

Sparse4DGS: 4D Gaussian Splatting for Sparse-Frame Dynamic Scene Reconstruction

Changyue Shi^{1,2}, Chuxiao Yang¹, Xinyuan Hu¹, Minghao Chen¹, Wenwen Pan¹,
Yan Yang¹, Jiajun Ding^{1*}, Zhou Yu¹, Jun Yu³

¹School of Computer Science and Technology, Hangzhou Dianzi University,

²School of AI for Science, Peking University,

³School of Computer Science and Technology, Harbin Institute of Technology
{shicy, chuxiao_yang, xinyuan, djj}@hdu.edu.cn

Abstract

Dynamic Gaussian Splatting approaches have achieved remarkable performance for 4D scene reconstruction. However, these approaches rely on dense-frame video sequences for photorealistic reconstruction. In real-world scenarios, due to equipment constraints, sometimes only sparse frames are accessible. In this paper, we propose **Sparse4DGS**, the first method for sparse-frame dynamic scene reconstruction. We observe that dynamic reconstruction methods fail in both canonical and deformed spaces under sparse-frame settings, especially in areas with high texture richness. Sparse4DGS tackles this challenge by focusing on texture-rich areas. For the deformation network, we propose Texture-Aware Deformation Regularization, which introduces a texture-based depth alignment loss to regulate Gaussian deformation. For the canonical Gaussian field, we introduce Texture-Aware Canonical Optimization, which incorporates texture-based noise into the gradient descent process of canonical Gaussians. Extensive experiments show that when taking sparse frames as inputs, our method outperforms existing dynamic or few-shot techniques on NeRF-Synthetic, HyperNeRF, NeRF-DS, and our iPhone-4D datasets.

Project Page — ChangyueShi.github.io/Sparse4DGS

1 Introduction

Modeling 4D scenes from 2D images and synthesizing photorealistic novel views is essential for computer vision and graphics (Pumarola et al. 2021; Yang et al. 2024b, 2023). This task has attracted significant attention from both industry and academia due to its great potential in augmented and virtual reality applications (Yang et al. 2024c; Jiang et al. 2024). The recent emergence of 3D Gaussian Splatting (Kerbl et al. 2023) has revolutionized dynamic scene reconstruction through its real-time rendering capabilities.

To reconstruct a dynamic scene, a variety of Gaussian Splatting methods (Huang et al. 2024; Yang et al. 2024b; Liang et al. 2023; Duisterhof et al. 2023; Wu et al. 2024a; Duan et al. 2024; Yang et al. 2023) have been proposed. The most commonly used approach explicitly represents the scene as a canonical Gaussian field and models Gaussian

variations at specific timestamps using a deformation network (Yang et al. 2024b; Wu et al. 2024a). However, existing methods depend on dense-frame video sequences as input to reconstruct high-fidelity dynamic scenes. In real-world applications, sometimes, only low-FPS video sequences are available due to device limitations.

This naturally raises the question: **Can high-quality 4D scenes be reconstructed from sparse frames?** We investigate the capability of 4D Gaussian Splatting for addressing sparse-frame inputs and demonstrates that it is possible to reconstruct a high-fidelity dynamic scene even from a limited sequence of frames. When the input image frames are sparse, existing methods (Yang et al. 2024b; Wu et al. 2024a) suffer a significant degradation. We observe that this degradation manifests in texture-rich regions, as these areas contain abundant high-frequency content that is challenging to preserve under deformation. In these areas, the geometry tends to collapse in deformed and canonical spaces, as shown in the second and fourth columns of Fig. 1.

In this work, we propose **Sparse4DGS**, a novel framework tailored for dynamic scene reconstruction from sparse-frame video sequences. Our intuition is that sparse-frame inputs inherently provide limited information, in which scenario, high-frequency texture signals become the primary source of rich detail and dynamic cues essential for accurate deformation modeling. Therefore, we encourage Gaussians to focus on these texture-rich regions, thereby modeling the underlying structure. Central to our approach is the *Texture Intensity (TI) Gaussian Field*: for each input frame, we compute per-pixel gradient magnitudes to derive 2D *TI* maps. We then embed this texture intensity into each 3D Gaussian by aligning rendered and extracted 2D *TI* maps.

We tackle the problems from the perspective of texture intensity consistency in both deformed and canonical spaces. For the Gaussian deformation, we propose *Texture-Aware Deformation Regularization (TADR)*. Apart from aligning the *TI* maps, we incorporate a texture-aware depth regularization to guide the geometric structure of the deformed space. Specifically, we extract texture intensity from depth maps to obtain *TI* maps of depth. The texture of rendered depth at a specific timestamp is aligned with the depth texture map extracted from the depth map predicted by Mono-Depth Estimator (Ranftl et al. 2021). For the canonical Gaussians, we propose *Texture-Aware Canonical Opti-*

*Corresponding Author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

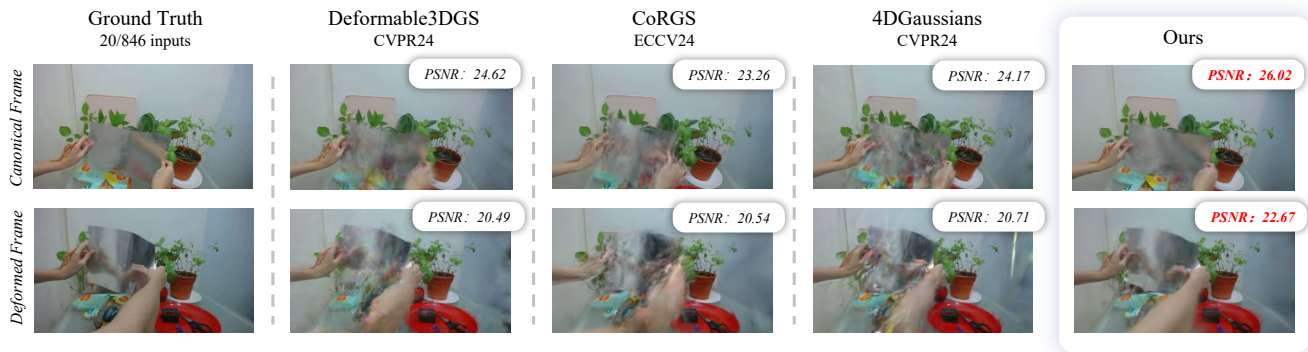


Figure 1: In this work, we introduce Sparse4DGS, a novel approach for dynamic scene reconstruction using sparse input frames. In the “Sheet” scene from the NeRF-DS (Yan, Li, and Lee 2023) dataset, when taking sparse frames as inputs, Sparse4DGS achieves high-quality novel view synthesis results in both canonical and deformed spaces.

mization (TACO). In *TACO*, the gradient descent of canonical Gaussians is reformulated based on Stochastic Gradient Langevin Dynamics (Welling and Teh 2011; Brosse et al. 2018; Kheradmand et al. 2024a,b) with a noise term based on the value of *TI* attribute. This noise term continuously perturbs the optimization of canonical Gaussians until they converge to texture-rich areas. When using inputs with either 5 FPS (low frame rate) or 30 FPS (high frame rate) for training, Sparse4DGS exhibits superior dynamic reconstruction, demonstrating its potential for a wide range of frame rate videos in real-world applications.

In summary, the major contributions of our work are:

- We demonstrate that high-quality dynamic scenes can be reconstructed from sparse-frame video sequences. To the best of our knowledge, this is the first work to focus on sparse-frame 4D scene reconstruction.
- We tackle the problems of optimization under sparse inputs from the perspective of high-frequent texture information and propose a texture-aware method (*TADR* and *TACO*) to align Gaussians with texture richness, thus preserving the underlying geometric structure.
- Extensive experiments demonstrate that Sparse4DGS outperforms previous dynamic and sparse NVS methods across multiple datasets, including NeRF-Synthetic, HyperNeRF, NeRF-DS, and our iPhone-4D dataset.

2 Related Work

Novel View Synthesis Novel view synthesis (NVS) aims to generate unseen views of a scene from a set of input images. Traditional methods, such as Structure-from-Motion (SfM) (Ullman 1979) and Multi-View Stereo (MVS) (Tomasi and Kanade 1992), rely on geometric reconstruction techniques. Neural Radiance Field (NeRF) (Mildenhall et al. 2021) introduces a learning-based approach, representing a scene as a volumetric radiance field parameterized by an MLP. While subsequent NeRF-based methods (Müller et al. 2022; Chen et al. 2022; Yu et al. 2021; Barron et al. 2021, 2022) have enhanced rendering quality and efficiency of the vanilla NeRF, they remain

constrained by the computational overhead of volumetric rendering. Recently, Gaussian Splatting (Kerbl et al. 2023) has emerged as an efficient alternative, leveraging rasterization to achieve real-time, high-fidelity scene reconstruction. It has inspired extensive researches across various domains, including dynamic scene reconstruction (Wu et al. 2024a; Yang et al. 2024b; Huang et al. 2024), few-shot learning (Zhu et al. 2025; Li et al. 2024; Zhang et al. 2025; Paliwal et al. 2025), and super-resolution reconstruction (Feng et al. 2024), among others (Shi et al. 2025a).

Gaussian Splatting for Dynamic Scenes Many studies have extended Gaussian Splatting to dynamic scene reconstruction. Research in this area can be broadly divided into two categories: deformation-based and 4D Gaussian-based methods. Deformation-based methods use an implicit structure (Huang et al. 2024; Yang et al. 2024b; Liang et al. 2023; Duisterhof et al. 2023; Wu et al. 2024a) to model the deformation from a static canonical Gaussian field. On the other hand, 4D Gaussian-based methods (Duan et al. 2024; Yang et al. 2023) introduce the time dimension in the original Gaussian Splatting formulation, which can be expressed with the 4D position and 4D covariance matrix. Many studies have now applied dynamic Gaussian Splatting to various fields. ZDySS (Saroha et al. 2025) proposes a zero-shot scene stylization method with dynamic Gaussian Splatting. Some methods (Shan et al. 2025) utilize it for reconstruction of surgical scenes. Deblur4DGS (Wu et al. 2024b) proposes a framework in dynamic Gaussian Splatting for blurry monocular video. Existing dynamic Gaussian Splatting methods rely on dense-frame video sequences as input for dynamic scene reconstruction. In this work, we propose Sparse4DGS and show that high-quality dynamic scenes can be reconstructed using sparse input frames.

Few Shot Gaussian Splatting Gaussian Splatting (Kerbl et al. 2023) has gained significant attention for its fast rendering speed and high-quality results across various applications (Feng et al. 2024; Qin et al. 2024; Lee et al. 2025; Gao et al. 2025; Choi et al. 2025; Shi et al. 2025b; Hu et al. 2025). Despite these advantages, novel view synthesis (NVS) still

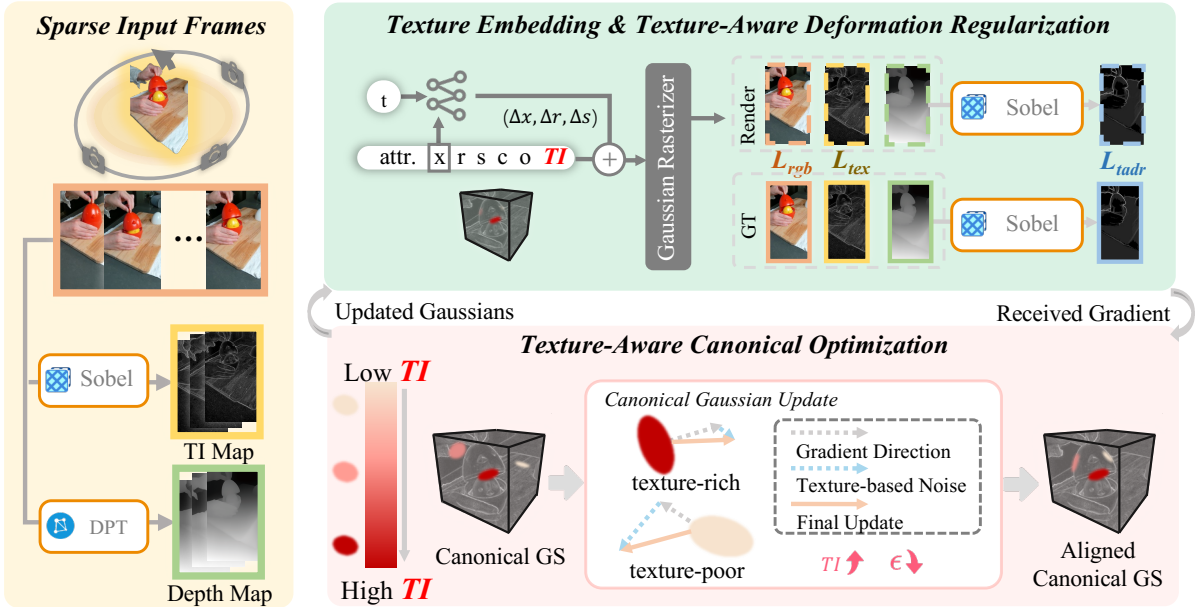


Figure 2: Overall pipeline of Sparse4DGS. *Left*: The Sobel operator and Mono-Depth Estimator (Ranftl et al. 2021) are employed to generate the texture intensity (TI) and depth maps from sparse frames. *Top right*: The TI attribute is embedded in each Gaussian via L_{tex} . *Texture-Aware Deformation Regularization* is employed to align the rendered and ground truth texture intensity of depth maps with L_{tadr} . *Bottom right*: After receiving the original gradient, *Texture-Aware Canonical Optimization* introduces an additional texture-based noise to each Gaussian, thereby improving their concentration on texture-rich regions.

requires hundreds of input images to achieve photo-realistic reconstructions, limiting its practicality in real-world scenarios. To address this, recent research has explored techniques for high-quality reconstruction from sparse inputs. GaussianObject (Yang et al. 2024a) leverages a diffusion model (Rombach et al. 2022) to synthesize photo-realistic novel views from just four input images. DNGaussian (Li et al. 2024) introduces a depth regularization module that enhances reconstruction efficiency and quality. CoRGS (Zhang et al. 2025) mitigates reconstruction inaccuracies by refining the training process. FSGS (Zhu et al. 2025) tackles the challenge of extremely sparse SfM initialization by introducing a Gaussian Unpooling mechanism. CoherentGS (Paliwal et al. 2025) proposes a structured Gaussian representation that enables explicit control in 2D image space. Additionally, generative approaches (Chen et al. 2025; Charatan et al. 2024) have been explored to predict Gaussians directly using feed-forward networks, further improving performance.

3 Preliminary

3.1 Dynamic Gaussian Splatting

Gaussian Splatting (Kerbl et al. 2023) represents scenes with anisotropic Gaussians $\{\mu_i, s_i, r_i, o_i, c_i\}$, where $\mu_i \in \mathbb{R}^3$ is the position, $s_i \in \mathbb{R}^3$ and $r_i \in \mathbb{R}^4$ define the covariance, $o_i \in \mathbb{R}$ is opacity, and $c_i \in \mathbb{R}^K$ encodes color. The rendering process follows alpha-blending:

$$c = \sum_{i=1}^n c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \quad (1)$$

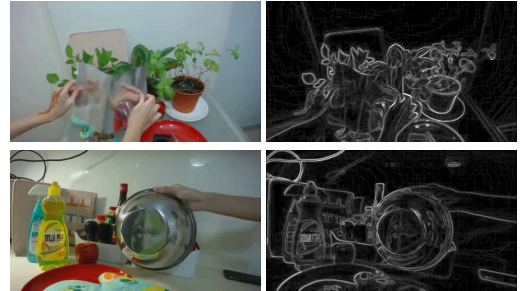


Figure 3: Visualization of texture intensity maps. *Left*: Input RGB images. *Right*: Extracted texture intensity maps.

where α_i is computed by projecting Gaussians to the image plane. To model dynamic scenes, recent methods (Yang et al. 2024b; Wu et al. 2024a) extend it with a deformation field. Given time t , the deformation MLP predicts offsets for each Gaussian:

$$(\delta x, \delta r, \delta s) = \mathcal{F}_\theta(\gamma(\text{sg}(x)), \gamma(t)), \quad (2)$$

where \mathcal{F}_θ is MLP, $\gamma(\cdot)$ denotes positional encoding (Mildenhall et al. 2021), and $\text{sg}(\cdot)$ stops gradient propagation. The deformed Gaussians $\{\mu_i + \delta x, s_i + \delta s, r_i + \delta r, o_i, c_i\}$ are rendered via Eq. 1. An RGB loss is utilized to optimize the canonical Gaussians and the deformation network:

$$L_{rgb} = (1 - \lambda)L_1(\hat{I}, I) + \lambda L_{SSIM}(\hat{I}, I), \quad (3)$$

where \hat{I} is the rendered image and I is the corresponding ground truth (GT). Here, L_1 represents the Mean Squared

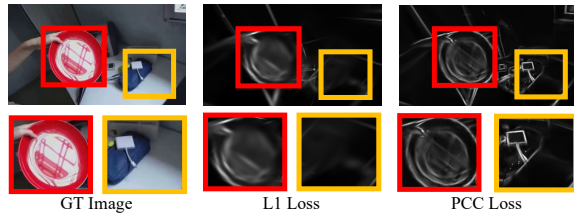


Figure 4: Comparison between L1 distance and PCC for the rendered texture map. L_{tex} with PCC achieves more precise texture embedding results.

Error (MSE) loss, while L_{SSIM} is the Structural Similarity Index Measure (SSIM) (Wang et al. 2004).

3.2 Stochastic Gradient Langevin Dynamics

The Stochastic Gradient Langevin Dynamics (SGLD) (Welling and Teh 2011; Brosse et al. 2018) method has been recently applied to novel view synthesis applications (Kheradmand et al. 2024a,b). The traditional SGD update rule (Amari 1993) in the original Gaussian Splatting can be described as:

$$g = g - \alpha \cdot \nabla_g \mathbb{E}_{I \sim \mathcal{I}} [L(g; I)], \quad (4)$$

where α is the learning rate, $\mathbb{E}_{I \sim \mathcal{I}}$ represents the expectation over the data distribution \mathcal{I} , and $L(g; I)$ is the total loss function that depends on the parameter g and the input image I . Meanwhile, the SGLD update takes the form as:

$$g = g + a \cdot \nabla_g \log \mathcal{P}(g) + b \cdot \epsilon, \quad (5)$$

where \mathcal{P} is the data-dependent probability density function of the target distribution. ϵ represents the noise distribution for exploration. Hyperparameters a and b control the trade-off between convergence speed and exploration. In Gaussian Splatting application (Kheradmand et al. 2024b), it can be rewritten as:

$$g = g - \alpha_g \cdot \nabla_g \mathbb{E}_{I \sim \mathcal{I}} [L(g; I)] + \alpha_{noise} \cdot \epsilon_o, \quad (6)$$

where the noise term ϵ_o can be described as follows:

$$\epsilon_o = \sigma(-k(o - t)) \cdot \sum \eta. \quad (7)$$

Here, $\eta \sim \mathcal{N}(0, I)$, and k and t are set to 100 and 0.995, respectively, to create a sharp transition function that shifts from zero to one. σ represents the sigmoid function and o is the opacity value of each Gaussian. In this work, we further extend this formulation to better suit dynamic scene reconstruction under sparse inputs.

4 Proposed Method

As illustrated in Fig. 2, we propose Sparse4DGS. In Sec. 4.1, we introduce *Texture Intensity Gaussian Field*. In Sec. 4.2, we introduce *Texture-Aware Deformation Regularization (TADR)*. In Sec. 4.3, we introduce the *Texture-Aware Canonical Optimization (TACO)*.

4.1 Representing 3D Texture Richness

To encourage texture consistency, we first propose the *Texture Intensity Gaussian Field*. We extract texture maps from input frames using a discrete differential operator. The texture information is then embedded into each Gaussian. **2D Texture Intensity Maps.** We represent texture richness using per-pixel gradients. These gradient values quantify the variation in colors between adjacent pixels and explicitly capture local texture richness.

Given an input RGB image $I \in \mathbb{R}^{H \times W \times 3}$, we first compute two gradient maps by convolving I with the standard horizontal and vertical Sobel operators, respectively:

$$TI_x = I * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, TI_y = I * \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \quad (8)$$

We then compute the per-pixel gradient magnitude:

$$TI_{gt}(i, j) = \sqrt{TI_x(i, j)^2 + TI_y(i, j)^2}, \quad (9)$$

where $TI(i, j)$ is the gradient magnitude at pixel (i, j) , which serves as an explicit measure of local texture richness. Fig. 3 presents the extracted texture intensity maps.

Texture Intensity in 3D Space. To represent the texture richness in 3D space, we introduce a new attribute *Texture Intensity (TI)* for each Gaussian. TI can be rendered into a texture map TI_{render} via differentiable rasterizer as shown in Eq. 1. Since the texture extractor is applied independently to each image, spatial inconsistencies may occur. The traditional L1 loss primarily focuses on absolute differences between pixels, disregarding spatial inconsistencies. To relax this issue, Sparse4DGS employs the Pearson Correlation Coefficient (PCC) (Cohen et al. 2009) to compute the relative variation rates between TI_{gt} and TI_{render} . The formulation of PCC can be described as:

$$\text{PCC}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}}, \quad (10)$$

where X and Y represent two distinct variables, $\text{Cov}(X, Y)$ denotes their covariance, and $\text{Var}(X)$ and $\text{Var}(Y)$ represent their variances. We optimize TI through the PCC loss between TI_{gt} and TI_{render} :

$$L_{tex} = 1 - \text{PCC}(TI_{gt}, TI_{render}). \quad (11)$$

This approach relaxes spatial inconsistencies and achieves superior texture embedding results (Fig. 4).

4.2 Texture-Aware Deformation Regularization

The optimization of Sparse4DGS builds upon previous dynamic Gaussian Splatting method (Yang et al. 2024b). The optimization performance under sparse frames is primarily influenced by the deformation network and the canonical Gaussian field. This section focuses on the regularization of the deformation field. We propose *Texture-Aware Deformation Regularization (TADR)* to guide the geometric structure of deformed Gaussians with depth-texture consistency.

Texture intensity is often correlated with changes in depth. For efficient geometric constraint, a previous method

Methods	NeRF-Synthetic (20 frames)			NeRF-DS (20 frames)			Hyper-NeRF (30 frames)			iPhone-4D (30 FPS)			iPhone-4D (5 FPS)		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Deform-3DGS	22.65	0.927	0.073	20.81	0.753	0.301	22.41	0.661	0.295	27.01	0.909	0.210	21.12	0.817	0.299
4DGaussians	22.47	0.931	0.071	19.70	0.697	0.350	20.64	0.637	0.414	28.79	0.896	0.246	16.37	0.678	0.421
CoRGS	20.15	0.920	0.089	19.86	0.746	0.319	20.50	0.638	0.364	21.58	0.851	0.266	16.81	0.737	0.374
Ours	25.31	0.944	0.056	22.34	0.801	0.233	23.91	0.711	0.294	29.81	0.929	0.177	27.51	0.910	0.205

Table 1: Quantitative comparison on multiple datasets. The metrics include PSNR \uparrow , SSIM \uparrow , and LPIPS \downarrow .

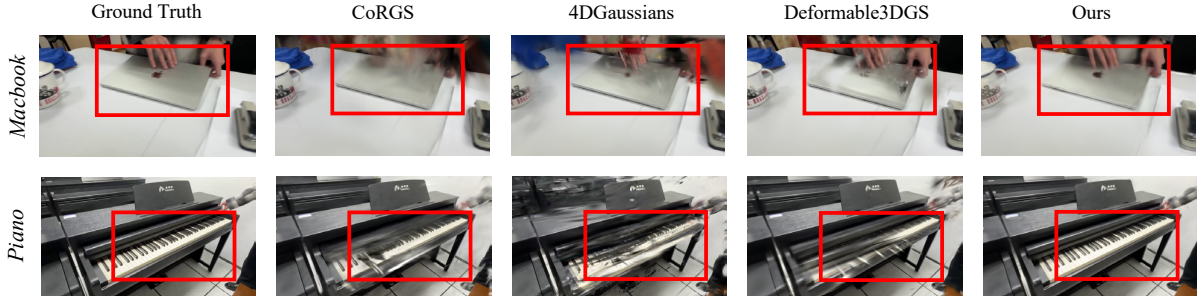


Figure 5: Visualization of various methods on our iPhone-4D dataset with 5 FPS video inputs.

(Zhu et al. 2025) computes the PCC between the rendered depth maps and the monocular depth estimated by DPT (Ranftl et al. 2021). This image-level regularization has limitations in capturing local depth variations (Fig. 8). To efficiently constrain the deformation network, Sparse4DGS develops a texture-based regularization method.

To begin with, TADR utilizes Sobel (Sec. 4.1) to extract texture intensity maps information of the rendered depth D_{render} and depth computed from DPT (Ranftl et al. 2021) D_{dpt} : $TI_{render}^{depth} = TE(D_{render})$, $TI_{gt}^{depth} = TE(D_{dpt})$. Therefore, the TADR loss can be formulated as:

$$L_{tadr} = 1 - \text{PCC}(TI_{gt}^{depth}, TI_{render}^{depth}). \quad (12)$$

Similar to L_{tex} , TADR employs PCC for supervision, since TI_{gt}^{depth} exhibits spatial inconsistency. Therefore, the overall training loss can be described as:

$$L = L_{rgb} + \lambda_1 \cdot L_{tex} + \lambda_2 \cdot L_{tadr}, \quad (13)$$

where L_{rgb} consists of an MSE loss and an SSIM loss (Kerbl et al. 2023), as shown in Eq. 3.

4.3 Texture-Aware Canonical Optimization

The supervision defined in Eq. 13 generates a per-Gaussian gradient $\nabla_g \mathbb{E}_{I \sim \mathcal{I}} [L(g; I)]$, which drives the canonical Gaussian to update its attributes using the SGD method, as formalized in Eq. 4. To help canonical Gaussian field concentrate more on texture-rich areas, we reformulate this update function based on SGLD (Welling and Teh 2011; Kheradmand et al. 2024a,b) and propose *Texture-Aware Canonical Optimization (TACO)*.

The main idea of TACO is to introduce stochastic noise in each iteration during the Gaussian update. Our update formulation for the canonical Gaussians can be described as:

$$g = g - \alpha_g \cdot \nabla_g \mathbb{E}_{I \sim \mathcal{I}} [L(g; I)] + \alpha_{noise} \cdot (\epsilon_{tex} + \epsilon_o), \quad (14)$$

where the ϵ_o is employed to reduce ambiguous Gaussians, as described in Eq. 7, which often manifest as floaters. TACO introduces an additional texture aware term ϵ_{tex} , which is defined as: $\epsilon_{tex} = \sigma(-k(TI - t)) \cdot \sum \eta$. Since the value range of TI is identical to that of $opacity$, we use the same hyperparameters as those in Eq. 7. A Gaussian with a greater TI value would receive a smaller noise term. When a Gaussian reaches texture-rich regions, its TI value approaches one, and ϵ_{tex} approaches zero. Consequently, ϵ_{tex} continuously influences the optimization process until the Gaussians attain high $opacity$ and Tex values, indicating that they have reached texture-rich regions.

5 Experiments

5.1 Experimental Settings

Dynamic datasets. The NeRF-Synthetic dataset (Pumarola et al. 2021) contains 8 synthetic scenes. The NeRF-DS dataset (Yan, Li, and Lee 2023) consists of 7 real-world scenes, each containing images captured using two monocular cameras. For NeRF-Synthetic and NeRF-DS datasets, we uniformly samples 20, 30, and 40 frames from their original training sets for model training and use the same evaluation sets (keeping the original number of frames) as previous works (Yang et al. 2024b). The Hyper-NeRF dataset (Park et al. 2021) includes 17 real-world scenes, with 5 typical scenes selected in this work for comparison. For Hyper-NeRF dataset, we uniformly samples 10, 20, and 30 frames from the original training set and adopt the same evaluation set as previous methods (Yang et al. 2024b).

iPhone-4D dataset. To demonstrate the capability of Sparse4DGS in various real-world frame-rate application, we introduce the iPhone-4D dataset, which consists of four real-world scenes. This dataset is captured using an iPhone at 30 FPS. We evaluate our method on this dataset using training inputs at 5 FPS and 30 FPS. The selection of evalu-

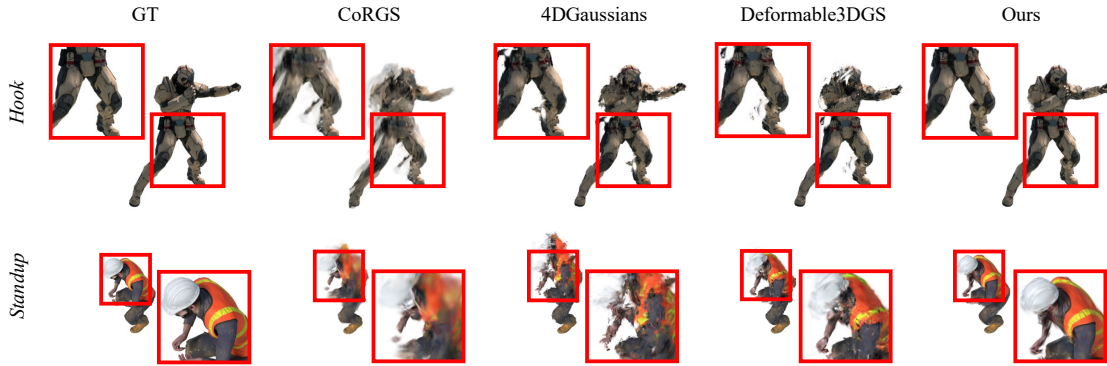


Figure 6: Qualitative results on the NeRF-Synthetic dataset with 20 inputs. The results of various dynamic reconstruction methods are demonstrated from new viewpoints and also in new frames.



Figure 7: Qualitative results on the NeRF-DS dataset with 20 inputs. Our method consistently outperforms other baselines.

ation set follows Hyper-NeRF dataset. We provide details of this dataset in the supplementary materials.

Baselines and metrics. We compare Sparse4DGS with typical dynamic and few-shot scene reconstruction methods including Deformable3DGS (CVPR24) (Yang et al. 2024b), 4DGaussians (CVPR24) (Wu et al. 2024a), and CoRGS (ECCV24) (Zhang et al. 2025). We run these methods on four datasets without modifying their original settings. In the following sections, we report the average PSNR, SSIM (Wang et al. 2004), and LPIPS (Zhang et al. 2018).

5.2 Main Results with Sparse Inputs

In this section, we evaluate our method and baselines using sparse-frame inputs to explore the potential of Sparse4DGS.

Quantitative results. A subset of the quantitative results is presented in Tab. 1. For most datasets, our method improves PSNR by more than 1 dB. Previous methods fail to reconstruct a high-fidelity 4D scene with sparse inputs. Sparse4DGS consistently achieves the best performance on three publicly available datasets. These results confirm the ability of Sparse4DGS to achieve photorealistic reconstruction even with sparse input frames.

Qualitative results. Fig. 6 and Fig. 7 present the visualization results of Sparse4DGS and other methods. The evaluation set consists of novel viewpoints (spatial interpolation) and novel temporal frames. Although CoRGS (Zhang et al.

2025) exhibits a strong ability to synthesize novel views and novel frames from sparse inputs, it fails to effectively handle dynamic objects. Deformable3DGS (Yang et al. 2024b) and 4DGaussians (Wu et al. 2024a) experience significant performance drops, particularly in regions around object edges. With *TACO* and *TADR*, the Gaussians in Sparse4DGS are rigorously constrained to align with object edges, thereby preserving detailed structural information.

5.3 Multi Frame Rate Application

In this section, to evaluate our method in the real-world application, we test Sparse4DGS on the iPhone-4D dataset using input sequences at 5 FPS and 30 FPS. We present the quantitative results in Tab. 1. Our method consistently outperforms previous baselines for input sequences both at 5 FPS and 30 FPS. In real-world frame-rate applications, Sparse4DGS demonstrates superior reconstruction performance, particularly with extremely low-frame-rate input videos. The qualitative results are shown in Fig. 5. Our method exhibits more accurate reconstruction of dynamic object edges. In the case of the “*MacBook*”, our method reconstructs a clear Apple logo, while other baselines suffer from blurry results. These results demonstrate that our method can achieve photo-realistic novel view and novel frame synthesis result with sparse input frame sequences. The results on iPhone-4D demonstrate the effectiveness of

Methods	PSNR	SSIM	LPIPS
Baseline	20.81	0.753	0.301
w/o TADR	21.89	0.792	0.245
w/o TACO	21.33	0.773	0.271
Ours	22.34	0.801	0.233

(a) Ablation Study on TADR and TACO.

L_{tex} weight	PSNR	SSIM	LPIPS
$\lambda_1 = 0.001$	21.80	0.789	0.249
$\lambda_1 = 0.01$	22.34	0.801	0.233
$\lambda_1 = 0.1$	21.79	0.788	0.254
$\lambda_1 = 1$	20.17	0.714	0.353

(b) Ablation study on L_{tex} weight λ_1 .

L_{tadr} weight	PSNR	SSIM	LPIPS
$\lambda_2 = 0.001$	22.04	0.797	0.240
$\lambda_2 = 0.01$	22.34	0.801	0.233
$\lambda_2 = 0.1$	22.12	0.797	0.241
$\lambda_2 = 1$	21.86	0.794	0.243

(c) Selection of hyper-parameter λ_2 .

Methods	PSNR	SSIM	LPIPS
TACO w/o ϵ_o	21.81	0.792	0.246
TACO w/o ϵ_{tex}	21.57	0.783	0.260
Ours	22.34	0.801	0.233

(d) Ablation study on ϵ_o or ϵ_{tex} in TACO.

Methods	PSNR	SSIM	LPIPS
L_{tex} w/o PCC	21.71	0.789	0.245
L_{tadr} w/o PCC	22.09	0.797	0.239
Ours	22.34	0.801	0.233

(e) PCC loss (replaced with L1 loss).

Methods	PSNR	SSIM	LPIPS
w/o texture-aware	21.46	0.775	0.277
Ours	22.34	0.801	0.233

(f) Texture-aware depth loss in TADR.

Table 2: Ablation study. We ablate our method on NeRF-DS dataset with 20 input views.

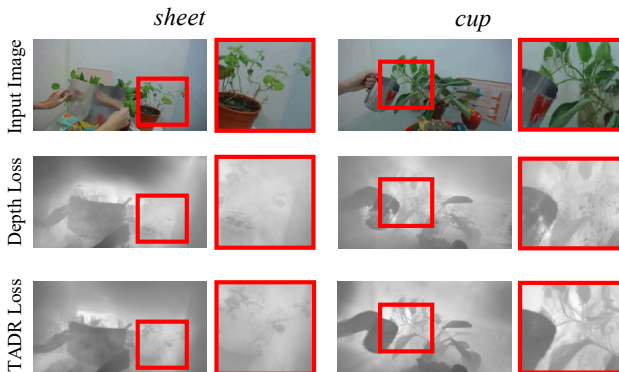


Figure 8: Illustration of TADR. TADR demonstrates superior performance in capturing local depth variations.

Sparse4DGS in real-world frame-rate applications. Detailed quantitative and qualitative results can be found in our supplementary materials.

5.4 Ablation Study

We ablate the effectiveness of Sparse4DGS on NeRF-DS dataset with 20 input images.

Architecture modules. In this section, we conduct a detailed analysis of the effectiveness of our proposed modules. The experimental results are summarized in Tab. 2a and Fig. 9. Removing TADR and TACO causes PSNR drops of 0.45 and 1.01, respectively, demonstrating their critical role in optimizing a dynamic scene.

Selection of λ_1 and λ_2 . Hyperparameter tuning results in Tab. 2b and Tab. 2c. Setting λ_1 and λ_2 to 0.01 yields the best performance for Sparse4DGS.

Noise terms of TACO. TACO consists of two noise terms: ϵ_o and ϵ_{tex} . Removing either causes a performance drop (Tab. 2d). Without ϵ_{tex} , Gaussians fail to concentrate in texture-rich areas, while ϵ_o helps converge ambiguous Gaussians that are too opaque for pruning.

Ablation study on PCC. Fig. 4 shows the effectiveness of PCC. Replacing PCC loss in L_{tex} and L_{tadr} with L1 loss results in performance drops of 0.6 and 0.3. This is because L1

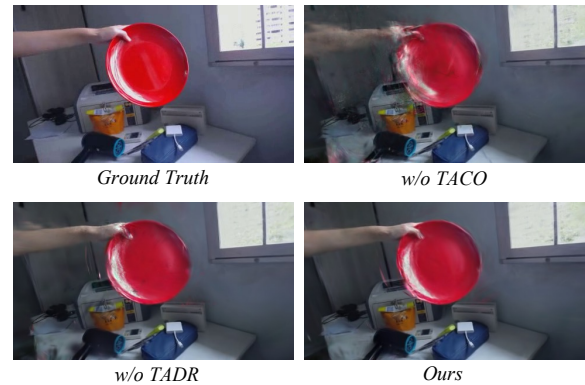


Figure 9: Illustration of our architecture modules. Removing TADR or TACO results in performance drop.

loss overlooks spatial inconsistencies in multi-view texture and depth maps (Tab. 2e).

Ablation study on texture-aware depth loss in TADR. Unlike conventional depth regularization methods (He et al. 2025; Zhu et al. 2025; Kumar and Vats 2024; Gao et al. 2024), this work introduces a texture-aware depth loss (Eq.12) in TADR. We compare our method with a conventional depth loss, which directly calculates PCC between D_{render} and D_{dpt} . As shown in Fig. 8 and Tab. 2f, both quantitative and qualitative results demonstrate the effectiveness of adopting texture in depth loss.

6 Conclusion

In this work, we introduce **Sparse4DGS**, the first method for dynamic scene reconstruction using sparse input frame sequences. We identify the challenges in optimizing deformation and canonical fields, particularly in regions with pronounced texture variations. To tackle this, we propose a texture-aware method (TADR and TACO) that aligns Gaussians with texture information. We evaluate **Sparse4DGS** on our iPhone-4D dataset with frame sequences captured at 5 FPS or 30 FPS, representing a significant advancement for real-world applications in dynamic scene reconstruction.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants (No. 62206082, 62422204, 62502135, 62402152, and 62406093), the Key Research and Development Program of Zhejiang Province (No. 2025C01026), the Zhejiang Provincial Natural Science Foundation of China under Grants (No. LDT23F02025F02, LQ24F020032, LQN25F030014, and LQN25F020017). This research was also supported by the National College Student Innovation and Entrepreneurship Training Program of China under Grants (No. 202410336018 and 202510336016).

References

- Amari, S.-i. 1993. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5): 185–196.
- Barron, J. T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; and Srinivasan, P. P. 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5855–5864.
- Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5470–5479.
- Brosse, N.; Durmus, A.; Moulines, E.; and and. 2018. The promises and pitfalls of stochastic gradient Langevin dynamics. *Advances in Neural Information Processing Systems*, 31.
- Charatan, D.; Li, S. L.; Tagliasacchi, A.; and Sitzmann, V. 2024. pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19457–19467.
- Chen, A.; Xu, Z.; Geiger, A.; Yu, J.; and Su, H. 2022. Tensorf: Tensorial radiance fields. In *European conference on computer vision*, 333–350. Springer.
- Chen, Y.; Xu, H.; Zheng, C.; Zhuang, B.; Pollefeys, M.; Geiger, A.; Cham, T.-J.; and Cai, J. 2025. Mvsplat: Efficient 3d gaussian splatting from sparse multi-view images. In *European Conference on Computer Vision*, 370–386. Springer.
- Choi, S.; Song, H.; Kim, J.; Kim, T.; and Do, H. 2025. Click-gaussian: Interactive segmentation to any 3d gaussians. In *European Conference on Computer Vision*, 289–305. Springer.
- Cohen, I.; Huang, Y.; Chen, J.; Benesty, J.; Benesty, J.; Chen, J.; Huang, Y.; and Cohen, I. 2009. Pearson correlation coefficient. *Noise reduction in speech processing*, 1–4.
- Duan, Y.; Wei, F.; Dai, Q.; He, Y.; Chen, W.; and Chen, B. 2024. 4d gaussian splatting: Towards efficient novel view synthesis for dynamic scenes. *arXiv preprint arXiv:2402.03307*.
- Duisterhof, B. P.; Mandi, Z.; Yao, Y.; Liu, J.-W.; Shou, M. Z.; Song, S.; and Ichnowski, J. 2023. Md-splatting: Learning metric deformation from 4d gaussians in highly deformable scenes. *arXiv preprint arXiv:2312.00583*.
- Feng, X.; He, Y.; Wang, Y.; Yang, Y.; Li, W.; Chen, Y.; Kuang, Z.; Fan, J.; Jun, Y.; et al. 2024. Srgs: Super-resolution 3d gaussian splatting. *arXiv preprint arXiv:2404.10318*.
- Gao, J.; Gu, C.; Lin, Y.; Li, Z.; Zhu, H.; Cao, X.; Zhang, L.; and Yao, Y. 2024. Relightable 3d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing. In *European Conference on Computer Vision*, 73–89. Springer.
- Gao, J.; Gu, C.; Lin, Y.; Li, Z.; Zhu, H.; Cao, X.; Zhang, L.; and Yao, Y. 2025. Relightable 3d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing. In *European Conference on Computer Vision*, 73–89. Springer.
- He, Z.; Xiao, Z.; Chan, K.-C.; Zuo, Y.; Xiao, J.; and Lam, K.-M. 2025. See In Detail: Enhancing Sparse-view 3D Gaussian Splatting with Local Depth and Semantic Regularization. *arXiv preprint arXiv:2501.11508*.
- Hu, X.; Shi, C.; Yang, C.; Chen, M.; Gu, X.; Ding, J.; He, J.; and Fan, J. 2025. Texture-aware 3D Gaussian Splatting for sparse view reconstructions. *Applied Soft Computing*, 113530.
- Huang, Y.-H.; Sun, Y.-T.; Yang, Z.; Lyu, X.; Cao, Y.-P.; and Qi, X. 2024. Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4220–4230.
- Jiang, Y.; Yu, C.; Xie, T.; Li, X.; Feng, Y.; Wang, H.; Li, M.; Lau, H.; Gao, F.; Yang, Y.; et al. 2024. Vr-gs: A physical dynamics-aware interactive gaussian splatting system in virtual reality. In *ACM SIGGRAPH 2024 Conference Papers*, 1–1.
- Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4): 139–1.
- Kheradmand, S.; Rebain, D.; Sharma, G.; Isack, H.; Kar, A.; Tagliasacchi, A.; and Yi, K. M. 2024a. Accelerating Neural Field Training via Soft Mining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20071–20080.
- Kheradmand, S.; Rebain, D.; Sharma, G.; Sun, W.; Tseng, J.; Isack, H.; Kar, A.; Tagliasacchi, A.; and Yi, K. M. 2024b. 3D Gaussian Splatting as Markov Chain Monte Carlo. *arXiv preprint arXiv:2404.09591*.
- Kumar, R.; and Vats, V. 2024. Few-shot Novel View Synthesis using Depth Aware 3D Gaussian Splatting. *arXiv preprint arXiv:2410.11080*.
- Lee, B.; Lee, H.; Sun, X.; Ali, U.; and Park, E. 2025. Deblurring 3d gaussian splatting. In *European Conference on Computer Vision*, 127–143. Springer.
- Li, J.; Zhang, J.; Bai, X.; Zheng, J.; Ning, X.; Zhou, J.; and Gu, L. 2024. Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20775–20785.
- Liang, Y.; Khan, N.; Li, Z.; Nguyen-Phuoc, T.; Lanman, D.; Tompkin, J.; and Xiao, L. 2023. Gaufre: Gaussian deformation fields for real-time dynamic novel view synthesis. *arXiv preprint arXiv:2312.11458*.

- Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.
- Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4): 1–15.
- Paliwal, A.; Ye, W.; Xiong, J.; Kotovenko, D.; Ranjan, R.; Chandra, V.; and Kalantari, N. K. 2025. Coherentgs: Sparse novel view synthesis with coherent 3d gaussians. In *European Conference on Computer Vision*, 19–37. Springer.
- Park, K.; Sinha, U.; Hedman, P.; Barron, J. T.; Bouaziz, S.; Goldman, D. B.; Martin-Brualla, R.; and Seitz, S. M. 2021. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *arXiv preprint arXiv:2106.13228*.
- Pumarola, A.; Corona, E.; Pons-Moll, G.; and Moreno-Noguer, F. 2021. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10318–10327.
- Qin, M.; Li, W.; Zhou, J.; Wang, H.; and Pfister, H. 2024. Langsplat: 3d language gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20051–20060.
- Ranftl, R.; Bochkovskiy, A.; Koltun, V.; and and. 2021. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, 12179–12188.
- Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.
- Saroha, A.; Hofherr, F.; Gladkova, M.; Curreli, C.; Litany, O.; and Cremers, D. 2025. ZDySS—Zero-Shot Dynamic Scene Stylization using Gaussian Splatting. *arXiv preprint arXiv:2501.03875*.
- Shan, J.; Cai, Z.; Hsieh, C.-T.; Cheng, S. S.; and Wang, H. 2025. Deformable Gaussian Splatting for Efficient and High-Fidelity Reconstruction of Surgical Scenes. *arXiv preprint arXiv:2501.01101*.
- Shi, C.; Chen, M.; Mao, Y.; Yang, C.; Hu, X.; Ding, J.; and Yu, Z. 2025a. REALM: An MLLM-Agent Framework for Open World 3D Reasoning Segmentation and Editing on Gaussian Splatting. *arXiv preprint arXiv:2510.16410*.
- Shi, C.; Yang, C.; Hu, X.; Yang, Y.; Ding, J.; and Tan, M. 2025b. MMSGs: Multi-Model Synergistic Gaussian Splatting for Sparse View Synthesis. *Image and Vision Computing*, 158: 105512.
- Tomasi, C.; and Kanade, T. 1992. Shape and motion from image streams under orthography: a factorization method. *International journal of computer vision*, 9: 137–154.
- Ullman, S. 1979. The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153): 405–426.
- Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.
- Welling, M.; and Teh, Y. W. 2011. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, 681–688. Citeseer.
- Wu, G.; Yi, T.; Fang, J.; Xie, L.; Zhang, X.; Wei, W.; Liu, W.; Tian, Q.; and Wang, X. 2024a. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20310–20320.
- Wu, R.; Zhang, Z.; Chen, M.; Fan, X.; Yan, Z.; and Zuo, W. 2024b. Deblur4DGS: 4D Gaussian Splatting from Blurry Monocular Video. *arXiv preprint arXiv:2412.06424*.
- Yan, Z.; Li, C.; and Lee, G. H. 2023. Nerf-ds: Neural radiance fields for dynamic specular objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8285–8295.
- Yang, C.; Li, S.; Fang, J.; Liang, R.; Xie, L.; Zhang, X.; Shen, W.; and Tian, Q. 2024a. Gaussianobject: High-quality 3d object reconstruction from four views with gaussian splatting. *ACM Transactions on Graphics (TOG)*, 43(6): 1–13.
- Yang, Z.; Gao, X.; Zhou, W.; Jiao, S.; Zhang, Y.; and Jin, X. 2024b. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20331–20341.
- Yang, Z.; Pan, Z.; Zhu, X.; Zhang, L.; Jiang, Y.-G.; and Torr, P. H. 2024c. 4D Gaussian Splatting: Modeling Dynamic Scenes with Native 4D Primitives. *arXiv preprint arXiv:2412.20720*.
- Yang, Z.; Yang, H.; Pan, Z.; and Zhang, L. 2023. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. *arXiv preprint arXiv:2310.10642*.
- Yu, A.; Fridovich-Keil, S.; Tancik, M.; Chen, Q.; Recht, B.; and Kanazawa, A. 2021. Plenoxels: Radiance fields without neural networks. *arXiv preprint arXiv:2112.05131*, 2(3): 6.
- Zhang, J.; Li, J.; Yu, X.; Huang, L.; Gu, L.; Zheng, J.; and Bai, X. 2025. CoR-GS: sparse-view 3D Gaussian splatting via co-regularization. In *European Conference on Computer Vision*, 335–352. Springer.
- Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.
- Zhu, Z.; Fan, Z.; Jiang, Y.; and Wang, Z. 2025. Fsgs: Real-time few-shot view synthesis using gaussian splatting. In *European conference on computer vision*, 145–163. Springer.