore keep the directional representation shared across time and place continuous temporal variation in the compact reflection feature attached to each primary Gaussian:

$$\mathbf{f}_i^s(t) = \tanh(\mathbf{z}_i) + 2\tanh(\mathbf{Z}_i)\boldsymbol{\gamma}(t), \tag{5}$$

where $\mathbf{z}_i$ is a base reflection code, $\mathbf{Z}_i$ contains temporal coefficients, and $\boldsymbol{\gamma}(t)$ is a Fourier time encoding.

The temporal gate and reflection feature consequently control different quantities: $m_i^p(t)$ determines *whether the physical carrier is present*, whereas $\mathbf{f}_i^s(t)$ determines *how its reflected appearance changes while that carrier remains present*. This separation allows substantial appearance dynamics without introducing corresponding changes in surface occupancy.

Trajectory-preserving temporal bifurcation. Continuous feature modulation may still be insufficient for a long-lived primitive whose observations require substantially different temporal states. Rather than assigning unrestricted temporal complexity to every Gaussian, we increase primitive-level capacity only where optimization indicates persistent temporal conflict.

For each visible primary Gaussian, we compute

$$G_i^\tau = \frac{1}{|\mathcal{O}_i|} \sum_{k \in \mathcal{O}_i} \left| \frac{\partial \mathcal{L}_k}{\partial \tau_i} \right|, \tag{6}$$

where $\mathcal{O}_i$ denotes observations in which Gaussian i contributes. The absolute gradient prevents conflicting temporal updates from cancelling.

When G_i^τ exceeds a fixed threshold, the primitive is divided into two temporally specialized children. Because the temporal center also anchors the linear trajectory, naively changing it would change the represented motion. We therefore initialize each child along the parent trajectory:

$$\boldsymbol{\mu}_{i,j} = \boldsymbol{\mu}_i + \mathbf{v}_i(\tau_{i,j} - \tau_i), \qquad \mathbf{v}_{i,j} = \mathbf{v}_i. \tag{7}$$

This preserves the parent motion at initialization while giving the children independent temporal capacity. The full split parameterization and trajectory-preservation proof are provided in the supplementary material.

Together, these mechanisms establish a hierarchy of temporal flexibility: *preserve the physical carrier, vary its reflected appearance, and expand the representation only when necessary*.

Progressive optimization. Flexible appearance components can otherwise absorb reconstruction errors before the physical surface is sufficiently constrained. We therefore first optimize primary motion, planar geometry, and base appearance; then activate the shared directional and auxiliary reflection components; and finally enable time-varying reflection features and temporal bifurcation.

The overall objective is

$$\mathcal{L} = \mathcal{L}_{\text{img}} + \mathcal{L}_{\text{geom}} + \lambda_{\text{aux}}\mathcal{L}_{\text{aux}} + \lambda_{\text{ent}}\mathcal{L}_{\text{ent}} + \lambda_{\text{edge}}\mathcal{L}_{\text{edge}}. \tag{8}$$

Here $\mathcal{L}_{\text{img}}$ provides image reconstruction supervision, $\mathcal{L}_{\text{geom}}$ constrains the primary surface, and the remaining terms regulate auxiliary usage and temporal support. Detailed loss definitions, training stages, architectures, and optimization hyperparameters are provided in the supplementary material.

4 Experiments

4.1 Experimental Settings

Datasets. We conduct our experiments using two datasets: Neural3DV (Li et al., 2022), and our self-collected SelfCap dataset. We selected four sequences from the Neural3DV (Li et al., 2022) dataset that feature object motion and specular highlights.

Compared with existing datasets, our dataset encompasses a diverse set of challenging dynamic scenarios, including moving objects with materials ranging from matte to highly specular, as well as appearance changes induced by dynamic illumination, such as moving shadows, time-varying reflections, and variations in environmental lighting.

The wrestling subset comprises two sequences featuring adult and child participants, captured using 60 cameras arranged along a 180° arc around the scene center. The indoor subset contains three sequences featuring a person walking on a reflective floor and reflective objects in motion, captured using 16 cameras spanning 60° horizontally and 45° vertically. For each scene, we hold out one camera view for testing.

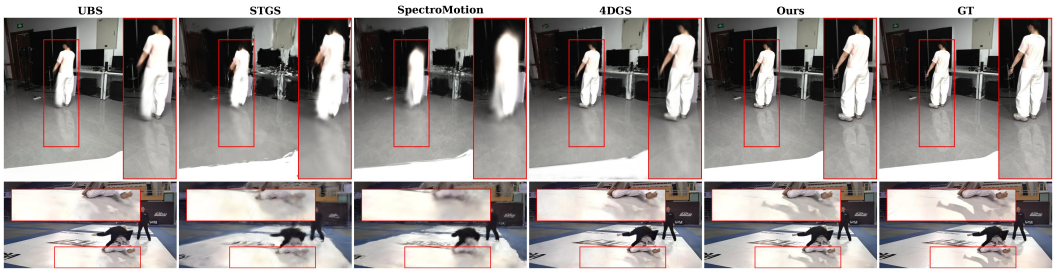

Figure 2: Qualitative comparison results on the SelfCap datasets under the novel-view. Compared with representative dynamic 4D reconstruction methods, our approach better preserves view-dependent reflective appearance under object motion and changing illumination.

Table 1: Average quantitative comparison on four Neural3DV (Li et al., 2022) scenes: *coffee*, *cut beef*, *sear steak*, and *cook spinach*.

Method	PSNR↑	SSIM↑	LPIPS↓
4DGS (Yang et al., 2024a)	31.15	**0.9482**	0.1506
STGS (Li et al., 2024)	30.94	0.9420	0.1825
UBS (Liu et al., 2025)	31.14	0.9407	0.1610
SpectroMotion (Fan et al., 2025)	30.79	0.9390	0.1888
Ours	**31.51**	0.9403	**0.1035**

All videos are captured at a resolution of 4096×3000 pixels and 30 FPS, with each sequence consisting of 60 frames. For processing and experiments, all images are downsampled using a spatial scaling factor of 0.3.

Evaluation Metrics. We evaluate rendering quality using three widely adopted image-based metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) (Wang et al., 2004), and Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al., 2018). PSNR and SSIM measure pixel-level fidelity and structural consistency, respectively, while LPIPS evaluates perceptual similarity using deep feature representations.

Baselines. We compare 4DRefGS with four representative methods for dynamic scene reconstruction: 4DGS (Yang et al., 2024a), STGS (Li et al., 2024), SpectroMotion (Fan et al., 2025), and UBS (Liu et al., 2025).

Implementation Details. All experiments were conducted on a single NVIDIA RTX 5090 GPU. For each 60-frame indoor sequence, we optimized our method for 30,000 iterations, requiring approximately 2 hours. For the wrestling sequence, which contains rapid motion, we increased the training schedule to 60,000 iterations, requiring approximately 3 hours.

4.2 Evaluation

Qualitative Evaluation. Qualitative comparisons in Fig. 2 show that our method reconstructs dynamic reflective scenes with high visual fidelity. Compared with deformation-based dynamic Gaussian methods, our approach better preserves view-dependent reflective appearance under object motion and changing illumination. In particular, the proposed reflection prediction module improves the modeling of glossy and mirror-like regions, producing sharper reflected details and more temporally consistent appearance.

4DGS (Yang et al., 2024a) uses spherical harmonics to model view-dependent appearance. In our experiments, it reconstructs predominantly diffuse surfaces well but struggles with glossy and mirror-like regions. In the wrestling subset, for example, the shadow near the leg varies in appearance and shifts slightly across neighboring viewpoints, yet is missing from the rendered test view.

Table 2: Quantitative comparison on our self-captured dynamic reflective dataset. The dataset contains fast-moving reflective objects and large environment illumination changes.

Method	PSNR↑	SSIM↑	LPIPS↓
4DGS (Yang et al., 2024a)	23.52	0.9000	0.1265
STGS (Li et al., 2024)	22.13	0.8780	0.1920
UBS (Liu et al., 2025)	26.50	0.9140	0.1152
SpectroMotion (Fan et al., 2025)	20.06	0.7090	0.2610
Ours	**27.68**	**0.9330**	**0.1128**

STGS (Li et al., 2024) models Gaussian motion using polynomial trajectories and renders appearance through feature splatting. In our experiments, it struggles to reconstruct rapidly moving regions, such as the leg in the wrestling subset and the hand in the indoor subset. It also fails to faithfully recover the leg shadow and the reflected virtual image on the indoor floor. Its feature-based appearance representation does not explicitly incorporate reflection-direction conditioning, in contrast to our reflection-aware shading formulation.

UBS (Liu et al., 2025) employs view-dependent opacity, allowing individual primitives to contribute selectively across viewpoints. Although this flexibility enables it to capture high-frequency view-dependent appearance, the leg shadow remains poorly reconstructed in the held-out wrestling view. On the indoor subset, however, its visual quality is close to that of our method.

Quantitative Evaluation. We evaluate our method on both the Neural3DV (Li et al., 2022) dataset and our self-captured dynamic reflective dataset. For Neural3DV, we use four scenes: *coffee*, *cut beef*, *sear steak*, and *cook spinach*. We report PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018). The average results on Neural3DV are shown in Tab. 1, and the results on our self-captured dataset are shown in Tab. 2.

On Neural3DV, our method achieves competitive reconstruction quality compared with existing dynamic reconstruction methods. Although Neural3DV mainly contains non-reflective or moderately view-dependent scenes and does not fully match our target setting, our method maintains comparable PSNR and SSIM while outperforming STGS and UBS in LPIPS. This indicates that the proposed reflection-aware representation does not sacrifice general dynamic scene reconstruction quality.

On our self-captured dynamic reflective dataset, our method achieves the best performance across all metrics, with an average PSNR of 27.6803, SSIM of 0.9330, and LPIPS of 0.1128. Compared with the strongest baseline, UBS, our method improves PSNR by 1.1704 dB and reduces LPIPS from 0.1152 to 0.1128. The improvement is more significant compared with STGS, 4DGS, and SpectroMotion, demonstrating the advantage of explicitly modeling time-varying reflective appearance under fast object motion and large environment illumination changes.

4.3 Ablation Study

We evaluate the contribution of each component through one-factor-at-a-time ablations, modifying only the component under evaluation while keeping all other training settings unchanged.

Flat-top temporal support. We replace the flat-top temporal window of the primary Gaussians with conventional Gaussian temporal support. The Gaussian profile continuously modulates each primitive's effective opacity, introducing undesired temporal variation into the alpha-composited screen-space features. As shown in Fig. 3 and Tab. 3, this coupling degrades reconstruction quality and produces visibly blurred moving shadows.

Local residual branch. We remove the auxiliary dynamic Gaussian set and retain only the global branch for reflection prediction. Without the spatially localized corrections provided by the residual branch, the global representation alone lacks sufficient capacity to capture sharp near-field reflections. As shown in Fig. 4 and Tab. 3, this variant fails to faithfully recover fine, mirror-like reflection details, resulting in blurred or incomplete reflections.



Figure 3: Ablation study of the proposed framework on challenging dynamic wrestling dataset with large motions and complex appearance variations.



Figure 4: Ablation study of the proposed framework on in-door SpecFloor dataset.

Time-varying Gaussian features. We replace the time-varying global reflection features with static per-Gaussian features. Without temporal feature modulation, each Gaussian must represent the reflective appearance throughout its lifetime using a single fixed feature vector. This restricts the model's ability to capture appearance changes induced by varying environmental illumination, leading to less accurate reflections and reduced temporal consistency.

Temporal splitting. We disable temporal splitting while retaining the time-varying Gaussian features. Without splitting, each long-lived primary Gaussian must model appearance changes over its existing temporal support, rather than distributing this task among temporally specialized primitives. Continuous feature modulation alone can be insufficient to capture complex appearance variations over extended intervals, particularly under rapidly changing illumination. As shown in Fig. 3 and Tab. 3, this variant struggles to accurately reconstruct the changing appearance of the leg-shadow region.

Table 3: One-factor-at-a-time ablation results at the final checkpoint. Values denote changes relative to each scene's reference configuration. For the first three ablations, LPIPS changes are averaged over Abuzabi-1 and Abuzabi-2, while PSNR and SSIM changes are reported for Abuzabi-1. The local residual branch ablation is evaluated on SpecFloor2.

Configuration	Δ PSNR $\uparrow$	Δ SSIM $\uparrow$	Δ LPIPS $\downarrow$
w/o flat-top temporal opacity	-0.18	-0.0008	$+0.0568$
w/o temporal dimension split	-0.06	-0.0002	$+0.0022$
w/o Fourier feature coefficients	-0.42	-0.0018	$+0.0047$
w/o local residual branch	-0.34	-0.0003	$+0.0018$

5 Conclusion

Our work suggests that the key challenge in dynamic reflective reconstruction is not simply to increase representation capacity, but to allocate that capacity to the right degrees of freedom. Physical geometry should remain constrained by the motion of the underlying surface, while reflected appearance requires additional spatial and temporal flexibility that should not directly alter the surface itself. 4DRefGS implements this principle through asymmetric representation roles, shared directional conditioning, and progressively introduced reflection capacity. The results show that this structured allocation of geometry and appearance freedom is effective for challenging dynamic reflective scenes. **Limitations.** Our method focuses on dynamic specular appearance and does not directly address general challenges such as very fast motion or sparse-view optimization, where inaccurate geometry or motion may degrade reflection reconstruction.

6 AI USE STATEMENT

In this work, we used generative AI tools to improve the grammar, clarity, and readability of the manuscript. All AI-assisted revisions were reviewed by the authors to ensure accuracy and preserve the intended scientific meaning. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Kara-Ali Aliev, Artem Sevastopolsky, Maria Kolos, Dmitry Ulyanov, and Victor Lempitsky. Neural point-based graphics. In *ECCV*, pp. 696–712, 2020.

Jeongmin Bae, Seoha Kim, Youngsik Yun, Hahyun Lee, Gun Bang, and Youngjung Uh. Per-gaussian embedding-based deformation for deformable 3d gaussian splatting. In *ECCV*, pp. 321–335. Springer, 2024.

Ang Cao and Justin Johnson. Hexplane: A fast representation for dynamic scenes. In *CVPR*, pp. 130–141, 2023.

Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14124–14133, 2021.

Danpeng Chen, Hai Li, Weicai Ye, Yifan Wang, Weijian Xie, Shangjin Zhai, Nan Wang, Haomin Liu, Hujun Bao, and Guofeng Zhang. Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. *TVCG*, 2024.

Yuanxing Duan, Fangyin Wei, Qiyu Dai, Yuhang He, Wenzheng Chen, and Baoquan Chen. 4d-rotor gaussian splatting: towards efficient novel view synthesis for dynamic scenes. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–11, 2024.

Cheng-De Fan, Chen-Wei Chang, Yi-Ruei Liu, Jie-Ying Lee, Jiun-Long Huang, Yu-Chee Tseng, and Yu-Lun Liu. Spectromotion: Dynamic 3d reconstruction of specular scenes. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 21328–21338, 2025.

Ningjing Fan and Yiqun Wang. Ssr-gs: Separating specular reflection in gaussian splatting for glossy surface reconstruction. *arXiv preprint arXiv:2603.05152*, 2026.

Ningjing Fan, Yiqun Wang, Dongming Yan, and Peter Wonka. Ref-dgs: Reflective dual gaussian splatting. *arXiv preprint arXiv:2603.07664*, 2026.

Zhongpai Gao, Benjamin Planche, Meng Zheng, Anwesa Choudhuri, Terrence Chen, and Ziyan Wu. 7dgs: Unified spatial-temporal-angular gaussian splatting. *arXiv preprint arXiv:2503.07946*, 2025.

Zhiyang Guo, Wengang Zhou, Li Li, Min Wang, and Houqiang Li. Motion-aware 3d gaussian splatting for efficient dynamic scene reconstruction. *TCSVT*, 2024.

Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5322–5332, 2024.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 2023.

Ye Keyang, Hou Qiming, and Zhou Kun. 3d gaussian splatting with deferred reflection. In *ACM SIGGRAPH Conference Proceedings, Denver, CO, United States*, pp. 3, 2024.

Mijeong Kim, Jongwoo Lim, and Bohyung Han. 4d gaussian splatting in the wild with uncertainty-aware regularization. *Advances in Neural Information Processing Systems*, 37:129209–129226, 2024.

Georgios Kouros, Minye Wu, and Tinne Tuytelaars. Rgs-dr: Deferred reflections and residual shading in 2d gaussian splatting. In *2026 International Conference on 3D Vision (3DV)*, pp. 1000–1009. IEEE, 2026.

Junoh Lee, ChangYeon Won, Hyunjun Jung, Inhwan Bae, and Hae-Gon Jeon. Fully explicit dynamic gaussian splatting. *Advances in Neural Information Processing Systems*, 37:5384–5409, 2024.

Tianye Li, Mira Slavcheva, Michael Zollhoefer, Simon Green, Christoph Lassner, Changil Kim, Tanner Schmidt, Steven Lovegrove, Michael Goesele, Richard Newcombe, et al. Neural 3d video synthesis from multi-view video. In *CVPR*, pp. 5521–5531, 2022.

Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. Spacetime gaussian feature splatting for real-time dynamic view synthesis. In *CVPR*, pp. 8508–8520, 2024.

Yiqing Liang, Numair Khan, Zhengqin Li, Thu Nguyen-Phuoc, Douglas Lanman, James Tompkin, and Lei Xiao. Gaufre: Gaussian deformation fields for real-time dynamic novel view synthesis. In *WACV*, pp. 2642–2652. IEEE, 2025.

Zhanfeng Liao, Jiajun Zhang, Hanzhang Tu, Zhixi Wang, Yunqi Gao, Hongwen Zhang, and Yebin Liu. Sharptimegs: Sharp and stable dynamic gaussian splatting via lifespan modulation. In *CVPR*, 2026.

Rong Liu, Zhongpai Gao, Benjamin Planche, Meida Chen, Van Nguyen Nguyen, Meng Zheng, Anwesa Choudhuri, Terrence Chen, Yue Wang, Andrew Feng, et al. Universal beta splatting. *arXiv preprint arXiv:2510.03312*, 2025.

Zhicheng Lu, Xiang Guo, Le Hui, Tianrui Chen, Min Yang, Xiao Tang, Feng Zhu, and Yuchao Dai. 3d geometry-aware deformable gaussian splatting for dynamic view synthesis. In *CVPR*, pp. 8900–8910, 2024.

Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *International Conference on 3D Vision (3DV)*, 2024.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: representing scenes as neural radiance fields for view synthesis. *Commun. ACM*, 65(1):99–106, December 2021. ISSN 0001-0782. doi: 10.1145/3503250. URL https://doi.org/10.1145/3503250.

Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *ICCV*, pp. 5845–5854, 2021.

LIU Qingming, Yuan Liu, Jiepeng Wang, Xianqiang Lyu, Peng Wang, Wenping Wang, and Junhui Hou. Modgs: Dynamic gaussian splatting from casually-captured monocular videos with depth priors. In *ICLR*, 2025.

Richard Shaw, Michal Nazarczuk, Jifei Song, Arthur Moreau, Sibi Catley-Chandar, Helisa Dhamo, and Eduardo Pérez-Pellitero. Swings: sliding windows for dynamic 3d gaussian splatting. In *ECCV*, pp. 37–54. Springer, 2024.

Ji Shi, Xianghua Ying, Bowei Xing, Ruohao Guo, and Wenzhen Yue. Rt-splatting: Joint reflection-transmission modeling with gaussian splatting. *arXiv preprint arXiv:2605.18263*, 2026.

Liangchen Song, Anpei Chen, Zhong Li, Zhang Chen, Lele Chen, Junsong Yuan, Yi Xu, and Andreas Geiger. Nerfplayer: A streamable dynamic scene representation with decomposed neural radiance fields. *TVCG*, 29(5):2732–2742, May 2023. ISSN 1077-2626. doi: 10.1109/TVCG.2023.3247082. URL https://doi.org/10.1109/TVCG.2023.3247082.

Jingxiang Sun, Xuan Wang, Yong Zhang, Xiaoyu Li, Qi Zhang, Yebin Liu, and Jue Wang. Fenerf: Face editing in neural radiance fields. In *CVPR*, pp. 7662–7672, 2022.

Zhe Jun Tang and Tat-Jen Cham. 3igs: Factorised tensorial illumination for 3d gaussian splatting. In *European Conference on Computer Vision*, pp. 143–159. Springer, 2024.

Yifan Wang, Peishan Yang, Zhen Xu, Jiaming Sun, Zhanhua Zhang, Yong Chen, Hujun Bao, Sida Peng, and Xiaowei Zhou. Freetimegs: Free gaussian primitives at anytime anywhere for dynamic scene reconstruction. In *CVPR*, pp. 21750–21760, 2025.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 13(4):600–612, 2004.

Jiawei Xu, Zexin Fan, Jian Yang, and Jin Xie. Grid4d: 4d decomposed hash encoding for high-fidelity dynamic gaussian splatting. *Advances in Neural Information Processing Systems*, 37: 123787–123811, 2024a.

Zhen Xu, Yinghao Xu, Zhiyuan Yu, Sida Peng, Jiaming Sun, Hujun Bao, and Xiaowei Zhou. Representing long volumetric video with temporal gaussian hierarchy. *ACM Transactions on Graphics*, 43(6):1–18, 2024b.

Zeyu Yang, Hongye Yang, Zijie Pan, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. In *ICLR*, 2024a.

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *CVPR*, 2024b.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pp. 586–595, June 2018.

Youjia Zhang, Anpei Chen, Yumin Wan, Zikai Song, Junqing Yu, Yawei Luo, and Wei Yang. Ref-gs: Directional factorization for 2d gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 26483–26492, 2025.

Ruijie Zhu, Yanzhe Liang, Hanzhi Chang, Jiacheng Deng, Jiahao Lu, Wenfei Yang, Tianzhu Zhang, and Yongdong Zhang. Motiongs: Exploring explicit motion guidance for deformable 3d gaussian splatting. In *NIPS*, volume 37, pp. 101790–101817, 2024.