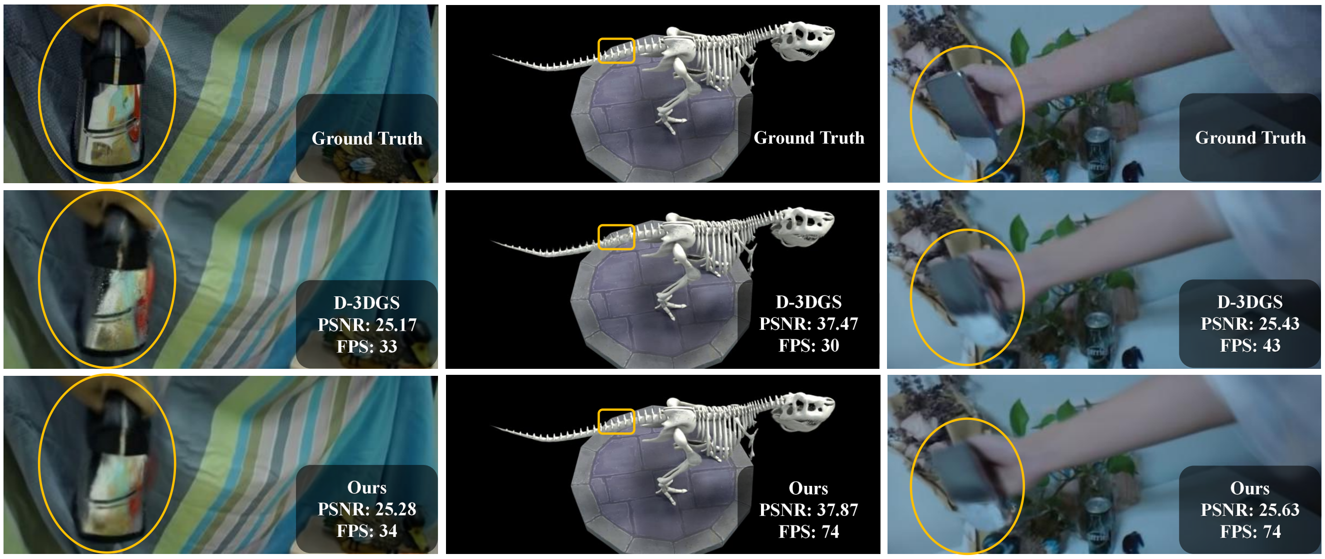
NVC-GS: Monocular Dynamic Scene Reconstruction via Normal-Regularized and Multi-View-Consistent 3D Gaussian Splatting

Huiwen Xue[1] Kaixing Zhao[1,*] Tingcheng Li[2] Tao Xu[1] Zuheng Ming[3,*]

[1]School of Software, Northwestern Polytechnical University
[2]School of Electronic Information and Engineering, Suzhou University of Science and Technology
[3]L2TI, Université Sorbonne Paris Nord



(a) Reducing Surface Noise (b) Correcting Structural Deformation (c) Enhancing Reflection Rendering

Figure 1. Our NVC-GS method develops Geometry-aware Normal Regularization (GeoNR) and Diffusion-based Multi-view Consistency (DMC) for high-quality monocular dynamic scene reconstruction, addressing three critical challenges: (a) Reducing Surface Noise, (b) Correcting Structural Deformation, and (c) Enhancing Reflection Rendering. In dynamic scenes with reflective surfaces, our method achieves higher consistency while maintaining real-time performance.

Abstract

High-quality, real-time dynamic scene reconstruction and rendering are essential for immersive applications. While techniques like 3D Gaussian Splatting (3DGS) succeed in static scenes, dynamic monocular scenarios still suffer from deformation, surface noise, and inconsistent view-dependent effects due to insufficient geometric constraints and inadequate multi-view consistency enforcement for monocular input. To address these challenges, we present Normal-Regularized and Multi-View-Consistent Gaussian Splatting (NVC-GS), a novel approach combining geometry-aware normal regularization with diffusion-based multi-view consistency. Our method explicitly preserves geometric consistency in dynamic objects through tailored normal constraints, while leveraging diffusion-driven latent space regularization to ensure cross-view rendering consistency, particularly for complex materials such as reflective surfaces. Experimental results demonstrate that our approach effectively improves geometric accuracy and visual quality in dynamic scenes while maintaining real-time capabilities, outperforming existing methods in terms of deformation handling, surface noise reduction, and rendering of reflective materials.

∗ Corresponding authors.

1. Introduction

Novel View Synthesis (NVS) involves generating images of a scene from arbitrary viewpoints using a limited set of known views. It has demonstrated significant value in fields such as virtual reality (VR) [23, 33], autonomous driving [30], and robotic manipulation [31, 45]. NVS techniques have evolved from early image-based rendering methods [4, 16, 24] to geometry-driven approaches [6], and more recently, to neural representations [8, 9, 25–28, 40]. Recently, 3D Gaussian Splatting (3DGS) [15] has made significant strides in 3D reconstruction, using explicit Gaussian distributions to represent geometric and appearance attributes, and synthesizing images via efficient rasterization, offering a balance between quality and efficiency.

3DGS has evolved from multi-view to single-view scenarios and expanded from static [5, 20, 42, 46] to dynamic [17, 22, 36, 41] scene modeling. Key improvements have been made in sparse view input processing [5], Gaussian kernel construction [46], and final rendered output [20]. Monocular dynamic scene reconstruction, known for its resource efficiency, versatility, and real-time rendering, faces critical challenges, particularly the lack of effective geometric constraints and multi-view consistency, which limit its application in complex scenarios.

The main challenge in monocular dynamic scene reconstruction with 3DGS is maintaining geometric consistency during deformation. Vanilla 3DGS [15] uses Gaussian kernels with smoothing properties, favoring gradual transitions over sharp details. Methods like Deformable 3DGS [41] optimize only position, rotation, and scaling, lacking surface continuity constraints, leading to high-frequency detail loss and artifacts. From a differential geometry perspective, normal vectors characterize surface properties such as orientation and curvature, and help address related challenges by promoting surface smoothness and preserving features. While methods like [13, 18] have integrated normal estimation to constrain geometry, they focus primarily on static scenes and overlook the relationship between deformation fields and normal updates, limiting their application to complex deformations. To address these challenges, we propose Geometry-aware Normal Regularization (GeoNR): 1) tightly coupling normal updates with deformation fields to ensure natural normal variation during motion; 2) introducing self-constrained normal smoothness regularization to promote geometric consistency.

Multi-view inconsistency is a key challenge in monocular dynamic scene reconstruction. In dynamic scenes, material appearance changes with viewpoint, but monocular reconstruction relies on single views at each time point, focusing on appearance matching and neglecting cross-view consistency. This lack of constraints causes artifacts during viewpoint transitions, particularly on reflective surfaces. Diffusion models [12, 29], trained on diverse datasets, provide rich visual priors and ensure semantic consistency in latent space, offering potential for addressing multi-view consistency. However, existing approaches face two limitations: implicit methods [10, 19, 38, 39] suffer from NeRF's volumetric rendering overhead, affecting real-time performance, while explicit methods [2, 21] improve efficiency for static scenes but fail to capture temporal variations, limiting their application to dynamic reconstruction. To address these issues, we propose Diffusion-based Multi-view Consistency (DMC): 1) a spatio-temporal constraint that samples images from different time points and viewpoints, ensuring consistency in latent space; 2) a regularization strategy employing optimal noise levels to prevent unrealistic scene generation.

Our Normal-Regularized and Multi-View-Consistent Gaussian Splatting (NVC-GS) method eliminates surface noise, artifacts, and unnatural deformations in dynamic scene reconstruction by integrating normal constraints with diffusion-based multi-view consistency (see Fig. 1). To summarize, our contributions are as follows:

- Geometry-aware Normal Regularization (GeoNR): A mechanism that couples normal updates with deformation fields, ensuring natural normal variation with object motion and maintaining sharp edges and smooth surfaces.
- Diffusion-based Multi-view Consistency (DMC): A lightweight framework that leverages diffusion model priors through efficient latent space projection to enforce cross-view consistency in dynamic scenes.
- Progressive three-stage training strategy: An effective pipeline that introduces constraints gradually through weight scheduling, ensuring stable optimization and high-quality reconstruction.

2. Related Work

2.1. Novel View Synthesis

Early novel view synthesis evolved through two main approaches. These include image-based techniques [4, 16, 24] requiring dense sampling, and geometry-aware methods [6] that struggled with complex lighting effects. NeRF [25] revolutionized the field with continuous volumetric modeling that achieved unprecedented visual quality. Dynamic scene extensions incorporated deformation fields [28], temporal embeddings [26], and surface constraints [27, 40], but remained computationally intensive for interactive applications. 3D Gaussian Splatting [15] addressed efficiency challenges by combining point-based representations with differentiable rasterization. Dynamic scene synthesis has since advanced through time-dependent modeling [17], deformation techniques [41], and 4D representations [36], with specialized approaches addressing sparse views [5], continuous transformation [46], multi-view generalization [20], and detail preservation [42].

Despite these advances, two challenges remain: surface distortions in deforming regions due to insufficient geometric constraints, and multi-view inconsistencies causing artifacts across viewpoints. Our work tackles these issues with geometry-aware normal regularization and diffusion-based multi-view consistency.

2.2. Normal-based Geometric Constraints

Surface normal constraints provide effective geometric regularization for 3D Gaussian representations. NeuS [32] utilizes SDF gradients as surface normals for high-quality static reconstruction, while recent work has integrated normal estimation into the 3DGS framework. GS-IR [18] and GaussianShader [13] estimate normals in different ways, the former for inverse rendering constraints, and the latter using the shortest axis of 3D Gaussians for reflective surface reconstruction. Wu et al. [37] address surface discontinuity using density consistency and normal-based reflection modeling, while Normal-GS [35] incorporates normals into physically-based rendering equations. Other geometry-aware methods, such as Gaussian Opacity Fields [43], calculate ray-Gaussian intersections to derive geometric priors, and VCR-GauS [3] optimizes Gaussian geometry via view-consistent depth-normal regularization.

Unlike previous methods that ignore normal constraints or require dense normal maps, our approach uses self-supervised normal regularization to address geometric inconsistencies while maintaining real-time performance.

2.3. Multi-view Consistency Constraint

The first enhances implicit representations (NeRF) with diffusion priors. Zero-1-to-3 [19] combines diffusion models with geometric priors for 3D reconstruction. DiffusioNeRF [39] uses denoising diffusion models to improve view consistency. Control3Diff [10] and ReconFusion [38] guide NeRF optimization with score distillation and conditional constraints, respectively. While these methods improve quality, efficiency bottlenecks limit their real-time applicability. The second paradigm focuses on enhancing explicit representations (3DGS) for static scenes. 3DGS-Enhancer [21] addresses sparse-view artifacts using video diffusion models. MV2MV [2] offers a unified framework for multi-view image translation with high quality and consistency.

However, these methods struggle to model temporal variations, limiting their use in dynamic scenes. Our DMC module overcomes these challenges with a lightweight diffusion prior projection mechanism, seamlessly integrating with deformation fields and normal constraints.

3. Preliminaries

3.1. 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [15] represents scene geometry and appearance using 3D Gaussian primitives. Each Gaussian is defined by a center position $\boldsymbol{\mu} \in \mathbb{R}^3$ and a covariance matrix $\Sigma \in \mathbb{R}^{3\times 3}$:

$$G(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \tag{1}$$

where Σ is optimized using a rotation matrix R and a scaling matrix S:

$$\Sigma = RSS^T R^T. \tag{2}$$

The optimization of 3DGS relies on a photometric loss that combines L1 distance and structural similarity:

$$\mathcal{L}_{\text{RGB}} = (1 - \lambda_{\text{DSSIM}})\mathcal{L}_1 + \lambda_{\text{DSSIM}}(1 - \text{SSIM}), \tag{3}$$

where $\mathcal{L}_1$ represents the L1 distance between rendered and ground truth images, while SSIM quantifies structural similarity.

3.2. Deformable 3D Gaussian Splatting

To model dynamic scenes, Deformable 3D Gaussian Splatting (D-3DGS) [41] introduces a time-conditional deformation field that adjusts Gaussian parameters over time. This approach decouples static and dynamic components using a deformation MLP network $\mathcal{F}_\theta$, which maps 3D Gaussians into canonical space and predicts position, rotation, and scaling offsets:

$$(\delta\mathbf{x}, \delta\mathbf{r}, \delta\mathbf{s}) = \mathcal{F}_\theta(\gamma(\mathbf{x}, t)), \tag{4}$$

where $\mathbf{x}$ represents the 3D Gaussian position, t is a temporal variable that extends the static 3DGS framework to dynamic scenes, and γ denotes the positional encoding. These outputs transform static Gaussian parameters into their dynamic counterparts. However, D-3DGS faces two key limitations: 1) geometric inconsistencies at object boundaries due to the absence of surface constraints, resulting in unnatural deformations; 2) view-dependent appearance instability, which causes flickering and inconsistent material rendering across different viewpoints. Our proposed method effectively addresses these issues, as demonstrated in the experimental results presented in Sec. 5.2.

4. Method

4.1. Overview

We propose Normal-Regularized and Multi-View-Consistent Gaussian Splatting (NVC-GS) to address geometric and temporal inconsistencies in dynamic scene reconstruction. As shown in Fig. 2, NVC-GS introduces Geometry-aware Normal Regularization (GeoNR) and Diffusion-based Multi-view Consistency (DMC), via a progressive three-stage training protocol.

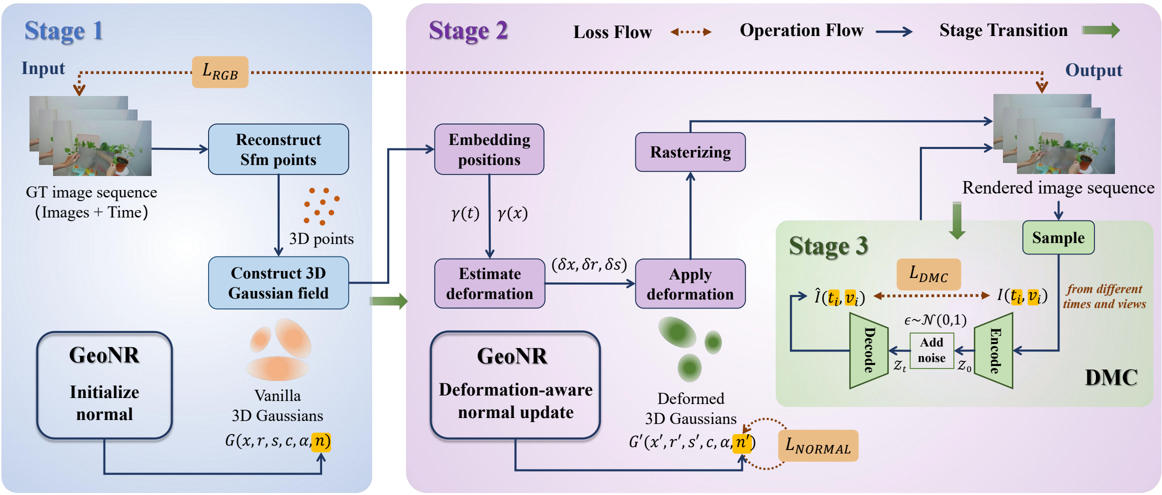


Figure 2. **Pipeline overview.** Our approach extends D-3DGS through a progressive three-stage training process: Stage 1 (blue region) performs static initialization where we optimize Gaussian properties and initialize surface normals while freezing the deformation network; Stage 2 (purple region) implements deformation-aware normal updates that maintain geometric consistency during object motion; and Stage 3 (green region) introduces diffusion-based multi-view consistency by encoding renders from different views and time steps into a shared latent space with calibrated noise.

Stage 1: Static Base Initialization. We freeze the deformation network, and optimize Gaussian properties through RGB reconstruction loss $\mathcal{L}_{\text{RGB}}$, while initializing normals using the shortest axis of each Gaussian ellipsoid.

Stage 2: Normal-guided Deformation. After static optimization, we activate the deformation network with deformation-aware normal updates and normal smoothness constraints to ensure geometric consistency during motion.

Stage 3: Diffusion Consistency Enforcement. We introduce DMC, which randomly samples rendered images from different time points and viewpoints, applies VAE encoding and decoding with calibrated noise to generate regularized representations, thereby enhancing multi-view consistency.

4.2. Geometry-aware Normal Regularization

Traditional 3D Gaussian Splatting relies solely on RGB reconstruction loss and does not provide explicit geometric constraints, resulting in inaccurate surface reconstruction at object edges and curved regions. To address this limitation, we propose Geometry-aware Normal Regularization (GeoNR) to enhance surface reconstruction accuracy.

4.2.1 Normal-enhanced Gaussian Representation.

Physically based rendering (PBR) [11] demonstrates that surface normal information is critical for both realistic rendering and accurate geometry representation. However, vanilla 3D Gaussians lack explicit surface orientation information, resulting in loss of geometric details. We therefore extend the standard 3D Gaussian representation by introducing normal attributes $\mathbf{n}_i \in \mathbb{R}^3$ for each Gaussian point:

$$\mathcal{G} = \{(\mathbf{x}_i, \mathbf{r}_i, \mathbf{s}_i, c_i, \alpha_i, \mathbf{n}_i) \mid i = 1, 2, \ldots, N\}, \quad (5)$$

where $\mathbf{x}_i$ is the center position, $\mathbf{r}_i$ is the rotation, $\mathbf{s}_i$ is the scaling, c_i is the color feature, α_i is the opacity, and $\mathbf{n}_i$ provides explicit surface orientation information essential for accurate geometric reconstruction.

4.2.2 Adaptive Normal Updating.

To ensure normals accurately reflect surface properties and maintain consistency during scene deformation, we design an adaptive normal update mechanism. As shown in Fig. 3, this mechanism adjusts computation according to training phases, with geometry-based initialization in stage 1 and deformation-aware updates in stages 2-3.

Normal initialization. We leverage the geometric properties of Gaussians to initialize normal vectors. 3D Gaussian representation is tied to Singular Value Decomposition (SVD) in Principal Component Analysis (PCA) [14], where data exhibits minimal variance along the eigenvector corresponding to the smallest eigenvalue. For 3D Gaussians, this implies their shortest axis aligns perpendicular to the local surface, yielding a suitable normal approximation:

$$\mathbf{n} = R[:, k] \in \mathbb{R}^3, k = \arg\min([s_1, s_2, s_3]), \quad (6)$$

where R is the rotation matrix, and k represents the index of the smallest scaling parameter, indicating the shortest axis.

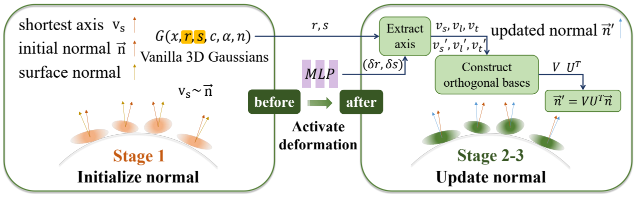


Figure 3. **Adaptive Normal Updating.** In Stage 1 (left), normals are initialized using the shortest axis of each Gaussian ellipsoid. In Stages 2-3 (right), our deformation-aware update mechanism adjusts normals based on the deformation field, using scaling and rotation parameters to construct orthogonal bases that ensure geometric consistency during dynamic motion.

Phase-based update strategy. We employ a staged normal update strategy. In stage 1, we use the shortest-axis-based method in Eq. (6) for initial geometry establishment. In stages 2-3, we use deformation-aware normal update strategy, which is based on the principle of coordinate system transformation in differential geometry [1].

$$\mathbf{n}' = \mathbf{V}\mathbf{U}^T\mathbf{n}, \tag{7}$$

where $\mathbf{n}' \in \mathbb{R}^3$ is the transformed normal vector. The matrices $\mathbf{U} = [\mathbf{v}_s, \mathbf{v}_l, \mathbf{v}_t]$ and $\mathbf{V} = [\mathbf{v}_s', \mathbf{v}_l', \mathbf{v}_t']$ are orthogonal bases before and after deformation, with $\mathbf{v}_s, \mathbf{v}_l$ being shortest and longest axes, and $\mathbf{v}_t$ their cross product.

4.2.3 Self-constrained Normal Regularization.

In contrast to methods requiring explicit normal supervision, we propose a self-constrained approach that preserves geometric properties without external normal maps. Our method employs smoothness constraints that encourage sequentially adjacent Gaussians to maintain continuously varying surface orientations:

$$\mathcal{L}_{\text{NORMAL}} = \frac{1}{N-1}\sum_{i=1}^{N-1} \|\mathbf{n}_{i+1} - \mathbf{n}_i\|_2^2, \tag{8}$$

where N is the total number of Gaussian points and $\mathbf{n}_i$ denotes the normal vector of point i. This constraint promotes local smoothness between adjacent normals without enforcing global consistency. Balancing with $\mathcal{L}_{\text{RGB}}$ prevents over-smoothing that would degrade rendering quality. Combined with deformation-aware normal updates, this approach significantly enhances geometric reconstruction in dynamic scenes.

4.3. Diffusion-based Multi-view Consistency

While GeoNR enhances geometric quality, monocular dynamic reconstruction suffers from appearance inconsistencies across temporal views, causing flickering and artifacts. We propose Diffusion-based Multi-view Consistency (DMC) that leverages pre-trained diffusion priors to harmonize visual appearance while preserving dynamic details. Unlike conventional iterative denoising approaches, our DMC introduces a lightweight diffusion-based VAE regularization with minimal computational overhead.

4.3.1 Multi-view Batch Sampling

To effectively address spatio-temporal inconsistencies, we design a periodic batch sampling strategy that collects rendered images from diverse temporal points and viewpoints during training:

$$\mathcal{B} = \{I(t_i, v_i) | i = 1, 2, ..., N\}, \tag{9}$$

where $I(t_i, v_i)$ represents an image rendered at time t_i from viewpoint v_i, with both randomly sampled to ensure broad spatio-temporal coverage, and N is the batch size.

4.3.2 Latent Noise Augmentation

After collecting the multi-view batch, we transform the images from pixel space to the diffusion model's latent space through a VAE encoder and apply a calibrated amount of noise:

$$z_t = \mathcal{E}(I) + \sigma \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, 1), \tag{10}$$

where $\mathcal{E}(\cdot)$ denotes the VAE encoder, σ controls the noise magnitude, and ϵ is random noise sampled from a standard normal distribution. Our approach encodes multi-view renders into a common VAE latent space with calibrated noise to create a shared optimization domain. This encoding process guides optimization toward more coherent representations across different viewpoints and time steps.

We then project the noise-augmented latent representations back to pixel space through direct decoding $\hat{I} = \mathcal{D}(z_t)$, where $\mathcal{D}(\cdot)$ is the VAE decoder. This operation leverages the learned priors of the VAE to regularize the latent space, preserving key structural and semantic information such as object positioning and color distribution while working with perturbed representations.

4.3.3 Multi-view Consistency Loss

Through this latent noise augmentation process, we define the DMC module's loss $\mathcal{L}_{\text{DMC}}$ as:

$$\mathcal{L}_{\text{DMC}} = \sum_{i \in \mathcal{B}} \|I(t_i, v_i) - \hat{I}(t_i, v_i)\|_2^2, \tag{11}$$

where $I(t_i, v_i)$ represents the originally rendered image and $\hat{I}(t_i, v_i)$ is its corresponding version reconstructed through VAE with calibrated latent noise. Although formulated as pixel-level MSE, this loss fundamentally differs from direct image comparison because the VAE bottleneck emphasizes

	As			Basin			Bell			Cup		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF-DS* [40]	25.34	0.8803	0.2150	20.23	0.8053	0.2508	22.57	0.7811	0.2921	24.51	0.8802	0.1707
HyperNeRF* [27]	17.59	0.8518	0.2390	22.58	0.8156	0.2497	19.80	0.7650	0.2999	15.45	0.8295	0.2302
4DGS* [36]	24.85	0.8632	0.2038	19.26	0.7670	0.2196	22.86	0.8015	0.2061	23.82	0.8695	0.1792
D-3DGS	26.12	0.8845	0.1806	19.61	0.7941	0.1915	25.17	0.8426	0.1619	24.66	0.8883	0.1589
Ours	26.22	0.8849	0.1817	19.65	0.7962	0.1890	25.28	0.8465	0.1583	24.82	0.8901	0.1585

	Plate			Press			Sieve			**Mean**		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF-DS* [40]	19.70	0.7813	0.2974	25.35	0.8703	0.2552	24.99	0.8705	0.2001	23.24	0.8384	0.2402
HyperNeRF* [27]	21.22	0.7829	0.3166	16.54	0.8200	0.2810	19.92	0.8521	0.2142	19.01	0.8167	0.2615
4DGS* [36]	18.77	0.7709	0.2721	24.82	0.8355	0.2255	25.16	0.8566	0.1745	22.79	0.8235	0.2115
D-3DGS	20.27	0.8096	0.2260	25.43	0.8623	0.1946	25.40	0.8726	0.1491	23.81	0.8506	0.1804
Ours	20.56	0.8161	0.2241	25.63	0.8655	0.1910	25.46	0.8726	0.1504	23.96	0.8531	0.1790

Table 1. **Quantitative comparison on the NeRF-DS dataset.** We report the average PSNR, SSIM, and LPIPS (VGG) of several previous models on test images. Best, second best, and third best results are highlighted in red, orange, and yellow, respectively. The results reported in [7] used the same dataset split as ours and conducted a unified evaluation of these methods. Note that the original 4DGS work did not test on the NeRF-DS dataset. The slight differences in our measurement metrics, compared to those in NeRF-DS and HyperNeRF, arise from their use of MS-SSIM instead of SSIM and LPIPS with an AlexNet backbone instead of VGG.

structural features that persist through encoding-decoding, rather than exact pixel alignment. This promotes cross-view consistency while preserving scene-specific details.

4.4. Training Loss

To balance RGB fidelity, geometric accuracy, and spatio-temporal consistency, we define a joint optimization loss:

$$\mathcal{L} = \lambda_{\text{rgb}}\mathcal{L}_{\text{RGB}} + \lambda_{\text{normal}}\mathcal{L}_{\text{NORMAL}} + \lambda_{\text{dmc}}\mathcal{L}_{\text{DMC}}, \qquad (12)$$

where $\mathcal{L}_{\text{RGB}}$ ensures rendering fidelity by enforcing pixel-level and perceptual consistency between rendered and reference images. $\mathcal{L}_{\text{NORMAL}}$ promotes geometric consistency by regularizing normal variations between adjacent Gaussian points, reducing surface artifacts. $\mathcal{L}_{\text{DMC}}$ leverages diffusion priors to enhance cross-view consistency, particularly for reflective surfaces and viewpoint transitions. Both the $\mathcal{L}_{\text{RGB}}$ and $\mathcal{L}_{\text{NORMAL}}$ are optimized across all three stages, while the $\mathcal{L}_{\text{DMC}}$ is optimized starting from stage 3. Through joint optimization of these complementary losses, NVC-GS achieves high-quality dynamic scene reconstruction, ensuring both realistic appearance and geometric accuracy while eliminating common artifacts such as ghosting, blurred edges, and deformation.

5. Experiments

5.1. Experimental Settings

5.1.1 Datasets and Metrics.

To validate our NVC-GS method, we conducted evaluations on two representative dynamic scene datasets: NeRF-DS [40] and D-NeRF [28]. The NeRF-DS dataset contains 7 real-world dynamic scenes featuring challenging specular objects with complex reflective properties, while D-NeRF consists of 8 synthetic dynamic scenes with varying camera poses at each timestep. We employ standard image quality metrics including PSNR, SSIM [34], and LPIPS [44] to assess rendering fidelity, along with frames per second (FPS) to evaluate computational efficiency.

5.1.2 Implementation.

Our PyTorch-based implementation runs a total of 20k iterations for NeRF-DS and 40k iterations for D-NeRF to accommodate their varying complexity levels. Progressive weight scheduling is applied to all regularization terms, gradually increasing their influence throughout the training process. All experiments were conducted on a single NVIDIA RTX 3090 GPU, using the original train-test splits of the respective datasets. D-3DGS results are reproduced using the official implementation, with minor variations due to framework updates and training randomness.

5.2. Comparisons

5.2.1 Real-World Dataset.

We compare against state-of-the-art approaches on the NeRF-DS dataset [40], including the NeRF-DS, HyperNeRF [27], 4DGS [36], and D-3DGS [41]. Quantitative results are shown in Tab. 1. Our method achieves an average PSNR of 23.96 dB (+0.63% over D-3DGS), SSIM of 0.8531 (+0.29%), and LPIPS of 0.1790 (a 0.78% im-

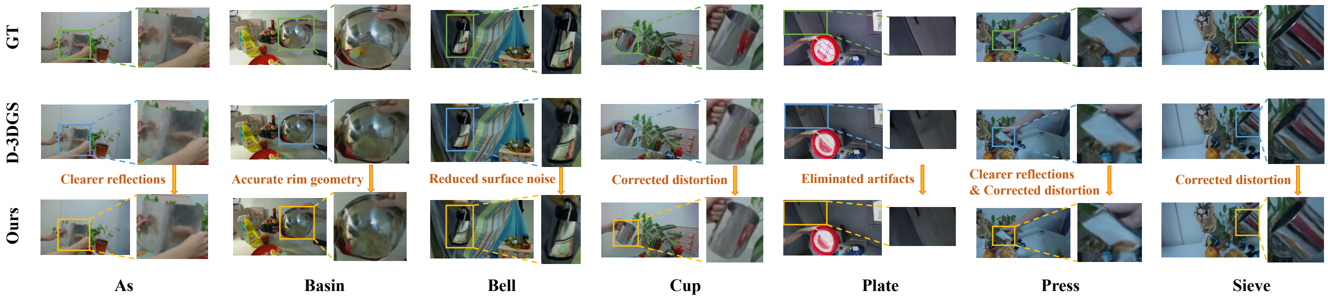


Figure 4. **Qualitative comparison on the NeRF-DS dataset between ground truth (top), D-3DGS (middle), and our NVC-GS (bottom).** Our approach demonstrates noticeable improvements including clearer reflections (As, Press), more accurate geometry in rim areas (Basin), reduced surface noise (Bell), corrected distortion (Cup, Sieve), and eliminated artifacts (Plate). These enhancements are particularly evident in challenging scenes with reflective materials.

Method	PSNR↑	SSIM↑	LPIPS↓	FPS↑
D-NeRF [28]	30.43	0.95	0.07	<1
TiNeuVox-B [8]	32.67	0.97	0.04	1.5
K-Planes [9]	31.61	0.97	-	0.97
4DGS [36]	34.05	0.98	0.02	82
D-3DGS	38.68	0.99	0.02	77
Ours	38.83	0.99	0.02	127

Table 2. **Quantitative comparison on the D-NeRF dataset.** We report the average PSNR, SSIM, and LPIPS (VGG) across all 8 scenes. '-' denotes that the metric is not reported in their works.

provement, lower is better). In the Plate scene with complex curved surfaces, our GeoNR reduces geometric distortion, delivering PSNR gains of +0.29 dB over D-3DGS. For scenes with reflective materials such as Bell and Press, the DMC module enhances consistency across viewpoints, driving perceptual quality improvements (LPIPS reduced by 2.22% and 1.85%, respectively).

The qualitative results presented in Fig. 4 further validate the visual advantages of our method. The comparisons reveal that NVC-GS effectively mitigates three key issues: geometric edge errors in dynamically deforming regions, noise artifacts on reflective surfaces, and blurring artifacts caused by inconsistent rendering.

5.2.2 Synthetic Dataset.

To verify the versatility of our approach across different scene types, we evaluate NVC-GS on the D-NeRF synthetic dataset [28]. As shown in Tab. 2, our method achieves a PSNR of 38.83 dB (+0.15 dB over D-3DGS) while maintaining equivalent SSIM (0.99) and LPIPS (0.02) scores against both neural implicit methods (D-NeRF [28], TiNeuVox-B [8], K-Planes [9]) and Gaussian-based approaches (4DGS [36], D-3DGS [41]). The method also demonstrates significant rendering efficiency improvements on these synthetic scenes, further verifying its generalizability to both real-world and synthetic dynamic scenes with different motion and appearance characteristics.

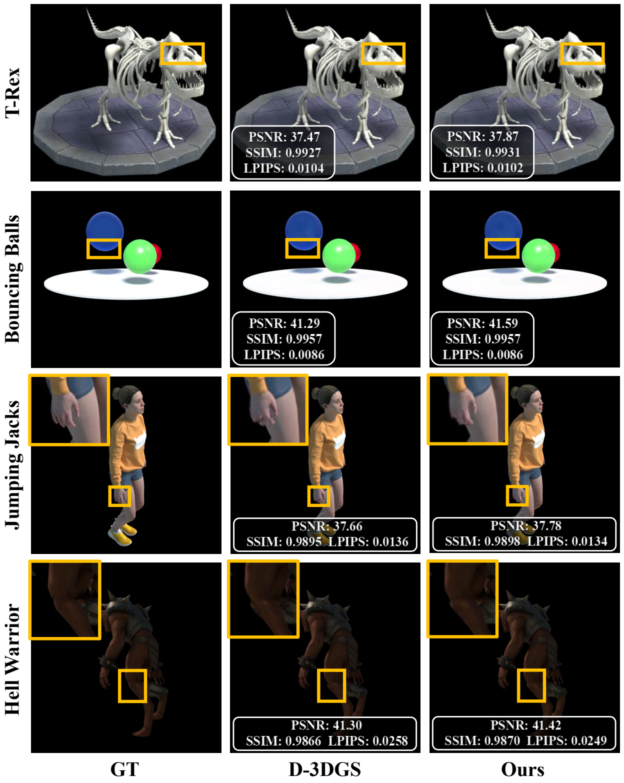


Figure 5. **Qualitative comparison on the D-NeRF dataset between ground truth (left), D-3DGS baseline (middle), and our NVC-GS method (right).** Yellow boxes highlight detailed regions. NVC-GS effectively reduces geometric deformation artifacts and improves surface consistency in complex areas.

As shown in Fig. 5, qualitative comparisons demonstrate that NVC-GS effectively enhances detail representation across diverse dynamic scenes, achieving superior

	NeRF-DS				D-NeRF			
Method	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑
D-3DGS	23.81	0.8506	0.1804	46	38.68	0.9854	0.0176	77
w/o DMC	23.91	0.8524	0.1787	49	38.77	0.9851	0.0177	98
w/o GeoNR	23.87	0.8512	0.1791	47	38.72	0.9854	0.0175	124
Full Model	23.96	0.8531	0.1790	57	38.83	0.9857	0.0173	127

Table 3. **Ablation Study of Our NVC-GS Components.** Starting from the baseline D-3DGS, we incrementally add components to evaluate the contribution of each module.

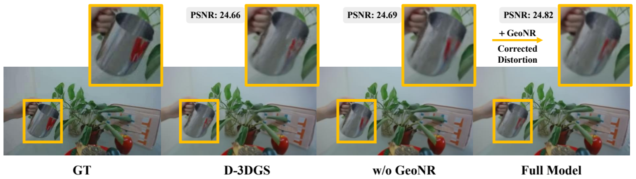


Figure 6. **Ablation study on reflective surface reconstruction in the Cup scene from NeRF-DS dataset.** From left to right: ground truth, D-3DGS baseline, our method without GeoNR, and our full model. The yellow boxes highlight the challenging reflective area with complex specular highlights. The GeoNR module improves geometric accuracy and surface detail preservation.

rendering quality to D-3DGS. Specifically, it preserves fine horn structures in T-Rex, enables more accurate reconstruction of the blue ball's boundary in Bouncing Balls, renders hand-region details with higher precision in Jumping Jacks, and achieves clearer knee geometry in Hell Warrior.

5.3. Ablation Studies

5.3.1 Quantitative results

To evaluate the contributions of key components, we perform ablation experiments on the NeRF-DS [40] and D-NeRF [28] datasets. Using D-3DGS as the baseline, we progressively incorporate the GeoNR and DMC modules, with results shown in Tab. 3.

Our ablation studies demonstrate individual module contributions. GeoNR provides quality improvements (PSNR +0.10 dB on NeRF-DS, +0.09 dB on D-NeRF) via effective geometric constraints. While DMC alone shows modest improvements, combining it with GeoNR creates synergistic effects, exceeding DMC alone by +0.09 dB and +0.11 dB on NeRF-DS and D-NeRF respectively. This synergy confirms that diffusion-based consistency constraints require accurate geometric foundations to realize their full potential in 3D reconstruction. Furthermore, the full model improves rendering efficiency, achieving 57 FPS on NeRF-DS and 127 FPS on D-NeRF.

5.3.2 Qualitative results

We present qualitative ablation results on representative scenes from both datasets. As illustrated in Fig. 6, the GeoNR module demonstrates effectiveness on the reflective Cup scene. While D-3DGS struggles with specular highlights and geometric distortions, integrating GeoNR with DMC further yields superior surface geometry and clearer reflection patterns in the red highlight region. Figure 7 shows the DMC module's benefits on the StandUp scene with large viewpoint variations. Beyond baseline inconsistencies and GeoNR's geometric improvements, our full model maintains coherent facial features across viewpoints through multi-view consistency optimization.

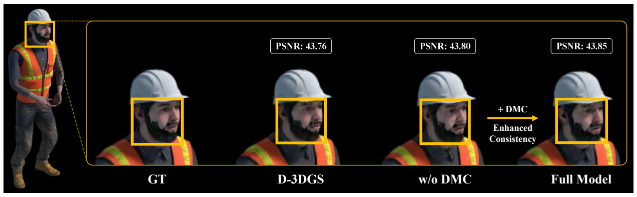


Figure 7. **Ablation study on multi-view consistency in the StandUp scene from D-NeRF dataset with large viewpoint variations.** From left to right: ground truth, D-3DGS baseline, our method without DMC, and our full model. The yellow boxes highlight facial details that benefit from consistent multi-view representation. The DMC module improves rendering consistency.

These results validate our architectural principle. GeoNR provides reliable geometric representations to enhance DMC's consistency constraints, while DMC boosts cross-view coherence. Their complementary effects jointly improve reconstruction quality beyond the performance of either module alone.

6. Conclusion

We present NVC-GS, a dynamic scene reconstruction approach that addresses insufficient geometric constraints and inadequate multi-view consistency in 3DGS. Our method combines geometry-aware normal regularization and diffusion-based multi-view consistency to address surface noise, structural deformation, and reflection rendering challenges, thereby improving overall rendering quality. Experiments demonstrate that these complementary constraints create a mutually reinforcing effect, where better geometry enables more consistent views, which in turn refines the underlying representation. Future work could explore adaptive regularization strategies and explicit reflectance modeling for more robust dynamic scene reconstruction in challenging environments.

Acknowledgments

The authors gratefully acknowledge the financial supports by research funding. Kaixing Zhao is supported by the National Natural Science Foundation of China (No. 62407035).

References

[1] Alan H. Barr. Global and local deformations of solid primitives. *ACM Siggraph Computer Graphics*, 18(3):21–30, 1984. 5

[2] Youcheng Cai, Runshi Li, and Ligang Liu. MV2MV: Multi-View Image Translation via View-Consistent Diffusion Models. *ACM Transactions on Graphics*, 43(6), 2024. 2, 3

[3] Hanlin Chen, Fangyin Wei, Chen Li, Tianxin Huang, Yunsong Wang, and Gim Hee Lee. VCR-GauS: View Consistent Depth-Normal Regularizer for Gaussian Surface Reconstruction. In *Advances in Neural Information Processing Systems*, pages 139725–139750, 2024. 3

[4] Shenchang Eric Chen and Lance Williams. View interpolation for image synthesis. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, pages 279–288. ACM, 1993. 2

[5] Yuedong Chen, Haofei Xu, Chuanxia Zheng, Bohan Zhuang, Marc Pollefeys, Andreas Geiger, Tat-Jen Cham, and Jianfei Cai. MVSplat: Efficient 3D Gaussian splatting from sparse multi-view images. In *Proceedings of the European Conference on Computer Vision*, pages 370–386, 2024. 2

[6] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, pages 11–20. ACM, 1996. 2

[7] 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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pages 21328–21338, 2025. 6

[8] Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast Dynamic Radiance Fields with Time-Aware Neural Voxels. In *Proceedings of SIGGRAPH Asia 2022 Conference Papers*, pages 1–9. ACM, 2022. 2, 7

[9] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-Planes: Explicit Radiance Fields in Space, Time, and Appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12479–12488, 2023. 2, 7

[10] Jiatao Gu, Qingzhe Gao, Shuangfei Zhai, Baoquan Chen, Lingjie Liu, and Josh Susskind. Control3Diff: Learning Controllable 3D Diffusion Models from Single-view Images. In *Proceedings of the International Conference on 3D Vision*, pages 685–696, 2024. 2, 3

[11] Stephen Hill, Stephen McAuley, Jonathan Dupuy, Yoshiharu Gotanda, Eric Heitz, Naty Hoffman, Sébastien Lagarde, Anders Langlands, Ian Megibben, Farhez Rayani, and Charles de Rousiers. Physically based shading in theory and practice. In *ACM SIGGRAPH 2014 Courses*, 2014. 4

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 2

[13] 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*, pages 5322–5332, 2024. 2, 3

[14] Ian T. Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016. 4

[15] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4):139:1–139:14, 2023. 2, 3

[16] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, pages 31–42. ACM, 1996. 2

[17] Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. Spacetime Gaussian Feature Splatting for Real-Time Dynamic View Synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8508–8520, 2024. 2

[18] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. GS-IR: 3D Gaussian splatting for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21644–21653, 2024. 2, 3

[19] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot One Image to 3D Object. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9298–9309, 2023. 2, 3

[20] Tianqi Liu, Guangcong Wang, Shoukang Hu, Liao Shen, Xinyi Ye, Yuhang Zang, Zhiguo Cao, Wei Li, and Ziwei Liu. MVSGaussian: Fast generalizable Gaussian splatting reconstruction from multi-view stereo. In *Proceedings of the European Conference on Computer Vision*, pages 37–53, 2024. 2

[21] Xi Liu, Chaoyi Zhou, and Siyu Huang. 3DGS-Enhancer: Enhancing unbounded 3D Gaussian splatting with view-consistent 2D diffusion priors. In *Advances in Neural Information Processing Systems*, pages 133305–133327, 2024. 2, 3

[22] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3D Gaussians: Tracking by persistent dynamic view synthesis. In *Proceedings of the International Conference on 3D Vision*, pages 800–809, 2024. 2

[23] Yikun Ma, Dandan Zhan, and Zhi Jin. FastScene: Text-Driven Fast 3D Indoor Scene Generation via Panoramic Gaussian Splatting. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, pages 1173–1181, 2024. 2

[24] Leonard McMillan and Gary Bishop. Plenoptic modeling: An image-based rendering system. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 39–46. ACM, 1995. 2

[25] 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 *Proceedings of the European Conference on Computer Vision*, pages 405–421. Springer, 2020. 2

[26] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021. 2

[27] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B. Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. Hypernerf: a higher-dimensional representation for topologically varying neural radiance fields. *ACM Trans. Graph.*, 40(6):238:1–238:12, 2021. 2, 6

[28] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10318–10327, 2021. 2, 6, 7, 8, 1

[29] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 2

[30] Shanlin Sun, Bingbing Zhuang, Ziyu Jiang, Buyu Liu, Xiaohui Xie, and Manmohan Chandraker. LidaRF: Delving into Lidar for Neural Radiance Field on Street Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19563–19572, 2024. 2

[31] Stephen Tian, Blake Wulfe, Kyle Sargent, Katherine Liu, Sergey Zakharov, Vitor Guizilini, and Jiajun Wu. View-Invariant Policy Learning via Zero-Shot Novel View Synthesis. In *Proceedings of the 8th Conference on Robot Learning*, pages 1173–1193, 2024. 2

[32] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Advances in Neural Information Processing Systems*, pages 27171–27183, 2021. 3

[33] Shuo Wang, Cong Xie, Shengdong Wang, and Shaohui Jiao. Geometry enhanced 3D Gaussian Splatting for high quality deferred rendering. In *ACM SIGGRAPH 2024 Posters*, pages 1–2. ACM, 2024. 2

[34] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 6

[35] Meng Wei, Qianyi Wu, Jianmin Zheng, Hamid Rezatofighi, and Jianfei Cai. Normal-GS: 3D Gaussian Splatting with Normal-Involved Rendering. In *Advances in Neural Information Processing Systems*, pages 76356–76379, 2024. 3

[36] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20310–20320, 2024. 2, 6, 7

[37] Qianyi Wu, Jianmin Zheng, and Jianfei Cai. Surface reconstruction from 3D Gaussian splatting via local structural hints. In *Proceedings of the European Conference on Computer Vision*, pages 441–458. Springer, 2024. 3

[38] Rundi Wu, Ben Mildenhall, Philipp Henzler, Keunhong Park, Ruiqi Gao, Daniel Watson, Pratul P. Srinivasan, Dor Verbin, Jonathan T. Barron, Ben Poole, and Aleksander Hołyński. ReconFusion: 3D Reconstruction with Diffusion Priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21551–21561, 2024. 2, 3

[39] Jamie Wynn and Daniyar Turmukhambetov. DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4180–4189, 2023. 2, 3

[40] Zhiwen Yan, Chen Li, and Gim Hee Lee. NeRF-DS: Neural radiance fields for dynamic specular objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8285–8295, 2023. 2, 6, 8, 1

[41] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20331–20341, 2024. 2, 3, 6, 7, 1

[42] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3D Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19447–19456, 2024. 2

[43] Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian Opacity Fields: Efficient Adaptive Surface Reconstruction in Unbounded Scenes. *ACM Transactions on Graphics*, 43 (6), 2024. 3

[44] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 6

[45] Allan Zhou, Moo Jin Kim, Lirui Wang, Pete Florence, and Chelsea Finn. NeRF in the palm of your hand: Corrective augmentation for robotics via novel-view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17907–17917, 2023. 2

[46] Junsheng Zhou, Weiqi Zhang, and Yu-Shen Liu. DiffGS: Functional Gaussian splatting diffusion. In *Advances in Neural Information Processing Systems*, pages 37535–37560, 2024. 2