SCENE FLOW CONSISTENT AND DEPTH-AWARE TEMPORAL REGULARIZATION FOR 4D GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review

ABSTRACT

3DGS represents scenes using Gaussian primitives, making it well-suited for a wide range of static scenarios. However, achieving high-quality dynamic video synthesis and physically realistic rendering requires accurate 3D motion modeling and temporal consistency to prevent physically implausible output across time. To address these challenges, we introduce PhysScene4DGS, a physics-based algorithm for 4D dynamic scene generation and rendering. By jointly optimizing 3D scene flow consistency and temporal coherence, PhysScene4DGS enables the physically realistic rendering of dynamic objects. PhysScene4DGS begins by computing 3D scene flow using depth and optical flow. Then, we introduce 4D temporal consistency regularization, which penalizes velocity discrepancies through inverse normalization between velocity and depth, promoting smooth and continuous motion. Furthermore, the scene flow consistency regularization aligns the predicted 3D motion with observed motion derived from back-projecting depth and optical flow, reinforcing physical accuracy. Experimental results demonstrate that PhysScene4DGS ensures both physical plausibility and visual realism in the real-time rendering of complex dynamic scenes, while maintaining comparable training time and FPS.

INTRODUCTION

Neural Radiance Field (NeRF) and 3D Gaussian Splatting (3DGS) Kerbl et al. (2023) have achieved impressive results in 3D reconstruction Mildenhall et al. (2020); Park et al. (2021a;b); Tretschk et al. (2021); Pumarola et al. (2021); Fang et al. (2022); Li et al. (2022b); Song et al. (2023), especially for static scenes. However, extending these methods to dynamic environments remains a substantial challenge. Unlike static scenes, dynamic scenes involve continuous temporal changes in geometry and motion, requiring models to account not only for spatial accuracy but also for the evolution of motion over time. Existing approaches often process each time step independently, lacking mechanisms to enforce temporal consistency or to model physically coherent motion across time. As a result, artifacts such as motion discontinuities are common. Robust dynamic scene rendering demands both physically plausible motion and smooth, temporally consistent trajectories. Addressing this problem introduces several key challenges. First, it is essential to fuse multimodal inputs, including depth and optical flow, while maintaining pixel-level precision and robustness to noise. Second, the design of a differentiable and adaptive temporal regularization is needed to penalize unrealistic acceleration and abrupt motion changes without suppressing natural dynamics. Third, for high-resolution scenes with dynamic content, scaling up point cloud size increases memory and processing demands, posing critical challenges for both training and real-time inference.

To address these challenges, PhysScene4DGS introduces 4D scene flow, temporal consistency, and physics-based constraints into a unified optimization framework grounded in physical plausibility, enabling high-quality rendering for dynamic video synthesis and 3D reconstruction that satisfies both visual realism and physical accuracy. Scene flow consistency ensures that the predicted motion of objects aligns with observed motion fields, addressing unrealistic artifacts during rendering. To preserve continuity, PhysScene4DGS combines scene flow prediction with temporal regularization, effectively eliminating abrupt jumps and correcting errors in motion trajectories across time. Furthermore, by incorporating both depth information and optical flow, the framework computes physically

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

grounded motion for each Gaussian primitive, maintaining coherence throughout the temporal domain. This joint modeling approach allows PhysScene4DGS to accurately predict 3D displacement while ensuring that temporal transitions conform to physical laws, resulting in smooth, realistic, and physically consistent dynamic scene rendering.

The core idea of 4D scene flow consistency is to estimate the motion of each Gaussian primitive by combining depth and optical flow features. This allows PhysScene4DGS to precisely capture object motion in dynamic scenes while enforcing alignment between predicted motion and observed scene flow. As a result, the framework strengthens the physical consistency of motion representation and enhances overall motion realism. To ensure smooth object trajectories, PhysScene4DGS introduces 4D temporal consistency regularization, which enables consistent velocity and depth evolution for each Gaussian primitive over time. This prevents rendering artifacts such as motion discontinuities and abrupt depth changes. However, enforcing constraints only at discrete time steps is insufficient for achieving true temporal coherence. To address this, the regularization is extended over unit time intervals, enabling continuity-aware supervision across temporal segments.

Ensuring coherence in 4D Gaussian representations presents notably greater challenges than in their static 3D counterparts, primarily due to the complexities introduced by temporal dynamics. To address this, we introduce a 4D temporal consistency regularization that jointly constrains velocity and depth within consecutive unit time intervals. This constraint suppresses inconsistencies between predicted depth and motion magnitude, enabling faithful modeling of object trajectories over time. By integrating this temporal regularization with scene flow consistency, PhysScene4DGS harnesses the complementary strengths of velocity, depth, and optical flow to capture physically plausible motion patterns. This joint optimization strategy enforces temporal coherence, reduces artifacts arising from unrealistic accelerations or depth discontinuities, and ultimately enables dynamic renderings that are both perceptually stable and physically grounded.

In dynamic scenes, the motion of each Gaussian primitive is characterized by its 4D scene flow, which encodes the primitive's displacement in 3D space over time. To accurately estimate this motion, PhysScene4DGS enforces 4D scene flow consistency by integrating optical flow and depth information across consecutive unit time intervals. Optical flow provides dense 2D motion features on the image plane, while depth enables the back-projection of these displacements into 3D space, resolving motion along the depth axis. By combining these complementary modalities, the framework reconstructs complete 3D motion trajectories for each primitive, capturing both lateral and longitudinal components of motion with high spatial fidelity. This framework allows PhysScene4DGS to robustly model complex object dynamics, enabling temporally coherent and visually accurate rendering of time-varying scenes.

To ensure that object motion remains smooth and physically coherent, we further introduce 4D temporal consistency regularization. This component normalizes velocity using depth, penalizing motion inconsistencies across consecutive unit time intervals. Specifically, the regularization models velocity as inversely proportional to depth, reflecting the real-world observation that objects farther from the camera should exhibit smaller relative displacements. This depth-aware normalization resolves the impact of perspective-induced motion distortion and ensures that velocity estimates remain consistent across varying depth layers. By focusing on relative motion rather than absolute positional differences, the regularization eliminates unrealistic variations and enforces temporal smoothness. Consequently, scene flow estimation and depth-normalized velocity constraints enable PhysScene4DGS to produce stable, coherent motion trajectories with physically plausible dynamics, ensuring the quality and realism of dynamic scene rendering. Our contributions are as follows:

- The introduction of 4D physical scene flow consistency regularization, which is the first rendering approach to ensure faithful physical realism, synergizing velocity, depth, and optical flow into a physically plausible framework.
- We propose 4D temporal consistency regularization, which leverages the inverse relationship between velocity and depth within the unit time interval based on physical assumptions, enabling realistic physical rendering.
- The introduction of the motion feature, scene flow, which combines depth and optical flow to ensure precise modeling and consistent rendering of complex physical motions.

- An optimization approach of 4D Gaussian properties that dynamically refines Gaussian distributions, capturing consistent, physically realistic representations of dynamic scenes without increasing the size of the point cloud.

RELATED WORK

We begin with traditional 3D reconstruction and view synthesis techniques, then transition to neural rendering methods. Finally, we examine point-based rendering methods.

From Geometric Reconstruction to Neural Representations Traditional rendering methods perform reliably in structured scenes with ample viewpoint coverage Levoy & Hanrahan (1996); Gortler et al. (1996); Buehler et al. (2001); Debevec et al. (1996); Riegler & Koltun (2020); Thies et al. (2019); Waechter et al. (2014); Wood et al. (2023); Kutulakos & Seitz (2000); Penner & Zhang (2017); Seitz & Dyer (1999); Flynn et al. (2019); Mildenhall et al. (2019); Srinivasan et al. (2019); Xu et al. (2022b); Zhou et al. (2018). They often rely on reprojection and image fusion but involve complex geometric modeling Broxton et al. (2020); Collet et al. (2015); Li et al. (2018); Guo et al. (2015); Su et al. (2020); Guo et al. (2019); Li et al. (2017); Flynn et al. (2019); Li et al. (2012); Kanade et al. (1997); Zitnick et al. (2004). Neural radiance field (NeRF) migrates these limitations by learning a volumetric scene representation via multi-layer perceptrons (MLPs) Mildenhall et al. (2020); Verbin et al. (2022); Kopanas et al. (2022); Hu et al. (2022); Bemana et al. (2022); Yan et al. (2023); Mildenhall et al. (2021); Barron et al. (2021); Zhang et al. (2020); Pumarola et al. (2021); Park et al. (2021a;b); Du et al. (2021). Volumetric rendering integrates these predictions along viewing rays to generate photorealistic images from novel viewpoints Fang et al. (2022); Gao et al. (2021); Li et al. (2021); Yi et al. (2023); Liu et al. (2023); Guo et al. (2023); Li et al. (2022b); Tian et al. (2023); Cao & Johnson (2023); Fridovich-Keil et al. (2023); Shao et al. (2023); Wang et al. (2023a); Gan et al. (2023); Zhou et al. (2024); Zhang et al. (2022); Xu et al. (2022a); Abou-Chakra et al. (2022). However, in dynamic scene rendering, maintaining motion consistency and physical plausibility remains challenging. Existing methods suffer from temporal artifacts and unrealistic motion, especially in complex, non-rigid environments. PhysScene4DGS addresses these challenges by effectively synchronizing 3D scene flow, temporal consistency, and physics-based constraints into the rendering process.

The Evolution of Point-Based Rendering Point-based rendering represents 3D geometry as discrete spatial samples but suffers from continuity artifacts due to its reliance on local surface attributes. Neural radiance field (NeRF) address these limitations by representing scenes as continuous volumetric functions parameterized by a fully-connected network Park et al. (2021a); Tretschk et al. (2021); Fang et al. (2022); Wang et al. (2023c); Gao et al. (2022); Peng et al. (2023); Lin et al. (2023); Wang et al. (2023b). Given a sparse set of posed images, NeRF learns to map 3D coordinates and viewing directions to density and view-dependent color via differentiable volume rendering Li et al. (2022a); Fridovich-Keil et al. (2022; 2023); Cao & Johnson (2023); Wang et al. (2023b); Fang et al. (2022); Gan et al. (2023); Shao et al. (2023); Attal et al. (2023). Its key innovation is the use of an implicit neural representation Müller et al. (2022); Chen et al. (2022); Li et al. (2021); Guo et al. (2023); Tian et al. (2023); Song et al. (2023). However, NeRF's high computational cost and its limitation to static content hinder practical deployment. To address this, 3D Gaussian Splatting (3DGS) Kerbl et al. (2023) represents scenes as anisotropic 3D Gaussians, enabling efficient GPU acceleration with high visual quality. However, 3DGS extensions lack explicit modeling of physical dynamics Yang et al. (2023); Wu et al. (2023); Yang et al. (2024); Duan et al. (2024), making it less effective in dynamic scenes where temporal coherence is critical. PhysScene4DGS overcomes this limitation by enforcing scene flow and temporal consistency constraints.

METHOD

PhysScene4DGS is a dynamic scene rendering framework that renders physically plausible results by synchronizing 4D scene flow, velocity, and depth, which effectively predicts motion through Gaussian distributions. To ensure temporal coherence, PhysScene4DGS introduces unit-time interval consistency regularization based on velocity and depth, enabling high-fidelity, physically consistent dynamic scene rendering aligned with real-world motion.

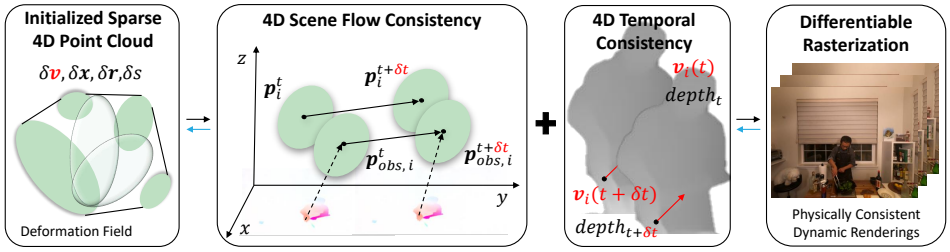


Figure 1: **Framework Overview.** 3D scene flow estimation computes per-pixel motion vectors, which are then refined by parallel consistency regularizations. The 4D temporal consistency enforces smooth, depth-normalized velocity changes across unit time intervals, and the scene flow consistency aligns predicted motion with observations derived from depth-flow backprojection.

Physical Scene 4D Gaussian Splatting

PhysScene4DGS is a physics-based framework for 4D dynamic scene rendering, designed to achieve high-fidelity rendering. Unlike traditional rendering methods, PhysScene4DGS integrates physical modeling and dynamic scene flow to capture not only the spatial motion of objects but also their temporal evolution. This enables the rendering of dynamic scenes that are both physically plausible and visually coherent. As illustrated in Figure 1, PhysScene4DGS departs from existing methods by introducing temporal and physical consistency at the input stage, operating over consecutive unit time intervals. On the output side, it explicitly predicts key dynamic physical features.

4D scene flow serves as a critical motion descriptor in dynamic scenes, capturing the three-dimensional displacement over time. To further improve physical plausibility, PhysScene4DGS integrates velocity and depth information into the PhysScene4DGS framework and imposes physics-based constraints on Gaussian distributions. These constraints ensure physically plausible motion by adapting to object depth and velocity, enabling temporal smoothness and consistency in dynamic scene rendering. PhysScene4DGS explicitly models object velocity and depth through the parameters of each Gaussian primitive, enforcing adherence to real-world dynamics.

Furthermore, to address temporal inconsistencies, our temporal consistency constraints ensure coherent motion across continuous unit time intervals. This regularization suppresses temporal discontinuities and motion jitter, improving both visual stability and motion continuity of dynamic scene rendering. By combining 4D scene flow with physics-aware modeling, PhysScene4DGS enforces adherence to physical principles, avoiding artifacts such as abrupt changes or unrealistic accelerations. It prioritizes accurate motion trajectories and object-environment interactions, resulting in natural, continuous, and physically grounded dynamic behavior. As a novel physics-aware approach to dynamic scene rendering, PhysScene4DGS ensures physical realism by leveraging multi-modal supervision and enforcing consistency across both spatial and temporal domains.

Optimization Scheme

4D Scene Flow Consistency This approach is an important component of the optimization framework that enforces the consistency between predicted and observed scene flow. Scene flow, as the three-dimensional motion of each pixel in a dynamic scene, is crucial for accurately understanding object motion and scene rendering in dynamic environments. Given an image with resolution $H \times W$, we define $N = H \times W$ as the total number of pixels. For each pixel $i \in \{1, \ldots, N\}$, we denote its 3D coordinates at time t by $\mathbf{p}_i^t \in \mathbb{R}^3$ and at time $t + \delta t$ by $\mathbf{p}_i^{t+\delta t} \in \mathbb{R}^3$. These coordinates are given by the Gaussian primitives at times t and $t + \delta t$, respectively, denoted as:

$$\mathbf{p}_i^t = \mathrm{G}_t(i), \quad \mathbf{p}_i^{t+\delta t} = \mathrm{G}_{t+\delta t}(i). \tag{1}$$

The predicted 4D scene flow is then computed as the difference between these two positions:

$$\delta_i^{\text{pred}} = \mathbf{p}_i^{t+\delta t} - \mathbf{p}_i^t. \tag{2}$$

Next, let the 2D pixel coordinates in the input image be denoted by:

$$\mathbf{xy}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \quad i = 1, \ldots, N. \tag{3}$$

Depth at times t and $t + \delta t$ are given as $d^t, d^{t+\delta t} \in \mathbb{R}^{H \times W}$. The observed optical flow is denoted as $\mathbf{f}_i \in \mathbb{R}^{H \times W}$, and from this optical flow, we compute the shifted pixel coordinates at time $t + \delta t$ as:

$$\mathbf{uv}_i^{t+\delta t} = \mathbf{xy}_i + \mathbf{f}_i. \tag{4}$$

The camera intrinsics matrix K, which contains the focal lengths f_x, f_y and the principal point coordinates c_x, c_y, is given by:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

To properly handle the pixel coordinates in the context of image sampling, we first normalize the pixel coordinates to the range $[-1, 1]$:

$$x_i^{\text{norm}} = \frac{x_i}{W-1} \times 2 - 1, \quad y_i^{\text{norm}} = \frac{y_i}{H-1} \times 2 - 1. \tag{6}$$

These normalized coordinates form the sampling grid, which is used to sample both d^t and the optical flow data. Similarly, the coordinates $\mathbf{uv}_i^{t+\delta t}$ are normalized for sampling the depth map $d^{t+\delta t}$. Using the camera intrinsics matrix K, we can backproject the 2D pixel coordinates and sampled depth values into 3D space. The backprojection function $\mathcal{B}$ for a pixel coordinate $\mathbf{x} = (x, y)^T$ and depth z is defined as:

$$\mathcal{B}(\mathbf{x}, z) = \begin{bmatrix} \frac{(x-c_x)\,z}{f_x} \\ \frac{(y-c_y)\,z}{f_y} \\ z \end{bmatrix}. \tag{7}$$

The backprojected 3D coordinates at times t and $t + \delta t$ are given by:

$$\mathbf{P}_{\text{obs},i}^t = \mathcal{B}(\mathbf{xy}_i, d_i^t), \quad \mathbf{P}_{\text{obs},i}^{t+\delta t} = \mathcal{B}(\mathbf{uv}_i^{t+\delta t}, d_i^{t+\delta t}). \tag{8}$$

The observed scene flow δ_i^{obs} is then calculated as the difference between the backprojected positions:

$$\delta_i^{\text{obs}} = \mathbf{P}_{\text{obs},i}^{t+\delta t} - \mathbf{P}_{\text{obs},i}^t. \tag{9}$$

Finally, the scene flow consistency regularization is defined as the mean squared error between the predicted scene flow and the observed scene flow:

$$L_{\text{Scene Flow}} = \frac{1}{N} \sum_{i=1}^{N} \left\| \delta_i^{\text{pred}} - \delta_i^{\text{obs}} \right\|^2. \tag{10}$$

To further refine the regularization calculation, we use a foreground mask m_i, where $m_i = 1$ indicates foreground and $m_i = 0$ indicates background. This mask allows the regularization to be computed only over the foreground points, reducing the impact of background regions:

$$L_{\text{Scene Flow}} = \frac{1}{\sum_i m_i} \sum_{i=1}^{N} m_i \left\| \delta_i^{\text{pred}} - \delta_i^{\text{obs}} \right\|^2. \tag{11}$$

By enforcing this scene flow consistency constraint, the predicted 3D motion is encouraged to be consistent with the observed multimodal data, improving the accuracy and realism of the motion prediction in dynamic scenes.

4D Temporal Consistency This approach is designed to enforce smoothness and consistency in the temporal evolution of the scene's 3D motion. This regularization ensures that the motion of each Gaussian primitive in the scene, as estimated by PhysScene4DGS, adheres to a physically plausible trajectory across consecutive unit time intervals. By using this regularization, we encourage the predicted velocities at time t to be consistent with the velocities at time $t - \delta t$, while accounting for changes in depth over time.

Let $\mathbf{v}_i(t)$ represent the velocity of the i-th Gaussian primitive at time t, which is computed by the rate of change of the 3D position with respect to time. Similarly, $d_i(t)$ represents the depth of the

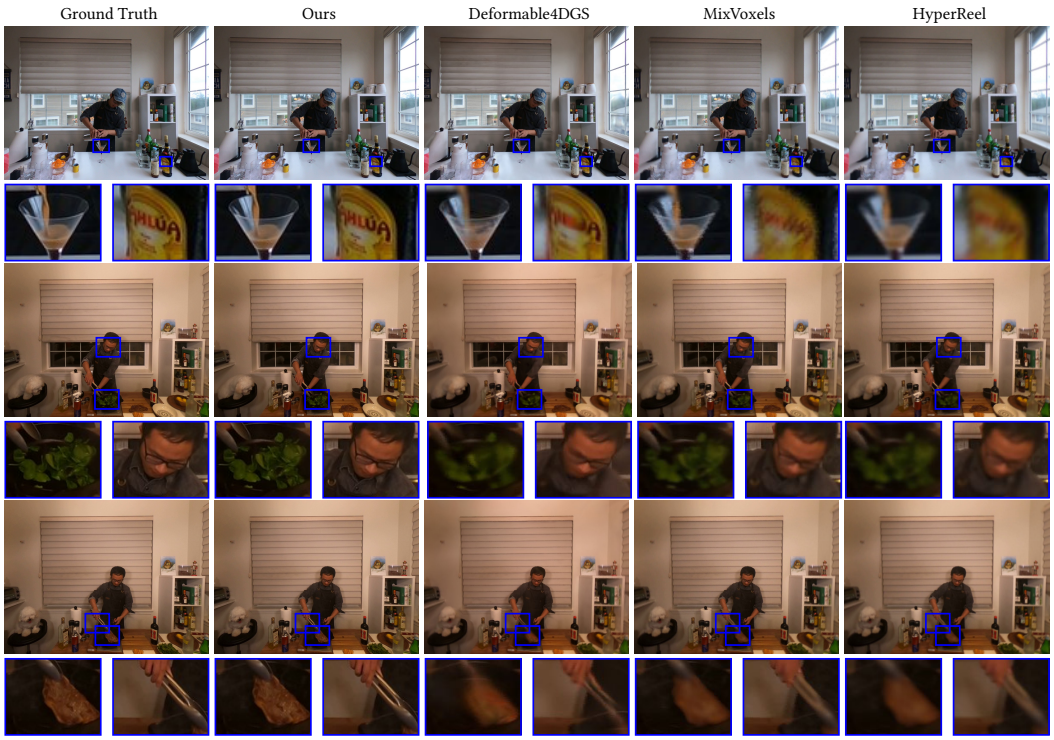


Figure 2: **Qualitative Comparison on the Plenoptic Video Dataset**. We show comparisons of ours to SOTA methods and highlight challenging details, such as the martini glass rim, window blind slats, foliage edges, and sizzling steak texture, where PhysScene4DGS consistently produces sharp geometry, accurate shading, and crisper fine structures.

i-th Gaussian primitive at time t, and N is the number of Gaussian primitives in the scene. The 4D temporal consistency regularization is defined as:

$$\mathcal{L}_{\text{4D Temporal Consistency}} = \sum_{i=1}^{N} \left(\frac{\mathbf{v}_i(t)}{d_i(t)} - \frac{\mathbf{v}_i(t-\delta t)}{d_i(t-\delta t)} \right)^2. \tag{12}$$

where $\mathbf{v}_i(t-\delta t)$ represents the velocity of the i-th Gaussian primitive at time $t-\delta t$, the previous time step. The depth values $d_i(t)$ and $d_i(t-\delta t)$ are used to normalize the velocities, ensuring that the motion is consistent despite changes in the depth information over time.

Temporal regularization penalizes large discrepancies between the velocities of the same Gaussian primitive within unit time intervals, taking into account the changes in depth within unit time intervals. This ensures that the motion of Gaussian primitives in the 3D scene remains physically plausible and consistent across time. It also prevents sudden, unrealistic jumps in motion, which arise due to inaccuracies in depth estimation or motion prediction.

By enforcing this temporal consistency constraint, we encourage PhysScene4DGS to produce smooth and coherent motion trajectories for each Gaussian primitive, which is essential for generating realistic dynamic scenes. This regularization is especially useful in dynamic scene rendering, where accurate motion prediction and depth estimation are crucial for maintaining the continuity of the scene over time.

4D Physical Scene Flow Consistency The final 4D physical scene flow consistency regularization is a crucial component for ensuring both the temporal consistency and scene flow consistency. It comprises two complementary terms: the scene flow consistency loss, which enforces alignment between predicted and observed 3D displacements, and the 4D temporal consistency loss, which regularizes the evolution of motion over time. The overall objective is to ensure that the motion of objects in the dynamic scene adheres to both realistic temporal changes and physically plausible scene flow (i.e., the spatial motion of Gaussian primitives). Formally, the regularization is expressed as:

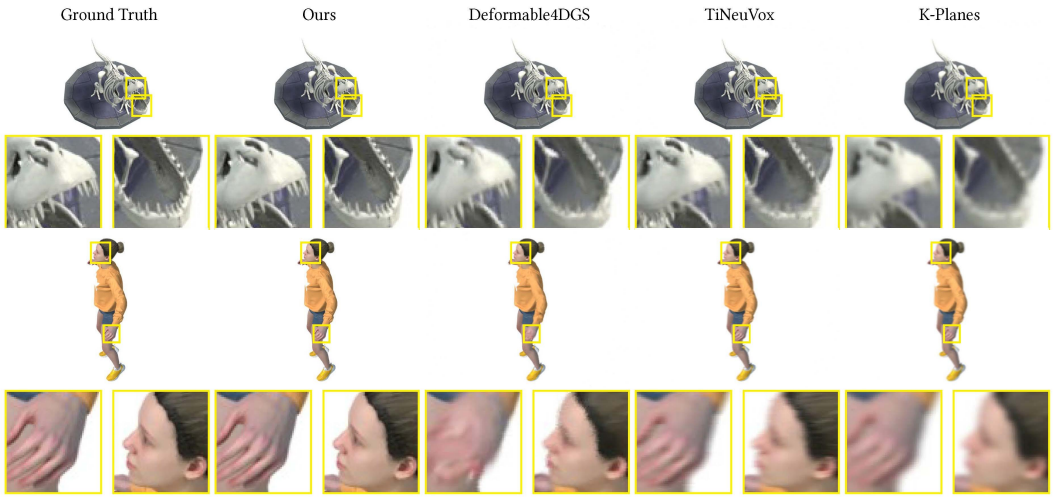


Figure 3: **Qualitative Comparison on the D-NeRF Dataset.** Our approach preserves fine-scale geometric structures and sharp object boundaries. Compared with the baseline methods, PhysScene4DGS exhibits reduced blurring and improved preservation of local details, demonstrating its effectiveness in high-fidelity dynamic scene rendering.

$$\mathcal{L}_{\text{4D PsysScene}} = L_{\text{Scene Flow}} + \lambda_1 \mathcal{L}_{\text{4D Temporal Consistency}}. \tag{13}$$

where λ_1 is the hyperparameter that controls the relative importance of the scene flow consistency and temporal consistency regularization. The 4D physical scene flow consistency regularization integrates temporal and scene flow consistency into a unified framework, balancing the PhysScene4DGS's ability to predict smooth temporal motion and physically plausible scene flow. Through this comprehensive optimization, PhysScene4DGS ensures that the predicted motion is consistent over time and aligns with the observed scene dynamics, enabling realistic and stable results for dynamic scene rendering.

EXPERIMENTS

IMPLEMENTATION DETAILS

We validate our PhysD-4DGS framework on three widely adopted dynamic-scene rendering benchmarks: **Plenoptic Video Dataset** Li et al. (2022b), **D-NeRF** Pumarola et al. (2021), and **HyperNeRF Dataset** Park et al. (2021b). The Plenoptic Video dataset employs 15–20 cameras to capture challenging real-world phenomena at 1352×1014 resolution, providing 17–20 training views and one held-out test view per sequence. We initialize structure-from-motion from the first frame and bound the point count to 100,000 to manage memory, yet our optimized four-dimensional Gaussian splatting and lightweight point-cloud representation converge to high-fidelity renderings in only 14,000 iterations. In contrast, the D-NeRF dataset comprises monocular, background-free video sequences resized to 800×800 pixels, with 50–200 frames for training, 10–20 for validation, and 20 for testing. For these scenes, we conduct 20,000 training iterations with an upsampling factor of two, pruning every 8,000 steps and halting Gaussian growth at 15,000. The HyperNeRF dataset captures 163–512 frames per scene with synchronized stereo cameras observing complex deformations and topology changes. We test on four scenes: 3D Printer, Chicken, Broom, and Banana, using

Table 1: **Quantitative Comparison on Plenoptic Video Dataset**. We compare our approach with both NeRF-based and Gaussian-based methods. Our approach significantly outperforms the baselines in PSNR, achieving the shortest training times and comparative rendering speeds. *: trained on 8 GPUs and tested only on the Flame Salmon scene.

Method	PSNR↑	SSIM↑	LPIPS↓	Train↓	FPS↑
DyNeRF Li et al. (2022b)*	29.58	-	0.08	1344 h	0.015
StreamRF Li et al. (2022a)	28.16	0.85	0.31	79 min	8.50
HyperReel Attal et al. (2023)	30.36	0.92	0.17	9 h	2.00
NeRFPlayer Song et al. (2023)	30.69	-	0.11	6 h	0.05
K-Planes Fridovich-Keil et al. (2023)	30.73	0.93	0.07	190 min	0.10
MixVoxels Wang et al. (2023b)	30.85	0.96	0.21	91 min	16.70
MSTH Wang et al. (2023a)	29.46	0.92	0.17	36 min	2.66
RealTime4DGS Yang et al. (2024)	29.95	0.92	0.16	8 h	72.80
Deformable4DGS Wu et al. (2023)	28.42	0.92	0.17	72 min	39.93
Ours	34.50	0.96	0.08	72 min	40.00

alternating frames from both viewpoints for evaluation and the remainder for training. All images are 960×540 pixels. All experiments adhere to the original 3DGS hyperparameter settings and are executed on a single NVIDIA RTX 3090 GPU.

RESULTS

EVALUATION ON PLENOPTIC VIDEO DATASET

Table 1 presents a comprehensive evaluation of PhysScene4DGS against leading NeRF-based and Gaussian-based methods on the Plenoptic Video Dataset. In terms of rendering fidelity, PhysScene4DGS achieves a PSNR of 34.50, exceeding competitive baselines such as MixVoxels and K-Planes by over 3 and outperforming HyperReel and NeRFPlayer. Furthermore, PhysScene4DGS enables these quality gains with higher efficiency: training converges in just 72 minutes, over 10× faster than HyperReel or NeRFPlayer, which require hours or even days of training. Our method enables real-time rendering, substantially outperforming most NeRF-based methods and even surpassing MixVoxels' 16.7 FPS. Figure 2 demonstrates that PhysScene4DGS ensures high rendering quality across diverse scenarios, including complex human motions. In Figure 6, flow vectors reliably capture dynamic motion. Figure 4 showcases photometrical editing with 4D Gaussians, enabling object transformation, scaling, and composition. These results underscore PhysScene4DGS's ability to ensure high-fidelity rendering with superior efficiency.

EVALUATION ON D-NERF DATASET

Table 2 presents a rigorous evaluation of PhysScene4DGS, demonstrating substantial advances in both rendering fidelity and runtime performance. PhysScene4DGS achieves a PSNR of 37.50, outperforming the leading method by more than 4, and exceeding the original D-NeRF by over 7. This improvement is further supported by a SSIM of 0.99, surpassing the 0.97 reported by TiNeuVox and K-Planes, and an LPIPS of 0.02, indicating excellent perceptual quality. Beyond image quality, PhysScene4DGS achieves these gains without sacrificing efficiency: it converges in just 12 minutes, substantially faster than D-NeRF. Moreover, PhysScene4DGS achieves a real-time rendering speed of 120 FPS, substantially outperforming all baselines. Figure 6 further shows that PhysScene4DGS captures coherent, depth-aware motion boundaries. Experimental results demonstrate its effectiveness in bridging high-fidelity and real-time performance for dynamic scene rendering.

Table 2: **Quantitative Comparison on D-NeRF Dataset**. Our approach surpasses both NeRF-based and Gaussian-based baselines in PSNR, achieving shorter training times and faster rendering speeds.

Method	PSNR↑	SSIM↑	LPIPS↓	Train↓	FPS↑
D-NeRF Pumarola et al. (2021)	29.17	0.95	0.07	24 h	0.13
TiNeuVox Fang et al. (2022)	32.87	0.97	0.04	28 min	1.60
K-Planes Fridovich-Keil et al. (2023)	31.07	0.97	0.02	54 min	1.20
FFDNeRF Guo et al. (2023)	31.70	0.96	0.05	-	<1.20
MSTH Wang et al. (2023a)	30.40	0.97	0.05	9.80 min	-
V4D Gan et al. (2023)	32.67	0.97	0.05	10.21 h	2.64
RealTime4DGS Yang et al. (2024)	29.95	0.92	0.16	8 h	72.80
Deformable4DGS Wu et al. (2023)	32.99	0.97	0.05	13 min	104.00
Ours	37.50	0.99	0.02	12 min	120.00

EVALUATION ON HYPERNERF DATASET

Table 3 compares PhysScene4DGS with leading methods. It achieves a PSNR of 27.5 dB, nearly 3 dB higher than HyperNeRF, demonstrating superior handling of complex non-rigid deformations. An SSIM of 0.90 surpasses TiNeuVox, indicating improved preservation of visual structure and appearance. PhysScene4DGS is also highly efficient: training converges in 28 minutes, orders of magnitude faster than HyperNeRF and V4D. At inference, it supports real-time rendering at 45 FPS with a compact 45 MB representation, outperforming Deformable4DGS. These results confirm that PhysScene4DGS enables high-fidelity, physically plausible renderings with real-time performance.

Table 3: **Quantitative Comparison on HyperNeRF Dataset**. Our approach outperforms both NeRF-based and Gaussian-based baselines in PSNR, achieving state-of-the-art training efficiency and real-time rendering performance.

Method	PSNR↑	SSIM↑	Times↓	FPS↑	Storage (MB)↓
Nerfies Park et al. (2021a)	22.18	0.80	$\sim$ h	<1	–
HyperNeRF Park et al. (2021b)	22.43	0.81	32 h	<1	–
TiNeuVox Fang et al. (2022)	24.26	0.84	30 mins	1	48
3D-GS Yang et al. (2023)	19.69	0.68	40 mins	55	52
FFDNeRF Guo et al. (2023)	24.24	0.84	–	0.05	440
V4D Gan et al. (2023)	24.83	0.83	5.5 hours	0.29	377
Deformable4DGS Wu et al. (2023)	25.19	0.85	30 mins	34	61
Ours	27.54	0.90	28 mins	45	45



Figure 4: **Interactive 4D Gaussian–Based Editing With PhysScene4DGS.**

Ablation Studies

4D Scene Flow Consistency. In the ablation setting *b*, we isolate the impact of the 4D scene flow consistency regularization. As shown in Table 4, this addition yields a notable boost in rendering fidelity: PSNR improves from 31.00 to 35.10 on Jumping Jacks. These gains indicate that enforcing alignment between predicted and observed 3D motion effectively reduces both geometric and photometric errors. Figure 5 further illustrates the importance of the scene flow constraint. When removing it, spatial coherence deteriorates severely. These artifacts underscore the effectiveness of the 4D scene flow regularization in rendering sharp, spatially consistent visual output.



Figure 5: **Qualitative Ablation of PhysScene4DGS.**

Table 4: **Ablation Study on D-NeRF Dataset (PSNR)**. We compare: (*a*) baseline without consistency terms; (*b*) 4D Scene Flow Consistency only; (*c*) 4D Temporal Consistency only; (*d*) full.

ID	4D Scene Flow Consistency	4D Temporal Consistency	D-NeRF (PSNR) Jumping Jacks	Stand Up	Trex
a			31.00	36.20	31.80
b	✓		35.10	36.50	32.80
c		✓	34.80	36.40	32.90
d	✓	✓	**35.70**	**38.00**	**35.79**

4D Temporal Consistency. As shown in Table 4, temporal regularization effectively improves rendering quality, such as increasing PSNR from 31.80 to 32.90 on Trex. This indicates that penalizing abrupt velocity changes enhances temporal coherence and reduces flickering artifacts. Figure 5 further highlights the effects of removing temporal consistency: while spatial geometry remains intact, instability appears as subtle jitter along glass edges. These results underscore the importance of temporal regularization for stable, artifact-free rendering in fast-moving scenes.



Figure 6: **Optical Flow Visualization.**

Conclusion

In this work, we present PhysScene4DGS, a physics-based framework for dynamic scene rendering that ensures both spatial and temporal consistency. Our approach incorporates 4D scene flow consistency and 4D temporal consistency regularizations to align predicted motion with observed trajectories and to suppress abrupt transitions through inverse velocity–depth normalization across consecutive unit time intervals. Experimental results demonstrate that PhysScene4DGS consistently outperforms state-of-the-art methods across a variety of complex dynamic scenes. By embedding physical consistency as an explicit constraint within both Gaussian representations and 4D scene modeling, our approach enables temporally coherent, physically plausible motion while maintaining computational efficiency. Consequently, PhysScene4DGS is a unified framework that simultaneously ensures high visual fidelity and realism grounded in physical dynamics.

REFERENCES

Jad Abou-Chakra, Feras Dayoub, and Niko Sünderhauf. Particlenerf: Particle based encoding for online neural radiance fields in dynamic scenes. *arXiv preprint arXiv:2211.04041*, 2022.

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*, pp. 16610–16620, 2023.

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5855–5864, 2021.

Mojtaba Bemana, Karol Myszkowski, Jeppe Revall Frisvad, Hans-Peter Seidel, and Tobias Ritschel. Eikonal fields for refractive novel-view synthesis. In *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–9, 2022.

Michael Broxton, John Flynn, Ryan Overbeck, Daniel Erickson, Peter Hedman, Matthew Duvall, Jason Dourgarian, Jay Busch, Matt Whalen, and Paul Debevec. Immersive light field video with a layered mesh representation. *ACM Transactions on Graphics (TOG)*, 39(4):86–1, 2020.

Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen. Unstructured lumigraph rendering. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 2001.

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*, pp. 130–141, 2023.

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision*, pp. 333–350. Springer, 2022.

Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. High-quality streamable free-viewpoint video. *ACM Transactions on Graphics (TOG)*, 34(4):1–13, 2015.

Paul E Debevec, Camillo J Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1996.

Yilun Du, Yinan Zhang, Hong-Xing Yu, Joshua B Tenenbaum, and Jiajun Wu. Neural radiance flow for 4d view synthesis and video processing. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14304–14314. IEEE Computer Society, 2021.

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

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*, pp. 1–9, 2022.

John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker. Deepview: View synthesis with learned gradient descent. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2367–2376, 2019.

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5501–5510, 2022.

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*, pp. 12479–12488, 2023.

Wanshui Gan, Hongbin Xu, Yi Huang, Shifeng Chen, and Naoto Yokoya. V4d: Voxel for 4d novel view synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 2023.

Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5712–5721, 2021.

Xiangjun Gao, Jiaolong Yang, Jongyoo Kim, Sida Peng, Zicheng Liu, and Xin Tong. Mps-nerf: Generalizable 3d human rendering from multiview images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1996.

Kaiwen Guo, Feng Xu, Yangang Wang, Yebin Liu, and Qionghai Dai. Robust non-rigid motion tracking and surface reconstruction using l0 regularization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3083–3091, 2015.

Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, et al. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (TOG)*, 38 (6):1–19, 2019.

Xiang Guo, Jiadai Sun, Yuchao Dai, Guanying Chen, Xiaoqing Ye, Xiao Tan, Errui Ding, Yumeng Zhang, and Jingdong Wang. Forward flow for novel view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16022–16033, 2023.

Tao Hu, Tao Yu, Zerong Zheng, He Zhang, Yebin Liu, and Matthias Zwicker. Hvtr: Hybrid volumetric-textural rendering for human avatars. In *2022 International Conference on 3D Vision (3DV)*, pp. 197–208, 2022.

Takeo Kanade, Peter Rander, and PJ Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE multimedia*, 4(1):34–47, 1997.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (TOG)*, 42(4):1–14, 2023.

Georgios Kopanas, Thomas Leimkühler, Gilles Rainer, Clément Jambon, and George Drettakis. Neural point catacaustics for novel-view synthesis of reflections. *ACM Transactions on Graphics (TOG)*, 41(6):1–15, 2022.

Kiriakos N Kutulakos and Steven M Seitz. A theory of shape by space carving. *International journal of computer vision*, 38:199–218, 2000.

Marc Levoy and Pat Hanrahan. Light field rendering. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1996.

Hao Li, Linjie Luo, Daniel Vlasic, Pieter Peers, Jovan Popović, Mark Pauly, and Szymon Rusinkiewicz. Temporally coherent completion of dynamic shapes. *ACM Transactions on Graphics (TOG)*, 31(1):1–11, 2012.

Lingzhi Li, Zhen Shen, Zhongshu Wang, Li Shen, and Ping Tan. Streaming radiance fields for 3d video synthesis. *Advances in Neural Information Processing Systems*, 35:13485–13498, 2022a.

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*, pp. 5521–5531, 2022b.

Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6498–6508, 2021.

Zhong Li, Yu Ji, Wei Yang, Jinwei Ye, and Jingyi Yu. Robust 3d human motion reconstruction via dynamic template construction. In *2017 International Conference on 3D Vision (3DV)*, pp. 496–505, 2017.

Zhong Li, Minye Wu, Wangyiteng Zhou, and Jingyi Yu. 4d human body correspondences from panoramic depth maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2877–2886, 2018.

Haotong Lin, Sida Peng, Zhen Xu, Tao Xie, Xingyi He, Hujun Bao, and Xiaowei Zhou. High-fidelity and real-time novel view synthesis for dynamic scenes. In *SIGGRAPH Asia Conference Proceedings*, 2023.

Yu-Lun Liu, Chen Gao, Andreas Meuleman, Hung-Yu Tseng, Ayush Saraf, Changil Kim, Yung-Yu Chuang, Johannes Kopf, and Jia-Bin Huang. Robust dynamic radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13–23, 2023.

B Mildenhall, PP Srinivasan, M Tancik, JT Barron, R Ramamoorthi, and R Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, 2020.

Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019.

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.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (TOG)*, 41(4):1–15, 2022.

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*, pp. 5865–5874, 2021a.

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 Transactions on Graphics (TOG)*, 40(6):1–12, 2021b.

Sida Peng, Yunzhi Yan, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Representing volumetric videos as dynamic mlp maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4252–4262, 2023.

Eric Penner and Li Zhang. Soft 3d reconstruction for view synthesis. *ACM Transactions on Graphics (TOG)*, 36(6):1–11, 2017.

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*, pp. 10318–10327, 2021.

Gernot Riegler and Vladlen Koltun. Free view synthesis. In *European Conference on Computer Vision*, pp. 623–640, 2020.

Steven M Seitz and Charles R Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35:151–173, 1999.

Ruizhi Shao, Zerong Zheng, Hanzhang Tu, Boning Liu, Hongwen Zhang, and Yebin Liu. Tensor4d: Efficient neural 4d decomposition for high-fidelity dynamic reconstruction and rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16632–16642, 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.

Pratul P Srinivasan, Richard Tucker, Jonathan T Barron, Ravi Ramamoorthi, Ren Ng, and Noah Snavely. Pushing the boundaries of view extrapolation with multiplane images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 175–184, 2019.

Zhuo Su, Lan Xu, Zerong Zheng, Tao Yu, Yebin Liu, and Lu Fang. Robustfusion: Human volumetric capture with data-driven visual cues using a rgbd camera. In *The European Conference on Computer Vision*, pp. 246–264, 2020.

Justus Thies, Michael Zollhöfer, and Matthias Nießner. Deferred neural rendering: Image synthesis using neural textures. *Acm Transactions on Graphics (TOG)*, 38(4):1–12, 2019.

Fengrui Tian, Shaoyi Du, and Yueqi Duan. Mononerf: Learning a generalizable dynamic radiance field from monocular videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17903–17913, 2023.

Edith Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Christoph Lassner, and Christian Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12959–12970, 2021.

Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5481–5490. IEEE, 2022.

Michael Waechter, Nils Moehrle, and Michael Goesele. Let there be color! large-scale texturing of 3d reconstructions. In *European Conference on Computer Vision*, pp. 836–850. Springer, 2014.

Feng Wang, Zilong Chen, Guokang Wang, Yafei Song, and Huaping Liu. Masked space-time hash encoding for efficient dynamic scene reconstruction. *Advances in Neural Information Processing Systems*, 2023a.

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*, pp. 19706–19716, 2023b.

Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3295–3306, 2023c.

Daniel N Wood, Daniel I Azuma, Ken Aldinger, Brian Curless, Tom Duchamp, David H Salesin, and Werner Stuetzle. Surface light fields for 3d photography. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pp. 487–496. 2023.

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. *arXiv preprint arXiv:2310.08528*, 2023.

Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5438–5448, 2022a.

Qingshan Xu, Weihang Kong, Wenbing Tao, and Marc Pollefeys. Multi-scale geometric consistency guided and planar prior assisted multi-view stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4945–4963, 2022b.

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 (CVPR)*, pp. 8285–8295, June 2023.

Zeyu Yang, Hongye Yang, Zijie Pan, Xiatian Zhu, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. In *International Conference on Learning Representations (ICLR)*, 2024.

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. *arXiv preprint arXiv:2309.13101*, 2023.

Taoran Yi, Jiemin Fang, Xinggang Wang, and Wenyu Liu. Generalizable neural voxels for fast human radiance fields. *arXiv preprint arXiv:2303.15387*, 2023.

Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020.

Qiang Zhang, Seung-Hwan Baek, Szymon Rusinkiewicz, and Felix Heide. Differentiable point-based radiance fields for efficient view synthesis. In *SIGGRAPH Asia 2022 Conference Papers*, pp. 1–12, 2022.

Kaichen Zhou, Jia-Xing Zhong, Sangyun Shin, Kai Lu, Yiyuan Yang, Andrew Markham, and Niki Trigoni. Dynpoint: Dynamic neural point for view synthesis. *Advances in Neural Information Processing Systems*, 36, 2024.

Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification. *ACM Transactions on Graphics*, 37(4):1–12, 2018.

C Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High-quality video view interpolation using a layered representation. *ACM transactions on graphics (TOG)*, 23(3):600–608, 2004.