

# ADC-GS: Anchor-Driven Deformable and Compressed Gaussian Splatting for Dynamic Scene Reconstruction

He Huang<sup>1</sup>, Qi Yang<sup>2</sup>, Mufan Liu<sup>1</sup>, Yiling Xu<sup>1</sup>, Zhu Li<sup>2</sup>

<sup>1</sup>Shanghai Jiao Tong University

<sup>2</sup>University of Missouri-Kansas City

{huanghe0429, sudo\_evan, yl.xu}@sjtu.edu.cn, {qiyang, lizhu}@umkc.edu

## Abstract

Existing 4D Gaussian Splatting methods rely on per-Gaussian deformation from a canonical space to target frames, which overlooks redundancy among adjacent Gaussian primitives and results in suboptimal performance. To address this limitation, we propose Anchor-Driven Deformable and Compressed Gaussian Splatting (ADC-GS), a compact and efficient representation for dynamic scene reconstruction. Specifically, ADC-GS organizes Gaussian primitives into an anchor-based structure within the canonical space, enhanced by a temporal significance-based anchor refinement strategy. To reduce deformation redundancy, ADC-GS introduces a hierarchical coarse-to-fine pipeline that captures motions at varying granularities. Moreover, a rate-distortion optimization is adopted to achieve an optimal balance between bitrate consumption and representation fidelity. Experimental results demonstrate that ADC-GS outperforms the per-Gaussian deformation approaches in rendering speed by 300%-800% while achieving state-of-the-art storage efficiency without compromising rendering quality. The code is released at <https://github.com/H-Huang774/ADC-GS.git>.

## 1 Introduction

Dynamic scene reconstruction from multi-view input videos has received significant attention due to its wide applications. Beyond methods based on 3D Gaussian Splatting (3DGS) [Kerbl *et al.*, 2023], 4D Gaussian Splatting (4DGS) has demonstrated substantial advances for dynamic scene reconstruction due to its impressive visual quality with ultra-fast training speed compared to neural radiance fields (NeRFs) based methods [Mildenhall *et al.*, 2021].

Recent research on 4DGS approaches has primarily focused on two categories. The first category employs 4D Gaussians to approximate the 4D volumes of scenes [Li *et al.*, 2024; Yang *et al.*, 2024b; Yan *et al.*, 2024] through temporal opacity and polynomial functions for each Gaussian. Another category focuses on temporally deforming a canonical space to target frames [Bae *et al.*, 2025; Wu *et al.*, 2024; Yang *et al.*, 2024c; Lu *et al.*, 2024a], utilizing a multilayer

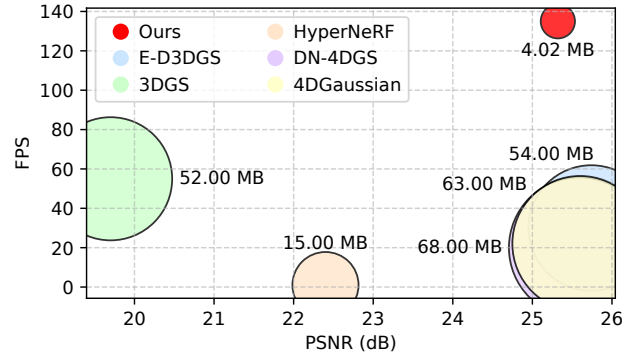


Figure 1: Comparison with concurrent dynamic scene reconstruction methods on the HyperNeRF dataset. Our method achieves the smallest storage size and the highest rendering speed while preserving excellent rendering quality.

perceptron (MLP) with latent embeddings to predict changes in 3D Gaussian attributes over time.

Directly optimizing 4D Gaussians provides higher rendering speeds but requires a large number of Gaussians to model the entire sequence, resulting in both significant training time and substantial storage requirements. For example, [Yan *et al.*, 2024] often leads to over 9 hours of training time and 2GB of data for a 10-second video. Although deformation-based methods address the above issues and reduce training time, they still require a large number of Gaussians to achieve high-quality rendering. Consequently, substantial storage and bandwidth requirements emphasize the need for more compact 4DGS representations and advanced compression techniques, which is the primary focus of this work.

Some recent methods have developed effective technologies to reduce 4DGS storage. For example, 4DGaussian [Wu *et al.*, 2024] maps the 4D space onto six orthogonal planes as latent embeddings, effectively mitigating per-frame training redundancy and lowering storage costs. Since spherical harmonic coefficients dominate storage consumption, MEGA [Zhang *et al.*, 2024] decomposes them into a per-Gaussian DC color component and a lightweight AC color predictor, eliminating redundant coefficients to reduce storage requirements. However, these methods compress each Gaussian independently in the canonical space, neglecting the strong similarity between local attributes (excluding opacity), as shown in Figure 2. Furthermore, existing methods rely

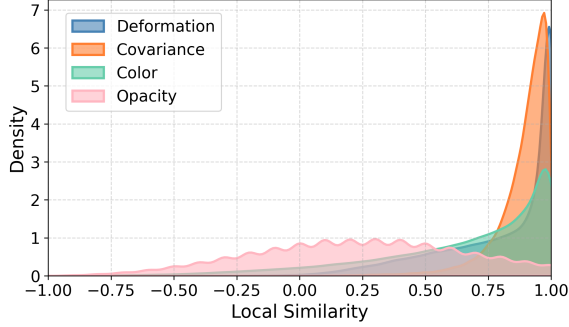


Figure 2: Illustration of local similarities of different features in [Bae *et al.*, 2025]. The local similarity is measured by the average cosine distances between a Gaussian primitive and its 20 neighbors with minimal Euclidean distance.

on per-Gaussian deformation for dynamic scene modeling, overlooking the consistent deformations among neighboring Gaussians since local scenes can often be approximated as rigid motion. These limitations hinder both the compactness of dynamic scene representations and rendering efficiency.

Inspired by deformation-based dynamic scene reconstruction, this paper proposes **Anchor-driven Deformable and Compressed Gaussian Splatting (ADC-GS)**, a novel framework for efficient and compact dynamic scene representation. Specifically, we organize Gaussian primitives into a sparse set of anchors, facilitating streamlined representation and processing. These anchors, with  $K$  neural Gaussian primitives predicted from each by shared MLPs, collectively construct the canonical space. To reconstruct dynamic scenes at any frame, we employ an anchor-based coarse-to-fine deformation strategy. In the coarse stage, the position, covariance, and color attributes of the anchors are deformed from the canonical space. Subsequently, the  $K$  Gaussian primitives are automatically updated based on the deformation of the associated anchor, notably reducing deformation redundancy. As the coarse stage focuses on capturing global deformation, a fine stage is introduced to dynamically refine the appearance of each primitive to recover fine-grained details. To further improve the compactness of the anchors, we implement a rate-distortion optimization scheme that incorporates a multi-dimension entropy model for accurate bitrate estimation and efficient compression. Besides, we propose a novel anchor refinement method that leverages the accumulated gradients of each Gaussian primitive’s temporal significance to guide anchor growing, while using accumulated opacity to determine anchor pruning. This adaptive strategy robustly addresses both under-reconstruction and over-reconstruction issues in dynamic scenes, ensuring a balanced and efficient anchor representation. As shown in Figure 1, ADC-GS achieves state-of-the-art (SOTA) storage size and rendering speed while preserving excellent rendering quality compared to prior works [Bae *et al.*, 2025; Park *et al.*, 2021b; Lu *et al.*, 2024a; Wu *et al.*, 2024; Kerbl *et al.*, 2023]. Our contributions can be summarized as follows:

- We propose anchor-driven deformable and compressed Gaussian Splatting (ADC-GS) for dynamic scene reconstruction. By leveraging compact anchors to efficiently

model 4D scenes, our approach achieves an extraordinary storage reduction of up to  $200\times$  over existing 4DGS methods.

- To accelerate rendering while preserving high reconstruction quality, we introduce a hierarchical coarse-to-fine deformation method. We also propose an adaptive anchor refinement strategy to address under-reconstruction and over-reconstruction issues.
- We develop a multi-dimension entropy model to enhance compactness through joint minimization of rendering distortion and bitrate consumption.

## 2 Related Work

### 2.1 Considering Spatial Relationships of Gaussian Primitives

In 3D space, Scaffold-GS [Lu *et al.*, 2024b] and CompGS [Liu *et al.*, 2024b] reconstruct scenes by synthesizing Gaussian primitives from anchors, exploiting the similarity of properties among neighboring Gaussians. HAC [Chen *et al.*, 2025] utilizes a structured hash grid to take advantage of inherent consistencies among unorganized 3D Gaussians. In 4D scenes, some approaches [Yang *et al.*, 2024c; Zhao *et al.*, 2024; Chen *et al.*, 2024a] introduce sparse control points combined with an MLP to model scene motion, based on the insight that motion can be effectively represented by a sparse set of basis. Inspired by the compact structures used in both 3D and 4D scenes, we design an anchor-driven deformable and compressed Gaussian splatting method to effectively eliminate intra-redundancy in both attributes and deformation among local Gaussian primitives.

### 2.2 Deforming 3D Canonical Space

Early research on deforming 3D canonical space to a target frame in both NeRF and 4DGS has explored various approaches. Nerfies [Mildenhall *et al.*, 2021] and HyperNeRF [Park *et al.*, 2021b] use per-frame trainable deformation rather than time-based conditions. D-NeRF [Pumarola *et al.*, 2021] reconstructs dynamic scenes by deforming ray samples over time, utilizing a deformation network that takes 3D coordinates and timestamps as inputs. E-D3DGS [Bae *et al.*, 2025] defines the deformation as a function of Gaussian and temporal embeddings, decomposing the deformation into coarse and fine stages to model slow and fast movements, respectively. Other methods [Yang *et al.*, 2024c; Wu *et al.*, 2024; Jiawei *et al.*, 2024] learn implicit position mappings to deform primitives using latent embeddings, such as positional embeddings. However, these methods overlook the similarity in deformation among neighboring Gaussian primitives, resulting in suboptimal rendering speeds. To address this issue, we propose an anchor-based coarse-to-fine deformation method that accounts for deformation redundancy, achieving exceptional rendering speed while maintaining high reconstruction quality.

### 2.3 3DGS Compression

Recent studies on 3DGS compression can be categorized into processing-based and context-based methods. Processing-based methods include significance pruning [Fan *et al.*,

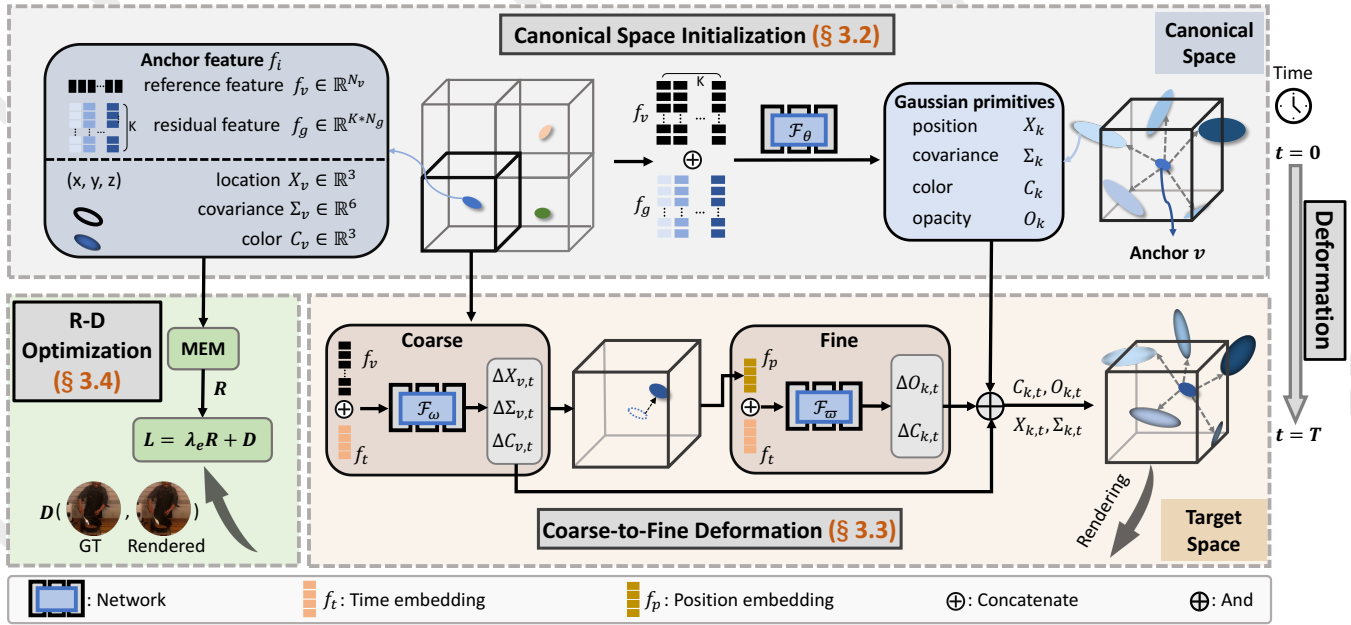


Figure 3: Overview of our ADC-GS framework. **Top:** ADC-GS organizes Gaussian primitives into a sparse set of anchors and compact residuals within canonical space. **Bottom right:** Gaussian primitives used for rendering are deformed from canonical space through a coarse-to-fine strategy based on anchors. **Bottom left:** Rendering distortion and estimated bitrates from the MEM are jointly minimized to balance rendering quality and storage efficiency.

2023], scalar quantization [Niedermayr *et al.*, 2024; Ma *et al.*, 2024], codebooks [Lee *et al.*, 2024]. Some point cloud compression techniques can also be adapted for Gaussian attributes due to their similar structure, such as graph-based signal transformation [Yang *et al.*, 2024a], region-adaptive hierarchical transform [Huang *et al.*, 2024], and octree encoding [Liu *et al.*, 2024a; Fan *et al.*, 2022]. Context-based methods leverage various contextual information to remarkably compress Gaussian attributes, including hash-grid features [Chen *et al.*, 2025], hyperpriors [Liu *et al.*, 2024b], and hybrid Gaussian features [Chen *et al.*, 2024b]. Building on these advancements, we propose a multi-dimension entropy model for bitrate estimation and compression in this work, further eliminating redundancies within the anchors. By integrating the estimated bitrate into a rate-distortion optimization framework, our method effectively balances the trade-off between rendering quality and bitrate costs.

### 3 Method

#### 3.1 Overview

As shown in Figure 3, ADC-GS begins with the initialization of a canonical space at  $t = 0$ , characterized by a sparse set of anchors with associated Gaussian primitives. Each anchor is equipped with latent features and explicit Gaussian attributes, including position, covariance and color. The explicit Gaussian attributes serve as the foundation for the associated primitives, while the latent features are responsible for predicting the residuals and deformation of each primitive. To reconstruct any subsequent frame  $t = T$ , we devise a coarse-to-fine strategy based on anchors. In the coarse stage, each anchor is deformed from the canonical space using lightweight multi-head MLPs, with the associated primitives subsequently up-

dated based on the coarse deformation. In the fine stage, the opacity and color of each primitive are further refined using the deformed position of the anchors and current timestamps information. By combining both coarse and fine deformations, all deformed attributes of Gaussian primitives are obtained to render the current frame using volume splatting [Kerbl *et al.*, 2023]. In the subsequent rate-distortion optimization, rendering distortion and estimated bitrates from the multi-dimension entropy model are jointly minimized to balance rendering quality and storage requirements. Moreover, the temporal significance-based strategy is employed to prune and grow anchors for mitigating both under-reconstruction and over-reconstruction issues in dynamic scenes.

#### 3.2 Canonical space initialization

Unlike prior deformation-based approaches [Yang *et al.*, 2024c; Lu *et al.*, 2024a] that directly use Gaussian primitives as the canonical space, we initialize our compact representation by employing a downsampled point cloud obtained via COLMAP [Snavely *et al.*, 2006] to serve as anchors. These anchors are the foundation for local regions that improve overall efficiency by minimizing unnecessary duplication of information across the dynamic scene.

In ADC-GS, each anchor  $v$  is associated with  $K$  neural Gaussian primitives  $\{g_1, \dots, g_K\}$ . The anchor consists of latent features  $\mathcal{A}_l = \{f_v \in \mathbb{R}^{N_v}, f_g \in \mathbb{R}^{K \times N_g}\}$  and explicit Gaussian attributes  $\mathcal{A}_e = \{X_v \in \mathbb{R}^3, \Sigma_v \in \mathbb{R}^6, C_v \in \mathbb{R}^3\}$ . Among the latent features, the reference feature  $f_v$  encapsulates the common characteristics between  $K$  associated primitives, while the residual feature  $f_g$  captures the variations specific to each primitive. The explicit Gaussian attributes serve as the foundation for the  $K$  associated primitives. To



generate associated primitives from anchors, we predict their explicit attributes in residual forms, as:

$$(\Delta X_k, \Delta \Sigma_k, \Delta C_k, O_k) = \mathcal{F}_\theta(f_v, f_g); \quad (1)$$

$$X_k = X_v + \Delta X_k, \Sigma_k = \Sigma_v \Delta \Sigma_k, C_k = C_v + \Delta C_k,$$

where  $\Delta X_k, \Delta \Sigma_k$  and  $\Delta C_k$  are residuals of position, covariance, and color.  $\mathcal{F}_\theta$  represents the prediction network, which is modeled using a residual neural network. Since opacity  $O_k$  exhibits no apparent spatial correlation as shown in Figure 2, it is directly predicted without residuals.

Anchors and their associated primitives collectively define the canonical space that captures the global structure of the entire scene. Subsequently, the reconstruction of the scene in any frame  $t = T$  is performed by deforming this canonical space through a coarse-to-fine pipeline.

### 3.3 Coarse-to-Fine Deformation

Previous methods [Wu *et al.*, 2024; Lu *et al.*, 2024a] deform the geometric properties of each Gaussian primitive individually from the canonical space to the target frame  $T$  individually. However, this per-Gaussian deformation suffers from low efficiency due to the similar deformation among adjacent primitives, leading to slow rendering speeds. To mitigate the limitation, we propose an anchor-driven coarse-to-fine deformation strategy. In the coarse stage, we deform the anchor explicit attributes  $\mathcal{A}_e$  by utilizing the concatenation of reference feature  $f_v$  and time embedding  $f_t$ , formulated as:

$$(\Delta X_{v,t}, \Delta \Sigma_{v,t}, \Delta C_{v,t}) = \mathcal{F}_\omega(f_v, f_t), \quad (2)$$

where  $\mathcal{F}_\omega$  denotes the anchor deformation network, implemented as a tiny MLP.  $\Delta X_{v,t}, \Delta \Sigma_{v,t}, \Delta C_{v,t}$  denote the position, covariance and color deformation of the anchors.  $f_t$  encodes temporal information for different frames,

$$f_t = \mathcal{F}_s(\text{Interp}(Z, t)), \quad (3)$$

where  $Z \in \mathbb{R}^{256}$  is a learnable weight for interpolation,  $t$  denotes the temporal index and  $\mathcal{F}_s$  represents the grid sampling function to encode temporal variations. By deforming a small number of anchors, the associated  $K$  Gaussian primitives are automatically updated, substantially diminishing deformation redundancy to enhance rendering efficiency.

While the coarse stage captures the global changes of the associated Gaussian primitives, it lacks the capability to address finer details. Considering that opacity exhibits minimal local similarity and color significantly influences scene reconstruction [Bae *et al.*, 2025], the fine stage dynamically refines these two attributes for each primitive, enhancing overall reconstruction quality. This refinement integrates the position embedding  $f_p$  [Vaswani, 2017] of the deformed anchors and the time embedding  $f_t$ ,

$$(\Delta O_{k,t}, \Delta C_{k,t}) = \mathcal{F}_\varpi(f_p, f_t), \quad (4)$$

where  $\mathcal{F}_\varpi$  represents the fine-stage deformation network similar to  $\mathcal{F}_\omega$  and  $\Delta O_{k,t}, \Delta C_{k,t}$  denote the per-Gaussian refinements of opacity and color, respectively. Therefore, associated Gaussian primitives are derived as a combination of coarse and fine-stage deformations:

$$X_{k,t} = X_k + \Delta X_{v,t}, \quad \Sigma_{k,t} = \Sigma_k \Delta \Sigma_{v,t}, \quad (5)$$

$$O_{k,t} = O_k + \Delta O_{k,t}, \quad C_{k,t} = C_k + \Delta C_{v,t} + \Delta C_{k,t},$$

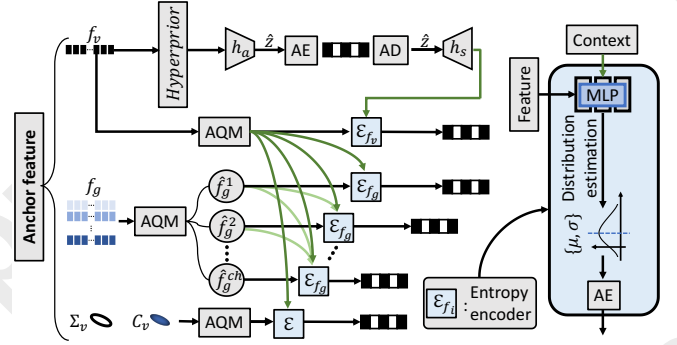


Figure 4: Illustration of the proposed MEM for accurate bitrates estimation. AQM refers to the adaptive quantization module.

where  $X_{k,t}, \Sigma_{k,t}, O_{k,t}$  and  $C_{k,t}$  are the position, covariance, opacity and color attributes used for volume splatting in the target frame  $T$ , respectively.

### 3.4 Rate-Distortion Optimization

The rate-distortion optimization scheme is designed to achieve a more compact representation by jointly minimizing both bitrate consumption and rendering distortion. While the anchor locations  $X_v$  are compressed using G-PCC [Liu *et al.*, 2024a], the remaining four anchor features are modeled for bitrate estimation through a Multi-dimension Entropy Model (MEM), as demonstrated in Figure 4. Specifically, scalar quantization is first applied to the feature  $f_i$ , where  $f_i \in \{f_v, f_g, \Sigma_v, C_v\}$ . However, conventional rounding essentially performs quantization with a fixed step size, which is less flexible for features with different scales. To address this issue, we design an adaptive quantization module that better accommodates the diverse feature value scales as follows,

$$\hat{f}_i = f_i + \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right) \times Q_i \times (1 + \tanh(\mathcal{F}_q(f_i))), \quad (6)$$

where  $\mathcal{F}_q$  indicates MLP-based model to adjust the predefined quantization step size  $Q_i$ . Note that  $Q_i$  varies for  $f_v, f_g, \Sigma_v$  and  $C_v$ . Uniform noise  $\mathcal{U}(\cdot)$  is injected during training to stimulate quantization loss in the testing stage as proposed in [Ballé *et al.*, 2018].

Subsequently, the probability distribution of  $f_i$  is estimated to calculate the corresponding bitrate through the MLP-based context model  $\mathcal{E}_{f_i}$ . The probability distribution  $p(\hat{f}_v)$  of reference feature is first parametrically formulated as a Gaussian distribution  $\mathcal{N}(\mu_{f_v}, \sigma_{f_v})$ , where the parameters  $\{\mu_{f_v}, \sigma_{f_v}\}$  are predicted based on hyperpriors [Ballé *et al.*, 2018] extracted from  $f_v$ ,

$$p(\hat{f}_v) = \mathcal{N}(\mu_{f_v}, \sigma_{f_v}), \quad \text{with} \quad \mu_{f_v}, \sigma_{f_v} = \mathcal{E}_{f_v}(\eta_{f_v}), \quad (7)$$

where  $\eta_{f_v}$  denotes the hyperpriors. Furthermore, the decoded reference feature  $\hat{f}_v$  is used as contexts to model the probability distributions of covariance  $\hat{\Sigma}_v$ , color  $\hat{C}_v$  and residual feature  $\hat{f}_g$ . The probability distribution of  $\hat{\Sigma}_v$  and  $\hat{C}_v$  is modeled similarly to  $p(\hat{f}_v)$ . The residual feature  $f_g$  constitutes the largest portion of total bitstream, which is divided into

$M$  chunks. The decoded reference features  $\hat{f}_v$  and chunks are organized into a multi-dimensional context to guide the encoding of remaining chunks. This strategy effectively reduces channel-wise redundancy and enhances compression efficiency, formulated as:

$$p(\hat{f}_g^{ch}) = \mathcal{N}(\mu_{f_g}, \sigma_{f_g}),$$

$$\text{with } \mu_{f_g}, \sigma_{f_g} = \mathcal{E}_{f_g} \left( \hat{f}_v \oplus \prod_{m=1}^{ch-1} \hat{f}_g^m \right), \quad (8)$$

where  $ch$  means the index of current encoding chunks and  $\oplus$  denotes the channel-wise concatenation, effectively integrating the decoded context for improved compression efficiency.

Consequently, the total estimated bitrate consumption  $R$  of anchors is calculated as:

$$R = -\log_2 p(\hat{f}_v) - \log_2 p(\eta_{f_v}) - \prod_{ch=1}^M \log_2 p(\hat{f}_g^{ch}) - \log_2 p(\hat{\Sigma}_v) - \log_2 p(\hat{C}_v). \quad (9)$$

Therefore, the rate-distortion optimization process is formulated as:

$$L_{loss} = (1 - \lambda_{ssim})L_1 + \lambda_{ssim}L_{ssim} + \lambda_e R. \quad (10)$$

where  $L_1$  and  $L_{ssim}$  represent the L1 loss and SSIM loss used in [Kerbl *et al.*, 2023].  $\lambda_{ssim}$  donates weighting coefficients for the SSIM loss and  $\lambda_e$  is the Lagrange multiplier to control the trade-off between rate and distortion.

### 3.5 Anchors Refinement with Temporal Significance

Anchors initialized from the sparse point cloud are often suboptimal, leading to under-reconstruction or over-reconstruction in certain areas. The existing strategies in 3DGS [Kerbl *et al.*, 2023; Lu *et al.*, 2024b] focus solely on static scenes and neglect the temporal significance of Gaussian primitives in dynamic frames, making them unsuitable for 4D scene reconstruction. To address this issue, we propose a temporal significance-aware anchor refinement method, including anchor growing and pruning. The importance of Gaussian primitives across frames is calculated by computing a weighted accumulation of their positional gradients,

$$\nabla_k = \frac{\sum^N \Psi(k, t) \|\nabla_{k,t}\|}{\sum^N \Psi(k, t)}, \quad (11)$$

where  $\nabla_{k,t}$  is the 2D position gradient of a Gaussian primitive and  $\nabla_k$  represents the accumulated gradients over  $N$  frames.  $\Psi(k, t)$  [Huang *et al.*, 2024] is the rendering weight of the  $\alpha$ -blending in the target frame  $T$ . Gaussian primitives with  $\nabla_k$  values exceeding a predefined threshold are added as new anchors. These new anchors inherit latent features from their associated anchors while preserving their own Gaussian attributes. Conversely, anchors associated with Gaussian primitives whose opacity falls below a predefined threshold are pruned to avoid over-reconstruction.

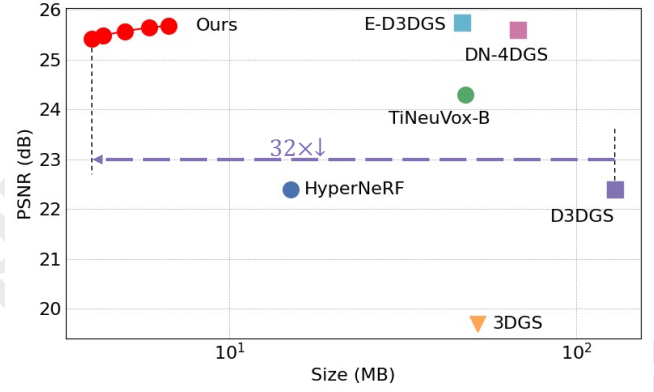


Figure 5: Rate-distortion curves of ADC-GS and comparison methods on HyperNeRF. We vary  $\lambda_e$  to achieve variable bitrates.

## 4 Experiments

In this section, we present the implementation details of ADC-GS, followed by quantitative comparisons with SOTA approaches on two commonly used datasets. Finally, we provide in-depth analysis and ablation studies to evaluate the performance and effectiveness of the proposed ADC-GS.

### 4.1 Experimental settings

Our proposed ADC-GS framework is implemented based on Pytorch and trained on a single NVIDIA RTX 3090 GPU. For hyperparameter settings, the dimension of the reference feature  $N_v$  is set to 32, while the residual feature  $N_g$  is 16. Each anchor is associated with  $K = 10$  neural Gaussian primitives. The Lagrange multiplier  $\lambda_e$  is varied from  $e - 2$  to  $e - 4$  for different bitrates, and  $\lambda_{ssim}$  is fixed at 0.2. Quantization step sizes  $Q_i$  are set to 0.1, 0.1, 0.01 and 0.01 for  $f_v$ ,  $f_g$ ,  $\Sigma_v$  and  $C_v$ , respectively. Meanwhile,  $f_g$  is divided into  $M = 4$  chunks. The prediction network  $\mathcal{F}_\theta$  and entropy estimation network  $\mathcal{E}_{f_i}$  are both implemented as 2-layer residual MLPs. Similarly, the deformation network  $\mathcal{F}_\omega$  and  $\mathcal{F}_\varpi$  are implemented as single 3-layer MLPs. The dimensions of time embedding  $f_t$  and position embedding  $f_p$  are 256 and 72, respectively.

**Dataset.** We evaluate our method on two multi-view video datasets: HyperNeRF [Park *et al.*, 2021b] and Neu3D [Li *et al.*, 2022]. HyperNeRF consists of videos captured using two phones rigidly mounted on a handheld rig, featuring four distinct scenes. Neu3D contains five distinct scenes, each comprising 20 multi-view videos. Data preprocessing follows the procedures outlined in [Wu *et al.*, 2024].

**Baselines.** We compare our method against SOTA approaches from four categories. For NeRF-based methods, we include Nerfies [Park *et al.*, 2021a], HyperNeRF [Park *et al.*, 2021b], TiNeuVox-B [Fang *et al.*, 2022], DyNeRF [Li *et al.*, 2022], NeRFPlayer [Song *et al.*, 2023], HexPlane [Cao and Johnson, 2023], MixVoxels [Wang *et al.*, 2023], and HyperReel [Attal *et al.*, 2023]. For 3DGS-based compression methods, we compare with 3DGS [Kerbl *et al.*, 2023] and CompGS [Liu *et al.*, 2024b]. In addition, we use 3DGStream [Sun *et al.*, 2024], Real-Time4DGS [Yan *et al.*, 2024], and STG [Li *et al.*, 2024] as baselines for direct modeling of 4DGS methods. Finally, for deformation-based 4DGS meth-

Model	PSNR↑	SSIM↑	LPIPS↓	FPS↑	Size (MB)↓
Nerfies	22.20	0.803	0.170	< 1	-
HyperNeRF	22.40	0.814	0.153	< 1	15
TiNeuVox-B	24.30	0.836	0.393	1	48
3DGS	19.70	0.680	0.383	55	52
D3DGS	22.40	0.612	0.275	22	129
DN-4DGS	25.59	0.861	-	20	68
4DGaussian	25.60	0.848	0.281	22	63
E-D3DGS	25.74	0.697	0.231	26	47
<b>Ours</b>	25.42	0.777	0.315	135	4.02
	25.53	0.791	0.278	117	5.20
	25.68	0.825	0.252	101	6.67

Table 1: Quantitative results on HyperNeRF. The best and 2nd best results are highlighted in red and yellow.

ods, we benchmark against D3DGS [Yang *et al.*, 2024c], DN-4DGS [Lu *et al.*, 2024a], 4DGaussian [Wu *et al.*, 2024], and E-D3DGS [Bae *et al.*, 2025].

**Metrics.** We report the quantitative results of rendered images using peak-signal-to-noise ratio (PSNR), structural similarity index (SSIM) [Wang *et al.*, 2004], perceptual quality measure LPIPS [Zhang *et al.*, 2018] and rendering speed (FPS). Model size is also reported to assess compression efficiency. Three bitrate points are displayed for each dataset due to space limitations. Moreover, rate-distortion (RD) performance is included to ensure a fair comparison.

## 4.2 Experimental Results

**Quantitative results.** The quantitative results presented in Tables 1, 2 and Figure 5 reveal consistent findings across all datasets: 1) our proposed ADC-GS achieves SOTA compression efficiency across all 4DGS methods, offering up to  $32\times$  storage reduction on the HyperNeRF dataset compared to deformation-based 4DGS while maintaining excellent rendering quality. This improvement stems from organizing Gaussian primitives into anchor-driven residual forms, effectively mitigating intra-redundancy among adjacent primitives. Moreover, entropy loss leverages the MEM model to compress anchors, further enhancing storage efficiency. 2) All NeRF-based methods render at extremely low speeds due to their computationally intensive volumetric rendering process. Although direct modeling 4DGS methods achieve the highest FPS, they demand  $204\times$  more storage compared to ADC-GS. Deformation-based methods demonstrate better convergence but still suffer from low FPS due to per-Gaussian deformation strategy. In contrast, our proposed anchor-driven deformation method delivers rendering speeds up to  $3\times$  faster than deformation-based methods, even surpassing some direct modeling 4DGS approaches. 3) ADC-GS only needs about 5MB of storage for both datasets, making it highly suitable for real-time transmission and practical applications.

Figure 7 presents a qualitative comparison between ADC-GS and other approaches. The rendered images produced by ADC-GS exhibit unperceived distortion while saving lots of storage. Moreover, our method demonstrates clearer textures in certain datasets, attributed to the temporal significance-based anchor refinement that captures finer details, such as Flame Steak in Neu3D.

Model	PSNR↑	SSIM↑	LPIPS↓	FPS↑	Size (MB)↓
DyNeRF	29.58	0.980	0.083	0.015	28
NeRFPlayer	30.69	0.966	0.111	0.045	-
HexPlane	31.70	0.987	0.075	0.2	250
Mixvoxels	31.73	-	0.064	4.6	500
HyperReel	31.10	0.964	0.096	2.0	360
CompGS	29.61	0.923	0.099	45	825
3DGStream	31.67	-	-	215	2340
Real-Time4DGS	32.01	0.986	0.055	114	1000
STG	32.04	0.974	0.044	110	175
DN-4DGS	32.02	0.984	0.043	15	112
4DGaussian	31.72	0.984	0.049	34	38
E-D3DGS	31.20	0.974	0.030	42	40
<b>Ours</b>	31.41	0.972	0.066	126	4.04
	31.55	0.975	0.065	116	5.32
	31.67	0.981	0.061	110	6.57

Table 2: Quantitative results on Neu3D. The best and 2nd best results are highlighted in red and yellow.

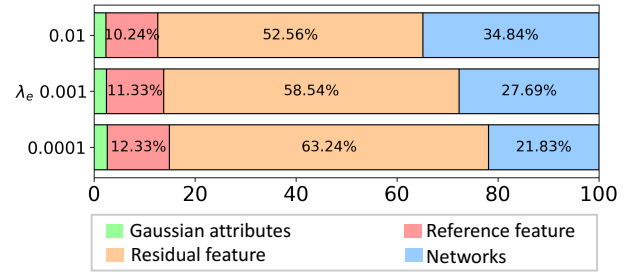


Figure 6: Bitstream analysis at multiple bitrate points on Neu3D.

**Bitstream.** Our bitstream is summarized as four components: reference feature  $f_v$ , residual feature  $f_g$ , explicit Gaussian attributes (comprising covariance  $\Sigma_v$ , color  $C_v$  and position  $X_v$ ) and networks  $\mathcal{F}$ . Among these, the networks are directly stored in 32 bits,  $X_v$  is compressed using G-PCC, and others are encoded using the AE entropy codec [Witten *et al.*, 1987] with estimated probabilities from our MEM model. As demonstrated in Figure 6,  $f_g$  constitutes the largest portion of the total bitstream, underscoring the importance of employing the MEM to effectively reduce channel-wise redundancy. The explicit Gaussian attributes account for a small portion of the bitstream, approximately 2%. When analyzing the bit allocation of each component specifically, they are 0.60MB, 3.10MB, 0.15MB (0.10MB, 0.03MB, 0.02MB) and 1.47MB on Neu3D with  $\lambda_e = e - 3$ , respectively. Notably, a smaller  $\lambda_e$  results in a larger bitstream allocation for features to enhance rendering quality.

## 4.3 Ablation Study

In this subsection, we evaluate the effectiveness of each component in our framework on the HyperNeRF dataset. As shown in Table 3 and Figure 8, the framework’s performance improves progressively with the addition of each module, demonstrating their individual and collective contributions.

The base configuration includes only the initialized canonical space, achieving a significant size reduction compared to existing 4DGS methods [Wu *et al.*, 2024; Lu *et al.*, 2024a]. While this setup provides high rendering speed due to the absence of deformation, the reliance on canonical space to



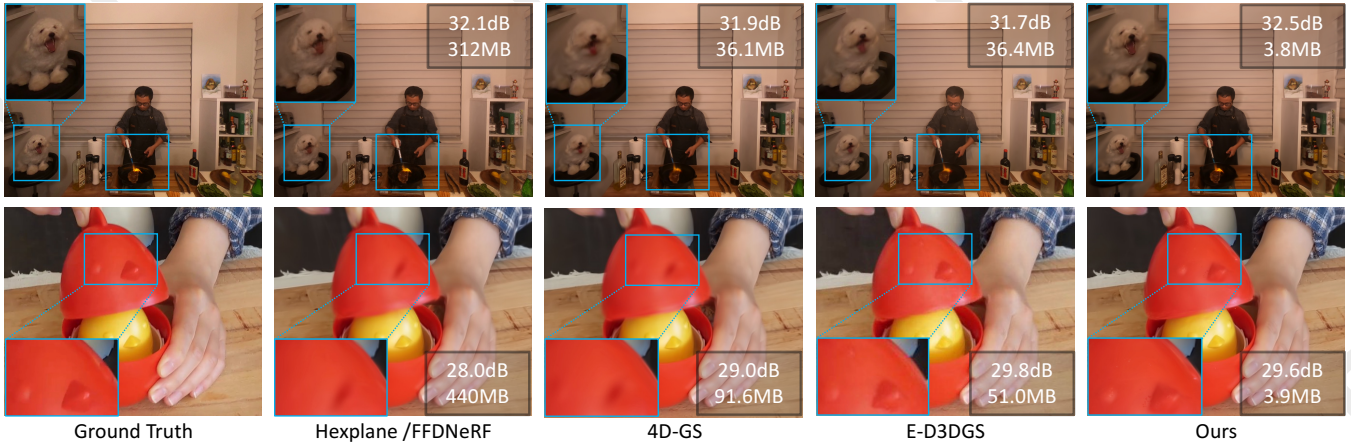


Figure 7: Qualitative quality comparisons of “Flame Steak” (Top) in Neu3D and “Chicken” (Bottom) in HyperNeRF dataset.

Stages	PSNR↑	SSIM↑	FPS↑	Size(MB)↓
Base	20.75	0.621	185	17.46
+ Coarse Deformation	24.83	0.676	119	17.77
+ Fine Deformation	25.03	0.689	117	18.11
+ R-D Optimization	25.00	0.687	112	4.49
+ Anchors Refinement	25.42	0.777	135	4.02

Table 3: Ablation study of different stages on “HyperNeRF”. “+” indicates adding current module to the previous stage. “Base” is the initialized canonical space.

represent the entire scene results in blurred rendering, particularly in motion-heavy regions. The addition of the coarse deformation module significantly improves the ability to capture the global motion structure. Although the rendering speed decreases compared to the baseline, it remains considerably faster than the per-Gaussian deformation approach in [Bae *et al.*, 2025], showcasing the efficiency of the anchor-driven strategy. The fine deformation module builds upon the coarse stage by refining appearance attributes, particularly in regions requiring higher detail. This refinement improves rendering quality by accurately modeling subtle temporal variations in color and opacity. The introduction of the rate-distortion optimization module leads to a remarkable reduction in bitstream size, from approximately 18 MB to 4.5 MB, without compromising rendering quality. This improvement can be attributed to the MEM module’s ability to learn compact anchor representations by effectively balancing bitrate consumption and reconstruction fidelity. Finally, the anchor refinement module further improves the representation by dynamically growing and pruning anchors to better capture fine details. This adaptive strategy ensures that the model maintains high-quality reconstruction in complex areas while optimizing the anchor distribution across the scene.

Overall, each module contributes to achieving a balanced trade-off between compression efficiency, rendering speed, and dynamic scene reconstruction quality, resulting in a compact and effective representation of the scene.

## 5 Conclusion

This paper introduces Anchor-Driven Deformable and Compressed Gaussian Splatting (ADC-GS), a novel method for

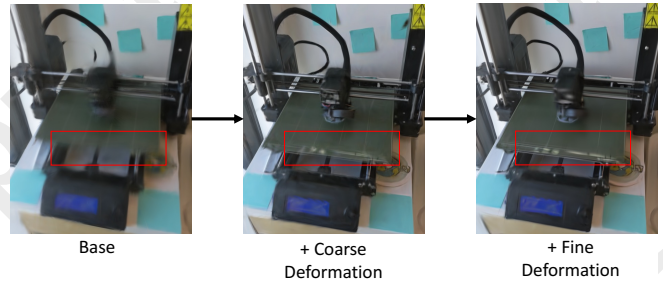


Figure 8: Ablation study of visual results for the coarse-to-fine deformation process.

dynamic scene reconstruction. By organizing Gaussian primitives in an anchor-based compact format within the canonical space, ADC-GS effectively minimizes the intra-redundancy between adjacent primitives. For dynamic scene modeling, we introduce a coarse-to-fine pipeline based on anchors instead of traditional per-Gaussian deformation, significantly reducing deformation redundancy. Meanwhile, a multi-dimension entropy model is employed to estimate the bitrate costs of anchors, serving as the factor for rate-distortion optimization. To further enhance details, we employ a temporal significance-based anchor refinement strategy. Extensive experiments demonstrate the effectiveness of ADC-GS and its technical components. Overall, our approach not only outperforms deformation-based 4DGS methods in rendering speed but also achieves SOTA storage efficiency without compromising rendering quality.

## Acknowledgments

This paper is supported in part by National Natural Science Foundation of China (62371290), National Key R&D Program of China (2024YFB2907204), the Fundamental Research Funds for the Central Universities of China, and STCSM under Grant (22DZ2229005). The corresponding author is Yiling Xu(e-mail: yl.xu@sjtu.edu.cn).

## Contribution Statement

He Huang and Qi Yang contributed equally to this work. Yiling Xu is the corresponding author.

## References

- [Attal *et al.*, 2023] Benjamin Attal, Jia-Bin Huang, Christian Richardt, Michael Zollhoefer, Johannes Kopf, Matthew O’Toole, and Changil Kim. Hyperreel: High-fidelity 6-dof video with ray-conditioned sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16610–16620, 2023.
- [Bae *et al.*, 2025] Jeongmin Bae, Seoha Kim, Youngsik Yun, Hahyun Lee, Gun Bang, and Youngjung Uh. Per-gaussian embedding-based deformation for deformable 3d gaussian splatting. In *European Conference on Computer Vision*, pages 321–335. Springer, 2025.
- [Ballé *et al.*, 2018] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. *arXiv preprint arXiv:1802.01436*, 2018.
- [Cao and Johnson, 2023] Ang Cao and Justin Johnson. Hexplane: A fast representation for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 130–141, 2023.
- [Chen *et al.*, 2024a] Tiejun Chen, Huabin Liu, Tianyao He, Yihang Chen, Chaofan Gan, Xiao Ma, Cheng Zhong, Yang Zhang, Yingxue Wang, Hui Lin, et al. MECD: Unlocking multi-event causal discovery in video reasoning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [Chen *et al.*, 2024b] Yihang Chen, Qianyi Wu, Mengyao Li, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. Fast feed-forward 3d gaussian splatting compression. *arXiv preprint arXiv:2410.08017*, 2024.
- [Chen *et al.*, 2025] Yihang Chen, Qianyi Wu, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. Hac: Hash-grid assisted context for 3d gaussian splatting compression. In *European Conference on Computer Vision*, pages 422–438. Springer, 2025.
- [Fan *et al.*, 2022] Tingyu Fan, Linyao Gao, Yiling Xu, Zhu Li, and Dong Wang. D-dpcc: Deep dynamic point cloud compression via 3d motion prediction. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 898–904. International Joint Conferences on Artificial Intelligence Organization, 7 2022. Main Track.
- [Fan *et al.*, 2023] Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, and Zhangyang Wang. Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps, 2023.
- [Fang *et al.*, 2022] 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 *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022.
- [Huang *et al.*, 2024] He Huang, Wenjie Huang, Qi Yang, Yiling Xu, et al. A hierarchical compression technique for 3d gaussian splatting compression. *arXiv preprint arXiv:2411.06976*, 2024.
- [Jiawei *et al.*, 2024] Xu Jiawei, Fan Zexin, Yang Jian, and Xie Jin. Grid4D: 4D decomposed hash encoding for high-fidelity dynamic scene rendering. *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [Kerbl *et al.*, 2023] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.
- [Lee *et al.*, 2024] Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3d gaussian representation for radiance field. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21719–21728, 2024.
- [Li *et al.*, 2022] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5521–5531, 2022.
- [Li *et al.*, 2024] 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.
- [Liu *et al.*, 2024a] Gexin Liu, Jiahao Zhu, Dandan Ding, and Zhan Ma. Encoding auxiliary information to restore compressed point cloud geometry. In Kate Larson, editor, *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pages 2189–2197. International Joint Conferences on Artificial Intelligence Organization, 8 2024. Main Track.
- [Liu *et al.*, 2024b] Xiangrui Liu, Xinju Wu, Pingping Zhang, Shiqi Wang, Zhu Li, and Sam Kwong. Compgs: Efficient 3d scene representation via compressed gaussian splatting. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 2936–2944, 2024.
- [Lu *et al.*, 2024a] Jiahao Lu, Jiacheng Deng, Ruijie Zhu, Yanzhe Liang, Wenfei Yang, Tianzhu Zhang, and Xu Zhou. Dn-4dgs: Denoised deformable network with temporal-spatial aggregation for dynamic scene rendering. *arXiv preprint arXiv:2410.13607*, 2024.
- [Lu *et al.*, 2024b] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffoldgs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20654–20664, 2024.
- [Ma *et al.*, 2024] Yikun Ma, Dandan Zhan, and Zhi Jin. Fastscene: Text-driven fast indoor 3d scene generation via panoramic gaussian splatting. In Kate Larson, editor, *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pages 1173–1181. International Joint Conferences on Artificial Intelligence Organization, 8 2024. Main Track.



- [Mildenhall *et al.*, 2021] 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. *Communications of the ACM*, 65(1):99–106, 2021.
- [Niedermayr *et al.*, 2024] Simon Niedermayr, Josef Stumpfegger, and Rüdiger Westermann. Compressed 3d gaussian splatting for accelerated novel view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10349–10358, 2024.
- [Park *et al.*, 2021a] 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.
- [Park *et al.*, 2021b] 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. *arXiv preprint arXiv:2106.13228*, 2021.
- [Pumarola *et al.*, 2021] 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.
- [Snaveley *et al.*, 2006] Noah Snaveley, Steven M. Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH ’06, page 835–846, New York, NY, USA, 2006. Association for Computing Machinery.
- [Song *et al.*, 2023] 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. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2732–2742, 2023.
- [Sun *et al.*, 2024] Jiakai Sun, Han Jiao, Guangyuan Li, Zhanjie Zhang, Lei Zhao, and Wei Xing. 3dstream: On-the-fly training of 3d gaussians for efficient streaming of photo-realistic free-viewpoint videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20675–20685, 2024.
- [Vaswani, 2017] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- [Wang *et al.*, 2004] 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.
- [Wang *et al.*, 2023] Feng Wang, Sinan Tan, Xinghang Li, Zeyue Tian, Yafei Song, and Huaping Liu. Mixed neural voxels for fast multi-view video synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19706–19716, 2023.
- [Witten *et al.*, 1987] Ian H Witten, Radford M Neal, and John G Cleary. Arithmetic coding for data compression. *Communications of the ACM*, 30(6):520–540, 1987.
- [Wu *et al.*, 2024] 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.
- [Yan *et al.*, 2024] Jinbo Yan, Rui Peng, Luyang Tang, and Ronggang Wang. 4d gaussian splatting with scale-aware residual field and adaptive optimization for real-time rendering of temporally complex dynamic scenes. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 7871–7880, 2024.
- [Yang *et al.*, 2024a] Qi Yang, Kaifa Yang, Yuke Xing, Yiling Xu, and Zhu Li. A benchmark for gaussian splatting compression and quality assessment study. In *Proceedings of the 6th ACM International Conference on Multimedia in Asia*, pages 1–8, 2024.
- [Yang *et al.*, 2024b] Zeyu Yang, Hongye Yang, Zijie Pan, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. In *International Conference on Learning Representations (ICLR)*, 2024.
- [Yang *et al.*, 2024c] 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.
- [Zhang *et al.*, 2018] 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018.
- [Zhang *et al.*, 2024] Xinjie Zhang, Zhening Liu, Yifan Zhang, Xingtong Ge, Dailan He, Tongda Xu, Yan Wang, Zehong Lin, Shuicheng Yan, and Jun Zhang. Mega: Memory-efficient 4d gaussian splatting for dynamic scenes. *arXiv preprint arXiv:2410.13613*, 2024.
- [Zhao *et al.*, 2024] Boming Zhao, Yuan Li, Ziyu Sun, Lin Zeng, Yujun Shen, Rui Ma, Yinda Zhang, Hujun Bao, and Zhaopeng Cui. Gaussianprediction: Dynamic 3d gaussian prediction for motion extrapolation and free view synthesis. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024.